Conservation of the folding mechanism between a thermostable $(\beta\alpha)_8$-barrel enzyme, its evolutionary relatives, and ancestral precursors generated by protein design

DISSERTATION ZUR ERLANGUNG DES DOKTORGRADES
DER NATURWISSENSCHAFTEN (DR RER. NAT.)
DER FAKULTÄT FÜR BIOLOGIE UND VORKLINISCHE MEDIZIN
DER UNIVERSITÄT REGENSBURG

vorgelegt von

Linn Carstensen

aus Köln

im Jahr 2012

Das Promotionsgesuch wurde eingereicht am: 22.05 2012

Die Arbeit wurde angeleitet von: Prof. Dr. Reinhard Sterner

Unterschrift:

Prüfungsausschuss:

Vorsitzender:	Prof. Dr. Christoph Oberprieler
Erster Prüfer:	Prof. Dr. Reinhard Sterner
Zweiter Prüfer:	Prof. Dr. Franz-Xaver Schmid
Dritter Prüfer:	Prof. Dr. Dr. Robert Kalbitzer
Ersatzprüfer:	PD Dr. Rainer Merkl

Index of contents

Abstract

The $(\beta\alpha)_8$-barrel is amongst the most ancient, frequent and versatile enzyme folds. Sequence and structure comparisons suggest that modern $(\beta\alpha)_8$-barrels have divergently evolved from a common ancestor, which raises the question whether they share the same folding mechanism. Here, this problem was appraoched by using the thermostable cyclase subunit HisF of imidazole glycerol phosphate (ImGP) synthase from *Thermotoga maritima* as a test case.

In the first part of the work, the folding mechanism of HisF was elucidated and compared to the folding mechanisms of other $(\beta\alpha)_8$-barrel proteins with low sequence identity. The equilibrium unfolding transition of HisF is reversible and adequately decribed by the two-state model without intermediates. Upon refolding at low denaturant concentrations a burst-phase intermediate is formed rapidly. It contains a significant amount of secondary structure but is probably an off-pathway species. In the further course of folding two phases are observed, which are associated with changes in both intrinsic protein fluorescence and far-UV circular dichroism (CD). Double mixing experiments revealed a sequential mechanism with an on-pathway intermediate which is formed in the first phase and converted into the native protein in the second phase. Refolding experiments in presence of the glutaminase subunit HisH of ImGP synthase showed that the formation of native HisF is a prerequisite for complex formation. The refolding of HisF grossly resembles other $(\beta\alpha)_8$-barrel proteins, suggesting that the general folding properties within this structural class of proteins have been conserved upon the vast sequence and functional diversification during evolution. The increased thermodynamic stability of HisF compared to $(\beta\alpha)_8$-barrels from mesophiles is entirely due to a drastically decreased unfolding rate.

The pronounced two-fold symmetry of the sequence and the $(\beta\alpha)_8$-barrel structure of HisF suggest that its fold has arisen by the duplication and fusion of a $(\beta\alpha)_4$-half-barrel. Primordial $(\beta\alpha)_8$-barrels mimicking the precursor of HisF were previously constructed by fusing two copies of its C-terminal half-barrel HisF-C. The resulting HisF-CC construct was optimized by rational protein design and directed laboratory evolution, yielding the artificial $(\beta\alpha)_8$-barrels Sym1 and Sym2.

In the second part of this thesis, the structures, stabilities, and folding mechanisms of Sym1 and Sym2 were analyzed and compared with HisF. Sym1 was shown to be less stable than HisF and its crystal structure shows disorder in the contact regions between the two half-barrels. The next-generation construct Sym2 turned out to be more stable than HisF, and the

contact regions are well resolved. However, important characteristics of the kinetic folding pathway are identical for the three proteins: At moderate denaturant conditions, unfolding and refolding are mono-exponential reactions, but at native-like conditions, refolding is complex. For all three proteins, HisF, Sym1, and Sym2, an off-pathway burst-phase folding intermediate I_{BP} is observed by far-UV CD, followed by a fast reaction that occurs with comparable rates ($\tau \sim 3$ s) and leads to the on-pathway intermediate I. The subsequent slow and rate-determining folding reaction has a similar rate for HisF and Sym2 ($\tau \sim 20$ s) but a lower rate for Sym1 ($\tau = 175$ s) and leads to the native state N. For Sym2, the additional on-pathway intermediate I' is formed in an extremely fast reaction ($\tau = 0.27$ s). These findings point to a similar sequential folding pathway for the three proteins ($I_{BP} \rightarrow U \rightarrow I' \rightarrow I \rightarrow N$) with I' being energetically favored in the case of Sym2 but too unstable to be populated in the case of HisF and Sym1. In contrast to the similarities in refolding, the unfolding kinetics of HisF, Sym1, and Sym2 are characterized by extremely different rates which are responsible for their different thermodynamic stabilities. The conservation of the folding mechanism further backs the hypothesis that $(\beta\alpha)_8$-barrels have evolved from an ancestral $(\beta\alpha)_4$-half-barrel and provide compelling evidence that Sym1 and Sym2 are realistic models for the early fusion protein product. Moreover, the different stabilities of Sym1 and Sym2 emphasize the significance of optimized contact regions of the fused half-barrels for the generation of a stable $(\beta\alpha)_8$-barrel with a well-defined folding landscape.

Kurzfassung der Arbeit

Das $(\beta\alpha)_8$-*barrel* gehört zu den ursprünglichsten, vielseitigsten und häufigsten Faltungstypen bekannter Proteine. Sequenz- und Strukturvergleiche legen nahe, dass $(\beta\alpha)_8$-*barrel* Enzyme das Produkt einer divergenten Evolution ausgehend von einem gemeinsamen Vorläufer sind. Daraus ergibt sich die Frage, ob alle Vertreter einen gemeinsamen Faltungsmechanismus aufweisen.

Die thermostabile Synthase-Untereinheit HisF der Imidazolglycerinphosphat (ImGP)-Synthase aus *Thermotoga maritima* ist ein sehr gutes Modellprotein, um diese Problemstellung zu untersuchen. Deshalb wurde im ersten Teil dieser Arbeit der Faltungsmechanismus von HisF aufgeklärt und mit anderen potentiell verwandten $(\beta\alpha)_8$-*barrel* Proteinen mit niedriger Sequenzidentität verglichen. Die GdmCl-induzierte Entfaltung von HisF ist reversibel und wird durch das apparente Zwei-Zustands-Modell ohne detektierbare Intermediate beschrieben. Bei der Rückfaltung unter nativ-ähnlichen Bedingungen wird schnell ein *burst-phase* Intermediat gebildet, das einen erheblichen Anteil an definierter Sekundärstruktur aufweist. Im Anschluss zur Bildung dieses vermutlich unproduktiven Intermediats lassen sich zwei Rückfaltungsreaktionen nachweisen, die jeweils zu einer Änderung der intrinsischen Proteinfluoreszenz und des Circulardichroismus- (CD) Signals führen. Anhand von Doppelsprungexperimenten konnte gezeigt werden, dass in der schnellen Phase ein produktives Intermediat gebildet wird, das in einem sequentiellen Mechanismus direkt im Rahmen der langsamen Faltungsphase in den nativen Zustand umgewandelt wird. Ferner belegten Rückfaltungsexperimente in Anwesenheit der Glutaminase-Untereinheit (HisH) der ImGP-Synthase, dass der native Zustand von HisF essentiell für die Komplexbildung ist. Insgesamt zeigt die Rückfaltung von HisF deutliche Übereinstimmungen zu anderen $(\beta\alpha)_8$-*barrel* Proteinen. Diese Ergebnisse stützen die These, dass grundlegende Eigenschaften der Faltungsmechanismen von $(\beta\alpha)_8$-*barrel* Enzyme trotz der Differenzierung der Aminosäuresequenzen und Funktionen im Laufe der Evolution konserviert wurden. Im Gegensatz zu $(\beta\alpha)_8$-*barrel* Proteinen aus mesophilen Organismen weist HisF eine höhere Thermostabilität auf, die auf einer extrem langsamen Entfaltungsrate begründet ist.

Die ausgeprägte zweifache Symmetrie der Aminosäuresequenz und der $(\beta\alpha)_8$-barrel Struktur von HisF weist auf die Entstehung der Faltung durch Genduplikation und Fusion ausgehend von $(\beta\alpha)_4$-Einheiten hin. In vorherigen Arbeiten wurde dieser Prozess nachgeahmt, indem ein

$(\beta\alpha)_8$-*barrel* Protein durch die Fusion zweier identischer Kopien der C-terminalen Hälfte von HisF (HisF-C) konstruiert wurde. Das resultierende Konstrukt HisF-CC wurde durch eine Kombination aus rationalem Proteindesign und gelenkter Laborevolution optimiert, was zu den stabilen artifiziellen $(\beta\alpha)_8$-*barrel* Proteinen Sym1 und Sym2 führte.

Im zweiten Teil dieser Arbeit wurden die Struktur, Stabilität und Faltung von Sym1 und Sym2 analysiert und mit dem natürlich evolvierten Enzym HisF verglichen. Sym1 ist weniger stabil als HisF und in der Kristallstruktur sind die Kontaktregionen zwischen den $(\beta\alpha)_4$-Einheiten nicht aufgelöst. Im Gegensatz dazu ist das stärker optimierte Sym2 stabiler als HisF und auch die Kontaktregionen sind in der Kristallstruktur klar erkennbar. Trotz dieser Stabilitätsunterschiede, weisen die Faltungsmechanismen der drei Proteine deutliche Ähnlichkeiten auf: Bei mittleren Denaturierungsmittelkonzentrationen ist die Ent- und Rückfaltungsreaktion mono-exponentiell, wohingegen die Rückfaltung unter nativ-ähnlichen Bedingungen komplexer ist. Bei allen drei Proteinen, HisF, Sym1 und Sym2, entsteht zunächst ein unproduktives *burst-phase* Intermediat, gefolgt von einer schnellen Faltungsphase mit ähnlicher Rate ($\tau \sim 3$ s), in der das produktive Intermediat I gebildet wird. In dem folgenden langsamen und raten-limitierenden Schritt wird der native Zustand von HisF und Sym2 mit ähnlicher Rate ($\tau \sim 20$ s) und der von Sym1 mit geringerer Rate ($\tau = 175$ s) gebildet. Zusätzlich zeigt Sym2 ein weiteres produktives Intermediat I‘, das extrem schnell ($\tau = 0.27$ s) gebildet wird. Diese Ergebnisse deuten auf einen übereinstimmenden sequentiellen Faltungsmechanismus für die drei Proteine hin ($I_{BP} \rightarrow U \rightarrow I' \rightarrow I \rightarrow N$), wobei I‘ bei Sym2 energetisch begünstigt, aber bei HisF und Sym1 hochenergetisch und dementsprechend nicht detektierbar ist. Im Gegensatz zu diesen Ähnlichkeiten bei der Rückfaltung, unterscheiden sich die Entfaltungsraten von HisF, Sym1 und Sym2 deutlich und sind für die Unterschiede der thermodynamischen Stabilität verantwortlich. Die konservierten Faltungsmechanismen von HisF, Sym1 und Sym2 stützen weiter die These der Entstehung von $(\beta\alpha)_8$-*barrel* Proteinen durch die Duplikation und Fusion von $(\beta\alpha)_4$-*barrel* Proteinen und zeigen, dass es sich bei Sym1 und Sym2 um realistische Modelle des ursprünglichen Fusionsproteins handelt. Ferner verdeutlichen die Unterschiede der Stabilität von Sym1 und Sym2 die Relevanz der optimierten Kontaktregionen zwischen den fusionierten $(\beta\alpha)_4$-Einheiten für die Entstehung eines stabilen $(\beta\alpha)_8$-*barrel* Proteins.

1 Introduction

1.1 Protein stability

Proteins are versatile macromolecules that are relevant for most cellular processes. They are involved in signaling and transport, have structural and defense functions, and act as enzymes that catalyze specific chemical reactions. The functionality of a protein is determined by its stable three-dimensional structure. However, many enzymes must also display a certain degree of local or global conformational flexibility to allow for efficient substrate binding and catalysis. Protein stability results from a variety of non-covalent forces. The main contributor is the hydrophobic effect, which is caused by the absence of hydrogen bonds between non-polar groups and water and results in the entropically disfavored solvation of hydrophobic residues. This leads to the transfer of hydrophobic residues from water to the protein core (1, 2), where they form enthalpically favored van der Waals interactions (3). Further stabilizing forces are hydrogen bonding interactions of the protein backbone and among polar amino acids, as well as salt brides between charged amino acid side chains (3-5). These favorable energetic contributions are largely compensated by the destabilizing loss of chain entropy during folding (3). As a consequence, most proteins from mesophilic organisms are only marginally stable. In order to determine the thermodynamic stability of a protein, the delicate balance between stabilizing and destabilizing contributions is being shifted towards unfolding by elevated temperature, high pressure, or denaturing agents such as urea or guanidinium HCl (GdmCl).

Small proteins commonly show a two-state equilibrium between the unfolded state U and the native state N, which is determined by the rate constants of folding k_f and unfolding k_u:

$$N \underset{k_f}{\overset{k_u}{\rightleftharpoons}} U \qquad (1)$$

In this simplest case the free-energy of unfolding (ΔG_D), describing the thermodynamic stability is calculated as follows:

$$\Delta G_D = - RT \cdot \ln (k_u/k_f) = - RT \cdot \ln K = - RT \cdot \ln \frac{[U]}{[N]} \qquad (2)$$

R, T and K represent the gas constant, the absolute temperature and the unfolding equilibrium constant. ΔG_D generally shows a linear dependency on denaturant concentration ([D]) with a slope m and an intersection point that corresponds to the free-energy change of unfolding in the absence of denaturant (ΔG_D^{H20}) (6):

$$\Delta G_D = \Delta G_D^{H20} - m \cdot [D] \qquad (3)$$

Numerous unfolding studies have shown that the ΔG_D^{H20} values of most proteins lie in the range between 10 and 70 kJ mol^{-1} (7).

In addition to thermodynamic stability, kinetic stability is an important feature which guarantees that a protein remains in its folded conformation on a biological relevant time scale (8). As unfolded and partially folded conformations are prone to aggregation and other irreversible alterations, a high free-energy barrier of unfolding is required to protect the native state at the harsh conditions of extracellular or crowded intracellular environments. In fact, decreased kinetic stability plays a crucial role in diseases related to misfolding and aggregation such as triosephosphate isomerase deficiency, amyotrophic lateral sclerosis, phenylketoenuria, and transthyretin amyloidoses (8-13). Even single mutations were shown to lead to a decreased kinetic stability, resulting in misfolding on a biological relevant time-scale (9).

1.2 Proteins from hyperthermophilic organisms show an increased stability

Hyperthermophilic organisms optimally grow at temperatures of 80 °C or above (14). Their proteins generally exhibit extreme thermal stabilities which often come along with a pronounced resistance against chemical denaturation and proteolysis. In order to identify the molecular basis of this high conformational and chemical robustness, the structures of homologous proteins from hyperthermophilic and mesophilic hosts were compared in a number of studies. The results suggested that an increase in stability can be achieved by a variety of strategies, including an elevated number of ionic interactions and H-bonds, improved core packing, greater rigidity, extended secondary structure, shorter surface loops, and higher states of oligomerization (14-18). However, it is often difficult to discern differences in amino acid sequence that are caused by the adaption to high temperatures from those ones that are the result of a selective pressure for an optimized function or natural drift occurring in the course of evolution (19). Whereas this phylogenetic noise has prevented the

deciphering of general structural rules determining high stability, a more uniform picture has emerged with respect to the kinetic basis for elevated ΔG_D values (20-25). In most analyzed cases stabilization was found to be caused by an increased kinetic energy barrier against unfolding resulting in a decreased value for k_u. On the contrary, the differences between the refolding rate constants k_f of homologous proteins from thermophilic and mesophilic hosts turned out to be rather small.

1.3 Protein folding

In the process of protein folding, the linear amino acid sequence of the polypeptide chain spontaneously adopts a well-defined and compact three-dimensional native structure (26). This structure is characterized by the lowest free energy of all possible conformations. Due to the multitude of statistical orientations of the chain during refolding, searching randomly all possible backbone conformations would take far longer than the typically milliseconds-to-seconds time scale observed for the folding of small proteins. Consequently, the folding process needs to be biased by a specific pathway and the free-energy as driving force (27). It has been postulated that the energy landscape resembles a global funnel which guides folding towards the native state through a stochastic process in which the free energy decreases spontaneously (28, 29). On these pathways the protein adopts well-defined intermediate states. Based upon this hypothesis, the unfolded state, possible intermediates, and the native state correspond to local minima and the transition states to saddle points in the free energy landscape. In early protein folding studies long-lived intermediates, which mainly undergo slow *cis*/*trans* isomerizations of peptidyl prolyl bonds or disulfide bond formation reactions, have been analyzed (30-33). In recent years, advances in experimental techniques such as nuclear magnetic resonance, mass spectrometry, hydrogen exchange, fluorescence resonance energy transfer, and atomic force microscopy have made it possible to gain detailed structural information also about transiently formed intermediate states (34-40).

Although protein folding pathways have been extensively studied, many of the basic principles that relate protein sequence with the structure of folding intermediates and the native structure are still unknown. For example, an important open question is, whether the folding mechanism of proteins is dictated by the topology of their polypeptide chains or by their individual amino acid sequences. The relevance of these factors has been evaluated by comparing the folding of proteins with similar folds but low sequence identity (41-45). These

studies suggest, that although local interactions determine the precise folding mechanism, the gross topology defines the structure adopted in the transition state and thus the repertoire of potential folding routes (46).

1.4 Natural selection for an optimal energy landscape of folding

Stability and folding are important factors in protein evolution, because conformational integrity is crucial for the preservation and advancement of molecular function. For example, a high thermodynamic stability efficiently buffers the often destabilizing effect of amino acid exchanges that are required for the generation of new enzymatic activities on an existing protein scaffold (47). A high kinetic barrier against unfolding prevents the formation of non-native conformational states that tend to aggregate or are proteolytically degraded. Remarkably, although only weak non-covalent interactions are being formed during the genesis of the three-dimensional structure, large energetic barriers have to be crossed also during protein folding. Thus, the cooperative folding/unfolding process of natural proteins is characterized by a smooth and funneled energy landscape with distinct energy barriers (28). On the contrary, proteins designed in the laboratory lacking an 'evolutionary history' are characterized by low-cooperativity equilibrium transitions, almost barrier-less rapid folding and unfolding kinetics, and the formation of non-native states during folding (48-51). Importantly, such ragged energy landscapes can readily lead to the fatal trapping of the polypeptide chain in non-native minima. The differences between natural and artificial proteins indicate that cooperative folding, comprising distinct energy barriers, is not a fundamental physicochemical requirement but the consequence of natural selection (8). It is plausible to assume that the selection pressure for efficient folding has been operative during the entire course of protein evolution. This notion raises the question to what extent the folding mechanisms of proteins and protein families that have evolved divergently from a common ancestor, have been conserved upon the diversification of their sequences.

1.5 The $(\beta\alpha)_8$-barrel fold

About 1200 different protein folds have been identified to date [Structural Classification of Proteins (SCOP) release 1.75, February 2009] (52), each of them being characterised by a distinct toplogical orientation of secondary structure elements. Whereas many folds are

represented by only a few members in the protein database, others have been recruited extensively in the course of evolution. A prominent example is the $(\beta\alpha)_8$-barrel, which is amongst the most ancient, frequent and versatile folds (18, 53, 54). About 10% of all proteins with known three-dimensional structure contain at least one $(\beta\alpha)_8$-barrel domain. With very few exceptions all known $(\beta\alpha)_8$-barrels are enzymes, and SCOP distuinguishes 33 superfamilies that catalyze more than 60 different reactions. With the exception of ligases, $(\beta\alpha)_8$-barrels occur in all Enzyme Commission (EC) classes and many of them are engaged in essential metabolic pathways (55, 56).

The canonical $(\beta\alpha)_8$-barrel fold contains at least 200 amino acid residues and is composed of eight units. A single unit consists of a β-strand and an α-helix, which are linked by a βα-loop, and the individual units are connected by αβ-loops (**Figure 1a**). The eight strands form a central parallel β-sheet, which is surrounded by an outer layer of eight α-helices (**Figure 1b**).

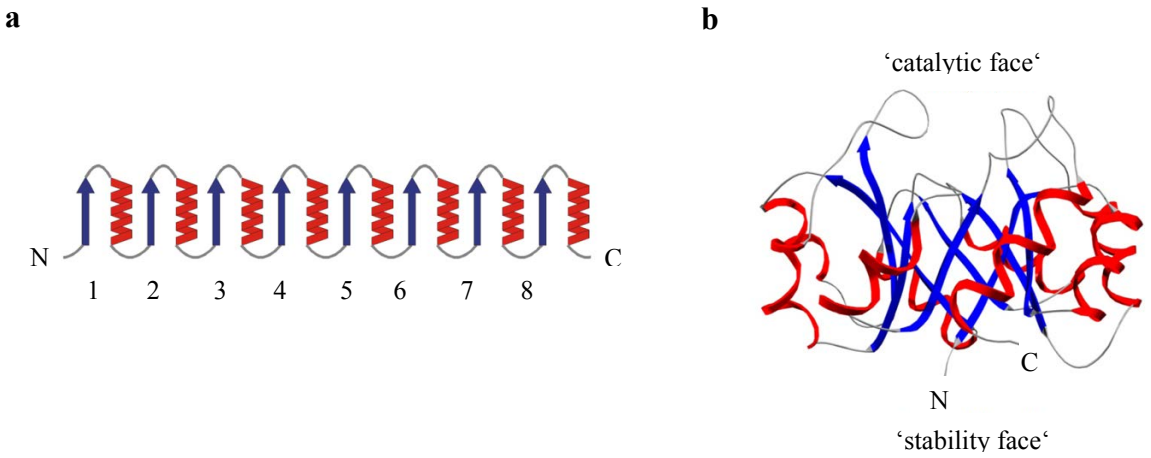


Figure 1: Schematic depiction of the $(\beta\alpha)_8$-barrel fold.
(a) The eight βα-units are connected by loops within the units (βα-loops) and between them (αβ-loops). (b) Three-dimensional structure of a specific protein in ribbon representation, with the central eight-stranded parallel β-sheet (the 'barrel') surrounded by the eight α-helices. The active site is formed by residues at the C-terminal ends of the β-strands and the βα-loops ('catalytic face'). The remainder of the fold, including the opposite face of the barrel with the αβ-loops, is important for stability ('stability face'). The N- and C-terminal ends of the polypeptide chain are labeled. α-helices and β-strands are shown in red and blue, loops are shown in gray.

1.6 Evolution of the $(\beta\alpha)_8$-barrel fold

1.6.1 Divergent *versus* convergent evolution of $(\beta\alpha)_8$-barrel proteins

In all known $(\beta\alpha)_8$-barrels the residues that are important for substrate specificity and catalytic activity are located at the C-terminal ends of the β-strands and in the subsequent βα-loops ('catalytic face') (**Figure 1b**). The remainder of the fold, including the opposite face of the barrel with the αβ-loops, is important for stability ('stability face') (18). However, in spite of the common chain topology and the conserved location of the active site, the geometry of the central barrel can differ considerably and many $(\beta\alpha)_8$-barrel enzymes contain extensions to the canonical topology, either at the N- or C-termini, or in loop segments. Moreover, the overall sequence identity between different $(\beta\alpha)_8$-barrel proteins is lower than 20 % (57, 58). For these reasons, it is not clear whether all $(\beta\alpha)_8$-barrels are the products of divergent evolution from a common ancestor or whether the fold has rather developed independently several times by convergent evolution (56, 59, 60).

More recent extensive sequence and structure comparisons are in favor of a common evolutionary origin, at least for most $(\beta\alpha)_8$-barrels (55, 57). A well studied example are phosphate-binding $(\beta\alpha)_8$-barrels such as phosphoribosylanthranilate (PRA) isomerase (TrpF), indole-3-glycerol phosphate synthase (TrpC), and α-subunit of tryptophan synthase (TrpA). TrpF, TrpC, and TrpA, which catalyze three successive steps in trytophan biosynthesis, share a common phosphate-binding site that is located in βα-loops 7 and 8 and serves to anchor their mono-phosphorylated substrates (61, 62). The enzymes N'-[(5'-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (ProFAR) isomerase (HisA) and imidazoleglycerol phosphate synthase (HisF), which catalyze two successive reactions within histidine biosynthesis, contain two symmetry-related phosphate binding sites to anchor their bi-phosphorylated substrates. Whereas the N-terminal binding site located in βα-loops 3 and 4 is unique to HisA and HisF, the C-terminal phosphate-binding site located in βα-loops 7 and 8 is shared with TrpF, TrpC, and TrpA (63-65). This notion has stimulated attempts to interconvert the catalytic activities between these enzymes (65). Indeed, a few amino acid exchanges turned out to be sufficient to establish PRA isomerisation activity on the scaffolds of TrpA, HisA, and HisF, as well as on a HisAF chimera (66-68), further supporting the view that the $(\beta\alpha)_8$-barrel enzymes from tryptophan and histidine biosynthesis have evolved from a common ancestor.

1.6.2 Mimicking the evolution of the $(\beta\alpha)_8$-barrel fold from a $(\beta\alpha)_4$-half-barrel

The striking two-fold symmetry of HisA and HisF is not restricted to the location of the two phosphate binding sites but is a feature of the entire $(\beta\alpha)_8$-barrel fold of both enzymes. For example, in HisF from the hyperthermophile *Thermotoga maritima*, the four N-terminal $(\beta\alpha)_{1\text{-}4}$ units (HisF-N) and the four C-terminal $(\beta\alpha)_{5\text{-}8}$ units (HisF-C) show sequence identities of about 20 % and r.m.s. deviations of their Cα-atoms of only 1.58 Å. When produced seperately, HisF-N and HisF-C form homodimers with well-defined secondary and tertiary structure elements. Following their joint expression in *Escherichia coli* or common refolding, the two half-barrels associate to the HisF-NC homodimer, which is catalytically as active as wild-type HisF (69, 70). These findings suggested that the $(\beta\alpha)_8$-barrel structure has evolved by the duplication and fusion of a $(\beta\alpha)_4$ half-barrel domain. This hypothetical evolutionary pathway was mimicked in the laboratory by constructing a stable $(\beta\alpha)_8$-barrel protein from HisF-C, which served as a model for the putative ancestral $(\beta\alpha)_4$-barrel. In a first step, two copies of HisF-C were fused in tandem to generate a complete $(\beta\alpha)_8$-barrel, termed HisF-CC. Using rational protein design, HisF-CC was stabilized by the reconstruction of a salt-bridge cluster present in HisF (71). The subsequent shortening of the linker region between the two fused halves HisF-C_N and HisF-C_C (72) and the introduction of beneficial mutations by library selection using the folding reporter chloramphenicolacetyl transferase yielded the monomeric, compact and stable protein Sym1 (**Figure 2**) (73).

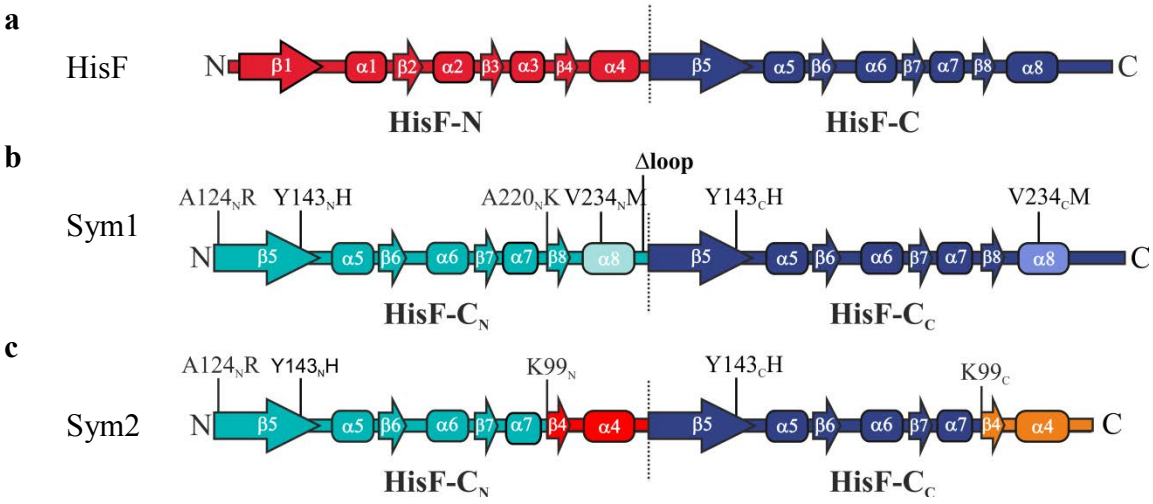


Figure 2: Design of the artificial $(\beta\alpha)_8$-barrel proteins Sym1 and Sym2 from two fused HisF-C half-barrels.
(a) Secondary structure elements of wild-type HisF. The N-terminal half-barrel HisF-N [modules $(\beta\alpha)_{1-4}$] is shown in red, the C-terminal half-barrel HisF-C [modules $(\beta\alpha)_{5-8}$] in blue. (b) Design of Sym1 (formerly denoted as HisF-C***C (73)). Two copies of HisF-C were fused and stepwise stabilized by a combination of rational design and library selection. The resulting protein Sym1 contains the indicated amino acid substitutions and a shortened loop connecting the two fused half-barrels HisF-C_N and HisF-C_C (71-73). Transparent boxes indicate the parts that were not resolved in the crystal structure. (c) Design of Sym2. Sym2 was derived from Sym1 by replacing module βα8 with βα4, in both HisF-C_N and HisF-C_C (red and orange). (Taken from ***Figure 1, publication B***)

Although the crystal structure of Sym1 revealed the expected $(\beta\alpha)_8$-barrel fold, the βα-modules 8 at the C-termini of HisF-C_N and HisF-C_C are not well defined, indicating that these regions are flexible (**Figures 2, 3**).

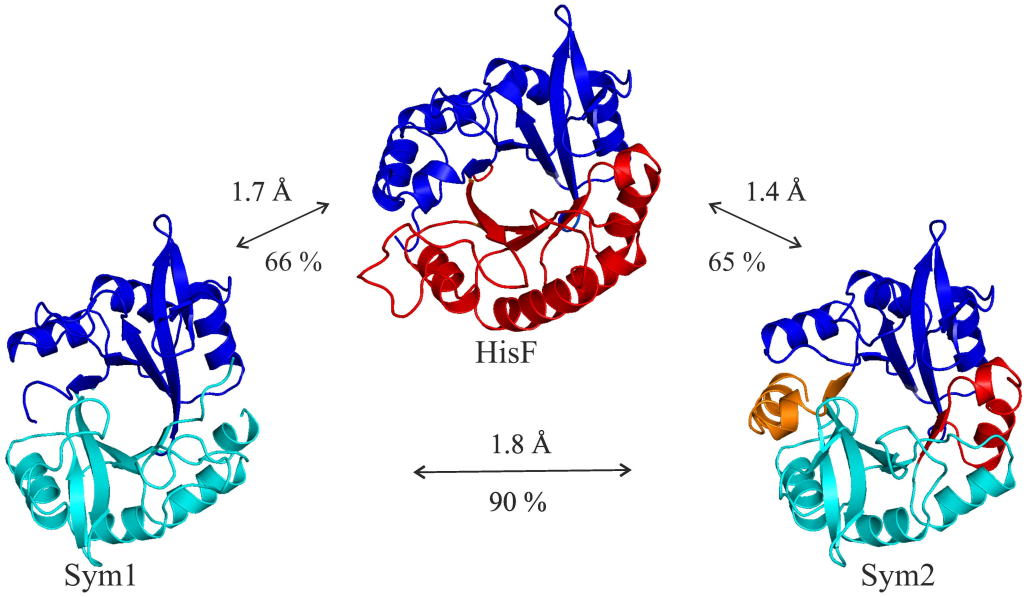


Figure 3: Ribbon diagrams of the crystal structures of HisF (1thf.pdb) (69), Sym1 (2w6r.pdb) (73), and Sym2 (3og3.pdb).
Colors represent the origin and location of secondary structure elements as described in **Figure 1**. The invisible parts in the crystal structure of Sym1 are fully resolved in Sym2. Amino acid sequence identities (%) and r.m.s. deviation values (Å) of corresponding Cα-atoms as deduced from pair-wise structure-based sequence alignments are indicated (HisF-Sym1: 184 superimposed Cα atoms; HisF-Sym2: 198 superimposed Cα atoms; Sym1-Sym2: 191 superimposed Cα atoms). (Taken from ***Figure 2, publication B***)

In order to further stabilize Sym1, the putatively labile βα-modules 8 at the C-termini of HisF-C_N and HisF-C_C were replaced by the corresponding native sequence stretch stemming from HisF-N (74). The exact intersection point was chosen on the basis of a structural superposition between HisF and Sym1, which revealed a high similarity between the regions linking the βα3 and βα4 units in HisF and the βα7 and βα8 units in Sym1. As a consequence, the βα-modules 8 in HisF-C_N and HisF-C_C of Sym1 were replaced by the βα-module 4 of HisF, leading to the new construct Sym2 (**Figure 2c**). Crystallization and structure determination of Sym2 at a resolution of 2.08 Å revealed a $(\beta\alpha)_8$-barrel with high similarity to HisF and Sym1 (**Figure 3**). The Cα-atoms of the three proteins, which show sequence identities between 65 % and 90 % superpose with r.m.s.d. values of 1.4 – 1.8 Å. Importantly, the regions at the termini of HisF-C_N and HisF-C_C, which could not be resolved in Sym1, are well defined in the structure of Sym2.

1.7 Folding mechanisms of $(\beta\alpha)_8$-barrel proteins

Early insights into the folding of $(\beta\alpha)_8$-barrel proteins were obtained by fragmentation experiments performed with TrpF and triosephosphate isomerase (TIM) (75-77). These studies provided evidence that both TrpF and TIM form stable folding intermediates consisting of four or six structured N-terminal ($\beta\alpha$) units, followed by the association of the unstructured segment to finally form the native $(\beta\alpha)_8$-barrel. Recent in-depth kinetic studies with TrpA from *E. coli* (78, 79), TrpC from *Sulfolobus solfataricus* (80-82), and a $(\beta\alpha)_8$-barrel protein of unknown function encoded by the *Bacillus subtilis iol*I gene (IOLI) (58) have afforded deeper insights into the folding of these $(\beta\alpha)_8$-barrels. All three proteins apparently form an off-pathway burst-phase intermediate, which must unfold to enter productive folding via one or more on-pathway intermediates. Although these similarities argue for conserved properties of the folding mechanisms of $(\beta\alpha)_8$-barrel proteins, the native secondary structural elements formed in the on-pathway intermediates seem to differ between TrpA and TrpC. Moreover, whereas the slow *cis*/*trans* isomerization of peptide bonds involving proline residues results in several parallel folding pathways in the case of TrpA, TrpC and IOLI contain only *trans* prolyl peptide bonds and show simpler folding mechanisms (82).

1.8 Aim of this work

Recent work suggests that a large fraction of the contemporary $(\beta\alpha)_8$-barrel proteins have evolved divergently from a common ancestor. However, it is still unclear to what extent their folding mechanism has been conserved upon the vast diversification of sequences and functions in the course of evolution. Thus, in the first part of this work the folding mechanism of the $(\beta\alpha)_8$-barrel protein HisF was elucidated and compared with the folding mechanisms of putative evolutionary relatives with different function and low sequence identity. Actually, HisF is an interesting protein for several reasons. First, unlike most other $(\beta\alpha)_8$-barrel proteins, HisF shows a striking two-fold symmetry and contains phosphate binding sites in both the N-terminal and the C-terminal half-barrel, which anchor its bi-phosphorylated substrate (69). The previously characterized TrpC and TrpA are evolutionarily related to HisF but lack such a clear two-fold symmetry and contain only a single conserved C-terminal phosphate binding site (65), which makes a comparison of their folding mechanisms particularly attractive. Second, HisF is active in its hyperthermophilic host at temperatures

above 80 °C and provides an excellent model to test the hypothesis that high thermodynamic stability is correlated with a large kinetic barrier to unfolding (20, 24, 25, 83). Finally, as HisF forms a 1:1 complex with the glutaminase subunit HisH of the imidazole glycerol phosphate synthase (64, 84), it offers an opportunity to study the relationship between protein folding and association.

Moreover, based on the hypothesis of the evolution of HisF by duplication and fusion of an ancestral $(\beta\alpha)_4$-half-barrel, two artificial $(\beta\alpha)_8$-barrel proteins (Sym1, Sym2) have been designed from HisF-C (65, 71-73). These artificial proteins should mimic evolutionary precursors of HisF. Based on the assumption, that the selection pressure for efficient folding has been operative during the entire course of protein evolution, ancestral precursors of natural proteins should show a similar folding mechanism as their contemporary descendants. In order to test this hypothesis, in the second part of this work the stability and folding of the artificial $(\beta\alpha)_8$-barrels Sym1 and Sym2 were analyzed and compared with wild-type HisF.

2 Summary and discussion

2.1 Stability and folding of HisF, Sym1, and Sym2

2.1.1 Unfolding equilibrium of HisF is attained slowly and adequately described by the two-state model

The stability and folding mechanism of HisF from the hyperthermophilic bacterium *T. maritima* was compared with other $(\beta\alpha)_8$-barrel proteins. For this purpose, the denaturant-induced loss of secondary and tertiary structure was monitored via the far-UV circular dichroism (CD) signal caused by the backbone of the polypeptide backbone chain and the intrinsic fluorescence of tryptophan (Trp) and tyrosine (Tyr) residues. The cooperative GdmCl-induced equilibrium unfolding could be adequately described by the two-state model, meaning that no significant amounts of stable equilibrium intermediates are populated (***Figure 2, publication A***). Remarkably, the time to reach equilibrium when starting from the folded state was about three weeks at 45 °C and about six weeks at 25 °C, which is reflected in the excessively slow unfolding kinetics observed in the transition region of the Chevron diagram (***Figure 5a, b, publication A***). Importantly, HisF unfolds also very slowly in the physiological temperature range of *T. maritima*, which allows for the purification of the recombinant protein by heat incubation of the *E. coli* host cell extract (85). Extremely slow unfolding kinetics has also been observed for other thermostable proteins (20, 24, 25, 83), supporting the notion that a high kinetic unfolding barrier efficiently protects proteins from hyperthermophiles against thermal denaturation (14).

2.1.2 The different thermodynamic stabilities of HisF, Sym1, and Sym2 are caused by their different unfolding rates

The evolutionary pathway from a $(\beta\alpha)_4$-half-barrel to a stable $(\beta\alpha)_8$-barrel was previously reconstructed by duplicating and fusing the C-terminal half-barrel of HisF, HisF-C. The resulting HisF-CC construct was then stabilized by several rounds of rational design and library selection (71, 73, 74, 86), yielding the artificial proteins Sym1 and Sym2 (**Figures 2, 3**). Thermodynamic stabilities of Sym1 and Sym2 were compared with the stability of HisF by GdmCl-induced unfolding monitored by Trp/Tyr fluorescence (**Figure 4a**).

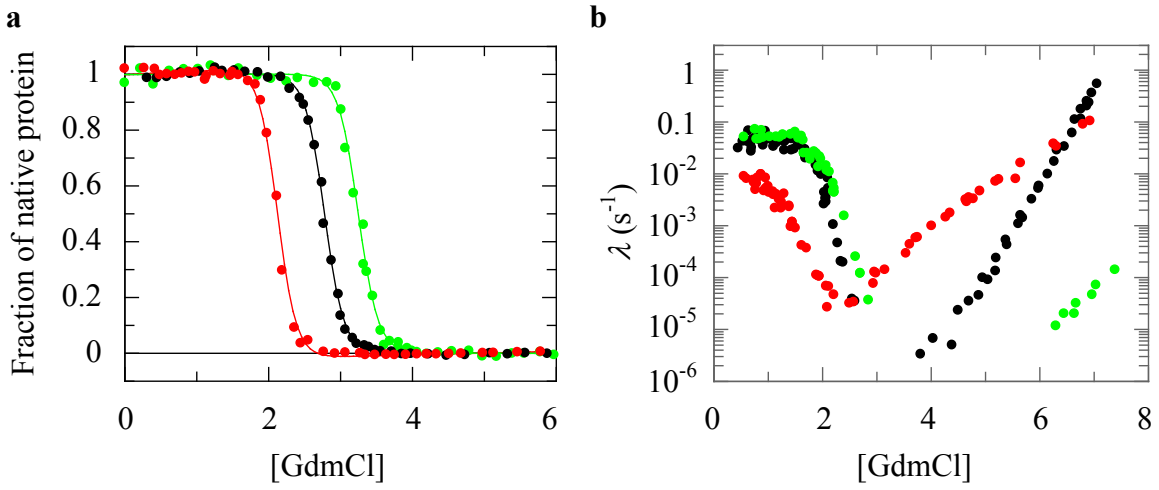


Figure 4: Equilibrium unfolding transitions and kinetic folding and unfolding of HisF (black), Sym1 (red), and Sym2 (green).
(a) GdmCl-induced unfolding was followed in 50 mM Tris/HCl buffer (pH 7.5) at 25 °C by monitoring Tyr/Trp fluorescence (excitation: 280 nm; emission: 320 nm) (Taken from ***Figure 3a, publication B***). The continuous lines represent the fit of the two-state model to the normalized unfolding transitions, yielding the thermodynamic parameters listed in **Table 1**. (b) The dependence on GdmCl concentration of the apparent rate constant (λ) is shown for the rate-limiting unfolding and refolding phases (Data taken from ***Figure 4, publication B***). Chevron diagrams of the individual proteins including all kinetic phases are shown in ***Figure S5, publication B***.

The analysis of the transitions with the two state-model showed that the three proteins unfold with similar cooperativity (m = 19.1-23.0 kJ mol^{-1} M^{-1}), which indicates that they have a comparably compact structure. However, compared to HisF, the transition midpoint ($[D]_{1/2}$) of Sym1 and Sym2 was shifted to lower and higher GdmCl concentrations, respectively, yielding values for the Gibbs free energy of denaturation (ΔG_D) of 42.8 kJ mol^{-1} for Sym1, of 53.5 kJ mol^{-1} for HisF, and of 62.2 kJ mol^{-1} for Sym2 (**Table 1**).

Table 1: **Thermodynamic stability parameters for HisF, Sym1, and Sym2.**

	N ↔ U			I_{BP} ↔ U		
	ΔG_D (kJ mol^{-1})	m (kJ mol^{-1} M^{-1})	$[D]_{1/2}$ (M)	ΔG_D (kJ mol^{-1})	m (kJ mol^{-1} M^{-1})	$[D]_{1/2}$ (M)
HisF	53.5 ± 1.7	19.1 ± 0.6	2.8	15.2 ± 2.4	7.9 ± 0.9	1.9
Sym1	42.8 ± 0.2	20.2 ± 0.4	2.1	13.2 ± 0.4	5.9 ± 0.2	2.2
Sym2	62.2 ± 3.9	19.2 ± 1.2	3.2	35.7 ± 6.5	21.4 ± 3.7	1.7

For the N ↔ U transition, the Gibbs free energy of denaturation (ΔG_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 4a**) with the two-state model. For the I_{BP} ↔ U transition, the parameters were obtained by analyzing the initial values of the refolding kinetics as monitored by the CD signal at 225 nm (***Figure S4, publication B***).

Unfolding of Sym1 reached equilibrium already after three days of incubation at 25 °C. This result is reflected in the relatively fast denaturation of the protein, which allowed for the determination of the complete unfolding limb of the chevron diagram (**Figure 4b**). However, similarly as observed for HisF, the unfolding equilibrium of Sym2 was reached only after several weeks of incubation at 45 °C and 25 °C (***Figure S3, publication B***), which precluded the determination of the unfolding limb of the two proteins at moderate concentrations of GdmCl. Remarkably, the unfolding kinetics of Sym2 is even further slowed down by three orders of magnitude compared to HisF (**Figure 4b**). Since the refolding rates of HisF, Sym1, and Sym2 are similar (**Figure 4b**) the different unfolding rates are the main cause for the varying thermodynamic stabilities of the three proteins (**Figure 4, Table 1**).

The incomplete unfolding limbs of HisF and Sym2 did not allow for a precise extrapolation of the unfolding rate constants in the absence of denaturant. However, a rough estimation yielded $k_u^{H_2O}$ values of about 10^{-11} and 10^{-13} s^{-1} for HisF and Sym2, corresponding to time constants for unfolding ($\tau = 1/k_u^{H_2O}$) of around 600 and 200,000 years, respectively. These values testify to an extremely high kinetic stability which efficiently protects the proteins from unfolding on a biologically relevant time scale. Interestingly, even Sym1 shows a considerable kinetic stability with an extrapolated unfolding time constant of 17 years.

Even more extreme activation barriers for unfolding with $k_u^{H_2O}$ values of 10^{-12} - 10^{-15} s^{-1}, corresponding to time constants up to two million years, have been observed for domains of

type1 pilin (87). These individual pilus subunits associate through donor strand complementation to form filamentous, adhesive pili, which are essential for many Gram-negative pathogens for adhesion to the host tissue. As a consequence of the high kinetic barriers for unfolding and dissociation the assembly of pilus subunits becomes practically irreversible under physiological conditions.

The extremely different unfolding rates of Sym1 and Sym2 are remarkable given their high overall amino acid sequence identity of 90 %. Apparently, they are caused by the replacement of the two βα modules 8 in Sym1 by the two βα modules 4 in Sym2 (**Figure 2**). These contact regions between the fused $(\beta\alpha)_4$-half-barrels are not resolved and therefore flexible in the crystal structure of Sym1 but clearly visible and therefore rigid in the structure of Sym2 (**Figure 3**). These findings emphasize that optimized contact regions between half-barrels are crucial for the thermodynamic and kinetic stability of $(\beta\alpha)_8$-barrel proteins. The significance of 'locking' the structure by forming tight contacts between the N- and C-terminal ends of the polypeptide chain was further tested by mutational analysis. For this purpose, Trp35 which is located in α-helix 1 of Sym2 and whose bulky indole side chain forms extensive van der Waals interactions with hydrophobic C-terminal residues, was replaced with Tyr. Remarkably, this single exchange led to a 40-fold increase of the unfolding rate (**Figure 5**).

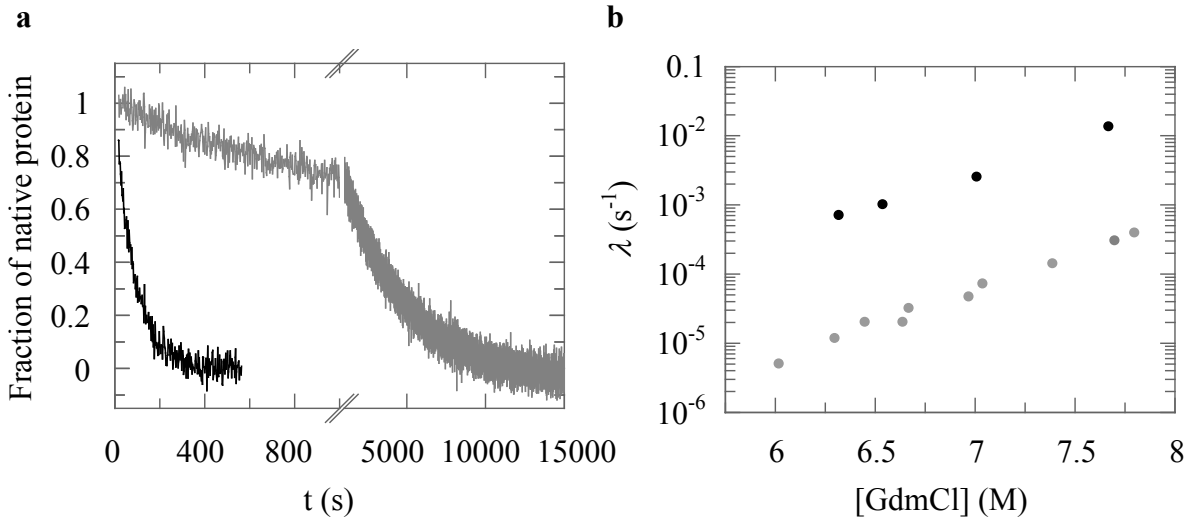


Figure 5: Unfolding of Sym2 (gray) and Sym2-W35Y (black).
(a) Following manual mixing, the unfolding traces in 7.7 M GdmCl were monitored by the far-UV CD signal at 225 nm in 50 mM Tris/HCl buffer (pH 7.5) at 25 °C. (b) GdmCl-dependence of the apparent rate constants (λ) of unfolding obtained by fitting mono-exponential functions to the unfolding traces.

2.1.3 HisF folds via a sequential mechanism

To elucidate the folding mechanism of HisF, un- and refolding kinetics were followed by far-UV CD and Trp/Tyr fluorescence. The analysis of the refolding reaction revealed a large change of the far-UV CD signal within the dead-time of stopped-flow mixing, indicating the formation of a burst-phase intermediate (***Figure 4a, publication A***). The subsequent productive folding occurs in two kinetic phases with time constants τ of about 3 s and 20 s (***Figure 5a, publication A***). The biphasic refolding kinetics and the reciprocal change of the amplitudes of the two phases (***Figure 5c, publication A***) suggested that they reflect the rapid formation of an intermediate and its subsequent slow transformation to the native state. To examine whether such an intermediate is in fact formed as a transient species during refolding of HisF, stopped-flow double-mixing approaches were performed. In these experiments, refolding was interrupted after various times, and the concentrations of species present at the time of sampling were determined from the rates and amplitudes of their unfolding reaction.

First, unfolding of the putative intermediate, which is formed within the fast 3 s refolding reaction, was monitored. For this purpose, refolding was interrupted after 10 s, a time point at which the putative intermediate should be well populated whereas the concentration of fully folded molecules should still be insignificantly low. Then the sample was transferred to unfolding conditions to monitor the rate and the amplitude of the unfolding reaction. These experiments showed that within the 3 s refolding phase an intermediate I is formed, which unfolds three orders of magnitude faster than the native protein N (***Figure 6, publication A***).

Second, another set of interrupted refolding experiments was performed to follow the time course of the population of the intermediate I and the accumulation of the native state N. In these experiments, samples of HisF were allowed to refold for various times before being transferred to unfolding conditions. The large difference in the unfolding rates of I and N allows one to detect their unfolding reactions simultaneously and with high precision. The fluorescence amplitudes of these unfolding reactions are proportional to the concentration of I and N present at the time when refolding was interrupted and thus trace the time course of I and N during refolding (***Figure 8, publication A***). Importantly, the observed rates for the formation of I coincided with the fast folding phase observed by conventional refolding kinetics. Moreover, the rates observed for the depletion of I and the formation of N were identical and coincided with the slow folding phase observed by conventional refolding kinetics. These findings, together with a lag in the time-course of N, complete the evidence that folding is a sequential process with I as an obligatory on-pathway intermediate (***Scheme***

1, publication A). I_{BP} is formed within the dead time of refolding. The pronounced roll-over observed for the fast folding phase below 1.0 M GdmCl (***Figure 5a, publication A***), resulting in a positive *m*-value of refolding, indicates that I_{BP} is off-pathway and contains non-native interactions that have to be broken before I can form. However, the alternative of I_{BP} being on-pathway cannot be definitely excluded. At concentrations > 2 M GdmCl both folding and unfolding can be described by a single phase, probably because I_{BP} and I are too unstable under these conditions to accumulate to detectable concentrations (***Figure 4 and 5, publication A***).

2.1.4 Sym1 and Sym2 show an identical folding mechanism as HisF

For a comparison with wild-type HisF, un- and refolding kinetics of Sym1 and Sym2 were followed by far-UV CD and Trp/Tyr fluorescence. Despite the large differences in stability and unfolding rate constants, the refolding mechanisms of the Sym1 and Sym2 are remarkable similar to HisF. As observed for HisF, also Sym1 and Sym2 form burst-phase off-pathway intermediates with a high content of secondary but no detectable tertiary structure (***Figure S3, publication B***). The further course of folding is characterized by a fast reaction, which has a similar time constant (τ = 3 s) for all three proteins. This reaction is followed by a slower phase, with similar time constants for HisF and Sym2 (τ = 20 s) but a lower one for Sym1 (τ = 175 s). Sym2 shows an additional very rapid folding reaction (τ = 0.27 s), which indicates the population of an additional intermediate I' that could not be detected for the other two proteins (**Figure 6**).

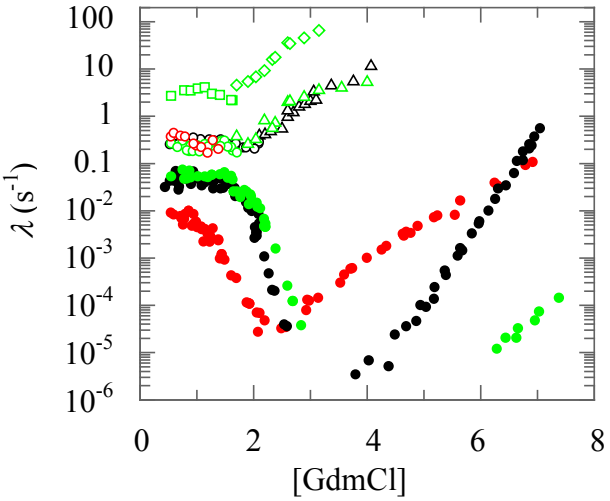


Figure 6: GdmCl-dependence of the apparent rate constants (λ) of folding and unfolding of HisF (black), Sym1 (red), and Sym2 (green).
The rate constants for unfolding of I' and I were determined by double-jump experiments (***Figure 6a, publication A; Figure 5, publication B***), and the other constants were determined by single-jump experiments (***Figure 5a, publication A; Figure 4, publication B***).

For HisF, the biphasic refolding kinetics with a reciprocal change of the amplitudes of the fast and the slow refolding reaction have suggested the formation of an on-pathway folding intermediate I and its subsequent transformation to the native state in a sequential folding mechanism. This presumption was confirmed by interrupted refolding experiments (2.1.3). For Sym2, the amplitudes of the three consecutive folding phases also change in a reciprocal manner with denaturant concentration (***Figure S6c, publication B***), suggesting that folding occurs sequentially via two productive intermediates, I' and I. In order to substantiate this notion, the time courses of the formation and decay of the putative intermediates were monitored by interrupted refolding experiments. These experiments showed that I' and I, which form within the very rapid and the fast refolding reaction, respectively, unfold much faster than the native protein (**Figure 6**). Moreover, the experiments provided compelling evidence for a sequential folding mechanism where I' is directly transformed to I and then further on to native Sym2 (***Figure 5b, c, publication B***). Interestingly, although I' is formed very rapidly, the overall folding process of Sym2 is not accelerated compared to HisF as the overall rate-limiting formation of I and N occurs equally fast for the two proteins (**Figure 6**). 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 (***Figure S6a, publication B***). Nevertheless, the observation of two refolding phases (**Figure 6**) and the amplitude profiles (***Figure S6a, publication B***) point to a similar sequential mechanism for Sym1 as well. Taken together, these findings suggest a common folding mechanism for HisF, Sym1, and Sym2 (**Figure 7a**).

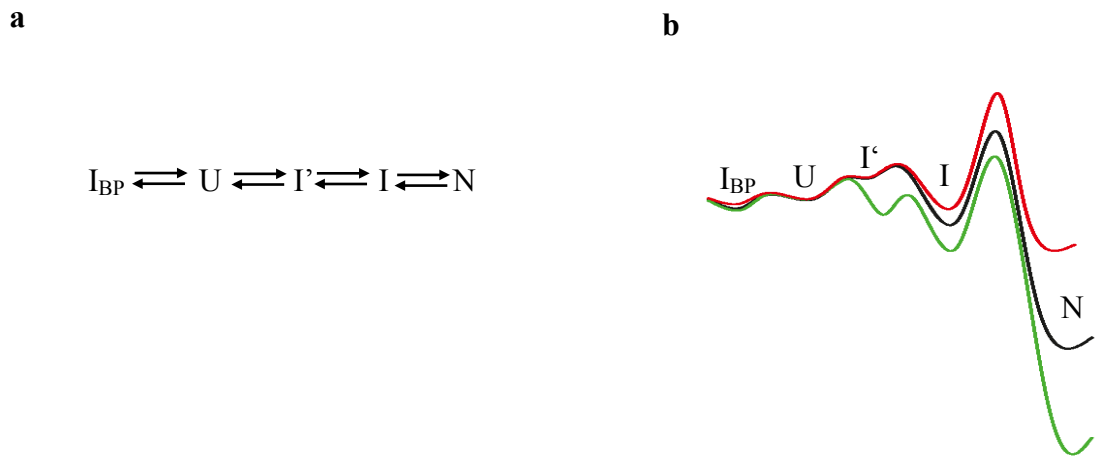


Figure 7: Unifying folding mechanism for HisF, Sym1, and Sym2.
(a) Folding model compatible with the experimental data. An off-pathway equilibrium between the unfolded state (U) and the burst-phase intermediate (I_{BP}) precedes the formation of the on-pathway intermediates I' and I. For HisF and Sym1, I' is a high energy intermediate and not kinetically detectable. At concentrations > 2 M GdmCl, I' and I are too instable to accumulate to detectable concentrations. (b) Energy diagram for the folding of HisF (black), Sym1 (red), and Sym2 (green). The different heights of the energy barriers indicate different folding/unfolding rates but are not at scale. (Taken from ***Figure 6, publication B.***)

All three proteins form a burst-phase intermediate I_{BP}, which is most likely off-pathway, as suggested by the positive *m*-value of the fast refolding phases of HisF (2.1.3) and Sym1, and of the very rapid refolding phase of Sym2. Then, rapidly the on-pathway intermediate I' is formed, which is probably an undetectable high energy intermediate for HisF and Sym1 (88). I' is directly converted to a second on-pathway intermediate I with similar rates for all three proteins. In the final rate-limiting step of folding, I is converted into the native state. At concentrations > 2 M GdmCl, both folding and unfolding of HisF, Sym1 and Sym2 can be described by a single phase, most probably because I' and I are destabilized by the denaturant

and do not accumulate. Thus, although the sequences between HisF and Sym1 / Sym2 differ by about 50 %, the folding mechanisms of the three $(\beta\alpha)_8$-barrel proteins are largely identical. The energy barriers for folding and unfolding as deduced from the determined rate constants are depicted in **Figure 7b**. For all three proteins, folding is synchronized by an activation barrier of similar height between I and N. The large differences in equilibrium stability develop only late in folding, after the passage over this critical activation barrier. Compared to HisF and especially Sym1, the native state of Sym2 is drastically stabilized with respect to the transition state. Furthermore, a stabilizing effect in Sym2 is already reflected in I' which, other than in Sym1 and HisF, is detectable due to its lower free-energy. The destabilization of Sym1 compared to HisF and Sym2 is already observed for I as the amplitude of the fast refolding phase is reduced and vanishes already in the presence of moderate concentrations of GdmCl (***Figure S6a, publication B***).

2.1.5 Structural properties of the refolding intermediates of HisF, Sym1, and Sym2

The experimental data presented in the previous chapters provide information about the structural properties of the folding intermediates I_{BP}, I', and I. For HisF, Sym1, and Sym2 the formation of the burst-phase intermediate I_{BP} within the first few milliseconds of refolding is accompanied by a large change of the far-UV CD signal. This CD signal decreases in a cooperative manner with increasing concentration of GdmCl, and the resulting unfolding curve suggests that I_{BP} is significantly less stable than the native protein (**Table 1**). Albeit the pronounced far-UV CD signal of I_{BP} indicates a high content of secondary structure, the identity of its Trp/Tyr fluorescence signal with the unfolded state U suggests the absence of a defined tertiary structure (***Figure 3a, publication A***). The formation of the on-pathway intermediates I' and I is accompanied by changes of both the far-UV CD signal and Trp/Tyr fluorescence, indicating that they contain more secondary and tertiary structure than their respective precursors on the folding pathway.

The structural properties of the folding intermediate states of HisF were further evaluated by four different experimental approaches: i) binding of the dye 1-anilino-8-naphthalenesulfonate (ANS), ii) Förster resonance energy transfer (FRET) between strategically positioned donor and acceptor chromophores, iii) intrinsic fluorescence of single Trp residues placed at various parts of the protein, and iv) nuclear magnetic resonance (NMR).

Binding of ANS was used to assay the accessibility of hydrophobic areas in the folding intermediates. ANS is known to bind with high affinity to solvent-exposed non-polar patches, leading to an increase of its fluorescence intensity (89, 90). Stopped-flow refolding experiments in the presence of ANS yielded a strong fluorescence increase within the dead time of mixing (***Figure S3a, publication A***), indicating that the dye efficiently binds to I_{BP}. Thus, I_{BP} is probably a molten globule-like state (91) that forms rapidly when the unfolded protein is exposed to native conditions. In the further course of folding the ANS fluorescence decreased, indicating that the transition from I_{BP} to N leads to a continuous decrease of the exposed hydrophobic surface of HisF. Since ANS binds with lower affinity to I than to I_{BP}, (***Figure S3b, publication A***), the hydrophobic side chains are shielded more efficiently from the solvent in the productive than in the off-pathway intermediate.

FRET experiments can be used as a tool to follow distance changes between individual amino acid residues during folding (39, 92-102). Depending on the donor and acceptor fluorophores, FRET can probe distances in a range of about 10 Å to 100 Å. Importantly, the transfer efficiency (E) depends on the sixth power of the distance between donor and acceptor, which allows for the detection of even small conformational rearrangements (93, 97). The introduced chromophores should be reporters only and not affect the stability and the folding kinetics of the protein, a condition that is met best when both donor and acceptor remain mobile and exposed during the folding process.

Based on these considerations, the exposed N- and C-termini of HisF, which are in close contact in the native protein but far remote in its unfolded state, were chosen for modification. Moreover, FRET measurements are most sensitive close to the Förster radius R_0, which is the donor-acceptor distance at which the transfer efficiency is equal to 50 %. Therefore, Trp residues and 5-(((acetylamino)ethyl)amino)naphthalene-1-sulfonate (AEDANS) were used as donor and acceptor. This pair of chromophores shows an R_0 value of about 22 Å (93), which is close to the 20 Å distance of the N-and C-terminal residues as observed in the crystal structure of HisF.

Single Trp residues and single cysteine (Cys) residues required for the coupling of the AEDANS group (100, 103) were introduced by site-directed mutagenesis. Wild-type HisF contains a single Cys at position 9 and a single Trp at position 156, which are not suitable for FRET experiments and were therefore replaced by alanine (Ala) and Tyr, respectively. Moreover, in order to be able to replace the essential methionine (Met) 1 by Trp and Cys, an additional Met residue was inserted at position 0. The resulting pseudo wild-type protein

(HisF*) served as the parent for the variants that carried individual pairs of donor and acceptor chromophores (**Table 2**).

Table 2: Distances between donor and acceptor sites of the AEDANS-labeled HisF* variants.

HisF* variant	Native protein (0 M GdmCl)			Unfolded protein (6.5 M GdmCl)		
	E	*R* (Å)	$R_{(crystal)}$ (Å)	*E*	*R* (Å)	$R_{(coill)}$ (Å)
HisF*-M1W/E251C	0.64	19.9	16.2	0.35	24.4	92.0
HisF*-M1C/E253W	0.63	20.1	20.6	0.45	22.7	92.4

HisF* corresponds to HisF with an additional N-terminal Met residue and the C9A and W156Y exchanges. The FRET efficiency (*E*) and the apparent distance between the fluorophores (*R*) are given for the HisF*-variants in the absence and presence of 6.5 M GdmCl. *R* was determined from *E* with $R=R_0((1/E)-1)^{1/6}$, where R_0 is the Förster distance of 22 Å. *E* was obtained from the emission spectra of the donor in the donor-acceptor-labeled (F_{DA}) and the donor-only protein (F_D) by $E=1-(F_{DA}/F_D)$ (100, 103). Exemplary emission spectra are shown in **Figure 8**. The distance between C_α of Trp and C_α of Cys deduced from the crystal structure (*R*(crystal)) and the calculated distance between the donor and acceptor amino acids in a polypeptide chain with non-repetitive structure (*R*(coil)) are given. The distance in the unfolded polypeptide chain was calculated by the expression $R(coil)=5.45(aa_1-aa_2)^{1/2}$, where aa_1 and aa_2 are the positions of the respective amino acids (98).

The FRET efficiency depends not only on the donor-acceptor distance but also on the relative orientation of the chromophores, the quantum yield of the donor, and the overlap integral between the fluorescence spectrum of the donor and the absorption spectrum of the acceptor. Ideally, these properties should remain constant when FRET is used to follow distance changes during folding. In addition, the introduced chromophores should not affect the stability and the folding kinetics of the protein (100). However, the variants of HisF were slightly destabilized (ΔG_D around 46 kJ mol^{-1} versus 54 kJ mol^{-1} for wild-type HisF). Moreover, the emissions of the introduced Trp residues in the donor-only proteins are blue shifted upon refolding. This finding indicates that they are at least partly immobilized, which is unfavorable for FRET experiments (**Figure 8a**).

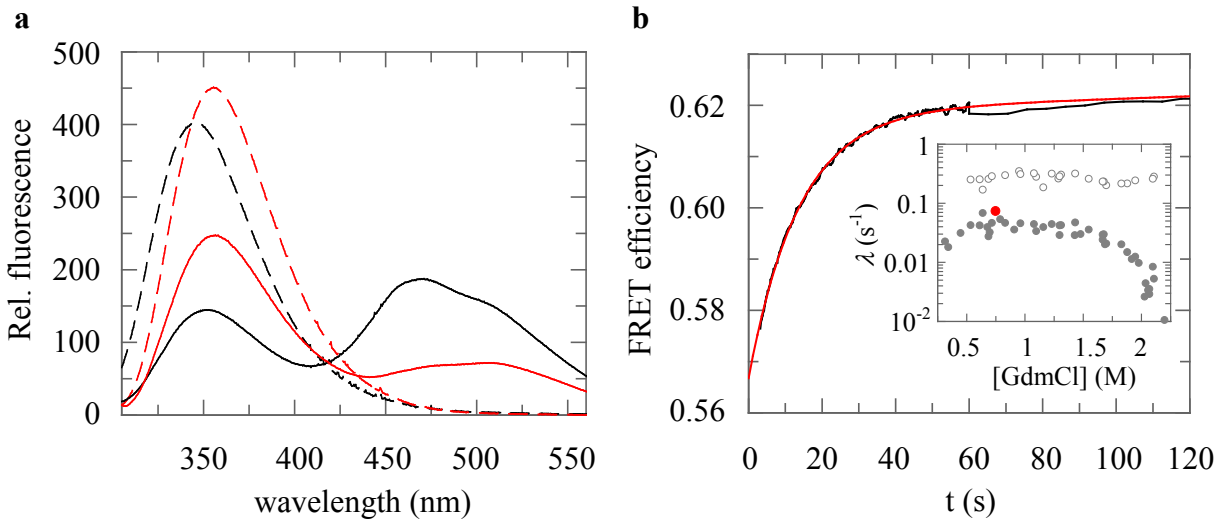


Figure 8: Fluorescence spectra and refolding kinetics monitored by FRET efficiency of HisF*-M1C/E253W.
(a) Fluorescence spectra of the donor–acceptor-labeled protein HisF*-M1C/E253W-AEDANS (continuous line), and the donor-only protein HisF*-M1C/E253W (broken line). Fluorescence emission spectra for the folded state (0 M GdmCl) are shown in black, those in the unfolded state (6.5 M GdmCl) are shown in red. The spectra were measured after excitation at 295 nm with 2 µM protein in 50 mM Tris/HCl (pH 7.0) at 25 °C. The calculated values for *E* and *R* are given in **Table 2**. (b) Change of transfer efficiency during refolding obtained from the fluorescence change of the donor in the donor-accepted-labeled and the donor only protein determined as outlined in **Table 2**. A mono-exponential function in combination with a linear term was fit to the experimental curve (red line), yielding a time constant τ of 13.7 s. Inset: This time constant (red circle) is compared with the refolding rates determined by single-jump experiments (gray circles, data taken from ***Figure 5a, publication A***).

Nevertheless, the distances *R* between the chromophores were investigated in the unfolded and native variants of HisF (**Figure 8a, Table 2**) and compared with values obtained from the crystal structure and a model of a unfolded polypeptide chain with non-repetitive structure (98). The calculated distances of about 20 Å deduced from the FRET experiments performed with the native proteins are in good agreement with the crystal structure, whereas the calculated distances of 23 Å for the unfolded state are far shorter than the distance of 92 Å, estimated by the chain-like model. The reasons for this deviance, which could for example be misplaced acceptor molecules, a limited mobility of the chromophores, or local structural regions in the unfolded state, were not explored. The kinetics of the FRET efficiency change corresponding to the distance change of about 3 Å during folding was mono-exponential. The deduced time constant was identical to the time constant of the slow folding reaction

determined by single-jump experiments (**Figure 8b**). This finding shows that detectable rearrangements in the terminal region of HisF 'locking' the barrel do not occur during the formation of I but in the subsequent rate-limiting step leading to N.

Intrinsic protein fluorescence was used to determine which parts of I adopt a native-like structure with shielded Trp residues. For this purpose, HisF variants with single Trp residues placed in different regions of the protein were produced, and their refolding reactions were analyzed. The results suggest that I has a native-like structure in the central region of the protein, whereas its N-terminal region is apparently unstructured (***Figure 11, publication A***). This finding is consistent with the previously discussed FRET experiments and indicates that rearrangements in the terminal region occur only in the late stages of folding. This assumption was further supported by an experiment in which refolding of HisF was combined with its complex formation with HisH. The results showed that only native HisF but not I can associate with HisH (***Figure 10, publication A***), indicating that the binding interface is not yet formed in I. In accordance with this notion, the crystal structure of the complex shows that the N-terminal region of HisF, which is unstructured in I, forms intimate contacts with residues from HisH (84).

Finally, to obtain detailed structural information on the folding intermediates of HisF refolding was analyzed by solution NMR spectroscopy using ^{1}H-^{15}N HSQC experiments. The spectral assignment of HisF (104, 105) and the slow formation of the native state at 2.4 M GdmCl (τ = 2.8 h) allow for the recording of NMR spectra during the early phase of folding. Refolding of 1 mM unfolded ^{1}H-^{15}N-HisF was initiated by diluting the protein into 2.4 M GdmCl (native conditions, see **Figure 4a**), and spectra were recorded by Dr. Christoph Liebold (University of Regensburg, Institute of Biophysics and Physical Biochemistry) as described previously (106). However, even after a long incubation time of about 20 h the obtained chemical shifts were typical of a random coil and did not indicate the formation of any defined structure (data not shown). Possibly, the high protein concentration promoted the formation of aggregates, which however did not reach a visible size due to the presence of 2.4 M GdmCl. It was not possible to work with lower protein concentrations for technical reasons.

2.2 Implications for the evolution of $(\beta\alpha)_8$-barrels

2.2.1 The folding of HisF resembles other $(\beta\alpha)_8$-barrels

Extensive sequence and structure comparisons suggest a common evolutionary origin for most $(\beta\alpha)_8$-barrels (55, 57). However, it is still unclear to what extent their folding mechanism has been conserved upon the vast diversification of sequences and functions in the course of evolution. In order to evaluate the conservation of the folding mechanism of $(\beta\alpha)_8$-barrels, the folding properties of HisF were compared with those of bona fide evolutionary related $(\beta\alpha)_8$-barrel proteins, namely TrpA from *E. coli* (78, 79), TrpC from the hyperthermophile *S. solfataricus* (80-82), and IOLI from *B. subtilis* (58).

Despite the low overall sequence identity of less than 20 % (as detected by a pair-wise comparison with DALI-lite; http://www.ebi.ac.uk/Tools/dalilite/), the kinetic folding mechanism of HisF resembles the three other analyzed $(\beta\alpha)_8$-barrel proteins. In all cases, a burst-phase intermediate I_{BP} is formed (58, 79, 82), which is marginally stable but contains a high fraction of the far-UV CD signal of the native state (HisF: 74%; TrpA: 47%; TrpC: 46%; IOLI: 37%). Global folding analysis or Gō-model simulation has suggested that the I_{BP} of TrpA (79, 107), TrpC (82), and IOLI (58) are off-pathway species, meaning that the unfolding of their non-native structural elements controls access to the productive on-pathway intermediates. Substantial burst-phase reactions have also been observed for the $(\beta\alpha)_8$-barrel proteins fructose-1,6-bisphosphate aldolase (108) and TrpF (109), as well as for the $(\beta\alpha)_5$-barrel proteins CheY (110) and Spo0F (111). The latter were proposed to be evolutionarily related to $(\beta\alpha)_8$-barrels via a common $(\beta\alpha)_4$-half-barrel precursor (112, 113).

In addition to the off-pathway intermediate I_{BP}, one or two on-pathway kinetic intermediates are formed on the folding pathways of HisF, TrpA, TrpC, and IOLI (58, 79, 80, 114). However, the native secondary structural elements formed in the on-pathway intermediates seem to differ between the individual proteins (75, 81, 82, 115) and the actual kinetics appear to be determined by protein stability and amino acid sequence. A direct comparison of the unfolding and refolding rates of the four $(\beta\alpha)_8$-barrel proteins HisF, TrpA, TrpC and IOLI is complicated by the fact that different denaturants, GdmCl and urea, were used. Nevertheless, at low denaturant concentration all four proteins show similar rate-limiting steps with time constants of about 10 s. These findings suggest that for refolding, the $(\beta\alpha)_8$-barrel proteins from hyperthermophiles use the same mechanism as their counterparts from mesophilic organisms. Thus, despite the strong differences in stability and amino acid sequence

diversities of more than 80 %, general properties of folding are conserved. This is in line with the divergent evolution of $(\beta\alpha)_8$-barrels from a common ancestor.

In contrast to the refolding reaction, unfolding of HisF from the hyperthermophilic organism *T. maritima* is at least 1000-fold slower than unfolding of TrpA and IOLI, which are from mesophiles. Remarkably, unfolding of HisF is also about 100-fold slower than the unfolding of truncated TrpC, which originates also from the hyperthermophilic organism *S. solfataricus*, but shows a lower thermodynamic stability (80). Thus, adaption to high temperatures essential for proteins of hyperthermophilic organisms is attained by an extreme deceleration of the unfolding rate without changing the kinetic folding mechanism. The stabilizing amino acid substitutions exert their influence late on the folding pathway, by selectively stabilizing the native state relative to the transition state.

2.2.2 Sym1 and Sym2 are models for an evolutionary precursor of HisF

Remarkable differences have been observed between the folding of proteins engineered in the laboratory and arisen by natural evolution. Whereas billions of years of natural selection has afforded significant barriers for un- and refolding to avoid the formation of partially unfolded conformations prone to aggregation, artificial proteins are generally characterized by rugged folding energy landscapes and almost barrier-less folding and unfolding (8, 48-51). On the contrary, all essential elements of the folding of native HisF are also observed for the artificial proteins Sym1 and Sym2. All three proteins show a linear five-state mechanism with one off-pathway and two on-pathway intermediates and comparable refolding energy barriers (**Figure 7a, b**). This conservation of folding provides further compelling evidence that $(\beta\alpha)_8$-barrels have evolved by the duplication and fusion of an ancestral $(\beta\alpha)_4$-half-barrel and that Sym1 and Sym2 are realistic models for an early fusion protein. Since the primary fusion event must have occurred in an early phase of evolution, the folding mechanism has been conserved over very long periods of time. Moreover, the differences between Sym1 and Sym2 emphasize the significance of optimized contact regions of the fused half-barrels for the generation of a stable $(\beta\alpha)_8$-barrel with high thermodynamic and kinetic stability and a smooth and well-defined folding landscape.

2.3 Comparison with the folding mechanism of other characterized proteins

Driven by natural selection, sequences have evolved that fold as efficiently as possible, to protect the functional native state against unfolding and aggregation. However, evolutionary pressure on a fast folding reaction has a cutoff, as proteins have to fold only fast enough to be produced without undesirable side reactions. Any further increase in the folding rate is no apparent advantage to the cell, leading to folding processes occurring typically on a milliseconds-to-seconds time scale. Albeit the folding reaction of $(\beta\alpha)_8$-barrel proteins with time constants of several seconds is much slower than the folding of small proteins, it still fits this time range for efficient folding.

The occurrence of on- and off-pathway intermediates is a common characteristic of all hitherto analyzed $(\beta\alpha)_8$-barrel proteins, as well as calmodulin (116), UMP/CMP kinase (117) and the $(\beta\alpha)_5$-barrel proteins CheY (110) and Spo0F (111). This property also contrasts with small proteins, which commonly fold via two-state mechanisms avoiding intermediates. In recent years, however, partially unfolded intermediates have been identified also for typical two-state proteins (118-123). These intermediates are usually weakly populated or short-lived and not amendable to direct detection in kinetic experiments. Even though energetically less favorable, some intermediates are important for the physiological function of proteins by exposing ligand binding sites or allowing for post-translational modifications. However, the folding intermediates of $(\beta\alpha)_8$-barrels do not have an obvious biological relevance. Their elaborate folding mechanism might rather be owed to the complex native topology, which is formed by at least 200 amino acids, in contrast to the commonly less than 100 amino acids of apparent two-state folders. Interestingly, even the apparently 'misfolded' species I_{BP} is a conserved feature of $(\beta\alpha)_8$-barrels, probably as a consequence of the opportunity for regional folding. It has been argued for TrpC that the formation of this sub-millisecond off-pathway species will not interfere with productive folding in the cell, because it would appear and dissipate far faster than the rate of synthesis on a ribosome or the subsequent productive folding reaction (82). Thus, the complex and slow folding of the $(\beta\alpha)_8$-barrel seems not to be a biological relevant disadvantage, in accordance with its frequent occurrence.

3 Abbreviations

Å	Ångström (10^{-10} m)
AEDANS	5-(((acetylamino)ethyl)amino)naphthalene-1-sulfonate
Ala	alanine
ANS	1-anilino-8-naphthalenesulfonate
B. subtilis	*Bacillus subtilis*
°C	degree celsius
CD	circular dichroism
C-terminal	carboxyterminal end of a polypeptide chain
Cys	cysteine
[D]	denaturant concentration
$[D]_{1/2}$	transition midpoint of denaturant-induced unfolding
E	FRET efficiency
EC	Enzyme Commission
E. coli	*Escherichia coli*
F_D	fluorescence of the donor in the donor–only protein
F_{DA}	fluorescence of the donor in the donor–acceptor-labeled protein
FRET	Förster resonance energy transfer
ΔG_D	Gibbs free-energy of unfolding
GdmCl	guanidinium chloride
HisA	*N*′-[(5′-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide isomerase
HisF	synthase subunit of imidazole glycerol phosphate synthase from *Thermotoga maritima*
HisF*	HisF pseudo wild-type protein containing the C9A and W156Y exchanges and an additional Met at position 0
HisF-N	N-terminal half of HisF
HisF-C	C-terminal half of HisF
HisF-CC	two in tandem fused copies of HisF-C
HisF-C_N	N-terminal half of HisF-CC
HisF-C_C	C-terminal half of HisF-CC

HisH	glutaminase subunit of imidazole glycerol phosphate synthase from *T. maritima*
HSQC	heteronuclear single quantum coherence
I	kinetic on-pathway intermediate
I'	further kinetic on-pathway intermediate
I_{BP}	burst-phase intermediate
ImGP	imidazole glycerol phosphate
IOLI	protein of unknown function encoded by the *Bacillus subtilis iol*I gene
K	unfolding equilibrium constant
k_f	rate constant of folding
k_u	rate constant of unfolding
λ	apparent rate constant of a reaction
m	parameter defining the cooperativity in denaturant-induced unfolding
Met	methionine
N	native state of a protein
NMR	nuclear magnetic resonance
PRA	phosphoribosylanthranilate
R	apparent distance between two fluorophores
R_0	Förster distance at which the FRET efficiency is 50 %
R(coil)	calculated distance between the FRET donor and acceptor amino acids in a polypeptide chain with non-repetitive structure
R(crystal)	distance between C_α of Trp and C_α of Cys from the crystal structure of HisF
r.m.s.d.	root mean square deviation
SCOP	structural classification of proteins
S. sulfataricus	*Sulfolobus solfataricus*
Sym1	HisF-CC stabilized by shortening the loop connecting HisF-C_N and HisF-C_C, as well as the A124R and A220K exchanges in HisF-C_N and the Y143H & V234M exchanges in both HisF-C_N and HisF-C_C
Sym2	Sym1 stabilized by replacing βα-module 8 in HisF-C_N and HisF-C_C with the stable βα-module 4 stemming from the N-terminal half of HisF
T	temperature
t	time
τ	time constant of a reaction

T. maritima	*Thermotoga maritima*
TIM	triosephosphate isomerase
Tris	2-amino-2-hydroxymethyl-propane-1,3-diol
Trp	tryptophan
TrpA	α-subunit of tryptophan synthase
TrpC	indole-3-glycerolphosphate synthase
TrpF	phosphoribosylanthranilate isomerase
Tyr	tyrosine
U	unfolded state of a protein

4 Literature

1. Kauzmann W (1959). Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **14**, 1-63.
2. Dill KA, Bromberg S, Yue K, S. CH, Fiebig KM, Yee DP, Thomas PD & Chan HS (1995). Principles of protein folding - A perspective from simple exact models. *Protein Sci.* **4**, 561-602.
3. Makhatadze GI & Privalov PL (1995). Energetics of protein structure. *Adv. Protein Chem.* **47**, 307-425.
4. Pace C, Shirley B, McNutt M, & Gajiwala K (1996). Forces contributing to the conformational stability of proteins. *FASEB J.* **10**, 75-83.
5. Honig B & Yang AS (1995). Free energy balance in protein folding. *Adv. Protein Chem.* **46**, 27-58.
6. Santoro MM & Bolen DW (1988). Unfolding free energy changes determined by the linear extrapolation method 1. Unfolding of phenylmethanesulfonyl α-chymotrypsin using different denaturants. *Biochemistry* **27**, 8063-8068.
7. Privalov PL (1979). Stability of proteins. *Adv. Protein Chem.* **33**, 167-241.
8. Sanchez-Ruiz JM (2010). Protein kinetic stability. *Biophys. Chem.* **148**, 1-15.
9. Rodríguez-Almazán C, Arreola R, Rodríguez-Larrea D, Aguirre-López B, de Gómez-Puyou MT, Pérez-Montfort R, Costas M, Gómez-Puyou A, & Torres-Larios A (2008). Structural basis of human triosephosphate isomerase deficiency. *J. Biol. Chem.* **283**, 23254-23263.
10. Hammarström P, Wiseman RL, Powers ET, & Kelly JW (2003). Prevention of transthyretin amyloid disease by changing protein misfolding energetics. *Science* **299**, 713-716.
11. Lynch SM, Boswell SA, & Colón W (2004). Kinetic stability of Cu/Zn superoxide dismutase is dependent on its metal ligands: Implications for ALS. *Biochemistry* **43**, 16525-16531.
12. Rumfeldt JAO, Lepock JR, & Meiering EM (2009). Unfolding and folding kinetics of amyotrophic lateral sclerosis-associated mutant Cu,Zn superoxide dismutases. *J. Mol. Biol.* **385**, 278-298.
13. Erlandsen H, Pey AL, Gámez A, Pérez B, Desviat LR, Aguado C, Koch R, Surendran S, Tyring S, Matalon R, Scriver CR, Ugarte M, Martínez A, & Stevens RC (2004). Correction of kinetic and stability defects by tetrahydrobiopterin in phenylketonuria patients with certain phenylalanine hydroxylase mutations. *Proc. Natl. Acad. Sci. USA* **101**, 16903-16908.
14. Sterner R & Liebl W (2001). Thermophilic adaptation of proteins. *Crit. Rev. Biochem. Mol. Biol.* **36**, 39-106.
15. Jaenicke R & Böhm G (1998). The stability of proteins in extreme environments. *Curr. Opin. Struct. Biol.* **8**, 738-748.
16. Dams T, Auerbach G, Bader G, Jacob U, Ploom T, Huber R, & Jaenicke R (2000). The crystal structure of dihydrofolate reductase from Thermotoga maritima: Molecular features of thermostability. *J. Mol. Biol.* **297**, 659-672.

17. Zhang X, Meining W, Fischer M, Bacher A, & Ladenstein R (2001). X-ray structure analysis and crystallographic refinement of lumazine synthase from the hyperthermophile *Aquifex aeolicus* at 1.6 Å resolution: Determinants of thermostability revealed from structural comparisons. *J. Mol. Biol.* **306**, 1099-1114.
18. Sterner R & Höcker B (2005). Catalytic versatility, stability, and evolution of the $(\beta\alpha)_8$-barrel enzyme fold. *Chem. Rev.* **105**, 4038-4055.
19. Kimura M (1968). Evolutionary rate at the molecular level. *Nature* **217**, 624-626.
20. Dams T & Jaenicke R (1999). Stability and folding of dihydrofolate reductase from the hyperthermophilic bacterium *Thermotoga maritima*. *Biochemistry* **38**, 9169-9178.
21. Ratcliff K, Corn J, & Marqusee S (2009). Structure, stability, and folding of ribonuclease H1 from the moderately thermophilic *Chlorobium tepidum*: Comparison with thermophilic and mesophilic homologues. *Biochemistry* **48**, 5890-5898.
22. Mukaiyama A, Takano K, Haruki M, Morikawa M, & Kanaya S (2004). Kinetically robust monomeric protein from a hyperthermophile. *Biochemistry* **43**, 13859 - 13866.
23. Cavagnero S, Debe D, Zhou Z, Adams M, & Chan S (1998). Kinetic role of electrostatic interactions in the unfolding of hyperthermophilic and mesophilic rubredoxins. *Biochemsitry* **37**, 3369-3376.
24. Perl D, Welker C, Schindler T, Schroder K, Marahiel M, Janicke R, & Schmid FX (1998). Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. *Nat. Struct. Biol.* **5**, 229 - 235.
25. Deutschman W & Dahlquist F (2001). Thermodynamic basis for the increased thermostability of CheY from the hyperthermophile *Thermotoga maritima*. *Biochemistry* **40**, 13107-13113.
26. Anfinsen CB (1973). Principles that govern the folding of protein chains. *Science* **181**, 223-230.
27. Levinthal CJ (1968). Are there pathways for protein folding? *J. Chim. Phys.* **65**, 44-45.
28. Onuchic JN & Wolynes PG (2004). Theory of protein folding. *Curr. Opin. Struct. Biol.* **14**, 70-75.
29. Shea J-E & Brooks III CL (2001). From folding theories to folding proteins: A review and assessment of simulation studies of protein folding and unfolding. *Annu. Rev. Phys. Chem.* **52**, 499-535.
30. Ikai A & Tanford C (1971). Kinetic evidence for incorrectly folded intermediate states in the refolding of denatured proteins. *Nature* **230**, 100-102.
31. Tsong TY, Baldwin RL, & Elson EL (1971). The sequential unfolding of ribonuclease A: detection of a fast initial phase in the kinetics of unfolding. *Proc. Natl. Acad. Sci. USA* **68**, 2712-2715.
32. Balbach J & Schmid FX (2000). Prolyl isomerization and its catalysis in protein folding. In *Mechanisms of Protein Folding.* (Pain RH, ed.), pp. 212-237, Oxford University Press, Oxford.
33. Creighton TE (2000). Protein folding coupled to disulfide-bond formation. In *Mechanisms of Protein Folding*. (Pain RH, ed.), pp. 250-278, Oxford University Press, Oxford.

34. Meinhold DW & Wright PE (2011). Measurement of protein unfolding/refolding kinetics and structural characterization of hidden intermediates by NMR relaxation dispersion. *Proc. Natl. Acad. Sci. USA* **108**, 9078-9083.
35. Nabuurs SM, Westphal AH, & van Mierlo CPM (2008). Extensive formation of off-pathway species during folding of an αβ parallel protein is due to docking of (non)native structure elements in unfolded molecules. *J. Am. Chem. Soc.* **130**, 16914-16920.
36. Roder H, Elöve GA, & Englander SW (1988). Structural characterization of folding intermediates in cytochrome c by hydrogen exchange labelling and proton NMR. *Nature* **335**, 700-704.
37. Roder H (1989). Structural characterization of protein folding intermediates by proton magnetic resonance and hydrogen exchange. *Methods Enzymol.* **176**, 446-473.
38. Englander SW (2000). Protein folding intermediates and pathways studied by hydrogen exchange. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 213-238.
39. Nishimura C, Riley R, Eastman P, & Fink AL (2000). Fluorescence energy transfer indicates similar transient and equilibrium intermediates in staphylococcal nuclease folding. *J. Mol. Biol.* **299**, 1133-1146.
40. Marszalek PE, Lu H, Li H, Carrion-Vazquez M, Oberhauser AF, Schulten K, & Fernandez JM (1999). Mechanical unfolding intermediates in titin modules. *Nature* **402**, 100-103.
41. Wensley BG, Gärtner M, Choo WX, Batey S, & Clarke J (2009). Different members of a simple three-helix bundle protein family have very different folding rate constants and fold by different mechanisms. *J. Mol. Biol.* **390**, 1074-1085.
42. Friel CT, Capaldi AP, & Radford SE (2003). Structural analysis of the rate-limiting transition states in the folding of lm7 and lm9: Similarities and differences in the folding of homologous proteins. *J. Mol. Biol.* **326**, 293-305.
43. Olofsson M, Hansson S, Hedberg L, Logan DT, & Oliveberg M (2007). Folding of S6 structures with divergent amino acid composition: Pathway flexibility within partly overlapping foldons. *J. Mol. Biol.* **365**, 237-248.
44. Lappalainen I, Hurley MG, & Clarke J (2008). Plasticity within the obligatory folding nucleus of an immunoglobulin-like domain. *J. Mol. Biol.* **375**, 547-559.
45. Lam AR, Borreguero JM, Ding F, Dokholyan NV, Buldyrev SV, Stanley HE, & Shakhnovich E (2007). Parallel folding pathways in the SH3 domain protein. *J. Mol. Biol.* **373**, 1348-1360.
46. Nickson AA & Clarke J (2010). What lessons can be learned from studying the folding of homologous proteins? *Methods* **52**, 38-50.
47. Tokuriki N & Tawfik DS (2009). Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* **19**, 596-604.
48. Sadqi M, de Alba E, Pérez-Jiménez R, Sanchez-Ruiz JM, & Muñoz V (2009). A designed protein as experimental model of primordial folding. *Proc. Natl. Acad. Sci. USA* **106**, 4127-4132.
49. Gillespie B, Vu DM, Shah PS, Marshall SA, Dyer RB, Mayo SL, & Plaxco KW (2003). NMR and temperature-jump measurements of de novo designed proteins

demonstrate rapid folding in the absence of explicit selection for kinetics. *J. Mol. Biol.* **330**, 813-819.

50. Scalley-Kim M & Baker D (2004). Characterization of the folding energy landscapes of computer generated proteins suggests high folding free energy barriers and cooperativity may be consequences of natural selection. *J. Mol. Biol.* **338**, 573-583.
51. Watters AL, Deka P, Corrent C, Callender D, Varani G, Sosnick T, & Baker D (2007). The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection. *Cell* **128**, 613-624.
52. Murzin AG, Brenner SE, Hubbard T, & Chothia C (1995). SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536-540.
53. Caetano-Anollés G, Kim HS, & Mittenthal JE (2007). The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture. *Proc. Natl. Acad. Sci. USA* **104**, 9358-9363.
54. Wierenga RK (2001). The TIM-barrel fold: A versatile framework for efficient enzymes. *FEBS Lett.* **492**, 193-198.
55. Nagano N, Orengo CA, & Thornton JM (2002). One fold with many functions: The evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J. Mol. Biol.* **321**, 741-765.
56. Pujadas G & Palau J (1999). TIM barrel fold: Structural, functional and evolutionary characteristics in natural and designed molecules. *Biologica (Bratislava)* **54**, 231-253.
57. Copley RR & Bork P (2000). Homology among $(\beta\alpha)_8$-barrels: Implications for the evolution of metabolic pathways. *J. Mol. Biol.* **303**, 627-641.
58. Forsyth WR, Bilsel O, Gu Z, & Matthews CR (2007). Topology and sequence in the folding of a TIM barrel protein: Global analysis highlights partitioning between transient off-pathway and stable on-pathway folding intermediates in the complex folding mechanism of a $(\beta\alpha)_8$-barrel of unknown function from *B. subtilis*. *J. Mol. Biol.* **372**, 236-253.
59. Raine AR, Scrutton NS, & Mathews FS (1994). On the evolution of alternate core packing in eightfold (βα)-barrels. *Protein Sci.* **3**, 1889-1892.
60. Reardon D & Farber G (1995). The structure and evolution of αβ-barrel proteins. *FASEB J.* **9**, 497-503.
61. Wilmanns M, Hyde CC, Davies DR, Kirschner K, & Jansonius JN (1991). Structural conservation in parallel (βα)-barrel enzymes that catalyze three sequential reactions in the pathway of tryptophan biosynthesis. *Biochemistry* **30**, 9161-9169.
62. Henn-Sax M, Höcker B, Wilmanns M, & Sterner R (2001). Divergent evolution of $(\beta\alpha)_8$-barrel enzymes. *Biol. Chem.* **382**, 1315-1320.
63. Henn-Sax M, Thoma R, Schmidt S, Hennig M, Kirschner K, & Sterner R (2002). Two $(\beta\alpha)_8$-barrel enzymes of histidine and tryptophan biosynthesis have similar reaction mechanisms and common strategies for protecting their labile substrates. *Biochemistry* **41**, 12032-12042.

64. Beismann-Driemeyer S & Sterner R (2001). Imidazole glycerol phosphate synthase from *Thermotoga maritima.* Quaternary structure, steady-state kinetics, and reaction mechanism of the bienzyme complex. *J. Biol. Chem.* **276**, 20387-20396.
65. List F, Sterner R, & Wilmanns M (2011). Related $(\beta\alpha)_8$-barrel proteins in histidine and tryptophan biosynthesis: A paradigm to study enzyme evolution. *ChemBioChem* **12**, 1487-1494.
66. Jürgens C, Strom A, Wegener D, Hettwer S, Wilmanns M, & Sterner R (2000). Directed evolution of a $(\beta\alpha)_8$-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc. Natl. Acad. Sci. USA* **97**, 9925-9930.
67. Leopoldseder S, Claren J, Jürgens C, & Sterner R (2004). Interconverting the catalytic activities of $(\beta\alpha)_8$-barrel enzymes from different metabolic pathways: Sequence requirements and molecular analysis. *J. Mol. Biol.* **337**, 871-879.
68. Evran S, Telefoncu A, & Sterner R (2012). Directed evolution of $(\beta\alpha)_8$-barrel enzymes: Establishing phosphoribosylanthranilate isomerisation activity on the scaffold of the tryptophan synthase α-subunit. *Protein Eng. Des. Sel*. PMID: 22490958.
69. Lang D, Thoma R, Henn-Sax M, Sterner R, & Wilmanns M (2000). Structural evidence for evolution of the β/α barrel scaffold by gene duplication and fusion. *Science* **289**, 1546-1550.
70. Höcker B, Beismann-Driemeyer S, Hettwer S, Lustig A, & Sterner R (2001). Dissection of a $(\beta\alpha)_8$-barrel enzyme into two folded halves. *Nat. Struct. Biol.* **8**, 32-36.
71. Höcker B, Claren J, & Sterner R (2004). Mimicking enzyme evolution by generating new $(\beta\alpha)_8$-barrels from $(\beta\alpha)_4$-half-barrels. *Proc. Natl. Acad. Sci. USA* **101**, 16448-16453.
72. Seitz T, Bocola M, Claren J, & Sterner R (2007). Stabilisation of a $(\beta\alpha)_8$-barrel protein designed from identical half barrels. *J. Mol. Biol.* **372**, 114-129.
73. Höcker B, Lochner A, Seitz T, Claren J, & Sterner R (2009). High-resolution crystal structure of an artificial $(\beta\alpha)_8$-barrel protein designed from identical half-barrels. *Biochemistry* **48**, 1145-1147.
74. Sperl JM (2010) Analyse der Faltung und Erhöhung der Stabilität künstlich erzeugter $(\beta\alpha)_8$-Barrel Proteine. *Diploma thesis,* University Regensburg.
75. Eder J & Kirschner K (1992). Stable substructures of eightfold $(\beta\alpha)_8$-barrel proteins: Fragment complementation of phosphoribosylanthranilate isomerase. *Biochemistry* **31**, 3617-3625.
76. Soberón X, Fuentes-Gallego P, & Saab-Rincón G (2004). In vivo fragment complementation of a $(\beta\alpha)_8$-barrel protein: Generation of variability by recombination. *FEBS Lett.* **560**, 167-172.
77. Bertolaet BL & Knowles JR (1995). Complementation of fragments of triosephosphate isomerase defined by exon boundaries. *Biochemistry* **34**, 5736-5743.
78. Gualfetti PJ, Bilsel O, & Matthews CR (1999). The progressive development of structure and stability during the equilibrium folding of the α-subunit of tryptophan synthase from *Escherichia coli*. *Protein Sci.* **8**, 1623-1635.

79. Bilsel O, Zitzewitz JA, Bowers KE, & Matthews CR (1999). Folding mechanism of the α-subunit of tryptophan synthase, an (βα)–barrel protein: Global analysis highlights the interconversion of multiple native, intermediate, and unfolded forms through parallel channels. *Biochemistry* **38**, 1018-1029.
80. Forsyth WR & Matthews CR (2002). 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.* **320**, 1119-1133.
81. Gu Z, Zitzewitz JA, & Matthews CR (2007). Mapping the structure of folding cores in TIM barrel proteins by hydrogen exchange mass spectrometry: the roles of motif and sequence for the indole-3-glycerol phosphate synthase from *Sulfolobus solfataricus*. *J. Mol. Biol.* **368**, 582-594.
82. Gu Z, Rao MK, Forsyth WR, Finke JM, & Matthews CR (2007). Structural analysis of kinetic folding intermediates for a TIM barrel protein, indole-3-glycerol phosphate synthase, by hydrogen exchange mass spectrometry and Gō model simulation. *J. Mol. Biol.* **374**, 528-546.
83. Mukaiyama A, Takano K, Haruki M, Morikawa M, & Kanaya S (2004). Kinetically robust monomeric protein from a hyperthermophile. *Biochemistry* **43**, 13859-13866.
84. Douangamath A, Walker M, Beismann-Driemeyer S, Vega-Fernandez MC, Sterner R, & Wilmanns M (2002). Structural evidence for ammonia tunneling across the $(\beta\alpha)_8$-barrel of the imidazole glycerol phosphate synthase bienzyme complex. *Structure* **10**, 185-193.
85. Thoma R, Obmolova G, Lang DA, Schwander M, Jenö P, Sterner R, & Wilmanns M (1999). Efficient expression, purification and crystallisation of two hyperthermostable enzymes of histidine biosynthesis. *FEBS Lett.* **454**, 1-6.
86. Seitz T, Bocola M, Claren J, & Sterner R (2007). Stabilisation of a $(\beta\alpha)_8$-barrel protein designed from identical half barrels. *J. Mol. Biol.* **372**, 114-129.
87. Puorger C, Eidam O, Capitani G, Erilov D, Grütter MG, & Glockshuber R (2008). Infinite kinetic stability against dissociation of supramolecular protein complexes through donor strand complementation. *Structure* **16**, 631-642.
88. Bachmann A & Kiefhaber T (2001). Apparent two-state tendamistat folding is a sequential process along a defined route. *J. Mol. Biol.* **306**, 375-386.
89. Stryer L (1965). The interaction of a naphthalene dye with apomyoglobin and apohemoglobin. A fluorescent probe of non-polar binding sites. *J. Mol. Biol.* **13**, 482-495.
90. Jones BE, Jennings PA, Pierre RA, & Matthews CR (1994). Development of nonpolar surfaces in the folding of *Escherichia coli* dihydrofolate reductase detected by 1-anilinonaphthalene-8-sulfonate binding. *Biochemistry* **33**, 15250-15258.
91. Ptitsyn OB (1995). Molten globule and protein folding. *Adv. Protein Chem.* **47**, 83-229.
92. Bilsel O & Matthews CR (2006). Molecular dimensions and their distributions in early folding intermediates. *Curr. Opin. Struct. Biol.* **16**, 86-93.
93. Wu P & Brand L (1994). Resonance energy transfer: Methods and applications. *Anal. Biochem.* **218**, 1-13.

94. Ratner V, Sinev M, & Haas E (2000). Determination of intramolecular distance distribution during protein folding on the millisecond timescale. *J. Mol. Biol.* **299**, 1363-1371.
95. Navon A, Ittah V, Landsman P, Scheraga HA, & Haas E (2001). Distributions of intramolecular distances in the reduced and denatured states of bovine pancreatic ribonuclease A. Folding initiation structures in the C-terminal portions of the reduced protein. *Biochemistry* **40**, 105-118.
96. Lillo MP, Szpikowska BK, Mas MT, Sutin JD, & Beechem JM (1997). Real-time measurement of multiple intramolecular distances during protein folding reactions: A multisite stopped-flow fluorescence energy-transfer study of yeast phosphoglycerate kinase. *Biochemistry* **36**, 11273-11281.
97. Lakowicz JR (1999). *Principles of fluorescence spectroscopy*. Kluwer Academic/Plenum Publishers, New York.
98. Haas E (2005). The study of protein folding and dynamics by determination of intramolecular distance distributions and their fluctuations using ensemble and single-molecule FRET measurements. *Chemphyschem* **6**, 858-870.
99. Kuzmenkina EV, Heyes CD, & Nienhaus GU (2005). Single-molecule Förster resonance energy transfer study of protein dynamics under denaturing conditions. *Proc. Natl. Acad. Sci. USA* **102**, 15471-15476.
100. Magg C & Schmid FX (2004). Rapid collapse precedes the fast two-state folding of the cold shock protein. *J. Mol. Biol.* **335**, 1309-1323.
101. Magg C, Kubelka J, Holtermann G, Haas E, & Schmid FX (2006). Specificity of the initial collapse in the folding of the cold shock protein. *J. Mol. Biol.* **360**, 1067-1080.
102. Huang F, Settanni G, & Fersht AR (2008). Fluorescence resonance energy transfer analysis of the folding pathway of Engrailed Homeodomain. *Protein Eng. Des. Sel.* **21**, 131-146.
103. Carstensen L (2008) Energietransfermessungen zur Analyse der Stabilität und Funktion des Faltungsenzyms SlyD aus *Escherichia coli*. *Diploma thesis,* University Bayreuth.
104. Lipchock JM & Loria PJ (2008). ^{1}H, ^{15}N and ^{13}C resonance assignment of imidazole glycerol phosphate (IGP) synthase protein HisF from *Thermotoga maritima*. *Biomol. NMR Assign.* **2**, 219-221.
105. Liebold C (2011) NMR-Spektroskopische Untersuchungen der Imidazolglycerin-phosphat-Synthase HisF aus *Thermotoga maritima*: Studien zur Bindung von Effektormolekülen und Xenon sowie zum pKa-Wert katalytisch aktiver Aminosäurereste. *Dissertation thesis,* University Regensburg.
106. Liebold C, List F, Kalbitzer HR, Sterner R, & Brunner E (2010). The interaction of ammonia and xenon with the imidazole glycerol phosphate synthase from *Thermotoga maritima* as detected by NMR spectroscopy. *Protein Sci.* **19**, 1774-1782.
107. Wu Y, Vadrevu R, Kathuria S, Yang X, & Matthews CR (2007). A tightly packed hydrophobic cluster directs the formation of an off-pathway sub-millisecond folding intermediate in the α-subunit of tryptophan synthase, a TIM barrel protein. *J. Mol. Biol.* **366**, 1624-1638.

108. Rudolph R, Siebendritt R, & Kiefhaber T (1992). Reversible unfolding and refolding behavior of a monomeric aldolase from *Staphylococcus aureus*. *Protein Sci.* **1**, 654-666.
109. Jasanoff A, Davis B, & Fersht AR (1994). Detection of an intermediate in the folding of the $(\beta\alpha)_8$-barrel N-(5'-phosphoribosyl)anthranilate isomerase from *Escherichia coli*. *Biochemistry* **33**, 6350-6355.
110. Kathuria SV, Day IJ, Wallace LA, & Matthews CR (2008). Kinetic traps in the folding of βα-repeat proteins: CheY initially misfolds before accessing the native conformation. *J. Mol. Biol.* **382**, 467-484.
111. Hills RD, Jr., Kathuria SV, Wallace LA, Day IJ, Brooks CL, 3rd, & Matthews CR (2010). Topological frustration in βα-repeat proteins: Sequence diversity modulates the conserved folding mechanisms of α/β/α sandwich proteins. *J. Mol. Biol.* **398**, 332-350.
112. Höcker B, Schmidt S, & Sterner R (2002). A common evolutionary origin of two elementary enzyme folds. *FEBS Lett.* **510**, 133-135.
113. Bharat TAM, Eisenbeis S, Zeth K, & Höcker B (2008). A βα-barrel built by the combination of fragments from different folds. *Proc. Natl. Acad. Sci. USA* **105**, 9942-9947.
114. Saab-Rincon G, Froebe CL, & Matthews CR (1993). Urea-induced unfolding of the α-subunit of tryptophan synthase: One-dimensional proton NMR evidence for residual structure near histidine-92 at high denaturant concentration. *Biochemistry* **32**, 13981-13990.
115. Rojsajjakul T, Wintrode P, Vadrevu R, Robert Matthews C, & Smith DL (2004). 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.* **341**, 241-253.
116. Stigler J, Ziegler F, Gieseke A, Gebhardt JCM, & Rief M (2011). The complex folding network of single calmodulin molecules. *Science* **334**, 512-516.
117. Lorenz T & Reinstein J (2008). The influence of proline isomerization and off-pathway intermediates on the folding mechanism of eukaryotic UMP/CMP kinase. *J. Mol. Biol.* **381**, 443-455.
118. Crespo MD, Simpson ER, & Searle MS (2006). Population of on-pathway intermediates in the folding of ubiquitin. *J. Mol. Biol.* **360**, 1053-1066.
119. Chen Y, Ding F, Nie H, Serohijos AW, Sharma S, Wilcox KC, Yin S, & Dokholyan NV (2008). Protein folding: Then and now. *Arch. Biochem. Biophys.* **469**, 4-19.
120. Bai Y, Sosnick TR, Mayne L, & Englander SW (1995). Protein folding intermediates: Native-state hydrogen exchange. *Science* **269**, 192-197.
121. Feng H, Vu N-D, & Bai Y (2005). Detection of a hidden folding intermediate of the third domain of PDZ. *J. Mol. Biol.* **346**, 345-353.
122. Fersht AR (2000). A kinetically significant intermediate in the folding of barnase. *Proc. Natl. Acad. Sci. USA* **97**, 14121-14126.
123. Kato H, Vu ND, Feng H, Zhou Z, & Bai Y (2007). The folding pathway of T4 lysozyme: An on-pathway hidden folding intermediate. *J. Mol. Biol.* **365**, 881-891.

5 List of publications

A. **Carstensen, L.**, Zoldák, G., Schmid, F.X. and Sterner, R. (2012). Folding mechanism of an extremely thermostable $(\beta\alpha)_8$-barrel enzyme: A high kinetic barrier protects the protein from denaturation. *Biochemistry* 51, 3420−3432.

B. **Carstensen, L.**, Sperl, J., List, F., Bocola, M., Schmid, F.X. and Sterner, R.. Conservation of the folding mechanism between a natural $(\beta\alpha)_8$-barrel protein and putative evolutionary precursors generated by protein design.
Submitted for publication

In the course of this work, I contributed to a further publication, which is not part of the dissertation:

C. Richter, M., Bosnali, M., **Carstensen, L**., Seitz, T., Durchschlag, H., Blanquart, S., Merkl, R., Sterner, R. (2010). Computational and experimental evidence for the evolution of a $(\beta\alpha)_8$-barrel protein from an ancestral quarter-barrel stabilized by disulfide bonds. *J. Mol. Biol.* 398, 763-773.

6 Personal contribution

A. The experimental work was performed by myself. The work was supervised by Reinhard Sterner and Franz X. Schmid, and the publication was written by myself, Reinhard Sterner, Franz X. Schmid, and Gabriel Zoldák.

B. Josef Sperl designed and produced Sym2. Felix List and Marco Bocola crystallized Sym2 and solved its X-ray structure. All other experiments were performed by myself. The publication was written by myself, Reinhard Sterner, and Franz X. Schmid.

7 Publications

Publication A

Folding mechanism of an extremely thermostable $(\beta\alpha)_8$-barrel enzyme: A high kinetic barrier protects the protein from denaturation

Linn Carstensen, Gabriel Zoldák, Franz-Xaver Schmid and Reinhard Sterner (2012).

Biochemistry, 51, 3420−3432.

DOI: 10.1021/bi300189f

BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Folding Mechanism of an Extremely Thermostable $(\beta\alpha)_8$-Barrel Enzyme: A High Kinetic Barrier Protects the Protein from Denaturation

Linn Carstensen,[†] Gabriel Zoldák,[‡] Franz-Xaver Schmid,[§] and Reinhard Sterner*[,†]

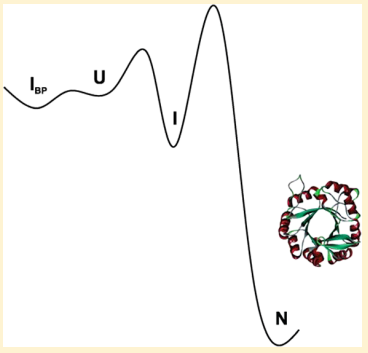
[†]Universität Regensburg, Institut für Biophysik und physikalische Biochemie, Universitätsstrasse 31, D-93053 Regensburg, Germany

[‡]Physik Department E22, Technische Universität München, D-85748 Garching, Germany

[§]Universität Bayreuth, Laboratorium für Biochemie, D-95440 Bayreuth, Germany

Supporting Information

ABSTRACT: HisF, the cyclase subunit of imidazole glycerol phosphate synthase (ImGPS) from *Thermotoga maritima*, is an extremely thermostable $(\beta\alpha)_8$-barrel protein. We elucidated the unfolding and refolding mechanism of HisF. Its unfolding transition is reversible and adequately described by the two-state model, but 6 weeks is necessary to reach equilibrium (at 25 °C). During refolding, initially a burst-phase off-pathway intermediate is formed. The subsequent productive folding occurs in two kinetic phases with time constants of ~3 and ~20 s. They reflect a sequential process via an on-pathway intermediate, as revealed by stopped-flow double-mixing experiments. The final step leads to native HisF, which associates with the glutaminase subunit HisH to form the functional ImGPS complex. The conversion of the on-pathway intermediate to the native protein results in a 10^6-fold increase of the time constant for unfolding from 89 ms to 35 h (at 4.0 M GdmCl) and thus establishes a high energy barrier to denaturation. We conclude that the extra stability of HisF is used for kinetic protection against unfolding. In its refolding mechanism, HisF resembles other $(\beta\alpha)_8$-barrel proteins.



Approximately 1200 different protein folds have been identified to date [Structural Classification of Proteins (SCOP) release 1.75, February 2009],[1] each of them being characterized by a distinct toplogical orientation of secondary structure elements. Whereas many folds are represented by only a few members of the protein database, others have been recruited extensively in the course of evolution. A prominent example is the $(\beta\alpha)_8$-barrel, which is among the most ancient, frequent, and versatile folds.[2−4] Approximately 10% of all proteins with known three-dimensional structure contain at least one $(\beta\alpha)_8$-barrel domain. With very few exceptions, all known $(\beta\alpha)_8$-barrels are enzymes, and SCOP distinguishes 33 superfamilies that catalyze more than 60 different reactions. They occur in five of the six Enzyme Commission (EC) classes, acting as oxidoreductases, transferases, lyases, hydrolases, and isomerases, and many of them are engaged in essential metabolic pathways.[5,6]

The canonical $(\beta\alpha)_8$-barrel fold is composed of eight units, each of which contains at least 25 residues. A single unit consists of a β-strand and an α-helix, which are linked by a $\beta\alpha$-loop, and the individual units are connected by $\alpha\beta$-loops. The eight strands form a central parallel β-sheet, which is surrounded by an outer layer of eight α-helices (Figure 1). Many $(\beta\alpha)_8$-barrel enzymes contain extensions to the canonical topology, at the N- or C-terminus or in loop segements. Remarkably, in all known $(\beta\alpha)_8$-barrels, the residues that are important for substrate specificity and catalytic activity are located at the C-terminal ends of the β-strands and in the subsequent $\beta\alpha$-loops. However, in spite of the common chain topology and the conserved location of the active site, the geometry of the central barrel can differ considerably between individual $(\beta\alpha)_8$-barrel proteins and the overall level of sequence identity is low. The question of whether all $(\beta\alpha)_8$-barrels are the products of divergent evolution from a common ancestor or whether the fold has developed independently several times by convergent evolution is therefore unresolved.[6−8] In any case, extensive sequence and structure comparisons suggest a common evolutionary origin for most $(\beta\alpha)_8$-barrels,[5,9] in line with the notion that conservation of structure exceeds conservation of amino acid sequence.

This observation raises the question of the extent to which the folding mechanism of $(\beta\alpha)_8$-barrel proteins has been conserved upon the vast diversification of their sequences in the course of evolution. Early insights into the folding of $(\beta\alpha)_8$-barrel proteins were obtained by fragmentation experiments performed with phosphoribosyl anthranilate isomerase (PRAI) and triosephosphate isomerase (TIM).[10−12] These studies provided evidence that both PRAI and TIM form stable folding intermediates consisting of four or six structured $(\beta\alpha)$ units, followed by the association of the unstructured segment to

Received: February 10, 2012
Revised: March 27, 2012
Published: March 28, 2012

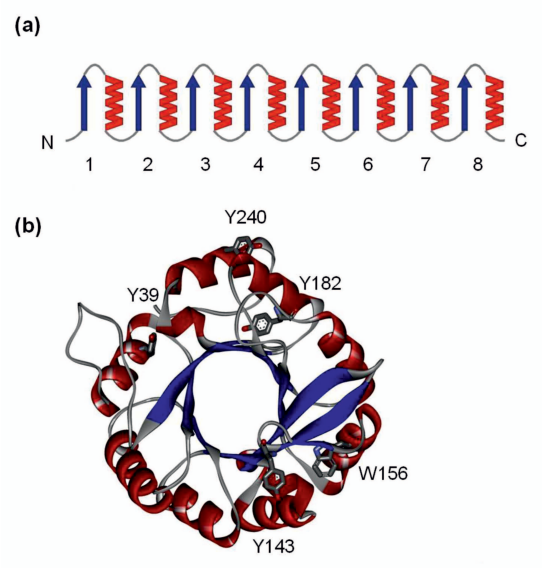


Figure 1. (a) Schematic depiction of the $(\beta\alpha)_8$-barrel fold and (b) ribbon diagram of the crystal structure of HisF (PDB entry 1THF).[23] α-Helices and β-strands are colored red and blue, respectively; loops are colored gray. Tyr and Trp residues are displayed as sticks. The figure was generated using Accelrys DS.

finally form the native $(\beta\alpha)_8$-barrel. Recent in-depth kinetic studies with the α-subunit of tryptophan synthase from *Escherichia coli* (αTS),[13,14] the indole-3-glycerol-phosphate synthase from *Sulfolobus solfataricus* (IGPS),[15−17] and a $(\beta\alpha)_8$-barrel protein of unknown function encoded by the *Bacillus subtilis iol*I gene (IOLI)[18] have provided more insights into the folding of $(\beta\alpha)_8$-barrels. All three proteins apparently form an off-pathway burst-phase intermediate, which must unfold to enter productive folding via one or more on-pathway intermediates. Although these similarities have suggested that the common topology is the main determinant for the folding mechanism of $(\beta\alpha)_8$-barrel proteins, the native secondary structural elements formed in the on-pathway intermediates seem to differ between αTS and IGPS. Moreover, whereas the slow cis−trans isomerization of peptide bonds involving proline residues results in several parallel folding pathways in the case of αTS, IGPS and IOLI contain only trans prolyl peptide bonds and show simpler folding mechanisms.[17]

We have studied the folding mechanism of the cyclase subunit HisF of the imidazole glycerol phosphate synthase from *Thermotoga maritima* (Figure 1b). HisF is an interesting protein for several reasons. First, it is active in its hyperthermophilic host above 80 °C and provides an excellent model for testing the hypothesis that high thermodynamic stability is related with a large kinetic barrier to unfolding.[19−22] Second, unlike other $(\beta\alpha)_8$-barrel proteins, HisF shows a striking 2-fold symmetry and contains phosphate binding sites in both the N-terminal and C-terminal half-barrel, which anchor its biphosphorylated substrate.[23] IGPS and αTS are evolutionarily related to HisF but lack such a clear 2-fold symmetry and contain only a single conserved C-terminal phosphate binding site,[24] which makes a comparison of their folding mechanisms particularly attractive. Finally, as HisF forms a 1:1 complex with the glutaminase subunit HisH of the imidazole glycerol phosphate synthase,[25,26] it offers an opportunity to study the relationship between protein folding and association.

MATERIALS AND METHODS

Mutagenesis, Cloning and Expression of Genes, and Purification of Recombinant Proteins. Point mutations were introduced into the *his*F gene by overlap extension polymerase chain reaction,[27] followed by cloning of the genes into plasmid pET24a(+) via the NdeI and BamHI restriction sites. Wild-type *his*F and its mutants 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. The recombinant proteins with molecular masses of ~27.7 kDa were purified as described previously.[28] According to sodium dodecyl sulfate−polyacrylamide gel electrophoresis, all HisF variants were >95% pure. Approximately 100 mg of protein per liter of culture was obtained and dialyzed against 50 mM Tris-HCl buffer (pH 7.5).

Equilibrium Unfolding Transitions. Samples with 4 μM protein were prepared in 50 mM Tris-HCl buffer (pH 7.5) containing various concentrations of GdmCl. GdmCl (ultrapure) was purchased from MP Biomedicals (Illkirch, France), and its concentration was determined by the refractive index of the solution.[29] The samples were incubated at 25 and 45 °C for 6 weeks and 19 days, respectively. The far-UV circular dichroism (CD) signal at 225 nm was monitored with a JASCO model J815 CD spectrophotometer, using a 5 mm cuvette and a bandwidth of 1 nm. Following excitation at 280 nm (bandwidth of 3 nm), the fluorescence emission signal at 320 nm (bandwidth of 5 nm) was monitored in a JASCO model FP-6500 spectrofluorimeter. The transitions were analyzed according to the two-state equilibrium model. The free energy change of unfolding in the absence of denaturant ($\Delta G_D°$) was obtained assuming a linear dependency of ΔG_D on GdmCl concentration.[30]

Kinetics of Unfolding and Refolding. All experiments were performed in 50 mM Tris-HCl buffer (pH 7.5) at 25 or 45 °C. Unfolding was initiated by dilution of the native protein to various denaturing concentrations of GdmCl. Refolding was measured after dilution of the denatured (in 6.0 M GdmCl) protein to folding conditions.

Manual mixing experiments monitored by the far-UV CD signal were performed with a JASCO model J815 CD spectrophotometer. Following a 10-fold dilution, the CD signal of 4 μM of protein (final concentration) was recorded at 225 nm, using a 5 mm cuvette, a bandwidth of 1 nm, and an averaging time of 1 s. Manual mixing experiments followed by Trp/Tyr fluorescence were performed with a JASCO model FP-6500 spectrofluorometer using a 1 cm cuvette. The fluorescence signal of 4 μM protein (final concentration) was followed using an excitation wavelength of 280 nm (bandwidth of 3 nm) and an emission wavelength of 320 nm (bandwidth of 5 nm). Monoexponential or biexponential functions were fit to the data using GraFit6 from Erithacus (West Sussex, U.K.).

Stopped-flow fluorescence experiments were performed using a SX.20MV spectrometer from Applied Photophysics (Leatherhead, U.K.). After rapid mixing, the fluorescence signal of 2 μM protein (final concentration) was followed using a 320 nm emission cutoff filter after excitation at 280 nm (Tyr and Trp) or 290 nm (Trp). The path length of the observation cell was 2 mm, and the bandwidth was 10 nm. The kinetics was measured eight times under identical conditions, averaged, and analyzed using exponential functions.

The time-dependent formation of I from U was followed by stopped-flow fluorescence interrupted refolding experiments.

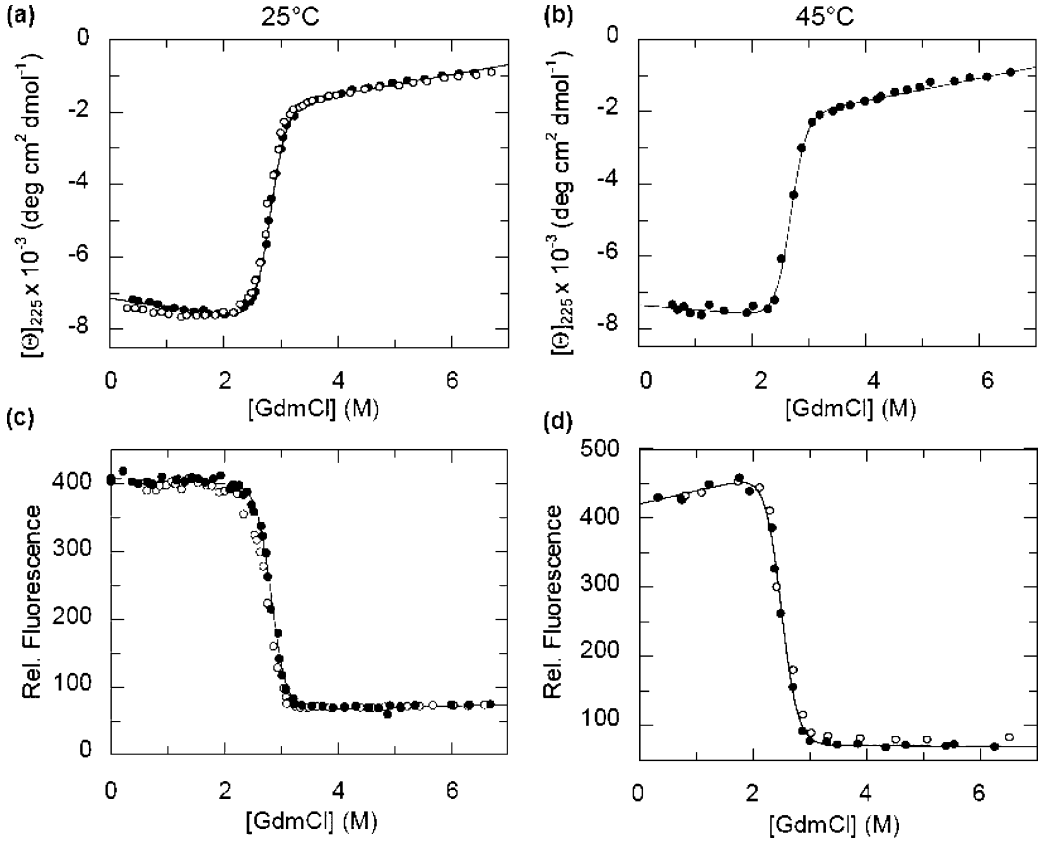


Figure 2. GdmCl-induced equilibrium unfolding of HisF at 25 °C (a and c) and 45 °C (b and d). The transitions were followed by the CD signal at 225 nm (a and b) and Trp/Tyr fluorescence (excitation at 280 nm; emission at 320 nm) (c and d) in 50 mM Tris-HCl buffer (pH 7.5). Filled symbols represent data from unfolding experiments, started with folded protein, and empty symbols represent data from refolding experiments, started with protein that was previously unfolded in 6.0 M GdmCl. The solid lines represent fits to the unfolding transitions on the basis of the two-state model; the parameters derived from the fit are listed in Table 1.

Unfolded HisF (132 μM in 6.0 M GdmCl) was diluted 6-fold to 1.0 M GdmCl and incubated for 10 s to allow for the conversion of U to I. The sample was then diluted 11-fold to final GdmCl concentrations between 1.9 and 7.0 M, which resulted in the rapid conversion of I to U. A monoexponential equation with a linear factor was fit to the unfolding traces to determine the rate constant for the I → U reaction at the respective denaturant concentration.

The kinetics of the sequential U → I → N reaction was monitored in further stopped-flow fluorescence interrupted refolding experiments. Unfolded HisF (132 μM in 6.0 M GdmCl) was diluted 6-fold to 1.0 M GdmCl and incubated for various times to allow for refolding to I and N. The sample was then diluted 11-fold to 2.3 or 6.9 M GdmCl, which resulted in the rapid unfolding of I and the slow unfolding of N. Monoexponential or double-exponential equations with a linear factor were fit to the kinetic traces. An equation describing the consecutive model was fit to the obtained amplitudes.[31,32]

The recovery of native HisF molecules during refolding was probed by double-jump stopped-flow fluorescence experiments combining refolding of HisF with binding to HisH. Unfolded HisF-W156Y (12 μM in 6.0 M GdmCl) was diluted 6-fold to 1.0 M GdmCl and refolded for variable times. The sample was then mixed with an equal volume of a solution containing the same concentration of HisH in buffer without GdmCl. A monoexponential equation was fit to the obtained amplitudes.

RESULTS

Equilibrium Transitions. The thermodynamic stability of HisF was characterized by GdmCl-induced equilibrium unfolding transitions at 25 and 45 °C. The loss of secondary structure was monitored by far-UV circular dichroism (CD) at 225 nm, and the loss of tertiary structure was followed by tryptophan and tyrosine fluorescence. HisF contains a single tryptophan residue (W156) located in α-helix 5 and four tyrosine residues (Y39, Y143, Y182, and Y240), which are distributed across the protein (Figure 1b). Their fluorescence served as an indicator for global folding.

HisF is a thermostable protein [T_M > 90 °C (data not shown)], and the folding equilibrium in the transition region is attained very slowly. After incubation at 45 °C for 10 days, the unfolding curve was still shifted to higher denaturant concentrations with respect to the refolding curve, indicating that the unfolding reaction had not yet reached equilibrium (Figure S1 of the Supporting Information). However, after 19 days, the unfolding and refolding curves coincided, indicating that equilibrium had been reached. Remarkably, unfolding remained fully reversible even after this excessively long period of incubation at intermediate denaturant concentrations. At 25 °C, unfolding was even slower, and an incubation time of 6 weeks was required to reach equilibrium.

The folding reaction of HisF is cooperative, and the transitions determined by CD and fluorescence superimpose well (Figure 2), suggesting that folding intermediates are not

Table 1. Thermodynamic Parameters for the GdmCl-Induced Unfolding of HisF and I_{BP}[a]

	25 °C			45 °C		
protein, signal	ΔG_D° (kJ mol^{-1})	m (kJ mol^{-1} M^{-1})	$[D]_{1/2}$ (M)	ΔG_D° (kJ mol^{-1})	m (kJ mol^{-1} M^{-1})	$[D]_{1/2}$ (M)
HisF, far-UV CD	53.5 ± 1.7	19.1 ± 0.6	2.80	52.7 ± 3.3	19.7 ± 1.2	2.68
HisF, fluorescence	55.6 ± 2.3	19.9 ± 0.9	2.79	47.9 ± 2.8	18.0 ± 1.1	2.66
I_{BP}, far-UV CD	15.2 ± 2.4	7.9 ± 0.9	1.90			

[a]Values are for 50 mM Tris-HCl buffer (pH 7.5) at the respective temperature. The thermodynamic parameters were calculated from the data shown in Figures 2 and 4.

populated at equilibrium. The data were analyzed on the basis of the two-state model, yielding the thermodynamic parameters listed in Table 1. At 25 and 45 °C, HisF unfolds with a similar cooperativity ($m \sim 19$ kJ mol^{-1} M^{-1}), which is close to the value of 24 kJ mol^{-1} M^{-1} that is expected for a protein of its size (253 residues).[33] The differences between the two temperatures with respect to the transition midpoint ($[D]_{1/2} \sim 2.8$ M vs ~2.7 M) and the Gibbs free energy of unfolding ($\Delta G_D^\circ \sim 54$ kJ mol^{-1} vs ~50 kJ mol^{-1}) are small. Thus, the thermodynamic stability of HisF seems to be only weakly dependent on temperature.

Kinetics of Unfolding and Refolding. Unfolding and refolding kinetics of the secondary structure of HisF were measured at 25 and 45 °C by following the far-UV CD signal of the protein after manual mixing at various concentrations of GdmCl. Unfolding was a monoexponential reaction, as shown in Figure 3a for the reaction in 6.0 M GdmCl, and the amplitude accounted for the entire CD change as observed in the equilibrium unfolding transition. Refolding at low GdmCl concentrations was biexponential, as shown in Figure 3b for the reaction at 0.7 M GdmCl. The faster refolding phase vanished above 2 M GdmCl. The CD change observed in the refolding kinetics was significantly smaller than that expected from the equilibrium transitions (Figure 4a), indicating that the major part of the CD change occurred within the dead time of the manual mixing experiment. The kinetics of this CD change could also not be resolved after stopped-flow mixing, indicating that a burst-phase intermediate (I_{BP}) formed in <5 ms. This CD change is independent of protein concentration and thus probably not caused by rapid reversible aggregation.[34] The amplitude of the burst-phase refolding reaction decreased with an increasing concentration of GdmCl in a sigmoidal manner (Figure 4a). This cooperative unfolding process and the high magnitude of the relative CD signal of I_{BP} compared to N (71% at 1 M GdmCl and 36% at 2 M GdmCl) suggest that I_{BP} contains a high content of compact secondary structure. A tentative two-state analysis of the unfolding transition of I_{BP} yielded an apparent ΔG_D° of 15.2 kJ mol^{-1} and an m value of 7.9 kJ mol^{-1} M^{-1}. Thus, I_{BP} exhibits ~30% of the stability and ~40% of the compactness of the native state (Table 1).

Unfolding and refolding kinetics of the tertiary structure of HisF were measured at 25 and 45 °C by following the fluorescence changes of the single tryptophan and the four tyrosine residues (Figure 1b). Again, the unfolding kinetics were monophasic, and the refolding kinetics at low GdmCl concentrations were biphasic (Figure S2 of the Supporting Information). In both unfolding and refolding, the fluorescence amplitudes accounted for the entire signal change as expected from the equilibrium transitions (Figure 4b), which indicates that I_{BP} lacks significant tertiary structure.

The dye 1-anilino-8-naphthalenesulfonate (ANS) binds to hydrophobic regions of folding intermediates, which leads to an increase in its fluorescence intensity.[35,36] Stopped-flow

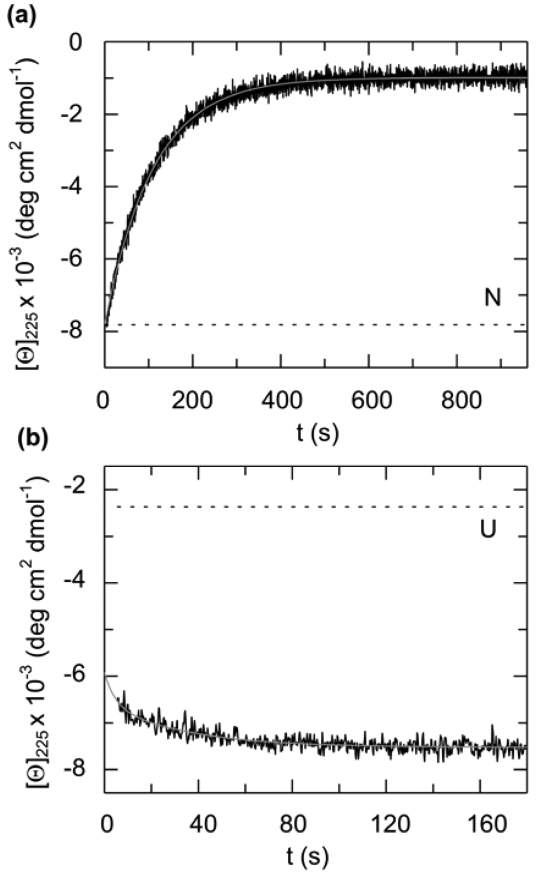


Figure 3. (a) Unfolding kinetics of HisF in 6.0 M GdmCl and (b) refolding kinetics in 0.7 M GdmCl as followed by far-UV CD. Following manual mixing, the traces of the CD signal at 225 nm were monitored in 50 mM Tris-HCl buffer (pH 7.5) at 25 °C. Gray solid lines represent fits of monoexponential and biexponential functions to the unfolding ($\tau = 110$ s) and refolding curves ($\tau_1 = 3.1$ s, and $\tau_2 = 34$ s), respectively. Dashed lines indicate the initial ellipticities characterizing the native (N) state (a) and the unfolded (U) state (b).

refolding experiments in the presence of ANS yielded a strong fluorescence increase within the dead time of mixing (Figure S3a of the Supporting Information), indicating that I_{BP} contains hydrophobic surfaces that are capable of ANS binding. The ANS fluorescence then decreased in two exponential phases with rates that correspond to the refolding rates detected by CD and by Trp/Tyr fluorescence (Figure S3b of the Supporting Information). These results show that the multistep transition from I_{BP} to N leads to a continuous decrease in the exposed hydrophobic surface area of HisF.

The observed rate constants of unfolding and refolding at 25 and 45 °C are shown in panels a and b of Figure 5 as a function of GdmCl concentration in the form of chevron plots. The two

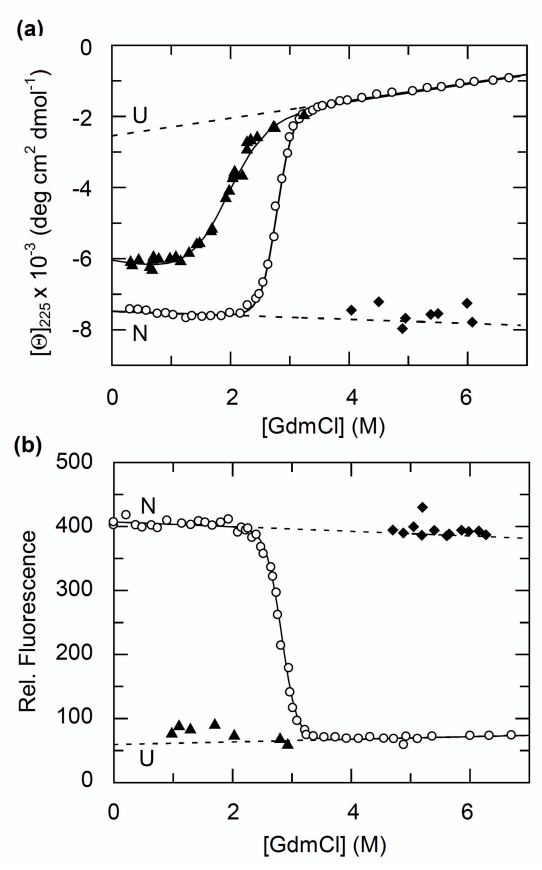


Figure 4. GdmCl-induced equilibrium unfolding transitions of HisF and initial values of refolding and unfolding kinetics. The equilibrium transitions (empty circles) were followed by (a) the CD signal at 225 nm and (b) Trp/Tyr fluorescence (excitation at 280 nm; emission at 320 nm) in 50 mM Tris-HCl buffer (pH 7.5) at 25 °C (data taken from Figure 2a,c) The dashed lines indicate the baselines for the pure N and U states. The initial values of the unfolding (filled diamonds) and refolding (filled triangles) kinetics were obtained by extrapolation to zero time (see Figure 3). The data show that a compact burst-phase refolding intermediate (I_{BP}) with a high content of secondary structure but no detectable tertiary structure is formed. The solid lines represent fits based on the two-state model to the data, yielding the thermodynamic parameters listed in Table 1.

refolding rate constants and the single unfolding rate constant as determined by CD and Trp/Tyr fluorescence superimpose well, indicating that, in these kinetic phases, secondary and tertiary structure formation and denaturation occur simultaneously.

The two refolding rates differ by ~10-fold at 25 °C and ~4-fold at 45 °C and are largely independent of denaturant concentration below 2 M GdmCl (Figure 5a,b). Denaturant-independent kinetics are often observed when folding is limited in rate by prolyl isomerization.[14,37] However, native HisF lacks cis prolines, and both refolding phases are insensitive to the cis/trans prolyl isomerase *Thermococcus* FKBP18 (data not shown). Decreased denaturant dependencies as in panels a and b of Figure 5 are usually called a "roll-over" and assumed to originate from the population of folding intermediates.[15] We suggest that the roll-over in the fast folding rate is caused by the burst-phase intermediate I_{BP}, which is populated at low GdmCl concentrations and shows a transition midpoint of 1.9 M GdmCl (see Figure 4a).

The roll-over observed for the slow folding rate might be caused by a folding intermediate I that is formed in the fast folding phase. In such a case, the two refolding phases observed between 0 and 2 M GdmCl (Figure 5a,b) should reflect two steps on a sequential pathway via a partially folded intermediate that is unstable above 2 M GdmCl. Such a sequential model is supported by the denaturant dependence of the refolding amplitudes (Figure 5c,d). These amplitudes change in a reciprocal fashion. The fast phase dominates at low GdmCl concentrations, but with an increasing GdmCl concentration, it loses amplitude and the amplitude of the slow phase increases in a reciprocal fashion, as expected for a sequential folding mechanism. At 45 °C, the decrease in the amplitude of the fast phase occurs at lower GdmCl concentrations than at 25 °C, suggesting that the intermediate is less stable at the higher temperature.

Above 2 M GdmCl, refolding and unfolding become monophasic at 25 and 45 °C, and the rates decrease and increase exponentially with denaturant concentration (Figure 5a,b). This apparent two-state refolding and unfolding behavior indicates that the putative folding intermediates are destabilized at high denaturant concentrations. The strong dependence of the observed refolding and unfolding rates on GdmCl concentration (corresponding to large m values) leads to extremely slow kinetics at intermediate denaturant concentrations, in line with the excessively long incubation time required to reach equilibrium in the unfolding transitions (Figure S1 of the Supporting Information). At 25 °C, the unfolding reaction exhibits a time constant ($\tau = 1/\lambda$) of 3.5 days at 4 M GdmCl, and between 4 and 2.5 M GdmCl, the kinetics were too slow for the determination of the involved rates.

Interrupted Refolding Assays. The refolding kinetics of HisF are biphasic at low denaturant concentrations (Figure 5), and the amplitude analysis (Figure 5c,d) suggested that the two phases might reflect the formation of an intermediate and its subsequent transformation to the native state. To examine whether such an intermediate is in fact formed as a transient species during refolding of HisF, we employed a stopped-flow double-mixing approach. In these experiments, refolding was interrupted after various times, and the concentrations of species present at the time of sampling were determined from the rates and amplitudes of their unfolding.

At 1.0 M GdmCl, the time constants of the two folding phases are 3.3 and 33 s [at 25 °C (Figure 5a)]. After refolding had been conducted for 10 s, a putative intermediate should thus be well populated and the concentration of fully folded molecules should still be very small. In the double-mixing experiment, refolding was therefore interrupted after 10 s, and the sample was transferred to unfolding conditions of 6.0 M GdmCl (Figure S4 of the Supporting Information), to monitor the rate and the amplitude of its unfolding reaction. Under these conditions, native HisF unfolds with a time constant of ~170 s (Figure 5a). The molecules formed in the fast refolding reaction, however, unfolded 3 orders of magnitude more rapidly with a time constant of 0.02 s. The product of the fast refolding reaction is thus not the native protein, and a parallel folding mechanism can be ruled out.

In further double-mixing experiments, unfolded HisF was also refolded in 1.0 M GdmCl for 10 s but then diluted to GdmCl concentrations between 1.9 and 7.5 M. In all cases, monoexponential unfolding kinetics were observed, and the rate constants were significantly higher (~10^5-fold at 5.0 M

dx.doi.org/10.1021/bi300189f | *Biochemistry* 2012, 51, 3420−3432

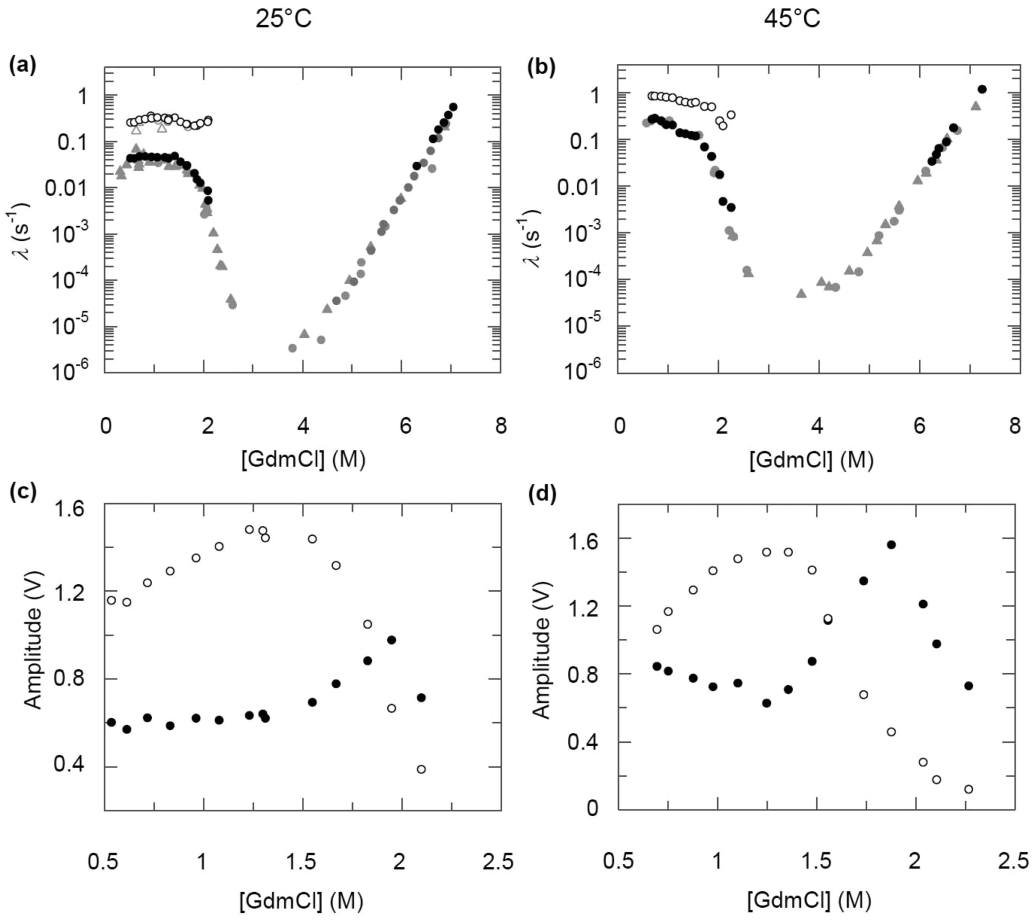


Figure 5. GdmCl dependence of the apparent rate constants (λ) of refolding and unfolding of HisF at (a) 25 °C and (b) 45 °C (chevron diagrams). Following manual mixing, the traces of the CD signal at 225 nm (gray triangles) and Trp/Tyr fluorescence (excitation at 280 nm; emission at 320 nm) (gray circles) were monitored in 50 mM Tris-HCl buffer (pH 7.5). Rate constants for the fast refolding phase are shown with empty symbols, and rate constants for the slow refolding and the single unfolding phases are shown with filled symbols. Reactions with λ values of >0.004 s^{-1} were also followed after stopped-flow mixing monitoring Trp/Tyr fluorescence (black circles). The observed amplitudes of the fast and slow refolding phases are shown (empty and filled black circles, respectively) at (c) 25 and (d) 45 °C.

GdmCl) than the rate constants of unfolding of the native protein, confirming that they reflect the unfolding of an intermediate (I) that formed during the fast folding phase. The measured unfolding rates of I connect smoothly with the previously determined fast refolding rates in the transition region around 2.1 M GdmCl (Figure 6a), indicating that the folding of I is reversible. The resulting chevron shows a minimum near 2.1 M GdmCl, in good agreement with the increase in the amplitude of unfolding of I (Figure 6b), which reflects its stability. Thus, I seems to be more stable than the burst-phase intermediate I_{BP} ($[D]_{1/2}$ = 1.9 M GdmCl) and much less stable than native HisF ($[D]_{1/2}$ = 2.8 M GdmCl). The unfolding rate of I increases between 2 and 4 M GdmCl but becomes denaturant-independent above 5.0 M GdmCl (Figure 6a). The reason for this behavior remains unclear. It might originate from a shift in the free energy of the transition state relative to the native state at high GdmCl concentrations (Hammond behavior)[38−40] or from the accumulation of an unfolding intermediate due to ground state effects.[41]

Interrupted refolding assays were also used to follow the time course of the accumulation and depletion of intermediate I. In these experiments, samples of HisF were allowed to refold in 1.0 M GdmCl for 1−200 s, and then they were transferred to 2.3 M GdmCl. At this denaturant concentration, native HisF molecules remain folded (Figure 2), but the intermediate I unfolds with a time constant of 2 s. The fluorescence amplitude of this unfolding reaction is proportional to the concentration of the intermediate present at the time when refolding was interrupted (Figure 7a) and thus traces the time course of the intermediate I during refolding (Figure 7b). The concentration of I first increased and reached a maximum at a refolding time of 10 s; then it decreased to almost zero after refolding had been conducted for ∼200 s. A fit of a double-exponential equation to these data yielded time constants of 2.9 and 40 s for the formation and depletion of I, respectively, which coincide with the time constants of the fast and the slow reaction, as determined in the conventional refolding kinetics (Figure 7b). This result demonstrates that these two phases reflect the formation and decay of intermediate I.

To examine whether the folding intermediate is directly transformed into the native protein, the unfolding assay in the second step of the double-mixing experiments was performed at 6.9 M GdmCl. At this high denaturant concentration, both the intermediate and the native state unfold, with time constants of 56 ms and 2.6 s, respectively. This large difference in rates allows one to trace the concentrations of I and N

dx.doi.org/10.1021/bi300189f | *Biochemistry* 2012, 51, 3420−3432

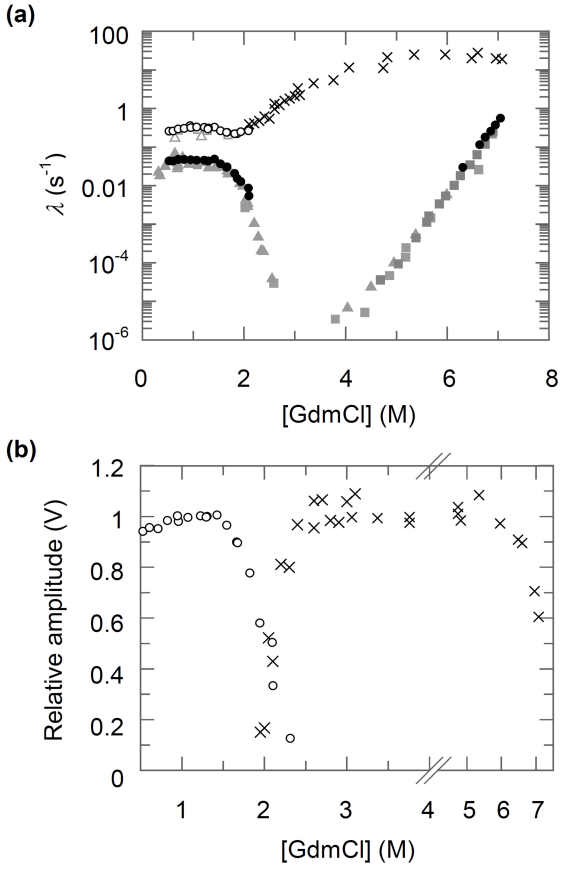


Figure 6. Analysis of the folding intermediate of HisF by interrupted refolding experiments. (a) Unfolding rate constants (λ) observed after refolding in 1.0 M GdmCl for 10 s in stopped-flow double-mixing experiments at 25 °C using Trp/Tyr fluorescence (excitation at 280 nm; emission at 320 nm) (crosses). The rate constants determined by single-mixing experiments (Figure 5a) are shown for comparison. (b) GdmCl dependence of the fluorescence amplitudes of the fast refolding phase (empty circles) derived from single-mixing experiments and of the unfolding phase (crosses) as observed in the double-mixing experiments. The amplitudes are constant between 0 and 1.4 M GdmCl for refolding and between 2.6 and 6.0 M GdmCl for unfolding. These amplitudes represent complete refolding from U to I and complete unfolding of I to U, respectively, and were therefore set to 1. Both amplitudes have a value of 0.5 at 2.1 M GdmCl, which is consistent with the transition midpoint ($[D]_{1/2}$) as estimated from the chevron diagram of I shown in panel a.

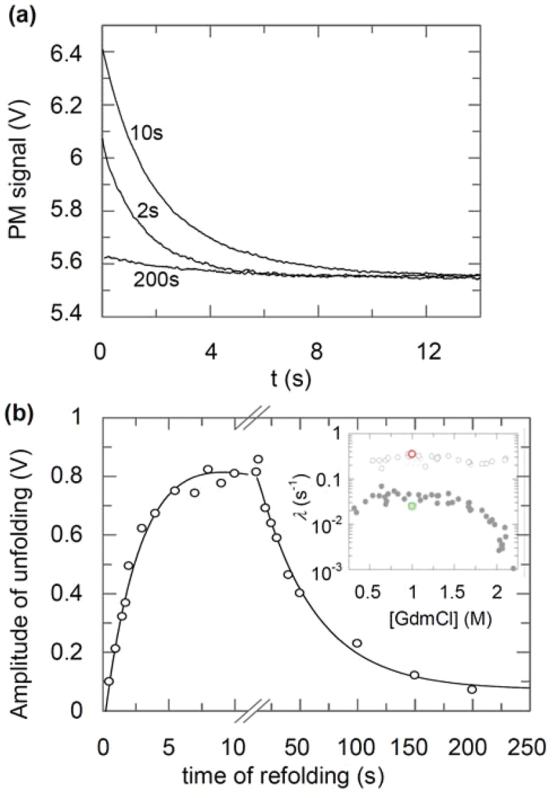


Figure 7. Time course of the folding intermediate (I) of HisF determined by stopped-flow double-mixing experiments. (a) Refolding of unfolded protein was initiated by dilution to 1 M GdmCl. After various refolding times (shown are 2, 10, and 200 s), the protein was diluted to 2.3 M GdmCl, and unfolding monitored by the decay of Trp/Tyr fluorescence (excitation at 280 nm; emission at >320 nm) was followed. A monoexponential function and a linear term, which represents the onset of the slow refolding reaction, were fit to the experimental curves. The time constant of unfolding of the intermediate ($\tau \sim 2$ s) was largely independent of the refolding time. (b) Relative amplitudes of unfolding in 2.3 M GdmCl as a function of refolding time in 1.0 M GdmCl. The solid line represents the fit of a double-exponential equation to the experimental points, yielding time constants of 2.9 s for the accumulation of the intermediate and 40 s for its depletion. In the inset, the observed rate constants for the accumulation (red circle) and depletion (green circle) of the intermediate are compared with the refolding rates determined by single-jump experiments (gray circles, data taken from Figure 5a). All experiments were conducted in 50 mM Tris-HCl buffer (pH 7.5) at 25 °C.

simultaneously and with high precision from the amplitudes of their unfolding reactions.

In a sequential three-state reaction, the formation of the native state should show a lag, because the concentration of the intermediate is zero at time zero. This hallmark of a sequential reaction is observed for the folding of HisF (Figure 8a). The lag in the time course for the native state is emphasized in the expanded view in Figure 8b. An equation describing a consecutive reaction was used for individual fits of the time courses of the intermediate and of the native state (Figure 8). The obtained rate constants were very similar to each other, showing that the intermediate is directly converted into the native form of HisF. Again, these rates coincided with those observed directly by fluorescence (Figure 8a). These findings complete the evidence that folding of HisF is sequential and that the intermediate is directly converted to the native protein.

Recovery of Protein Binding Competence during Refolding. The interrupted refolding experiments showed that molecules with the specific unfolding rate of the native state are formed in the slow folding phase. To confirm that correctly folded and functional HisF is recovered during this phase, refolding experiments were combined with a functional assay. In imidazole glycerol phosphate synthase, HisF (the cyclase subunit) and HisH (the glutaminase subunit) form a functional 1:1 complex.[25] The formation of this complex can be followed by a change in the fluorescence of the single Trp123 residue of HisH. This residue is solvent-exposed in the isolated protein but becomes buried upon binding to HisF (Figure 9a), which leads to a blue shift of its emission spectrum (Figure 9b). To avoid interference from the single Trp156 of HisF, the HisF-W156Y variant was used to follow the association of the

dx.doi.org/10.1021/bi300189f | *Biochemistry* 2012, 51, 3420−3432

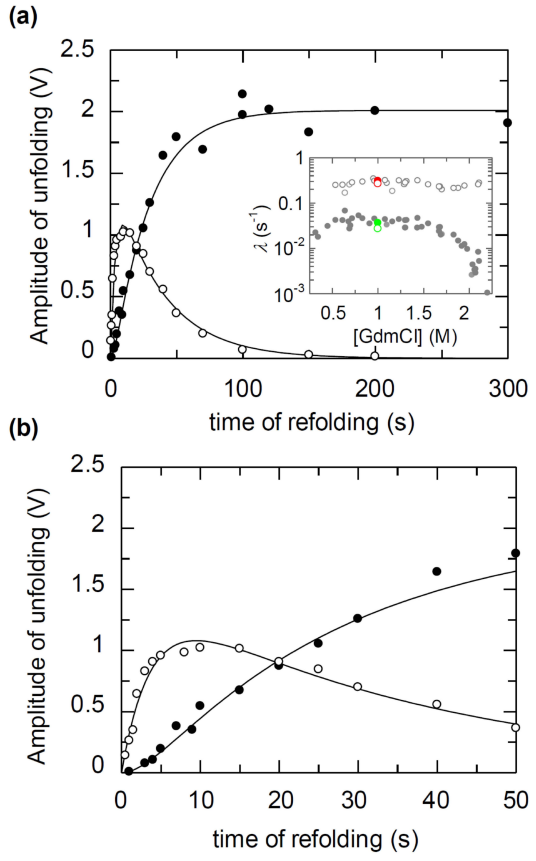


Figure 8. Time course of the unfolding intermediate and the native state of HisF determined by stopped-flow double-mixing experiments. Data were obtained as described in the legend of Figure 6, but unfolding was induced by dilution to 6.9 M GdmCl. A double-exponential function in combination with a linear term was fit to the observed fluorescence decays. The relative amplitudes of the fast [(empty circles) τ = 56 ms] and slow [(filled circles) τ = 2.6 s] phases, which characterize the unfolding of the intermediate and of the native state, respectively, are shown as a function of the duration of refolding in 1.0 M GdmCl (a) for the entire time range and (b) for the first 50 s of refolding. The solid lines represent a fit of an equation describing a consecutive process to the data.[31,32] In the inset of panel a, the obtained rate constants from the time course of I [(empty red circle) τ = 3.2 s, and (empty green circle) τ = 27 s] and from the time course of N [(filled red circle) τ = 3.7 s, and (filed green circle) τ = 36 s] are compared with the refolding rates determined by single-jump experiments (gray circles, data taken from Figure 5a).

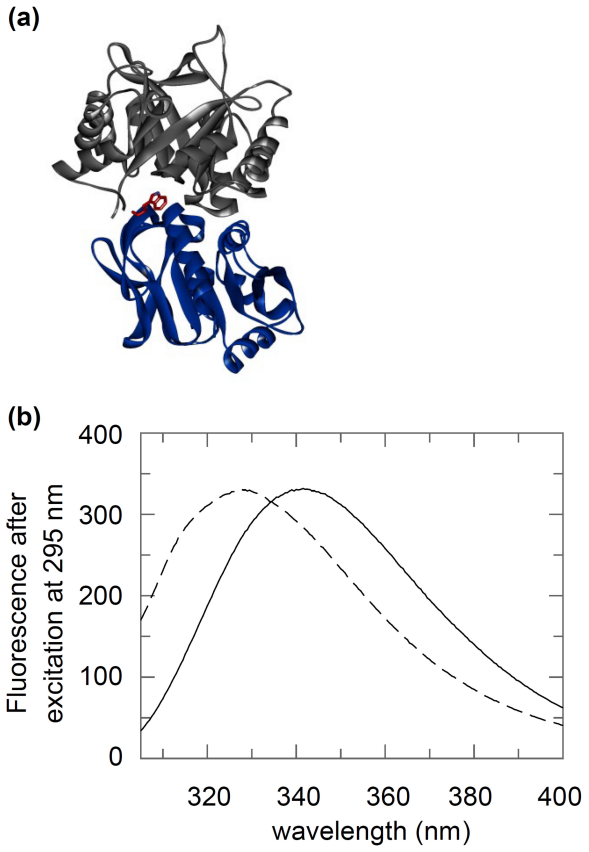


Figure 9. Stoichiometric complex of HisF and HisH that constitutes the imidazole glycerol phosphate synthase from *T. maritima*. (a) Ribbon diagram of the heterodimer (PDB entry 1GPW)[26] with HisH colored blue and HisF colored gray. The single Trp123 of HisH is displayed as red sticks. (b) The fluorescence emission of Trp123 from HisH (solid line, maximum at 341 nm) is blue-shifted upon binding to HisF-W156Y (dashed line, maximum at 327 nm). The fluorescence spectra of 2 μM protein were measured in 50 mM Tris-HCl buffer (pH 7.5) at 25 °C.

two proteins. The dissociation constant of the HisH−HisF-W156Y complex (K_D) was determined by equilibrium fluorescence titration to be ~5 nM (data not shown). The kinetics of association between 1 μM HisF and 1 μM HisH were measured in a stopped-flow apparatus. The apparent time constant of 0.07 s shows that the association is much more rapid than both the fast (50-fold) and slow (500-fold) refolding reaction of HisF. To determine the time course of the formation of binding-competent HisF upon refolding, double-mixing stopped-flow experiments were performed. Unfolded HisF-W156Y was first refolded in 1.0 M GdmCl for various times. Then the sample was mixed with HisH, and binding was followed by the decay of the fluorescence of Trp123 of HisH (Figure 10a). The observed amplitudes directly reflect the amount of binding-competent HisF present after various periods of refolding. A monoexponential equation was fit to the time-dependent increase in the amplitude of the binding reaction. The time constant of this reaction (τ = 40 s) is virtually identical with the time constant of the slow refolding phase (Figure 10b). These results show that the product of the rate-limiting slow folding step is native HisF. The fast folding phase (τ = 3.3 s) was not accompanied by a fluorescence decrease, suggesting that the folding intermediate I is not yet capable of binding to HisH. A lower limit for the dissociation constant (K_D) describing a putative binding of I to HisH can be estimated. After HisF had refolded for 2 s, half of the protein exists as the folding intermediate (Figure 7b), corresponding to 0.5 μM I in our binding assay. The maximal concentration of a potential complex between I and HisH can be estimated from the observation that the relative binding amplitude after refolding of HisF-W156Y for 2 s is <0.06 (Figure 10b), indicating that no more than 6% (<0.03 μM) of I could be involved in a complex with HisH. Consequently, the K_D for the binding of I to HisH must be larger than 17 μM, suggesting that the affinities of I and native HisF for HisH differ at least 1000-fold.

dx.doi.org/10.1021/bi300189f | *Biochemistry* 2012, 51, 3420−3432

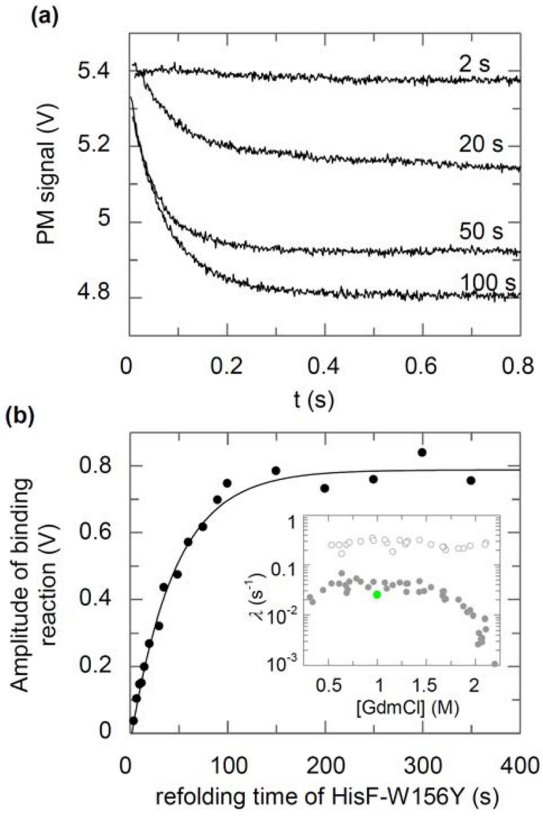


Figure 10. Kinetics of the formation of the HisF−HisH complex upon refolding of HisF determined by stopped-flow double-mixing experiments. (a) Refolding of unfolded HisF-W156Y was initiated by dilution to 1.0 M GdmCl. After various time intervals, HisF was mixed with HisH (final concentrations of 1.0 μM for each protein and 0.47 M GdmCl), and complex formation was followed by the decay of the fluorescence of W123 from HisH (excitation at 295 nm; emission at >320 nm). The experimental traces were fit with a monoexponential function. As expected, the time constant for complex formation (τ = 0.07 s) was independent of refolding time, whereas the binding amplitudes increased with time (shown are 2, 20, 50, and 100 s). (b) Amplitudes of the binding of HisF-W156Y to HisH as a function of refolding time in 1.0 M GdmCl. The solid line represents a fit of a single-exponential equation to the experimental points, yielding a time constant of 40 s. In the inset, this rate constant (green circle) is compared with the refolding rates determined by single-jump experiments (gray circles, data taken from Figure 5a). All experiments were conducted in 50 mM Tris-HCl buffer (pH 7.5) at 25 °C.

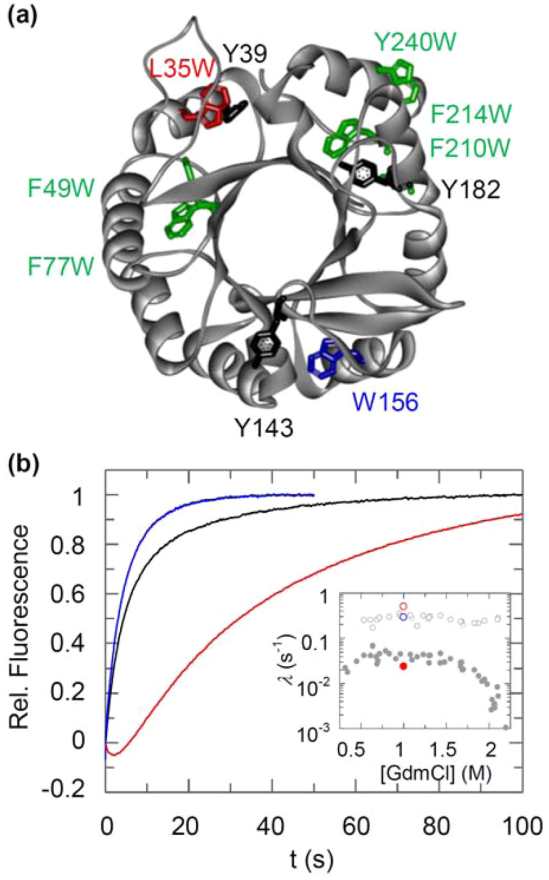


Figure 11. Monitoring refolding of HisF via individual Trp residues. (a) Ribbon diagram of the HisF structure (PDB entry 1THF)[23] with marked Tyr and Trp residues. The introduced Trp residues are colored red and green; native W156 is colored blue. (b) Refolding kinetics of wild-type W156-HisF and W35-HisF in 0.6 M GdmCl. Traces are shown for the Tyr/Trp fluorescence of wild-type W156-HisF (excitation at 280 nm; emission at >320 nm) (black), the Trp156 fluorescence of W156-HisF (excitation at 290 nm; emission at >320 nm) (blue), and the Trp35 fluorescence of W35-HisF (excitation at 290 nm; emission at >320 nm) (red). Monoexponential and biexponential functions were fit to the experimental curves. In the inset, the determined rate constants (λ) when monitoring the fluorescence of Trp156 [(empty blue circle) τ = 5.1 s] and Trp35 [(empty red circle) τ_1 = 2.2 s, and (filled red circle) τ_2 = 39 s] are compared with the refolding rates obtained by Tyr/Trp fluorescence [(empty gray circles) τ_1 = 4 s, and (filled gray circles) τ_2 = 25 s at 0.6 M GdmCl] taken from Figure 5a.

Following the Fluorescence of Individual Tryptophan Residues upon Refolding. To obtain the first evidence of the structure of the folding intermediate of HisF, we produced variants with Trp residues placed in different regions of the protein and analyzed their refolding reactions. Wild-type HisF contains only a single Trp (Trp156). It is located in α-helix 5 and monitors formation of tertiary structure in the central region of the protein chain. To follow formation of the tertiary structure in the N-terminal segment of HisF, we removed Trp from position 156 (by the W156Y substitution) and introduced a Trp into α-helix 1 at position 35 (by the L35W substitution). We call this variant "W35-HisF". Because HisF shows a pronounced 2-fold symmetry,[42] positions 35 and 156 in the N- and C-terminal half-barrels, respectively, are symmetry-related (Figure 11a). In the native state, Trp35 has a lower fluorescence quantum yield than Trp156 but the same emission maximum, and the spectrum becomes similarly red-shifted upon unfolding (Figure S5 of the Supporting Information). Equilibrium transitions at 45 °C showed that W35-HisF unfolds with cooperativity similar to that of wild-type HisF (W156-HisF) and is only slightly less stable (Figure S6 of the Supporting Information), suggesting that the folding mechanism is not changed by the shift of the reporter Trp from position 156 to position 35.

The kinetics of refolding of HisF, as monitored by the fluorescence of Trp156 in α-helix 5 or Trp35 in α-helix 1 (excitation at 290 nm; emission at >320 nm) at 0.6 M GdmCl, are compared in Figure 11b with the kinetics of refolding of wild-type HisF (W156-HisF) monitored by the fluorescence of both Trp156 and the four Tyr residues (excitation at 280 nm;

dx.doi.org/10.1021/bi300189f | *Biochemistry* 2012, 51, 3420−3432

emission at >320 nm). In accordance with Figure 5a, the Tyr/Trp fluorescence of W156-HisF monitors the formation of the intermediate (τ = 4 s) and its conversion to the native state (τ = 25 s). In contrast, when refolding of wild-type HisF was selectively followed by the fluorescence of Trp156, only the fast reaction with a time constant of 5.1 s was observed. Apparently, Trp156 in α-helix 5 is already in a native-like environment in the folding intermediate I and does not undergo further changes during the slow (rate-determining) transition from I to the native state N. This finding suggests that the central part of HisF adopts already a native-like tertiary structure in I, in which Trp156 is shielded from the solvent as efficiently as in N. When the refolding of W35-HisF was selectively followed by the fluorescence of Trp35, the signal was dominated by the slow reaction with a time constant of 39 s (Figure 11b). This result indicates that Trp35 in α-helix 1 is largely in an unfolded-like environment in the intermediate I and that the tertiary structure surrounding it is mainly formed upon transition from I to the native state N. Additionally, a fluorescence phase with a negative amplitude and a time constant of 2.2 s was observed. The latter agrees with the fast folding rate observed by CD and Tyr/Trp fluorescence (Figure 11b), but the relative amplitude is less than 5%. Taken together, our results suggest that the folding intermediate I of HisF contains native-like tertiary structure in its central part but is partly unfolded in its N-terminal region. To gain more insight into the tertiary structure content of I, we moved the single Trp residue also to positions 49, 77, 210, 214, and 240 (Figure 11a) and measured their refolding kinetics (excitation at 290 nm; emission at >320 nm). All these Trp residues detected both folding phases, suggesting that, at these positions, the folding intermediate is not yet nativelike (Table S1 of the Supporting Information).

DISCUSSION

Thermodynamic Stability. The GdmCl-induced equilibrium unfolding of HisF from the hyperthermophilic bacterium *T. maritima* is described well by the two-state mechanism. Remarkably, it took ~3 weeks at 45 °C and ~6 weeks at 25 °C to reach the folding equilibrium, which is reflected in excessively slow folding kinetics in the transition region. Importantly, HisF unfolds also very slowly in the physiological temperature range of *T. maritima*, which allows for the purification of the recombinant protein by heat incubation of the *E. coli* host cell extract.[28] Extremely slow unfolding kinetics have also been observed for other thermostable proteins,[19−22] and the properties of HisF further support the notion that a high kinetic barrier to unfolding provides a major mechanism for protecting proteins from hyperthermophiles against thermal denaturation.[43]

Folding Mechanism. In refolding, HisF forms a burst-phase intermediate I_{BP} with a significant amount of secondary structure. Then, in a sequential folding reaction, first a folding intermediate (I) is formed that is converted into the native state in the final, rate-limiting step of folding. Interrupted refolding experiments showed that the intermediate I unfolded much faster than the native protein and that folding is a sequential process with I as an obligatory on-pathway intermediate and a lag in the formation of the native protein. Both intermediates I_{BP} and I are less stable than N and not populated in the transition region. The on-pathway intermediate I cannot bind to HisH, which forms a functional complex with HisF. I_{BP} is formed within the dead time of refolding. The pronounced rollover observed for the fast folding phase below 1.0 M GdmCl (Figure 5a), resulting in a positive m value of refolding, indicates that I_{BP} is off-pathway and that non-native interactions within I_{BP} have to be broken before I can form. However, the alternative of I_{BP} being on-pathway cannot be excluded. At >2 M GdmCl, both folding and unfolding of HisF can be described by a single phase (Figure 5), most probably because I_{BP} and I are destabilized by the denaturant and do not accumulate.[44] As a consequence, our data are in accordance with a linear four-species model (Scheme 1).

Scheme 1. Linear Four-Species Model for HisF Folding[a]

$$I_{BP} \rightleftharpoons U \rightleftharpoons I \rightleftharpoons N$$

[a]An off-pathway equilibrium between the unfolded state (U) and the burst-phase intermediate (I_{BP}) precedes the formation of the on-pathway intermediate (I) and the native state (N). At >2 M GdmCl, I_{BP} and I are too unstable to accumulate to detectable concentrations. Although it is improbable, our data cannot unambiguously exclude the possibility that I_{BP} is an on-pathway intermediate that exists in an equilibrium with I.

Structural Properties of the Burst-Phase Intermediate I_{BP} and the On-Pathway Intermediate I. Our experimental data provide information about the structural properties of the folding intermediates I_{BP} and I. The formation of the burst-phase intermediate I_{BP} in the first few milliseconds of refolding is accompanied by a large change in the far-UV CD signal in the dead time of stopped-flow mixing. The magnitude of this CD signal decreases in a cooperative manner with an increasing concentration of GdmCl, and the resulting "unfolding" curve suggests that I_{BP} is significantly less stable than the native protein. I_{BP} contains a high content of secondary structure but not a defined tertiary structure, as evidenced by a fluorescence that resembles that of the unfolded state U. In line with this finding, I_{BP} efficiently binds the dye ANS (Figure S3a of the Supporting Information), which is known to have a high affinity for solvent-exposed hydrophobic patches.[35,36] I_{BP} is probably a molten globule-like state[45] that forms rapidly when the unfolded protein is exposed to native conditions.

The on-pathway intermediate I binds weaker to ANS than I_{BP} (Figure S3b of the Supporting Information), suggesting that in I the hydrophobic side chains are shielded more efficiently from the solvent. The formation of I is accompanied by further changes in far-UV CD and Trp/Tyr fluorescence, which indicates that it contains both secondary and tertiary structure. Analysis of refolding by following the fluorescence of engineered Trp residues at various positions in HisF shows that tertiary structure formation differs at different positions. It seems more complete in the central region of the protein chain than in the N-terminal region (Figure 11). Also, the HisF folding intermediate cannot associate with its functional partner HisH (Figure 10), indicating that the binding interface is not yet formed. This finding is consistent with the crystal structure of the HisH−HisF complex demonstrating that the N-terminal segment of HisF is required for docking to HisH.[26]

Comparison of the Folding Mechanisms of HisF and Other $(\beta\alpha)_8$-Barrel Proteins. It is assumed that the shape of the folding landscape of a protein is determined by its chain topology and by individual sequence patterns.[46−50] To evaluate the relevance of these factors for the folding of $(\beta\alpha)_8$-barrels, we compared the folding properties of HisF with those of other $(\beta\alpha)_8$-barrel proteins, αTS,[13,14] IGPS,[15−17] and IOLI.[18] Unlike HisF, which displays two-state equilibrium unfolding, these

dx.doi.org/10.1021/bi300189f | *Biochemistry* 2012, 51, 3420−3432

three proteins form one or two equilibrium intermediates that contain ~50% of the far-UV CD signal of the native state. These proteins are much less stable than HisF from *T. maritima* and can be unfolded by the weak denaturant urea. Obviously, folded HisF contains additional stabilizing interactions, and therefore, >2.5 M GdmCl is required for its unfolding, conditions under which partially folded forms of the protein are destabilized already. Thus, the strong difference in stability between HisF and the other $(\beta\alpha)_8$-barrel proteins explains probably why equilibrium unfolding intermediates are not populated in the unfolding transition of HisF.

Despite these differences in equilibrium unfolding and despite the low overall level of sequence identity of <20% [as detected by a pairwise comparison with DALI-lite (http://www.ebi.ac.uk/Tools/dalilite/)], HisF resembles the three other analyzed $(\beta\alpha)_8$-barrel proteins in the kinetic folding mechanism. In all cases, a burst-phase intermediate I_{BP} is formed,[14,17,18] which is marginally stable but contains a high fraction of the far-UV CD signal of the native state (HisF, 71%; IGPS, 46%; αTS, 47%; IOLI, 37%). Global folding analysis or Gō-model simulation has suggested that the I_{BP} forms of αTS,[14,51] IGPS,[17] and IOLI[18] are off-pathway species, meaning that the unfolding of their non-native structural elements controls access to the productive on-pathway intermediates. Our kinetic data for HisF suggest that its I_{BP} is also an off-pathway intermediate. Substantial burst-phase reactions have also been observed for the $(\beta\alpha)_8$-barrel proteins fructose-1,6-bisphosphate aldolase[52] and PRAI,[53] as well as for the $(\beta\alpha)_5$-barrel proteins CheY[54] and Spo0F.[55] The latter were proposed to be evolutionarily related to $(\beta\alpha)_8$-barrels via a common $(\beta\alpha)_4$-half-barrel precursor.[56,57] The accumulation of non-productive burst-phase intermediates in the folding of $(\beta\alpha)_8$-proteins might be linked with the complexity of this structure and the concomitant slowness of its formation. Thus, rapid alternative folding paths that lead to the exclusion of solvent from side chains and from the main chain and to secondary structure formation compete with productive folding.[16,17,51,54,55]

At first glance, the chevron diagrams show different degrees of complexity. In particular, αTS folds along four parallel folding channels, which was explained by the cis−trans isomerization reactions of three proline residues.[58−60] HisF, IGPS, and IOLI contain only trans prolyl peptide bonds and show simpler folding mechanisms.[17,18] As an explicit similarity, all four proteins show roll-over phenomena at low denaturant concentrations and a strong decrease of the folding rates at moderate denaturant concentrations.

In addition to the off-pathway intermediate I_{BP}, one or two on-pathway kinetic intermediates are formed on the folding pathways of these $(\beta\alpha)_8$-barrel proteins.[14,15,18,61] How well conserved are the structures of these productive folding intermediates? PRAI, IGPS, and αTS catalyze three consecutive steps in tryptophan biosynthesis and contain a conserved phosphate binding site in the C-terminal half of their protein chain, and most probably, they diverged from a common ancestor by several gene duplications.[24,62] Early fragment sudies with PRAI suggested a common "6+2" model for the folding for these proteins, in which a prefolded $(\beta\alpha)_{1-6}$ module serves as a template for the association and folding of the remaining $(\beta\alpha)_{7-8}$ moiety.[10] Hydrogen exchange mass spectrometry (HX-MS) experiments with IGPS indicated that productive folding begins with the formation of an intermediate (termed I_a) that comprises the central $\beta_2(\beta\alpha)_{3-5}\beta_6$ segment. Subsequently, the adjacent α_1, α_6 and β_7 segments fold, yielding the intermediate I_b, and finally, the terminal segments α_0, α_7, β_8, and α_8 are structured to form the complete $(\beta\alpha)_8$-barrel (N).[16,17] These findings point to a "6+2" mechanism for the formation of the $(\beta\alpha)_8$ structure. For αTS, however, HX protection experiments rather suggest that the N-terminal half-barrel $(\beta\alpha)_{1-4}$ is formed in the rapidly formed folding intermediate.[63] This is reminiscent of the "4+4" folding model that was suggested for HisF, based on the finding that a catalytically active HisF−NC complex could be reconstituted in the joint refolding of the isolated N-terminal (HisF-N) and C-terminal (HisF-C) half-barrels.[42] Our refolding experiments with the individual Trp variants of HisF, however, suggest that its intermediate rather consists of a folded central region and unfolded termini (Figure 11), which resembles more the intermediate of IGPS than that of αTS.

How well conserved are the rates for the formation and for the unfolding of the $(\beta\alpha)_8$-barrel structure in these proteins? A direct comparison of the unfolding and refolding rates of the four $(\beta\alpha)_8$-barrel proteins HisF, αTS, IGPS, and IOLI is complicated by the fact that different denaturants, GdmCl and urea, were used. However, the rate-limiting steps of refolding of the four proteins at low denaturant concentrations are very similar. Despite the low level of sequence similarity and the strong differences in stability, they all show time constants of ~10 s, supporting the view that chain topology is an important determinant for refolding of $(\beta\alpha)_8$-barrels.

Strong differences are observed, however, for the unfolding rates of the four $(\beta\alpha)_8$-barrel proteins when compared at equal distances from the midpoint of the transition curve, that is at equal stability. Unfolding of HisF from the hyperthermophilic organism *T. maritima* is at least 1000-fold slower than the unfolding of αTS and IOLI, which are from mesophilic organisms. Remarkably, unfolding of HisF is also ~100-fold slower than unfolding of truncated IGPS, which originates also from a hyperthermophilic organism (*S. solfataricus*) but shows a lower thermodynamic stability.[15]

These findings suggest that for refolding, the proteins from hyperthermophiles use the same mechanism as their counterparts from mesophilic organisms. The stabilizing amino acid substitutions exert their influence late on the folding pathway, after the passage through the transition state, by selectively stabilizing the native state relative to the transition state. The increased equilibrium stability of these proteins is thus used entirely for a maximal increase in their kinetic stability.

■ ASSOCIATED CONTENT

Supporting Information

Figures showing GdmCl-induced equilibrium unfolding and refolding transitions of HisF (Figure S1), unfolding and refolding kinetics followed by fluorescence of Trp/Tyr residues (Figure S2) and bound ANS (Figure S3), unfolding of the intermediate in an interrupted refolding experiment (Figure S4), fluorescence emission spectra of W156-HisF and W35-HisF (Figure S5), and the equilibrium refolding transition of W35-HisF (Figure S6) and relative amplitudes and rate constants of refolding calculated from the fluorescence signal of different single Trp residues (Table S1). This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Phone: +49-941-943 3015. Fax: +49-941-943 2813. E-mail: reinhard.sterner@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, Germany) for stopped-flow CD measurements.

ABBREVIATIONS

HisF, cyclase subunit of imidazole glycerol phosphate synthase (ImGPS) from *T. maritima*; HisH, glutaminase subunit of ImGPS; N, native state; U, unfolded state; I_{BP}, burst-phase intermediate; I, kinetic on-pathway intermediate; W156-HisF, wild-type HisF containing the single tryptophan residue at position 156; W35-HisF, HisF variant containing the W156Y and L35W amino acid exchanges; αTS, α-subunit of tryptophan synthase from *E. coli*; IGPS, indole-3-glycerol phosphate synthase from *S. solfataricus*; IOLI, protein of unknown function encoded by the *B. subtilis iol*I gene; PDB, Protein Data Bank.

REFERENCES

(1) Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol. 247*, 536−540.

(2) Caetano-Anollés, G., Kim, H. S., and Mittenthal, J. E. (2007) The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture. *Proc. Natl. Acad. Sci. U.S.A. 104*, 9358−9363.

(3) Wierenga, R. K. (2001) The TIM-barrel fold: A versatile framework for efficient enzymes. *FEBS Lett. 492*, 193−198.

(4) Sterner, R., and Höcker, B. (2005) Catalytic versatility, stability, and evolution of the $(\beta\alpha)_8$-barrel enzyme fold. *Chem. Rev. 105*, 4038−4055.

(5) Nagano, N., Orengo, C. A., and Thornton, J. M. (2002) One fold with many functions: The evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J. Mol. Biol. 321*, 741−765.

(6) Pujadas, G., and Palau, J. (1999) TIM barrel fold: Structural, functional and evolutionary characteristics in natural and designed molecules. *Biologica (Bratislava, Slovakia) 54*, 231−253.

(7) Raine, A. R., Scrutton, N. S., and Mathews, F. S. (1994) On the evolution of alternate core packing in eightfold $(\beta\alpha)$-barrels. *Protein Sci. 3*, 1889−1892.

(8) Reardon, D., and Farber, G. (1995) The structure and evolution of $\alpha\beta$-barrel proteins. *FASEB J. 9*, 497−503.

(9) Copley, R. R., and Bork, P. (2000) Homology among $(\beta\alpha)_8$-barrels: Implications for the evolution of metabolic pathways. *J. Mol. Biol. 303*, 627−641.

(10) Eder, J., and Kirschner, K. (1992) Stable substructures of eightfold $(\beta\alpha)_8$-barrel proteins: Fragment complementation of phosphoribosylanthranilate isomerase. *Biochemistry 31*, 3617−3625.

(11) Soberón, X., Fuentes-Gallego, P., and Saab-Rincón, G. (2004) In vivo fragment complementation of a $(\beta\alpha)_8$-barrel protein: Generation of variability by recombination. *FEBS Lett. 560*, 167−172.

(12) Bertolaet, B. L., and Knowles, J. R. (1995) Complementation of fragments of triosephosphate isomerase defined by exon boundaries. *Biochemistry 34*, 5736−5743.

(13) Gualfetti, P. J., Bilsel, O., and Matthews, C. R. (1999) The progressive development of structure and stability during the equilibrium folding of the α-subunit of tryptophan synthase from *Escherichia coli*. *Protein Sci. 8*, 1623−1635.

(14) Bilsel, O., Zitzewitz, J. A., Bowers, K. E., and Matthews, C. R. (1999) Folding mechanism of the α-subunit of tryptophan synthase, an $\beta\alpha$-barrel protein: Global analysis highlights the interconversion of multiple native, intermediate, and unfolded forms through parallel channels. *Biochemistry 38*, 1018−1029.

(15) Forsyth, W. R., and Matthews, C. R. (2002) 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. 320*, 1119−1133.

(16) Gu, Z., Zitzewitz, J. A., and Matthews, C. R. (2007) Mapping the structure of folding cores in TIM barrel proteins by hydrogen exchange mass spectrometry: The roles of motif and sequence for the indole-3-glycerol phosphate synthase from *Sulfolobus solfataricus*. *J. Mol. Biol. 368*, 582−594.

(17) Gu, Z., Rao, M. K., Forsyth, W. R., Finke, J. M., and Matthews, C. R. (2007) Structural analysis of kinetic folding intermediates for a TIM barrel protein, indole-3-glycerol phosphate synthase, by hydrogen exchange mass spectrometry and Gō model simulation. *J. Mol. Biol. 374*, 528−546.

(18) Forsyth, W. R., Bilsel, O., Gu, Z., and Matthews, C. R. (2007) Topology and sequence in the folding of a TIM barrel protein: Global analysis highlights partitioning between transient off-pathway and stable on-pathway folding intermediates in the complex folding mechanism of a $(\beta\alpha)_8$-barrel of unknown function from *B. subtilis*. *J. Mol. Biol. 372*, 236−253.

(19) Mukaiyama, A., Takano, K., Haruki, M., Morikawa, M., and Kanaya, S. (2004) Kinetically robust monomeric protein from a hyperthermophile. *Biochemistry 43*, 13859−13866.

(20) Perl, D., Welker, C., Schindler, T., Schroder, K., Marahiel, M., Janicke, R., and Schmid, F. X. (1998) Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. *Nat. Struct. Biol. 5*, 229−235.

(21) Dams, T., and Jaenicke, R. (1999) Stability and folding of dihydrofolate reductase from the hyperthermophilic bacterium *Thermotoga maritima*. *Biochemistry 38*, 9169−9178.

(22) Deutschman, W., and Dahlquist, F. (2001) Thermodynamic basis for the increased thermostability of CheY from the hyperthermophile *Thermotoga maritima*. *Biochemistry 40*, 13107−13113.

(23) Lang, D., Thoma, R., Henn-Sax, M., Sterner, R., and Wilmanns, M. (2000) Structural evidence for evolution of the $\beta\alpha$-barrel scaffold by gene duplication and fusion. *Science 289*, 1546−1550.

(24) List, F., Sterner, R., and Wilmanns, M. (2011) Related $(\beta\alpha)_8$-barrel proteins in histidine and tryptophan biosynthesis: A paradigm to study enzyme evolution. *ChemBioChem 12*, 1487−1494.

(25) Beismann-Driemeyer, S., and Sterner, R. (2001) Imidazole glycerol phosphate synthase from *Thermotoga maritima*: Quaternary structure, steady-state kinetics, and reaction mechanism of the bienzyme complex. *J. Biol. Chem. 276*, 20387−20396.

(26) Douangamath, A., Walker, M., Beismann-Driemeyer, S., Vega-Fernandez, M. C., Sterner, R., and Wilmanns, M. (2002) Structural evidence for ammonia tunneling across the $(\beta\alpha)_8$-barrel of the imidazole glycerol phosphate synthase bienzyme complex. *Structure 10*, 185−193.

(27) Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene 77*, 51−59.

(28) Thoma, R., Obmolova, G., Lang, D. A., Schwander, M., Jenö, P., Sterner, R., and Wilmanns, M. (1999) Efficient expression, purification and crystallisation of two hyperthermostable enzymes of histidine biosynthesis. *FEBS Lett. 454*, 1−6.

(29) Pace, C. N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol. 131*, 266−280.

(30) Santoro, M. M., and Bolen, D. W. (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl α-chymotrypsin using different denaturants. *Biochemistry 27*, 8063−8068.

(31) McDaniel, D. H., and Smoot, C. R. (1956) Approximations in the kinetics of consecutive reactions. *J. Phys. Chem. 60*, 966−969.
(32) Casado, J., González, J., and Moreno, M. (1987) Consecutive reactions in chemical kinetics: A method for the treatment of experimental data. *React. Kinet. Catal. Lett. 33*, 357−362.
(33) Myers, J. K., Pace, C. N., and Martin Scholtz, J. (1995) Denaturant *m*-values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci. 4*, 2138−2148.
(34) Silow, M., and Oliveberg, M. (1997) Transient aggregates in protein folding are easily mistaken for folding intermediates. *Proc. Natl. Acad. Sci. U.S.A. 94*, 6084−6086.
(35) Stryer, L. (1965) The interaction of a naphthalene dye with apomyoglobin and apohemoglobin. A fluorescent probe of non-polar binding sites. *J. Mol. Biol. 13*, 482−495.
(36) Jones, B. E., Jennings, P. A., Pierre, R. A., and Matthews, C. R. (1994) Development of nonpolar surfaces in the folding of *Escherichia coli* dihydrofolate reductase detected by 1-anilinonaphthalene-8-sulfonate binding. *Biochemistry 33*, 15250−15258.
(37) Lorenz, T., and Reinstein, J. (2008) The influence of proline isomerization and off-pathway intermediates on the folding mechanism of eukaryotic UMP/CMP kinase. *J. Mol. Biol. 381*, 443−455.
(38) Hammond, G. S. (1955) A correlation of reaction rates. *J. Am. Chem. Soc. 77*, 334−338.
(39) Oliveberg, M., Tan, Y.-J., Silow, M., and Fersht, A. R. (1998) The changing nature of the protein folding transition state: Implications for the shape of the free-energy profile for folding. *J. Mol. Biol. 277*, 933−943.
(40) Silow, M., and Oliveberg, M. (1997) High-energy channeling in protein folding. *Biochemistry 36*, 7633−7637.
(41) Zoldak, G., Carstensen, L., Scholz, C., and Schmid, F. X. (2009) Consequences of domain insertion on the stability and folding mechanism of a protein. *J. Mol. Biol. 386*, 1138−1152.
(42) Höcker, B., Beismann-Driemeyer, S., Hettwer, S., Lustig, A., and Sterner, R. (2001) Dissection of a $(\beta\alpha)_8$-barrel enzyme into two folded halves. *Nat. Struct. Biol. 8*, 32−36.
(43) Sterner, R., and Liebl, W. (2001) Thermophilic adaptation of proteins. *Crit. Rev. Biochem. Mol. Biol. 36*, 39−106.
(44) Bachmann, A., and Kiefhaber, T. (2001) Apparent two-state Tendamistat folding is a sequential process along a defined route. *J. Mol. Biol. 306*, 375−386.
(45) Ptitsyn, O. B. (1995) Molten globule and protein folding. *Adv. Protein Chem. 47*, 83−229.
(46) Wensley, B. G., Gärtner, M., Choo, W. X., Batey, S., and Clarke, J. (2009) Different members of a simple three-helix bundle protein family have very different folding rate constants and fold by different mechanisms. *J. Mol. Biol. 390*, 1074−1085.
(47) Friel, C. T., Capaldi, A. P., and Radford, S. E. (2003) Structural analysis of the rate-limiting transition states in the folding of lm7 and lm9: Similarities and differences in the folding of homologous proteins. *J. Mol. Biol. 326*, 293−305.
(48) Olofsson, M., Hansson, S., Hedberg, L., Logan, D. T., and Oliveberg, M. (2007) Folding of S6 structures with divergent amino acid composition: Pathway flexibility within partly overlapping foldons. *J. Mol. Biol. 365*, 237−248.
(49) Lappalainen, I., Hurley, M. G., and Clarke, J. (2008) Plasticity within the obligatory folding nucleus of an immunoglobulin-like domain. *J. Mol. Biol. 375*, 547−559.
(50) Lam, A. R., Borreguero, J. M., Ding, F., Dokholyan, N. V., Buldyrev, S. V., Stanley, H. E., and Shakhnovich, E. (2007) Parallel folding pathways in the SH3 domain protein. *J. Mol. Biol. 373*, 1348−1360.
(51) Wu, Y., Vadrevu, R., Kathuria, S., Yang, X., and Matthews, C. R. (2007) A tightly packed hydrophobic cluster directs the formation of an off-pathway sub-millisecond folding intermediate in the α subunit of tryptophan synthase, a TIM barrel protein. *J. Mol. Biol. 366*, 1624−1638.
(52) Rudolph, R., Siebendritt, R., and Kiefhaber, T. (1992) Reversible unfolding and refolding behavior of a monomeric aldolase from *Staphylococcus aureus*. *Protein Sci. 1*, 654−666.
(53) Jasanoff, A., Davis, B., and Fersht, A. R. (1994) Detection of an intermediate in the folding of the $(\beta\alpha)_8$-barrel N-(5′-phosphoribosyl)-anthranilate isomerase from *Escherichia coli*. *Biochemistry 33*, 6350−6355.
(54) Kathuria, S. V., Day, I. J., Wallace, L. A., and Matthews, C. R. (2008) Kinetic traps in the folding of $\beta\alpha$-repeat proteins: CheY initially misfolds before accessing the native conformation. *J. Mol. Biol. 382*, 467−484.
(55) Hills, R. D., Jr., Kathuria, S. V., Wallace, L. A., Day, I. J., Brooks, C. L., III, and Matthews, C. R. (2010) Topological frustration in $\beta\alpha$-repeat proteins: Sequence diversity modulates the conserved folding mechanisms of $\alpha/\beta/\alpha$ sandwich proteins. *J. Mol. Biol. 398*, 332−350.
(56) Höcker, B., Schmidt, S., and Sterner, R. (2002) A common evolutionary origin of two elementary enzyme folds. *FEBS Lett. 510*, 133−135.
(57) Bharat, T. A. M., Eisenbeis, S., Zeth, K., and Höcker, B. (2008) A $\beta\alpha$-barrel built by the combination of fragments from different folds. *Proc. Natl. Acad. Sci. U.S.A. 105*, 9942−9947.
(58) Wu, Y., and Matthews, C. R. (2002) Parallel channels and rate-limiting steps in complex protein folding reactions: Prolyl isomerization and the α-subunit of Trp synthase, a TIM barrel protein. *J. Mol. Biol. 323*, 309−325.
(59) Wu, Y., and Matthews, C. R. (2002) A *cis*-prolyl peptide bond isomerization dominates the folding of the α-subunit of trp synthase, a TIM barrel protein. *J. Mol. Biol. 322*, 7−13.
(60) Wu, Y., and Matthews, C. R. (2003) Proline replacements and the simplification of the complex, parallel channel folding mechanism for the α-subunit of Trp synthase, a TIM barrel protein. *J. Mol. Biol. 330*, 1131−1144.
(61) Saab-Rincon, G., Froebe, C. L., and Matthews, C. R. (1993) Urea-induced unfolding of the α-subunit of tryptophan synthase: One-dimensional proton NMR evidence for residual structure near histidine-92 at high denaturant concentration. *Biochemistry 32*, 13981−13990.
(62) Wilmanns, M., Hyde, C. C., Davies, D. R., Kirschner, K., and Jansonius, J. N. (1991) Structural conservation in parallel $(\beta\alpha)$-barrel enzymes that catalyze three sequential reactions in the pathway of tryptophan biosynthesis. *Biochemistry 30*, 9161−9169.
(63) Rojsajjakul, T., Wintrode, P., Vadrevu, R., Robert Matthews, C., and Smith, D. L. (2004) 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. 341*, 241−253.

Supporting Information for Publication A:

Figures S1-S6; Table S1.

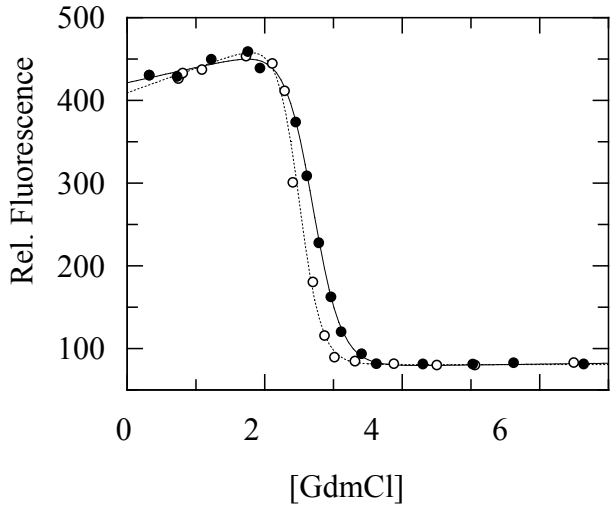


Figure S1. GdmCl-induced equilibrium unfolding and refolding transitions of HisF at 45 °C after an incubation time of 10 days. Closed symbols represent unfolding experiments started with folded protein, and open symbols represent refolding experiments started with protein that was previously unfolded in 6.0 M GdmCl. The transitions were followed by Trp/Tyr fluorescence (excitation at 280 nm; emission at 320 nm) in 50 mM Tris/HCl buffer (pH 7.5). The discrepancy between the traces is caused by the fact that unfolding has not yet reach equilibrium.

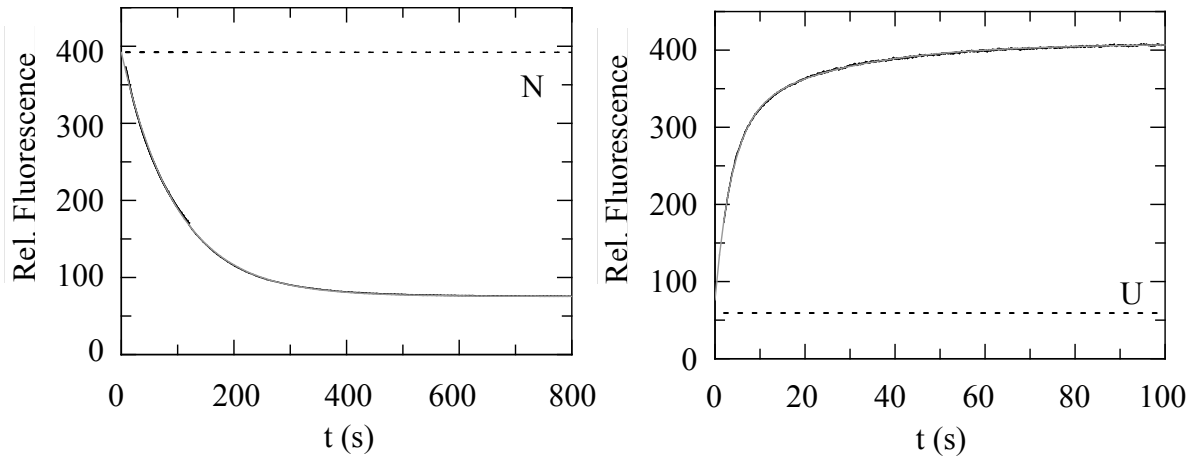


Figure S2. (a) Unfolding kinetics of HisF at 6.1 M GdmCl and (b) refolding kinetics at 0.96 M GdmCl. The traces of the Trp/Tyr fluorescence signal (excitation at 280 nm; emission at 320 nm) were monitored after manual mixing in 50 mM Tris/HCl buffer (pH 7.5) at 25 °C. Gray solid lines represent fits of mono-exponential and bi-exponential functions to the unfolding ($\tau = 98$ s) and refolding curves ($\tau_1 = 3.7$ s and $\tau_2 = 24$ s), respectively. Dashed lines indicate the initial fluorescence values of the native (N) and unfolded (U) state.

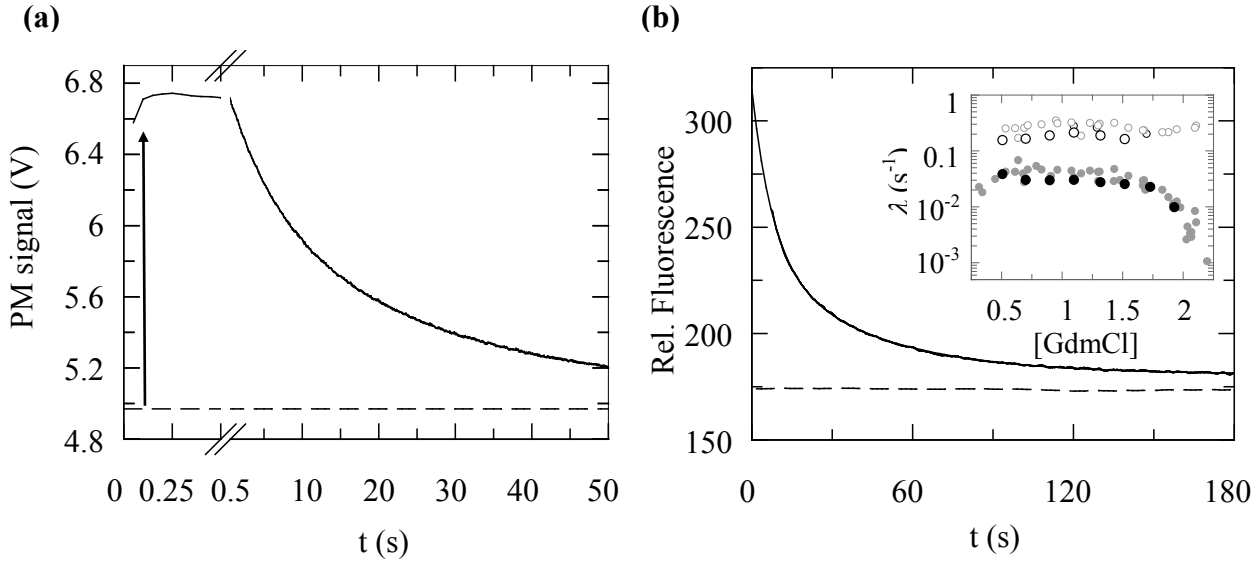


Figure S3. Refolding kinetics of HisF followed by the fluorescence signal of bound ANS. Unfolded HisF was refolded in 0.6 M GdmCl in the presence of 400 µM ANS. The ANS fluorescence in the absence of HisF is indicated as dashed lines. (a) Fluorescence traces (excitation at 350 nm emission at > 395 nm) after stopped-flow mixing. The rapid increase of the signal (arrow) within the dead time of mixing (5 ms) demonstrates binding of ANS to hydrophobic regions of the burst-phase intermediate I_{BP}. (b) Fluorescence traces (excitation at 350 nm; emission at 479 nm) after manual-mixing. The decrease of the signal demonstrates that ANS is displaced as folding proceeds to species with less hydrophobic surfaces. Inset: rate constants of folding of HisF determined as observed by ANS fluorescence (black circles) and by CD and Trp/Tyr fluorescence (gray circles, data taken from **Figure 5a**).

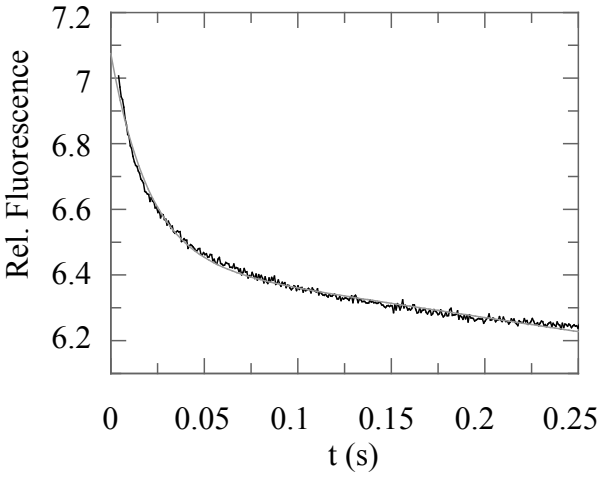


Figure S4. Analysis of the putative folding intermediate I of HisF by interrupted refolding experiments. Unfolded protein was refolded in 1.0 M GdmCl for 10 s to allow for the conversion of U to I. Then the protein was diluted to 6.0 M GdmCl, and the decay of Trp/Tyr fluorescence (excitation at 280 nm; emission at > 320 nm) was followed. The gray solid line represents a mono-exponential function and a linear term fitted to the experimental curve.

With a time constant of 0.02 s the unfolding reaction is three orders of magnitude faster than unfolding of N (**Figure 5a**).

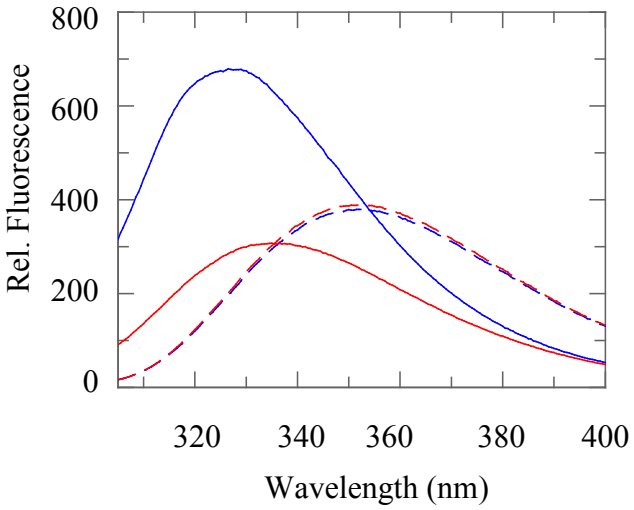


Figure S5. Fluorescence emission spectra of wild-type W156-HisF (blue) and W35-HisF (red). Spectra were recorded in 50 mM Tris/HCl buffer (pH 7.5) in absence (solid line) and presence of 6.0 M GdmCl (broken line). The fluorescence emission maxima of the native proteins (maxima: 327 and 335 nm) are red-shifted upon unfolding (maxima: 352 and 351 nm).

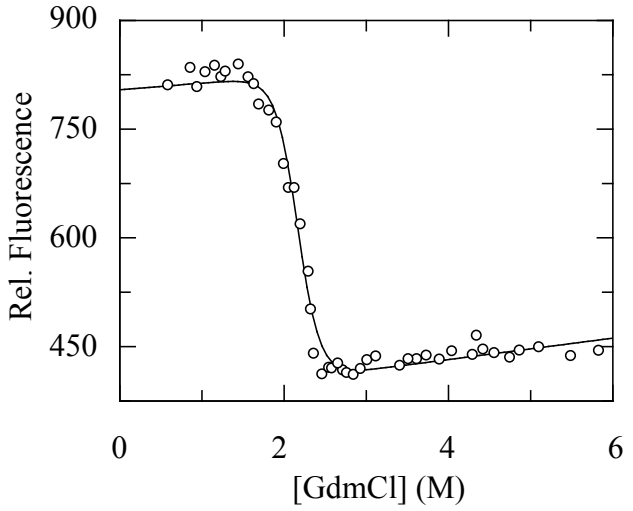


Figure S6. Equilibrium refolding transition of W35-HisF at 45 °C. The transition was followed by Tyr/Trp fluorescence (excitation at 280 nm; emission at 320 nm) in 50 mM Tris/HCl buffer (pH 7.5). The continuous line represents a fit of the two-state model to the data, yielding the following thermodynamic parameters: m = 19.7 kJ mol^{-1} M^{-1}, $\Delta G_D°$ = 42.6 kJ mol^{-1}. The corresponding values for wild-type HisF are m = 18.0 kJ mol^{-1} M^{-1} and $\Delta G_D°$ = 47.9 kJ mol^{-1} (see **Table 1**).

Table S1. Rate constants and relative amplitudes of refolding kinetics at 0.6 M GdmCl followed by fluorescence of single Trp residues at different positions in HisF.

	λ_1 (s-1)	λ_2 (s^{-1})	A_1	A_2	Position
W156-HisF	0.20	-	-1.31	0	α5
L35W-HisF	0.46	0.026	-0.28	1.97	α1
F49W-HisF	0.19	0.015	-0.52	1.68	β2
F77W-HisF	0.15	0.059	-0,26	-2.01	β3
F210W-HisF	0.33	0.057	-0,10	-0.11	α7
F214W-HisF	0.23	0.027	0.55	0.38	α7
F240W-HisF	0.17	0.015	0.16	0.34	α8

Wild-type HisF contains the single Trp156 (W156-HisF). The variants contain the indicated single Trp residues (**Figure 11a**), together with the W156Y exchange. Kinetic traces were monitored by fluorescence above 320 nm after excitation at 290 nm in 50 mM Tris/HCl buffer (pH 7.5) at 25 °C.

Publication B

Linn Carstensen, Josef Sperl, Felix List, Marco Bocola, Franz-Xaver Schmid and Reinhard Sterner.

Conservation of the folding mechanism between a natural $(\beta\alpha)_8$-barrel protein and putative evolutionary precursors generated by protein design.

Submitted to J. Am. Chem. Soc.

DOI: 10.1021/ja304951v

Conservation of the folding mechanism between a natural $(\beta\alpha)_8$-barrel protein and putative evolutionary precursors generated by protein design

Linn Carstensen[a], Josef M. Sperl[a], Marco Bocola[a], Felix List[a], Franz X. Schmid[b] and Reinhard Sterner[a,*]

[a]Universität Regensburg, Institut für Biophysik und physikalische Biochemie, D-93040 Regensburg, Germany.
[b]Laboratorium für Biochemie und Bayreuther Zentrum für Molekulare Biowissenschaften, Universität Bayreuth, D-95440 Bayreuth, Germany.

*Corresponding author
Reinhard Sterner: Phone: +49-941-943 3015; Fax: +49-941-943 2813.
Email: Reinhard.Sterner@biologie.uni-regensburg.de

Abstract

In evolution, $(\beta\alpha)_8$-barrel proteins presumably arose from the fusion of $(\beta\alpha)_4$-half-barrels. We constructed potential primordial $(\beta\alpha)_8$-barrel proteins 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, and apparently it remained conserved throughout the evolution of this 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 (1). 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 (2-5). 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 (6). 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 (7-9). The canonical $(\beta\alpha)_8$-barrel contains at least 200 amino acids and is composed of eight modules. Each module consists of a β-strand and an α-helix linked by a βα-loop; the individual modules are connected by αβ-loops. The eight strands form a closed parallel β-sheet, the barrel, which is surrounded by an outer layer of eight α-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 β-strands and in the subsequent βα-loops. This conserved location of the active site suggests that the contemporary $(\beta\alpha)_8$-barrels might have evolved from of a common ancestor (10-13). Moreover, the modular structure of the fold points to an evolutionary precursor that has emerged by the duplication and fusion of individual βα-entities. In fact, the $(\beta\alpha)_8$-barrel enzymes *N*′-[(5′-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide isomerase (HisA) and imidazole glycerol phosphate synthase (HisF) display a weak four-fold and a strong two-fold internal symmetry, suggesting that they evolved by the stepwise duplication and fusion of $(\beta\alpha)_2$-quarter and $(\beta\alpha)_4$-half-barrels (14-18). 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 (17) is renamed here as Sym1, and its two half-barrels are designated as HisF-C_N and HisF-C_C (**Figure 1**). Crystal structure analysis showed that the $(\beta\alpha)_8$-barrels of Sym1 and wild-type HisF are highly

similar (17, 19). 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**).

We have now stabilized Sym1 by replacing βα-module 8 from HisF-C with the stable βα-module 4 stemming from HisF-N, the N-terminal half of HisF (**Figure 1**). We solved the crystal structure of the resulting Sym2 protein (**Figures 2**) and compared it in its unfolding and refolding kinetics with Sym1 and the wild-type protein (20). 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 towards the native state. We conclude that this folding mechanism is an ancient property of $(\beta\alpha)_8$-barrel proteins.

Results and Discussion

Design of Sym2 and structural comparison with HisF and Sym1

In the crystal structure of Sym1, the βα-modules 8 at the C-termini of HisF-C_N and HisF-C_C are not well defined. We replaced these flexible and putatively labile regions by the stable βα-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 Cα-atoms of the three proteins, which show sequence identities between 65 % (Sym1-HisF; Sym2-HisF) and 90 % (Sym1-Sym2) superpose with an r.m.s.d. 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 β-barrel of HisF is also present in Sym1 and Sym2 (**Figure S2**).

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 fluorescence, the loss of secondary structure 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 (20) and Sym2 (**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 (ΔG_D). Sym2 is significantly more stable, Sym1 significantly less stable than HisF. The ΔG_D and *m* values derived from the CD-monitored unfolding transitions (**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.

Sym1, HisF, and Sym2 form burst-phase refolding intermediates

Folding and unfolding kinetics were followed by tryptophan and tyrosine fluorescence and by far-UV circular dichroism (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 fivefold 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 (**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 together, 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 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 βα module 8 in Sym1 with βα module 4 in Sym2 (**Figure 1, 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 $(\beta\alpha)_4$-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 τ of about 3 s is observed for all three proteins (**Figure 4**). This reaction is followed by a slower one, again with a common τ 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' behaviour is indicative of the formation of folding intermediates (21). 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 (**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 (20). For Sym2, the amplitudes of the three consecutive folding phases also change in a reciprocal manner with denaturant concentration (**Figure S6C**). This suggests that the folding of this stable artificial protein also occurs sequentially 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 towards 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 (20). 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 s 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 s or 7 s of refolding was mono- or biphasic, respectively, as shown for 2.9 M GdmCl in **Figure S7.** After 0.5 s of refolding, a single species, the intermediate I', was formed, which unfolded very rapidly (τ = 0.02 s). After 7 s of refolding, in addition to the 0.02-s unfolding reaction a second, 20-fold slower unfolding reaction (τ = 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**).

In order 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 s and 0.48 s (**Figure 5A, 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 (**Figure 5B**). 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. The time lag in the formation of I is emphasized in the expanded view in **Figure 5C**. 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. 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 (**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 together, our findings suggest a common folding mechanism for Sym1, HisF, and Sym2 (**Figure 6**). All three proteins form a burst-phase intermediate I_{BP} that contains a significant amount of secondary structure but is most likely off-pathway (20). 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 (22). 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 2M 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.

Conclusions

In evolution the present-day $(\beta\alpha)_8$-barrel proteins might have arisen by the duplication and subsequent combination of two $(\beta\alpha)_4$-half-barrels (15-18). 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 (20, 23, 24). 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 N 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 (2-6). These findings suggest that a synchronized and ordered sequential folding mechanism with a high energy barrier arose early and remained conserved during the evolution and diversification of $(\beta\alpha)_8$-barrel proteins.

Materials and Methods

Expression of genes and purification of recombinant proteins

Cloning of the genes *his*F and *sym*1 (formerly *his*F-C***C) into pET24a(+) was described previously (17, 20). The *sym*2 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(25). Sym1 and Sym2 were purified as described for Sym1 by metal chelate affinity chromatography using the C-terminal His_6-tag(17). 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 µM protein were measured at 25 °C in 50 mM Tris/HCl buffer (pH 7.5) containing different concentrations of GdmCl, basically as described(20). GdmCl (ultrapure) was purchased from MP Biomedicals (Illkirch, France), and its concentration was determined by the refractive index of the solution (26). 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 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 stopped-flow SX.20MV spectrometer from Applied Photophysics (Leatherhead, UK).

In order to reach equilibrium, HisF and Sym2 were pre-incubated 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 (27). The rates of conventional unfolding and refolding kinetics were determined by fitting mono-exponential 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 μM in 6 M GdmCl) was diluted 6-fold to 1 M GdmCl, and incubated for 0.5 s 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 → I' → I → N reaction was monitored in further stopped-flow fluorescence interrupted refolding experiments. Unfolded Sym2 (132 μM 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→I'→I→N) was fitted to the obtained amplitudes using DynaFit (BioKin Ltd.) (28).

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 (29).

Acknowledgements. 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 JN & Wolynes PG (2004) Theory of protein folding. *Curr Opin Struct Biol* 14:70-75.
2. Sadqi M, de Alba E, Pérez-Jiménez R, Sanchez-Ruiz JM, & Muñoz V (2009) A designed protein as experimental model of primordial folding. *Proc Natl Acad Sci USA* 106:4127-4132.
3. Gillespie B*, et al.* (2003) NMR and temperature-jump measurements of de novo designed proteins demonstrate rapid folding in the absence of explicit selection for kinetics. *J Mol Biol* 330:813-819.
4. Scalley-Kim M & Baker D (2004) Characterization of the folding energy landscapes of computer generated proteins suggests high folding free energy barriers and cooperativity may be consequences of natural selection. *J Mol Biol* 338:573-583.
5. Watters AL*, et al.* (2007) The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection. *Cell* 128:613-624.
6. Sanchez-Ruiz JM (2010) Protein kinetic stability. *Biophys Chem* 148:1-15.
7. Caetano-Anollés G, Kim HS, & Mittenthal JE (2007) The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture. *Proc Natl Acad Sci USA* 104:9358-9363.
8. Wierenga RK (2001) The TIM-barrel fold: A versatile framework for efficient enzymes. *FEBS Lett* 492:193-198.
9. Sterner R & Höcker B (2005) Catalytic versatility, stability, and evolution of the $(\beta\alpha)_8$-barrel enzyme fold. *Chem Rev* 105:4038-4055.
10. Copley RR & Bork P (2000) Homology among $(\beta\alpha)_8$-barrels: Implications for the evolution of metabolic pathways. *J Mol Biol* 303:627-641.
11. Nagano N, Orengo CA, & Thornton JM (2002) One fold with many functions: The evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J Mol Biol* 321:741-765.
12. Reardon D & Farber G (1995) The structure and evolution of αβ-barrel proteins. *FASEB J* 9:497-503.
13. Pujadas G & Palau J (1999) TIM barrel fold: Structural, functional and evolutionary characteristics in natural and designed molecules. *Biologica (Bratislava)* 54:231-253.
14. Richter M*, et al.* (2010) Computational and experimental evidence for the evolution of a $(\beta\alpha)_8$-barrel protein from an ancestral quarter-barrel stabilised by disulfide bonds. *J Mol Biol* 398:763-773.
15. Höcker B, Claren J, & Sterner R (2004) Mimicking enzyme evolution by generating new $(\beta\alpha)_8$-barrels from $(\beta\alpha)_4$-half-barrels. *Proc Natl Acad Sci USA* 101:16448-16453.
16. Seitz T, Bocola M, Claren J, & Sterner R (2007) Stabilisation of a $(\beta\alpha)_8$-barrel protein designed from identical half barrels. *J Mol Biol* 372:114-129.
17. Höcker B, Lochner A, Seitz T, Claren J, & Sterner R (2009) High-resolution crystal structure of an artificial $(\beta\alpha)_8$-barrel protein designed from identical half-barrels. *Biochemistry* 48:1145-1147.
18. List F, Sterner, R. and Wilmanns, M. (2011) Related $(\beta\alpha)_8$-barrel proteins in histidine and tryptophan biosynthesis: A paradigm to study enzyme evolution. *ChemBioChem* 12:1487–1494.

19. Lang D, Thoma R, Henn-Sax M, Sterner R, & Wilmanns M (2000) Structural evidence for evolution of the (βα)-barrel scaffold by gene duplication and fusion. *Science* 289:1546-1550.
20. Carstensen L, Zoldak G, Schmid FX, & Sterner R (2012) Folding mechanism of an extremely thermostable $(\beta\alpha)_8$-barrel enzyme: A high kinetic barrier protects the protein from denaturation. *Biochemistry* DOI: 10.1021/bi300189f.
21. Lorenz T & Reinstein J (2008) The influence of proline isomerization and off-pathway intermediates on the folding mechanism of eukaryotic UMP/CMP kinase. *J Mol Biol* 381:443-455.
22. Bachmann A & Kiefhaber T (2001) Apparent two-state tendamistat folding is a sequential process along a defined route. *J Mol Biol* 306:375-386.
23. Forsyth WR, Bilsel O, Gu Z, & Matthews CR (2007) Topology and sequence in the folding of a TIM barrel protein: global analysis highlights partitioning between transient off-pathway and stable on-pathway folding intermediates in the complex folding mechanism of a $(\beta\alpha)_8$-barrel of unknown function from *B. subtilis*. *J Mol Biol* 372:236-253.
24. Forsyth WR & Matthews CR (2002) 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* 320:1119-1133.
25. Thoma R*, et al.* (1999) Efficient expression, purification and crystallisation of two hyperthermostable enzymes of histidine biosynthesis. *FEBS Lett* 454:1-6.
26. Pace CN (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol* 131:266-280.
27. Santoro MM & Bolen DW (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl α-chymotrypsin using different denaturants. *Biochemistry* 27:8063-8068.
28. Kuzmič P (1996) Program DYNAFIT for the Analysis of Enzyme Kinetic Data: Application to HIV Proteinase. *Anal Biochem* 237:260-273.
29. Russell RB & Barton GJ (1992) Multiple protein sequence alignment from tertiary structure comparison: Assignment of global and residue confidence levels. *Proteins* 14:309-323.

Figures and tables with legends

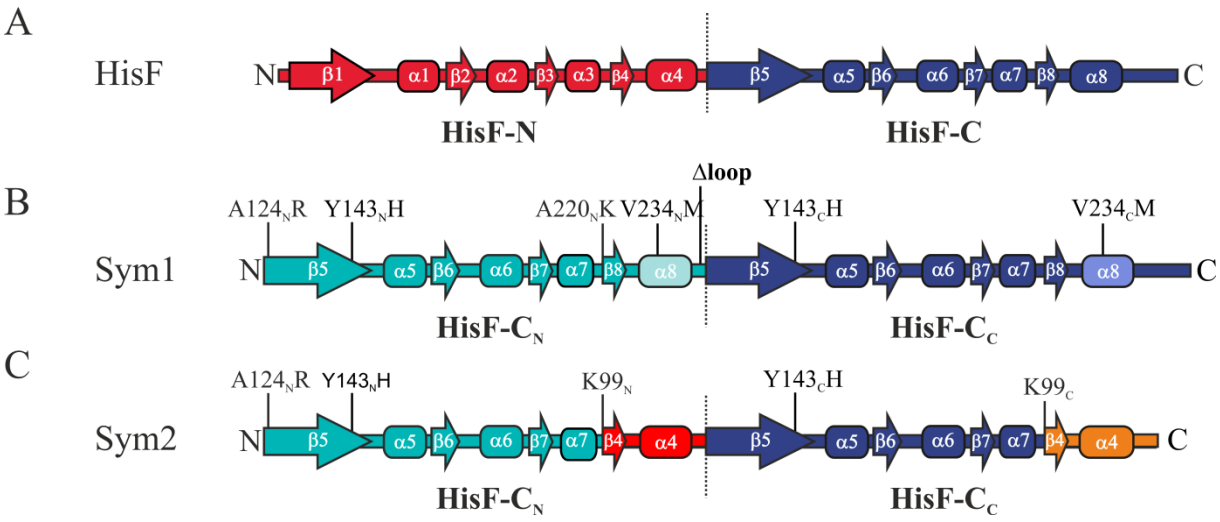


Figure 1: Design of the artificial $(\beta\alpha)_8$-barrel proteins Sym1 and Sym2 from two identical and fused half-barrels. (A) Secondary structure elements of wild-type HisF. The N-terminal half-barrel HisF-N [modules $(\beta\alpha)_{1-4}$] is shown in red, the C-terminal half-barrel HisF-C [modules $(\beta\alpha)_{5-8}$] 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 (15-17). Transparent boxes indicate the parts that were not resolved in the crystal structure. (C) Design of Sym2. Sym2 was derived from Sym1 by replacing βα module 8 with βα module 4 in both half-barrels (red and orange).

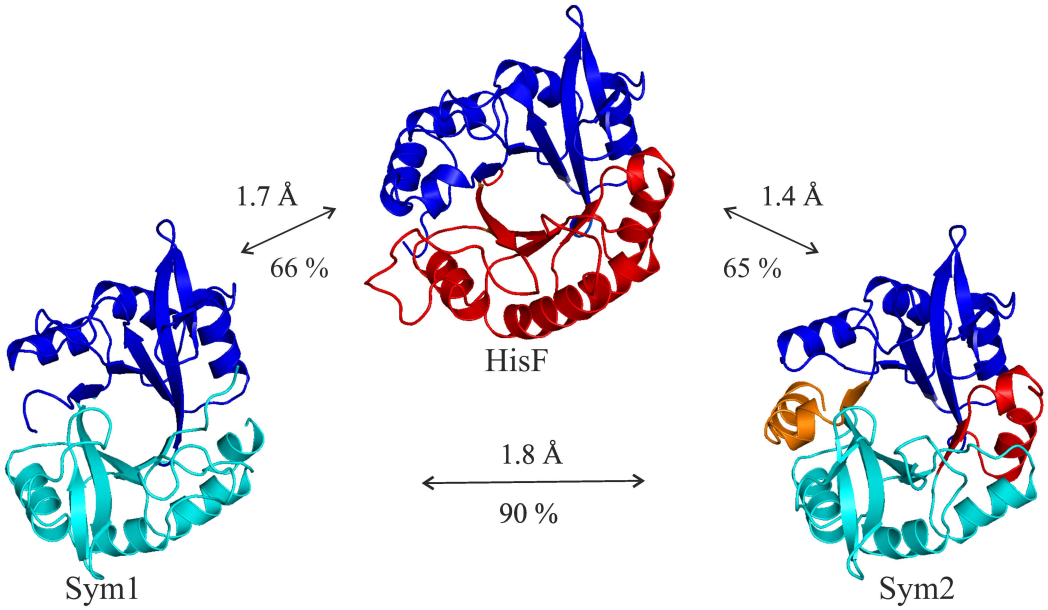


Figure 2. Ribbon diagrams of the crystal structures of Sym1 (2w6r.pdb) (17), HisF (1thf.pdb) (19) and Sym2 (3og3.pdb). The invisible parts in the crystal structure of Sym1 are fully resolved in Sym2. Colors represent the origin and location of secondary structure elements as described in **Figure 1**. Amino acid sequence identities (%) and r.m.s.d. values (Å) of corresponding Cα-atoms as deduced from pair-wise structure-based sequence alignments are indicated (HisF-Sym1: 184 superimposed Cα atoms; HisF-Sym2: 198 superimposed Cα atoms; Sym1-Sym2: 191 superimposed Cα atoms).

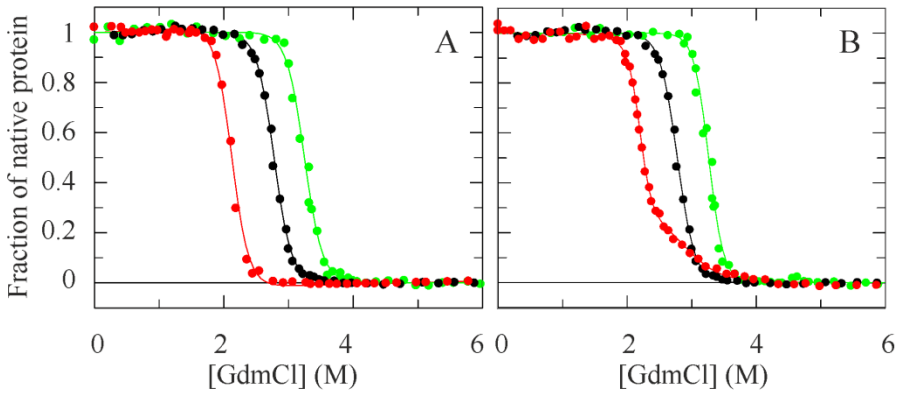


Figure 3. Equilibrium unfolding transitions of Sym1 (red), HisF (black), and Sym2 (green). GdmCl-induced unfolding of 4 µM protein was followed in 50 mM Tris/HCl buffer (pH 7.5) at 25 °C by (A) Tyr/Trp fluorescence (excitation: 280 nm; emission: 320 nm) and (B) far-UV CD at 225 nm. The continuous lines 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 **Table S1**.

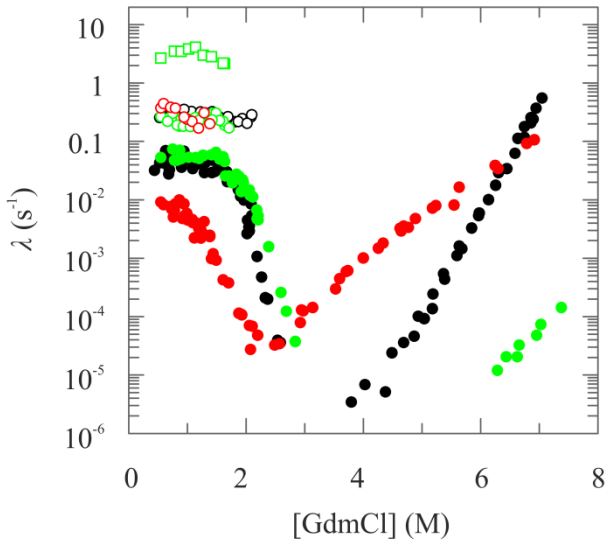


Figure 4. Folding kinetics of Sym 1 (red), HisF (black), and Sym2 (green). The dependence on GdmCl concentration of the apparent rate constant (λ) is shown for the slow unfolding and refolding phases (filled circles), for the fast refolding phase (open circles), and for the additional very rapid refolding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

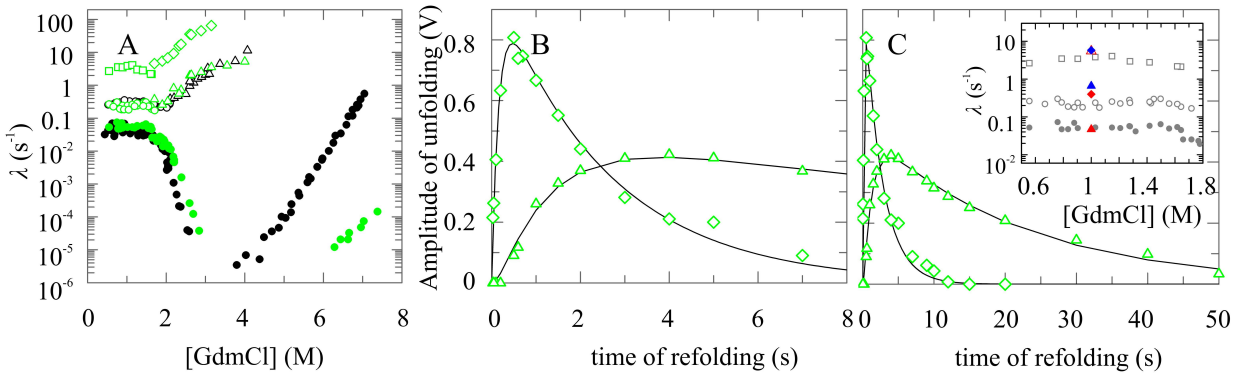


Figure 5. Kinetic intermediates during the folding of HisF (black) and Sym2 (green). (A) Chevron diagram. Symbols for rate constants determined by single-mixing experiments were taken from **Figure 4**. The unfolding limb of the on-pathway intermediate I of HisF (black triangles) was determined by interrupted refolding experiments as described (20). 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, τ = 56 ms) and slow (green triangle, τ = 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→I'→I→N). Inset to (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': τ = 0.17 s; red diamond: rate of depletion of I': τ = 2.5 s; red open triangle: lag in the formation of I: τ = 0.18 s; blue triangle: rate of formation of I: τ = 1.5 s; red filled triangle: rate of depletion of I: τ = 21.3 s) are compared with the refolding rates of Sym2 determined by conventional refolding kinetics (gray symbols, data taken from **Figure 4**).

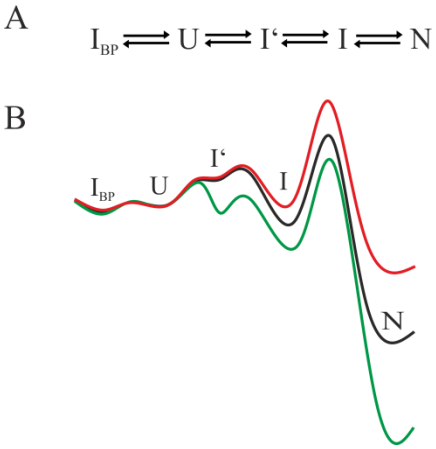


Figure 6. (A) Unifying folding mechanism for Sym1, HisF and Sym2. An off-pathway equilibrium between the unfolded state (U) and the burst-phase intermediate (I_{BP}) precedes the formation of the on-pathway intermediates I' and I. For HisF and Sym1, I' is assumed to be a high energy intermediate. (B) 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.

Table 1. Thermodynamic stability parameters for HisF, Sym1, and Sym2

	N ↔ U			I_{BP} ↔ U		
	ΔG_D (kJ mol^{-1})	m (kJ mol^{-1} M^{-1})	$[D]_{1/2}$ (M)	ΔG_D (kJ mol^{-1})	m (kJ mol^{-1} M^{-1})	$[D]_{1/2}$ (M)
HisF	53.5 ± 1.7	19.1 ± 0.6	2.8	15.2 ± 2.4	7.9 ± 0.9	1.9
Sym1	42.8 ± 0.2	20.2 ± 0.4	2.1	13.2 ± 0.4	5.9 ± 0.2	2.2
Sym2	62.2 ± 3.9	19.2 ± 1.2	3.2	35.7 ± 6.5	21.4 ± 3.7	1.7

For the N ↔ U transition, the Gibbs free energy of denaturation (ΔG_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_{BP} ↔ 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 **Figure S4**).

Supporting Information for Publication B:

SI Methods, SI References, Figures S1-S8; Table S1-S2.

SI Methods

Construction and cloning of the *sym*2 gene

Sym2 was constructed from Sym1 by replacing the labile C-termini of HisF-C_N and HisF-C_C with the stable C-terminus of HisF-N. The structural superposition of HisF with Sym1 revealed a high similarity between the parts linking βα module 3 and βα module 4 in HisF and the parts linking βα module 7 and βα module 8 in Sym1. This finding determined the exact intersection point for the construction of Sym2, which contains βα module 4 instead of βα module 8, both in HisF-C_N and HisF-C_C. For construction of the *sym2* gene, its 5'-half sequence *sym2-N* was amplified by PCR using the isolated *hisF-*C_N fragment of *sym1* (formerly denoted as *hisF*-C***C) fragment as template. The oligonucleotide 5'-TAA TAC GAC TCA CTA TAG GG-3' was used as 5'-primer, and the oligonucleotide 5'-CTG GGA TCC GAA GGT CTG TGC GAT TTG TGT GAT GAG GCT CGG GTT TTC GAC AGC GGC AGT ATT GAT AGA GAC CTT GTC AGC ACC TGC CAG GAA G-3'with a *BamH*I-site at the 3'-terminus (underlined) was used as 3'-primer. The sequence coding for βα module 4 was optimized with respect to codon usage (Gene designer) and template annealing. The amplified fragment *sym2-N* was cloned into pET24a(+) using *NdeI* and *BamHI* restriction sites. Next *sym2-C* was amplified using *sym2-N* as template. The oligonucleotide 5'- GGC GGA GGA TCC CAG GCC GTT GTC GTG GCG ATA GAT G-3' with a *BamHI*-site at the 5'-terminus (underlined) was used as 5'-primer, and the oligonucleotide 5'- GTG CTC GAG ACT CCC GAA GGT CTG TGC GAT TTG -3' with an *Xho*I-site at the 3'-terminus (underlined) was used as 3'-primer. The amplified fragment *sym2-C* was cloned into pET24a(+)-*sym-N* using *BamH*I and *Xho*I restriction sites, yielding pET24a(+)-*sym2*.

Crystallization and structure determination of Sym2

Initial crystallization trials were carried out using the sitting-drop vapor-diffusion method in 96-well plates (Greiner) at 292 K at the EMBL Hamburg High-throughput Crystallization Facility (1). Crystals were obtained under various conditions. Further optimization of these conditions was performed manually in 24-well plates (Qiagen) using the hanging-drop vapor-diffusion method at 292 K. Drops contained 1 µl of a Sym2 protein solution (10 mg/ml)

mixed with 2 µl of 0.2 M ammonium sulfate, 18% w/v PEG 4000, and were equilibrated against 1 ml of reservoir buffer. Crystals appeared after a few days. Single crystals were cryo-protected with 25% (v/v) glycerol before they were flash-cooled directly in the cryo stream. A data set was measured from a single crystal at 123 K on an Oxford Diffraction Gemini R Ultra diffractometer with a cryojet using $Cu_{K\alpha}$ radiation (λ = 1.54178 Å) equipped with a CCD detector KM4CCD/sapphire. The crystal to detector distance was 55 mm, a total of 332 images with an oscillation range of 0.75 ° were collected to a maximum resolution of 2 Å. The data were integrated with CrysalisPro (Oxford diffraction) and scaled with Scala (2). The recorded images were indexed, processed and integrated using the CrysAlisPro software package (Oxford Diffraction Ltd., Version 1.171.34.9). The data were scaled with Scala from the CCP4 program suite 6.1.1 (3). Empirical absorption correction was performed using spherical harmonics, implemented in SCALE3 ABSPACK scaling algorithm. The initial phases were obtained by molecular replacement using a half-barrel fragment from the HisF-structure (1thf.pdb) as search model employing PHASER software from the CCP4 program suite 6.1.1 (3). The structure was refined by alternating rounds of Refmac5 refinement and manual model building in COOT (4, 5).

SI References

1. Mueller-Dieckmann J (2006) The open-access high-throughput crystallization facility at EMBL Hamburg. *Acta Crystallogr D Biol Crystallogr* 62:1446-1452.
2. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62:72-82.
3. Collaborative Computational Project N (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50:760-763.
4. Murshudov GN, Vagin AA, & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53:240-255.
5. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66:486-501.
6. Lang D, Thoma R, Henn-Sax M, Sterner R, & Wilmanns M (2000) Structural evidence for evolution of the ($\beta\alpha$)-barrel scaffold by gene duplication and fusion. *Science* 289:1546-1550.
7. Höcker B, Lochner A, Seitz T, Claren J, & Sterner R (2009) High-resolution crystal structure of an artificial $(ba)_8$-barrel protein designed from identical half-barrels. *Biochemistry* 48:1145-1147.

SI Figures

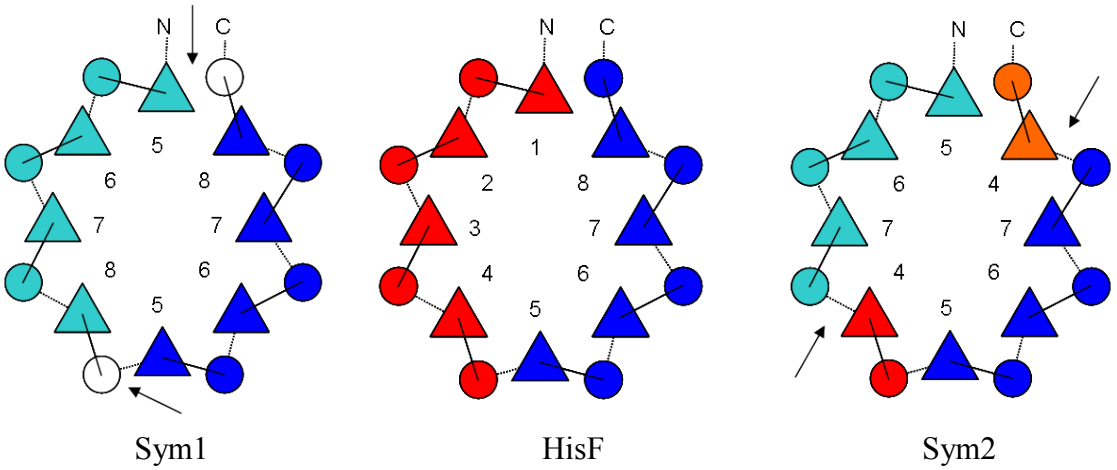


Figure S1: Schematic representation of the βα-composition of Sym1, HisF, and Sym2. Triangles and circles represent β-strands and α-helices, numbered according to HisF. αβ–loops are depicted as solid lines, and βα–loops are depicted as dotted lines. Colors represent the origin and location of secondary structure elements and are as in **Figure 1** (red: HisF-N and HisF-N_N, orange: HisF-N_C, blue: HisF-C and HisF-C_C, aquamarin: HisF-C_N). The arrows indicate artificial contacts between secondary structure elements, and the transparent circles show invisible parts of the structure of Sym1 due to high B-factors.

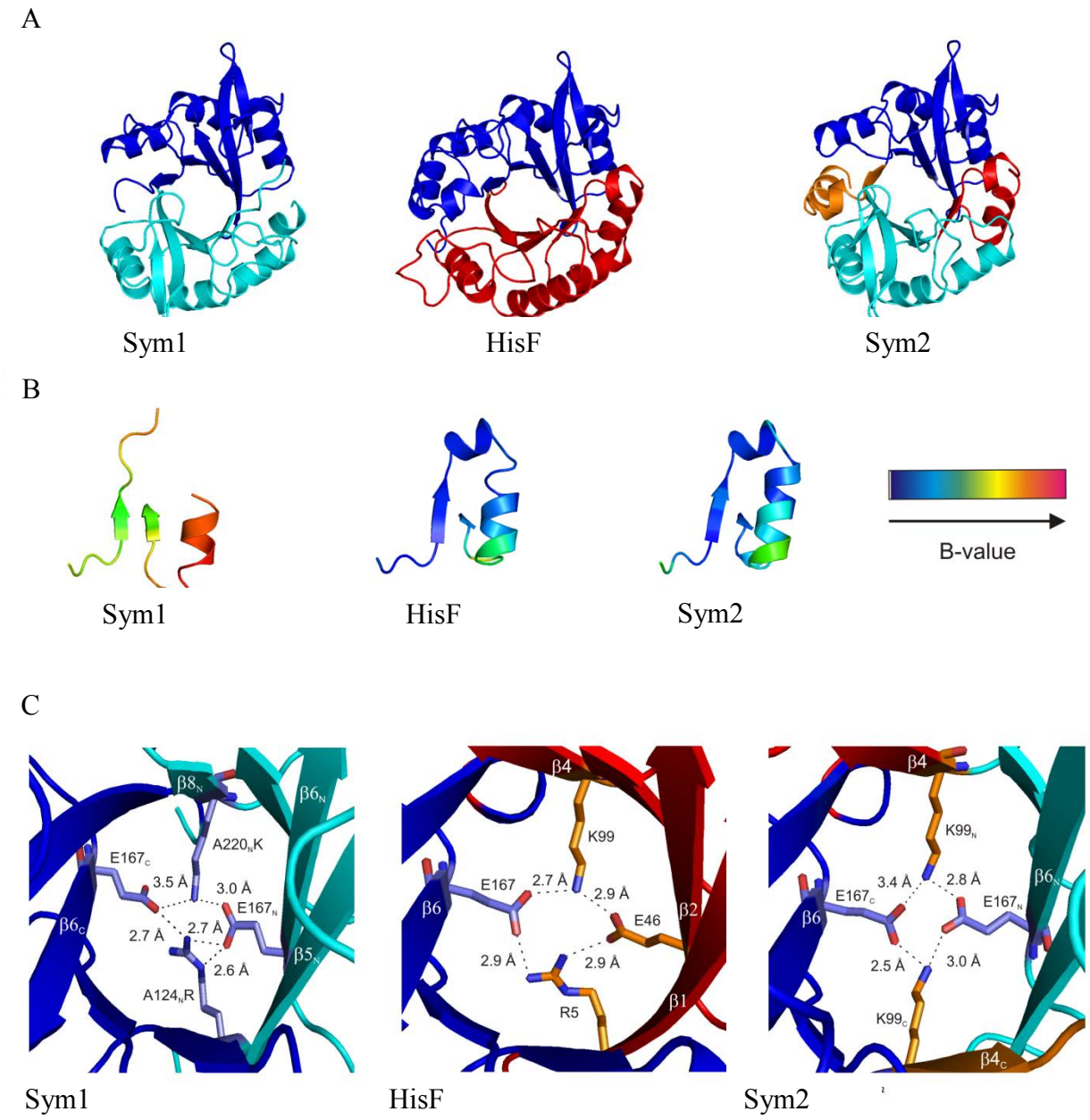


Figure S2: Important structural features of Sym2 in comparison to HisF and Sym1. (A) Ribbon diagrams of the crystal structures of Sym2 (3og3.pdb), HisF (1thf.pdb) (6), and Sym1 (2w6r.pdb) (7): Top view onto the catalytic face. Colors are as in **Figure 2**. (B) The partially unresolved regions in the crystal structure of Sym1 are visible in the crystal structure of Sym2 and have a high similarity to HisF. (C) The stabilizing salt bridge cluster at the N-terminal face of the central β-barrel of HisF is formed in a similar way in Sym1 through the exchanges $A124_NR$ and $A220_NK$. The salt bridge cluster in Sym2 is highly symmetrical through the introduction of βα-module 4 (residues $K99_N$ and $K99_C$). Side chains located in the N- and C-terminal halves are colored orange and blue, respectively. Hydrogen bonds between the involved side chain residues, which are shown as sticks, are indicated by dashed lines up to a distance cutoff of 3.5 Å.

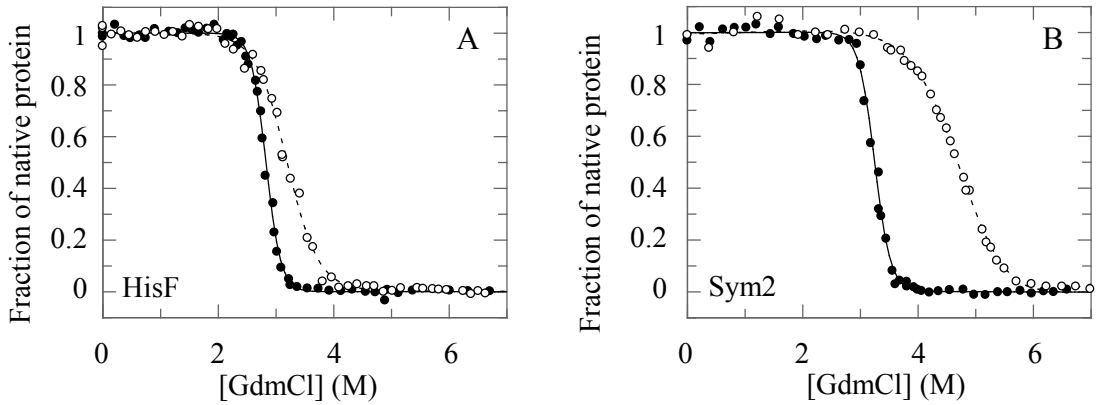


Figure S3. GdmCl-induced equilibrium unfolding transitions of (A) HisF and (B) Sym2. The transitions of 4 µM protein in 50 mM Tris/HCl buffer (pH 7.5) at 25 °C were followed by Tyr/Trp fluorescence (excitation: 280 nm; emission: 320 nm). Open and closed symbols represent unfolding experiments after an incubation time of two weeks and the time period required to reach equilibrium, respectively. Lines connecting the data points were drawn to guide the eye. After a pre-incubation time of 10 and 20 days at 45 °C, respectively, the samples of HisF and Sym2 had to be incubated for three weeks at 25 °C to reach equilibrium. (Unfolding of Sym1 reaches equilibrium after three days of incubation at 25 °C.)

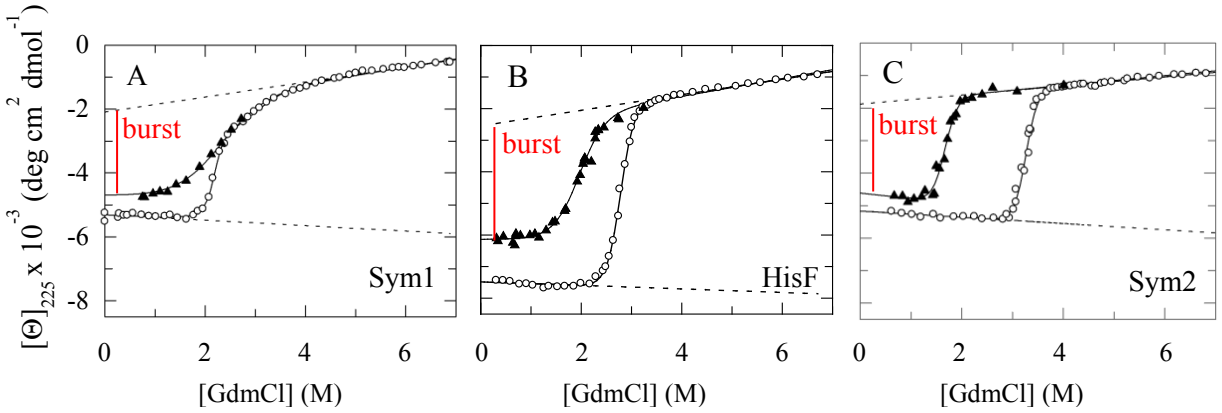


Figure S4. Equilibrium unfolding transitions of (A) Sym1, (B) HisF, and (C) Sym2 (open circles) in comparison with the unfolding transitions of their burst-phase intermediates (I_{BP}) (filled triangles). GdmCl-induced unfolding of the three proteins was followed in 50 mM Tris/HCl buffer (pH 7.5) at 25 °C by the far-CD signal at 225 nm. The unfolding transition of I_{BP} was calculated by plotting the initial CD signal upon refolding, which was obtained by extrapolation of the kinetics to zero time, as a function of the corresponding GdmCl concentration. The dashed lines indicate the signal of the pure N and U states. Continuous lines represent the fit of the data with the two-state model, yielding the thermodynamic parameters listed in **Table 1**.

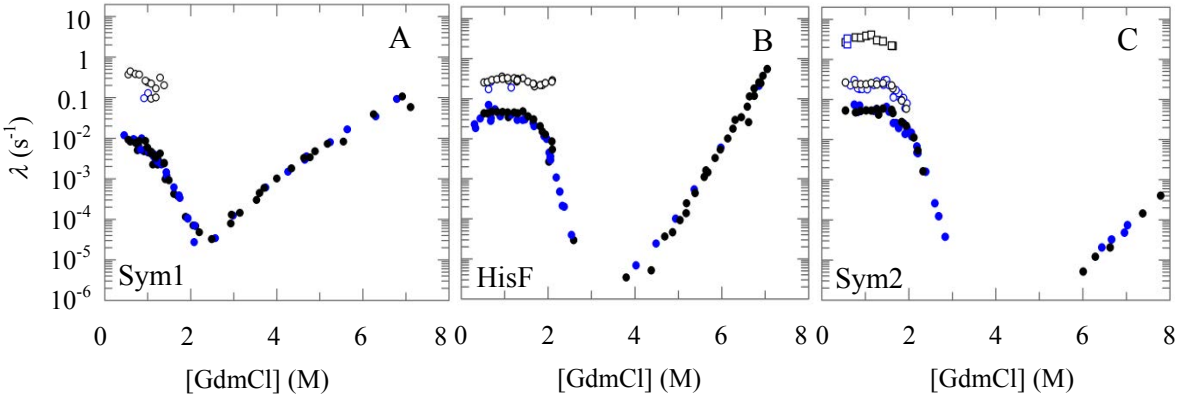


Figure S5. GdmCl-dependence of the apparent rate constants (λ) for refolding and unfolding of (A) Sym1, (B) HisF, and (C) Sym2 (chevron diagrams). Following manual mixing, the kinetics were monitored by the CD signal at 225 nm (blue circles) or fluorescence emission at 320 nm after excitation at 280 nm (black circles). Fast reactions were also followed after stopped-flow mixing. Rate constants of the slow unfolding and refolding phases are shown as filled circles, and rate constants of the fast refolding phase are shown as open circles. Rate constants of the additional very rapid refolding phases of Sym2 are shown as squares. A comparison of the rate constants of the three proteins is shown in **Figure 4**.

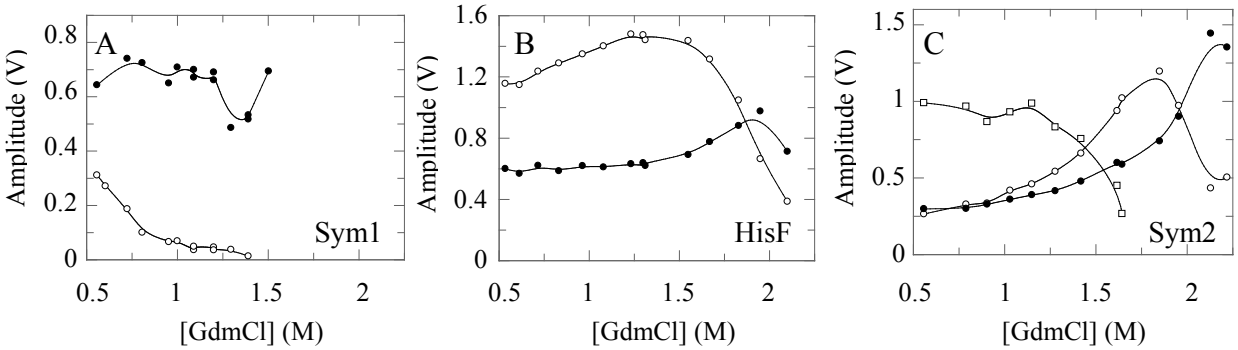


Figure S6. Amplitudes of the very rapid (squares), fast (open circles) and slow (closed circles) refolding phases of (A) Sym1, (B) HisF and (C) Sym2. Reactions were followed in 50 mM Tris/HCl buffer (pH 7.5) at 25 °C by stopped-flow mixing monitoring Trp/Tyr fluorescence (excitation: 280 nm; emission: 320 nm). The corresponding rate constants are shown in **Figure 4**. Lines connecting the data points were drawn to guide the eye.

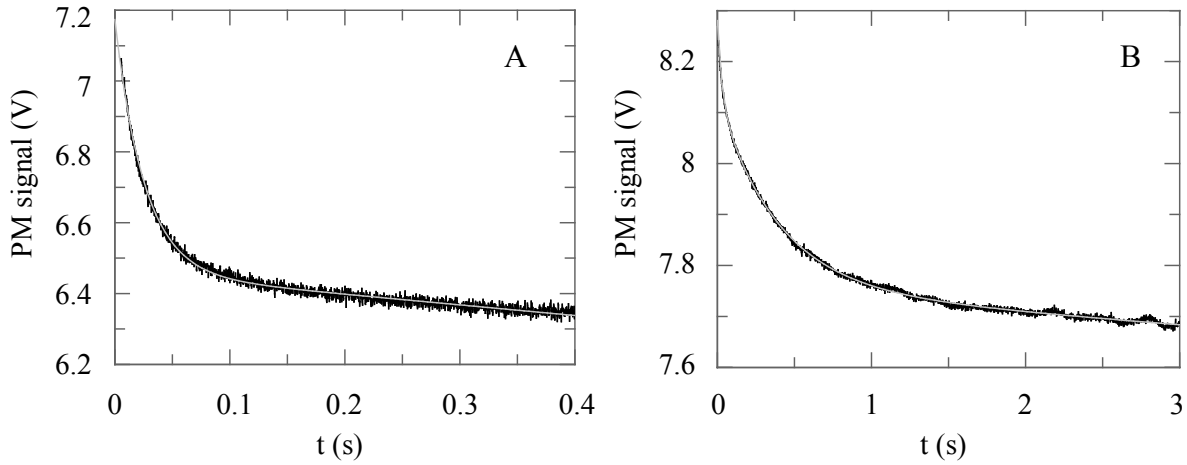


Figure S7. Unfolding kinetics of the putative folding intermediates I' and I of Sym2 after interrupted refolding. Unfolded protein was refolded in 1.0 M GdmCl for (A) 0.5 s and (B) 7 s to allow for the conversion of U to I' and I, respectively. Then the protein was diluted to 2.9 M GdmCl, and unfolding was followed by the decrease of Trp/Tyr fluorescence (excitation at 280 nm; emission at > 320 nm) was followed. At this denaturant concentration, native Sym2 molecules remain folded (**Figure 3**), and the unfolding reaction originates exclusively from folding intermediates. The gray solid lines represent (A) mono-exponential or (B) double-exponential equations including a linear term fitted to the experimental curves, yielding time constants for unfolding of I' (τ = 0.02 s) and I (τ = 0.39 s). These rates correspond to the unfolding rates of I and I' at 2.9 M GdmCl which are depicted in **Figure 5A**.

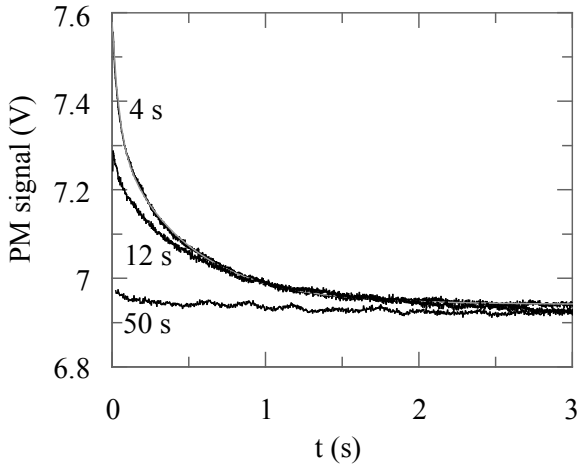


Figure S8. Detection of the folding intermediates I' and I of Sym2 by interrupted refolding experiments. Refolding of unfolded protein was initiated by dilution to 1.0 M GdmCl. After various times (shown are 4 s, 12 s, 50 s), the protein was diluted to 2.6 M GdmCl, and unfolding was monitored by the decrease of the Trp/Tyr fluorescence (excitation: 280 nm; emission: > 320 nm). A double exponential function and a linear term were fitted to the experimental curves. In the presence of 2.6 M GdmCl, native Sym2 molecules remain folded but the intermediates I' and I unfold with time constants of 0.05 s and 0.48 s, respectively. The amplitudes of the two phases were determined and used to calculate the accumulation and depletion of I' and I in the course of refolding as shown in **Figure 5B, C**.

SI Tables

Table S1. Thermodynamic parameters of Sym2, HisF, and Sym1 determined by far-UV CD.

	N↔U			I_{BP}↔U		
	ΔG_D (kJ mol^{-1})	m (kJ mol^{-1} M^{-1})	$[D]_{1/2}$ (M)	ΔG_D (kJ mol^{-1})	m (kJ mol^{-1} M^{-1})	$[D]_{1/2}$ (M)
HisF	55.6 ± 2.3	19.9 ± 0.9	2.8			
Sym1	45.0 ± 3.5	19.6 ± 1.6	2.3	8.0 ± 2.0	3.3 ± 0.9	2.4
Sym2	75.2 ± 4.7	23.0 ± 1.4	3.3			

The parameters were obtained by the two-state (HisF and Sym2) or three-state (Sym1) analysis of the unfolding transitions monitored by CD signal at 225 nm (**Figure 3B**).

Table S2. Data collection and refinement statistics

	Sym2 (3og3.pdb)
Data collection	
Space group	C 1 2 1
Cell dimensions	
a, *b*, *c* (Å)	31.37, 82.09, 74.51
α, β, γ (°)	90, 90.11, 90
Resolution (Å)	2.08 (2.08-2.133)
R_{sym} or R_{merge}	0.052 (0.168)
I / σI	18.58 (3.8)
Completeness (%)	99.6 (98.38)
Refinement	
Resolution (Å)	2.08
No. reflections	10719 (791)
R_{work} / R_{free}	0.152 (0.139) / 0.196 (0.190)
No. atoms	
Protein	2021
Water	173
B-factors	17.48
R.m.s. deviations	
Bond lengths (Å)	0.025
Bond angles (°)	2.006

Values in parentheses are for highest-resolution shell.

Acknowledgement

I would like to gratefully thank my supervisor Prof. Dr. Reinhard Sterner, who has been always there to listen and give advice. In addition to his guidance and encouragement, he provided me with an excellent atmosphere for doing research and his confidence gave me the opportunity to develop self-sufficiency.

I am deeply grateful to Prof. Dr. Franz Schmid for his expertise and support, in particular, the long discussions that helped me to evaluate details of my work. I am also thankful to him for carefully reading and improving our manuscripts.

I am grateful to Dr. Gabriel Zoldák for his encouragement, expertise and practical advice. I am also thankful to him for reading my reports, commenting on my views and helping me understand and enrich my ideas.

I would also like to thank all the current and former colleagues and friends at University of Regensburg, for their support and for some much needed humor and entertainment. In particular, I would like to thank Christiane Endres for her assistance and my former students Max Plach, Stefan Oehm and Josef Sperl for their work. Thanks also go to Dr. Jochen Reinstein for performing the CD Stopped-flow measurements and Dr. Christoph Liebold, who did the NMR experiments.
Additionally, I would like to thank the members of my former office Dr. Sandra Schlee, Dr. Thomas Schwab, Dr. Susanne Dietrich, Josef Sperl and Dr. Heike Stüer for discussions, as well as Jan-Oliver Janda and Dietmar Birzer for the current accommodation.

Finally, I would like to thank my friends and, most importantly, my parents for their faith in me and providing me with unending encouragement and support.