doi:10.1016/j.jmb.2007.06.036 J. Mol. Biol. (2007) 372, 114–129

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Stabilisation of a $(\beta\alpha)_8$-Barrel Protein Designed from Identical Half Barrels

Tobias Seitz, Marco Bocola, Jörg Claren and Reinhard Sterner*

Institute of Biophysics and Physical Biochemistry
University of Regensburg
Universitätsstrasse 31
D-93053 Regensburg, Germany

It has been suggested that the common $(\beta\alpha)_8$-barrel enzyme fold has evolved by the duplication and fusion of identical $(\beta\alpha)_4$-half barrels, followed by the optimisation of their interface. In our attempts to reconstruct these events *in vitro* we have previously linked in tandem two copies of the C-terminal half barrel HisF-C of imidazole glycerol phosphate synthase from *Thermotoga maritima* and subsequently reconstituted in the fusion construct HisF-CC a salt bridge cluster present in wild-type HisF. The resulting recombinant protein HisF-C*C, which was produced in an insoluble form and unfolded with low cooperativity at moderate urea concentrations has now been stabilised and solubilised by a combination of random mutagenesis and selection *in vivo*. For this purpose, *Escherichia coli* cells were transformed with a plasmid-based gene library encoding HisF-C*C variants fused to chloramphenicol acetyltransferase (CAT). Stable and soluble variants were identified by the survival of host cells on solid medium containing high concentrations of the antibiotic. The selected HisF-C*C proteins, which were characterised *in vitro* in the absence of CAT, contained eight different amino acid substitutions. One of the exchanges (Y143C) stabilised HisF-C*C by the formation of an intermolecular disulfide bond. Three of the substitutions (G245R, V248M, L250Q) were located in the long loop connecting the two HisF-C copies, whose subsequent truncation from 13 to 5 residues yielded the stabilised variant HisF-C*C Δ. From the remaining substitutions, Y143H and V234M were most beneficial, and molecular dynamics simulations suggest that they strengthen the interactions between the half barrels by establishing a hydrogen-bonding network and an extensive hydrophobic cluster, respectively. By combining the loop deletion of HisF-C*C Δ with the Y143H and V234M substitutions, the variant HisF-C**C was generated. Recombinant HisF-C**C is produced in soluble form, forms a pure monomer with its tryptophan residues shielded from solvent and unfolds with similar cooperativity as HisF. Our results show that, starting from two identical and fused half barrels, few amino acid exchanges are sufficient to generate a highly stable and compact $(\beta\alpha)_8$-barrel protein with wild-type like structural properties.

Keywords: protein design; enzyme evolution; $(\beta\alpha)_8$-barrel; half barrel; stability screen

*Corresponding author

Introduction

It has been suggested that complex contemporary protein folds have evolved by the association and fusion of small polypeptide fragments.[1,2] The internal symmetry observed in a number of enzymes suggests that identical polypeptide segments first assembled to homo-oligomers, followed by the covalent linkage of the fragments as a consequence of gene duplication and fusion.[3–5] The $(\beta\alpha)_8$ (or

Abbreviations used: AdoMet, *S*-adenosyl-L-methionine; CAT, chloramphenicol acetyltransferase; MD, molecular dynamics; GdmCl, guanidinium chloride.

E-mail address of the corresponding author: Reinhard.Sterner@biologie.uni-regensburg.de

TIM) barrel is the most common enzyme fold, which currently contains about 400 identified members listed in the SCOP database†, and therefore is ideally suited for the study of enzyme evolution.[6–11] The canonical $(\beta\alpha)_8$-barrel consists of eight $(\beta\alpha)$ units, each of which comprises a β-strand and a subsequent α-helix which are connected by a βα-loop; the individual units are linked by αβ-loops.[12–14] The eight β-strands form a central parallel β-sheet (the barrel), which is surrounded by the eight α-helices. Within the central β-barrel, the residues are forming four layers, which are ordered perpendicular to the barrel axis and are numbered from the C- to the N-terminal barrel face. Each layer consists of residues from the odd (1, 3, 5, 7) or the even (2, 4, 6, 8) numbered strands, which generates a 4-fold symmetry. The active site of all known $(\beta\alpha)_8$-barrel enzymes is harboured by the C-terminal ends of the β-strands and by the βα-loops.

The apparent modularity of the $(\beta\alpha)_8$-barrel suggests that it has evolved by the duplication, fusion and recombination of fragments, which comprised either a single primordial $(\beta\alpha)$ element, or an even numbered multiple thereof. This hypothesis has been supported by protein fragmentation and folding studies.[11] For example, the observation that many $(\beta\alpha)_8$-barrels contain a conserved Gly-X-Asp motif in the loops α1β2, α3β4, α5β6, and α7β8 but not in the loops α2β3, α4β5, and α6β7 implies that $(\beta\alpha)_2$-units might constitute the smallest evolving entity.[10] Along these lines, the members of the *S*-adenosyl-L-methionine (AdoMet) radical protein family have three putative architectures, which comprise four, six and eight $(\beta\alpha)$-units.[15] The four or six N-terminal units contain the elements for radical generation, whereas the remainder of the protein is crucial for substrate binding. These findings suggest that AdoMet proteins might have evolved by the stepwise fusion of $(\beta\alpha)_2$ units, starting from a $(\beta\alpha)_4$-half barrel to a $(\beta\alpha)_6$-three-quarter to a $(\beta\alpha)_8$-full barrel structure. Along the same lines, fragmentation studies with the α-subunit of tryptophan synthase are compatible with a 6 (4+2) + 2 folding mechanism.[16,17] Moreover, the phosphoribosyl anthranilate isomerase (PRAI) from yeast forms a stable fragment consisting of the N-terminal $(\beta\alpha)_{1\text{-}6}$ units, which associates with the unstructured $(\beta\alpha)_{7–8}$ unit to a functional complex *in vitro*.[18] However, out of three fragment combinations of PRAI from *Escherichia coli* co-expressed *in vivo*, only the non-covalent complex between $(\beta\alpha)_{1\text{-}4}$ and $(\beta\alpha)_{5\text{-}8}$ yielded a functional enzyme.[19] This finding is in favour of a "4+4" folding and evolution mechanism, which has also been supported by studies on triosephosphate isomerase from chicken and rabbit muscle.[20,21]

Although the mechanistic details appear to be more complex than suggested by such a simple model,[22,23] the analysis of the enzymes *N*′-[(5′-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide isomerase (HisA) and imidazole glycerol phosphate synthase (HisF) has provided ample evidence for the crucial role of $(\beta\alpha)_4$-half barrels in the folding and evolution of $(\beta\alpha)_8$-barrel enzymes.[11] The amino acid sequences and X-ray structures of HisA and HisF from *Thermotoga maritima*, which catalyse two successive transformations of similar bisphosphorylated substrates within histidine biosynthesis, show an internal 2-fold symmetry (Figure 1(a)).[24,25] The pairs of N-terminal halves (HisA-N and HisF-N) that consist of the first four units ($(\beta\alpha)_{1–4}$) and the pairs of the C-terminal halves (HisA-C and HisF-C) that consist of the last four units ($(\beta\alpha)_{5–8}$) show sequence identities between 16 and 26% and root-mean-square (rms) deviations of their main-chain non-hydrogen bonds between only 1.4 Å and 2.1 Å, respectively. When produced separately, HisF-N and HisF-C form folded and mainly homo-dimeric proteins, which upon co-expression *in vivo* or joint refolding *in vitro* assemble to the non-covalent hetero-dimeric HisF–NC complex that displays wild-type catalytic activity.[9] Remarkably, four of the five $(\beta\alpha)$ units of several members of the flavodoxin-like fold family show striking similarities to the four $(\beta\alpha)$ units of HisF-N and HisF-C.[26] Moreover, the catalytically crucial N-terminal $(\beta\alpha)_4$-half barrels of eukaryotic and prokaryotic phosphoinositide-specific phospholipases C (PI-PLC) superimpose with rms deviations of 1.85 Å for 104 equivalent C^{α}-atoms.[27] Taken together, these findings suggest that $(\beta\alpha)_4$-half barrels and presumably also $(\beta\alpha)_2$-quarter barrels are independently evolving units. This conclusion would be further strengthened if stable and functional $(\beta\alpha)_8$-barrels could be generated by the fusion and recombination of existing half barrels.

Along these lines, we recently generated and characterised the stable chimeric $(\beta\alpha)_8$-barrels HisAF and HisFA, in which the N-terminal half of HisA was linked to the C-terminal half of HisF, and *vice versa*.[28] Moreover, we duplicated and fused in tandem the *his*F-C gene, yielding the HisF-CC protein (Figure 1(b)). Using rational protein design, the interface between the two identical halves in HisF-CC was stabilised by reconstituting a salt bridge cluster, which is conserved in all known HisF proteins. Although the resulting construct HisF-C*C containing the mutations A124R and A220K (Figure 1(b)) is a predominantly monomeric and a more compact protein than HisF-CC, it was produced in *E. coli* as inclusion bodies, had a slight tendency to aggregate, and differed in its spectroscopic properties from wild-type HisF.[28] To further increase the stability and solubility of HisF-C*C, we have now applied random mutagenesis followed by selection *in vivo* using a chloramphenicol acetyltransferase (CAT) reporter system.[29] The combination of several beneficial mutations identified by this approach allowed us to generate the HisF-C**C protein (Figure 1(c)), which can be produced in soluble form in *E. coli*, forms a homogeneous and compact monomer in gel filtration experiments, shows an identical fluorescence spectrum as wild-

† scop.mrc-lmb.cam.ac.uk/scop/

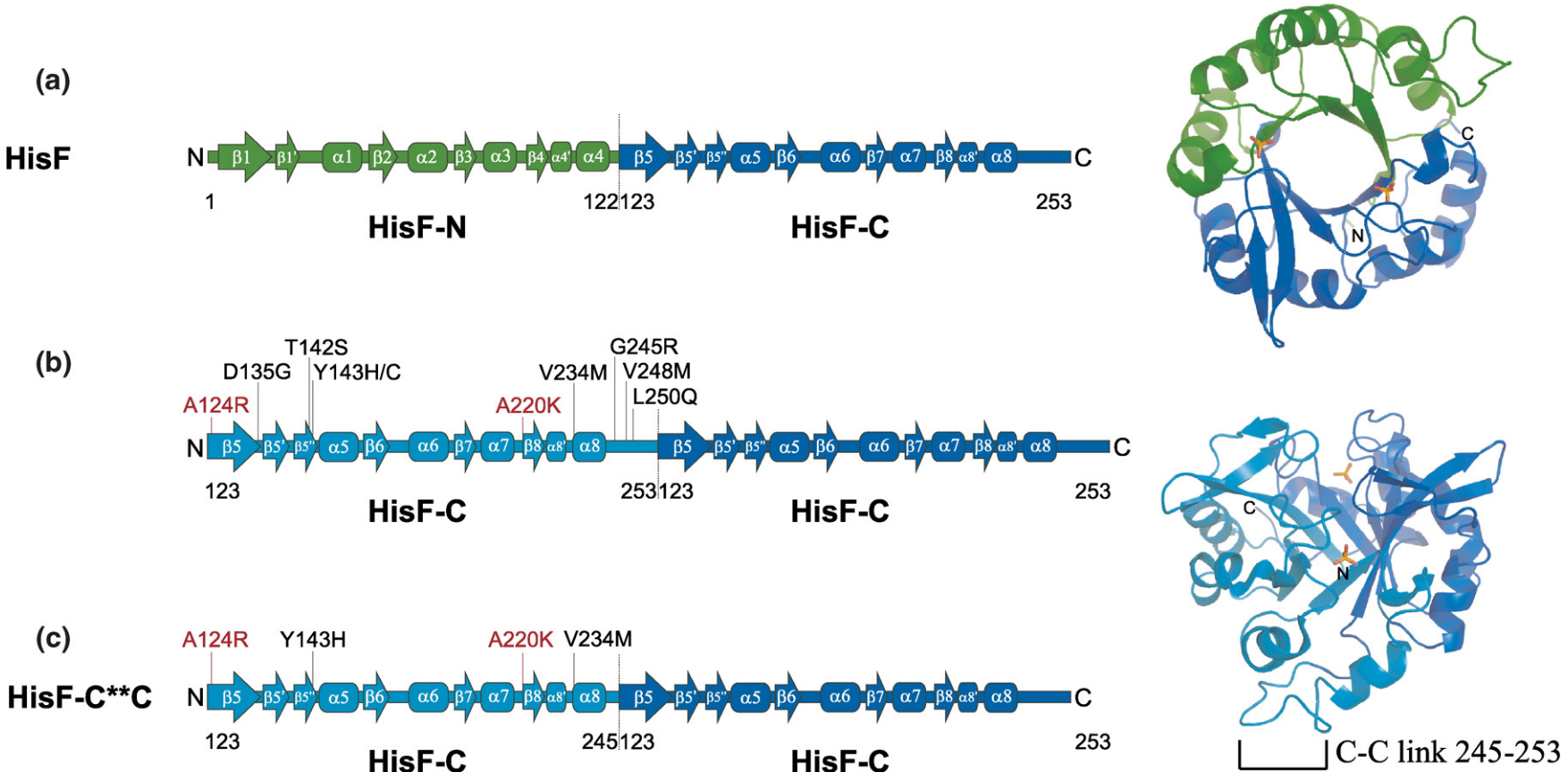

Figure 1. Design of the $(\beta\alpha)_8$-barrel protein HisF-C**C from two identical and fused HisF-C half barrels. (a) Secondary structure elements and tertiary structure (PDB code 1thf) of wild-type HisF (ribbon diagram showing a view onto the C-terminal face of the central β-barrel). The N-terminal half barrel HisF-N[$(\beta\alpha)_{1-4}$] is shown in green, and the C-terminal half barrel HisF-C[$(\beta\alpha)_{5-8}$] is shown in blue. The two phosphate ions bound to HisF-N and HisF-C are depicted as tetrahedrons. The N- and C-termini of the polypeptide chain are marked. (b) The fusion of two HisF-C units (the N-terminal one depicted in cyan, the C-terminal depicted in blue) yielded HisF-CC. The two exchanges A124R and A220K leading to the reconstitution of the native salt bridge cluster in HisF-C*C are shown in red,[28] and the exchanges introduced by random mutagenesis and selected *in vivo* in this work are shown in black. A model of the HisF-C*C scaffold (ribbon diagram showing a side view onto the central β-barrel) with the long loop linking the two HisF-C units being highlighted is shown on the right. The two phosphate ions bound to the N and C-terminal half barrels are depicted as tetrahedrons. The N- and C-termini of the polypeptide chain are marked. (c) The optimised variant HisF-C**C contains a shortened loop and four amino acid substitutions.

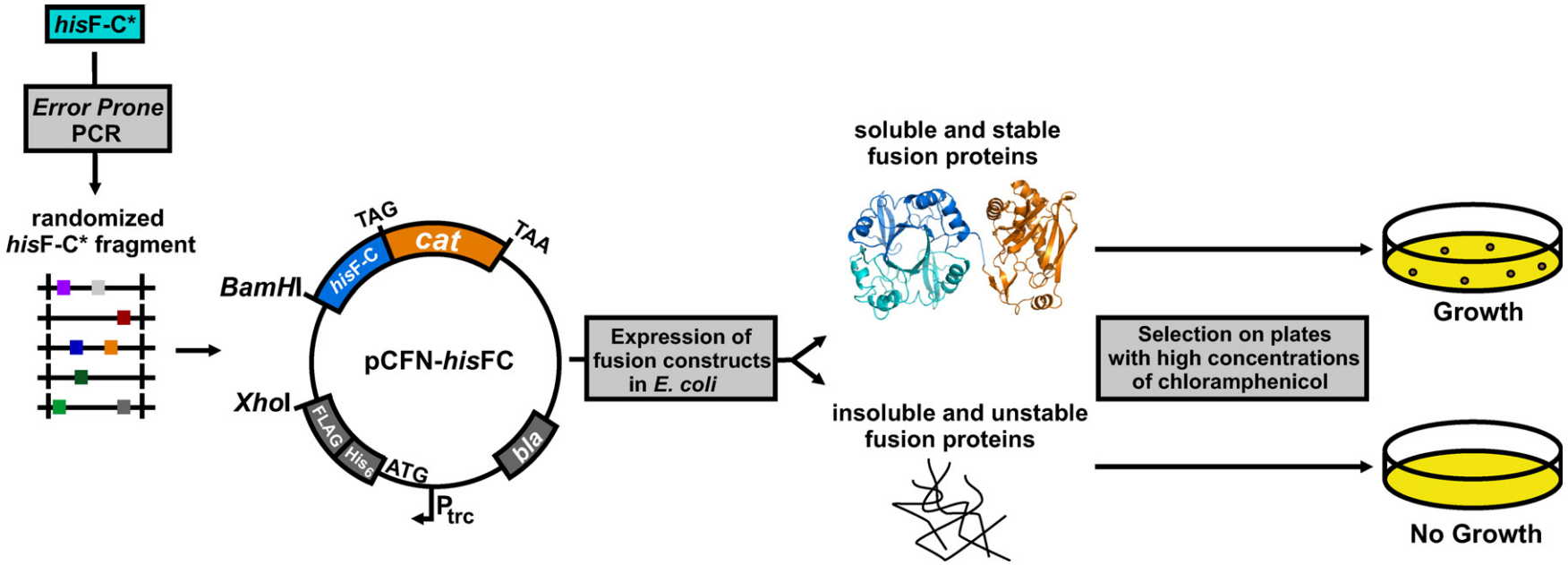

Figure 2. Scheme of the experimental protocol to isolate stabilized HisF-C*C-CAT fusion proteins by random mutagenesis and selection *in vivo*.[29] See the text for details.

type HisF, unfolds with high cooperativity, and contains an intact phosphate binding site. Molecular dynamics simulations provide a structural explanation for the observed stability/solubility increase (HisF-CC < HisF-C*C < HisF-C**C < HisF), which seems to be due to the formation of beneficial electrostatic interactions and the optimisation of van der Waals interactions at the interface of the two fused half barrels.

Results

Random mutagenesis and selection of improved HisF-C*C variants

The method used to increase the stability of HisF-C*C is based on the observation that cells expressing fusions of an unstable and insoluble protein to CAT exhibit decreased resistance to chloramphenicol compared to fusions with stable proteins (Figure 2).[29] To generate improved variants of HisF-C*C, the N-terminal half of the *his*F-C*C gene (*his*F-C*) was subjected to random mutagenesis using error-prone PCR. The resulting *his*F-C* mixture was ligated into a modified pCFN vector, upstream of an unchanged wild-type *his*F-C fragment fused to the *cat* gene. The resulting plasmid repertoire was first amplified in the *E. coli* strain XL1-BlueMrf', and then used to transform *E. coli* JM101 cells which were streaked on medium plates containing chloramphenicol at concentrations of 0 μg/ml (control), 300 μg/ml, 350 μg/ml and 400 μg/ml. The sequencing of ten *his*F-C*C gene variants isolated from the control plate showed that the error-prone PCR had, on average, introduced two nucleotide exchanges. However, colony PCR showed that all 35 clones grown on plates with the highest chloramphenicol concentration of 400 μg/ml were false-positives, which had lost the major part of their *his*F-C*C insert. Moreover, from the approximately 300 colonies grown in the presence of 300 and 350 μg/ml chloramphenicol, only 17 allowed to amplify a PCR product of the expected length. The sequencing of these clones yielded eleven different *his*F-C*C gene variants, which were sub-cloned without the *cat* gene into the expression plasmid pET24a (+) such that a C-terminal His_6-tag was added to the recombinant protein. The HisF-C*C variants were produced by heterologous gene expression in *E. coli* BL21 (DE3) and were found mainly in the insoluble fraction of the cell extract. Nine variants were solubilised in guanidinium chloride, refolded in phosphate buffer and characterised. Two variants, HisF-C*C 2 and HisF-C*C 4, could be purified from the soluble fraction of the cell extract using metal chelate affinity chromatography.

Stability and solubility of selected variants

From the 11 purified HisF-C*C variants, seven turned out to be no more stable in the absence of CAT than the parent HisF-C*C protein (data not shown) and were therefore not further investigated. The remaining four variants (HisF-C*C 1-4), which contained between one and four amino acid exchanges, unfolded at higher concentrations of urea and with a higher cooperativity than HisF-C*C (Table 1; Figure 3(a)). The stability of HisF-C*C 4, which carries, besides two other replacements, the Y143C exchange, was decreased in the presence of 1 mM of the reducing agent dithiothreitol (DTT). This finding suggested that HisF-C*C 4 can form a stabilizing intermolecular disulfide bond under oxidizing conditions. This hypothesis was confirmed by analytical gel filtration, which demonstrated that HisF-C*C 4 is mainly dimeric in the absence of DTT but forms a pure monomer in its presence (Figure 3(b)). The variant HisF-C*C 2 is exclusively monomeric and elutes from the gel filtration column as a symmetrical peak (Figure 3(b)). In contrast, the elution peaks of HisF-C*C, HisF-C*C 1 and HisF-C*C 3 are much less symmetrical, suggesting that a considerable fraction of these proteins forms ill-defined oligomers. In

Table 1. Results of urea induced unfolding of HisF-C*C variants

Variant	Amino acid exchanges in N-terminal half (HisF-C*)	$D_{1/2}$ (M)[a]	m^{app} (kJ/mol*M)[b]
HisF-C*C		4.3	3.7
HisF-C*C 1	V234M + G245R	4.3	5.0
HisF-C*C 2	D135G + T142S + Y143H + L250Q	4.7	6.3
HisF-C*C 3	V234M	4.4	4.1
HisF-C*C 4	Y143C + V234M + V248M	4.8 (5.4)[c]	6.7 (6.5)[c]
HisF-C*C Δ	Short $\alpha_8\beta_5$-loop	4.6	5.8
HisF-C*C Δ + D135G	Short $\alpha_8\beta_5$-loop + D135G	4.5	4.9
HisF-C*C Δ + T142S	Short $\alpha_8\beta_5$-loop + T142S	4.5	5.6
HisF-C*C Δ + Y143H	Short $\alpha_8\beta_5$-loop + Y143H	4.9	8.4
HisF-C*C Δ + V234M	Short $\alpha_8\beta_5$-loop + V234M	4.8	7.5
HisF-C**C	Short $\alpha_8\beta_5$-loop + Y143H + V234M	5.2 (4.8)[d]	11.6 (10.6)[d]

[a] $D_{1/2}$: urea concentration at which 50% of the protein is non-native.
[b] m^{app} (apparent *m*-value): operational measure for the cooperativity of unfolding, obtained by analyzing normalized unfolding traces with the two-state model.[66]
[c] The values of HisF-C*C 4 were determined in the presence of 1 mM DTT (and in its absence).
[d] The values of HisF-C**C were determined in 50 mM potassium phosphate (pH 7.5) (and in 50 mM Tris–HCl (pH 7.5)).

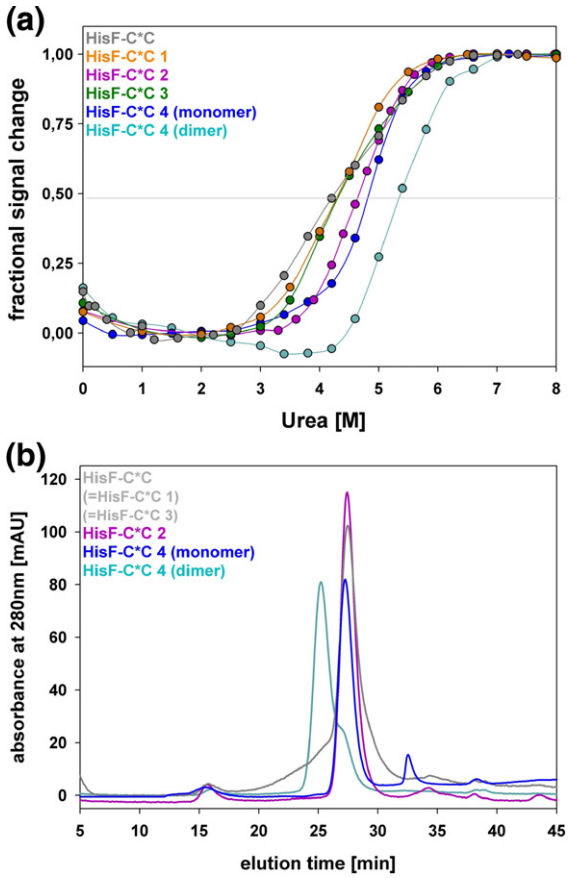

Figure 3. Stability and homogeneity of selected variants HisF-C*C 1–4. (a) Traces of reversal chemical unfolding. Proteins at a concentration of 5 μM were incubated with the given concentrations of urea in 50 mM potassium phosphate (pH 7.5) at 25 °C. Unfolding was followed by the decrease of the fluorescence emission at 320 nm after excitation at 280 nm, and the signals obtained after equilibration were normalized. The $D_{1/2}$ (urea) and apparent *m*-values of the variants are listed in Table 1. (b) Elution profiles of analytical gel filtration. Each protein (0.15–0.2 mg) was applied on a Superdex 75 (Amersham) column equilibrated with 50 mM potassium phosphate (pH 7.5), containing 300 mM potassium chloride, at 25 °C. The elution times of the main peaks of HisF-C*C and HisF-C*C 2 correspond to the monomers, and the shape of the flanking region of HisF-C*C indicates the population of ill-defined higher oligomers. The HisF-C*C 4 variant was assayed both in the presence of 1 mM DTT (monomer populated), and in its absence (disulfide-linked dimer populated).

accordance with these findings, more than 10% of recombinant HisF-C*C 2 and HisF-C*C 4 but less than 3% of recombinant HisF-C*C, HisF-C*C 1 and HisF-C*C 3 are produced in soluble form in *E. coli* (Figure 4). These data demonstrate that the HisF-C*C 2 and HisF-C*C 4 variants are significantly more stable and soluble than their parent HisF-C*C, whereas the substitutions present in the HisF-C*C 1 and HisF-C*C 3 result in only minor improvements.

Three out of the eight residue substitutions found in HisF-C*C 1-4 are located in the loop connecting the two half barrels, which corresponds to the C-terminal extension of HisF-C (Figure 1(b)). This finding suggested that this long stretch of wild-type sequence is destabilising. The loop was therefore shortened from 13 to five residues (Figure 1(c)), and the stability and solubility of the resulting variant HisF-C*C Δ was investigated. Moreover, HisF-C*C Δ was used as a scaffold to test the effect of the substitutions D135G, T142S, Y143H and V234M individually. As deduced from urea-induced unfolding, the stabilizing effect of the exchanges is as follows: Δ+D135G ≈ Δ + T142S < Δ < Δ + V234M < Δ + Y143H (Table 1; Figure 5(a)). The effect of these substitutions on the percentage of recombinant protein found in the soluble fraction of the cell extract is somewhat different: Δ + D135G < Δ ≈ Δ + T142S ≈ Δ + V234M < Δ + Y143H (Figure 4). These results show that the shortening of the loop between the two HisF-C half barrels improves both the stability and solubility of HisF-C*C. From the individual substitutions, D135G and T142S are neutral at best and were isolated only accidentally in the context of HisF-C*C 2. The exchanges Y143H and V234M are increasing the stability, whereas Y143H is by far the most crucial substitution for the improvement of solubility.

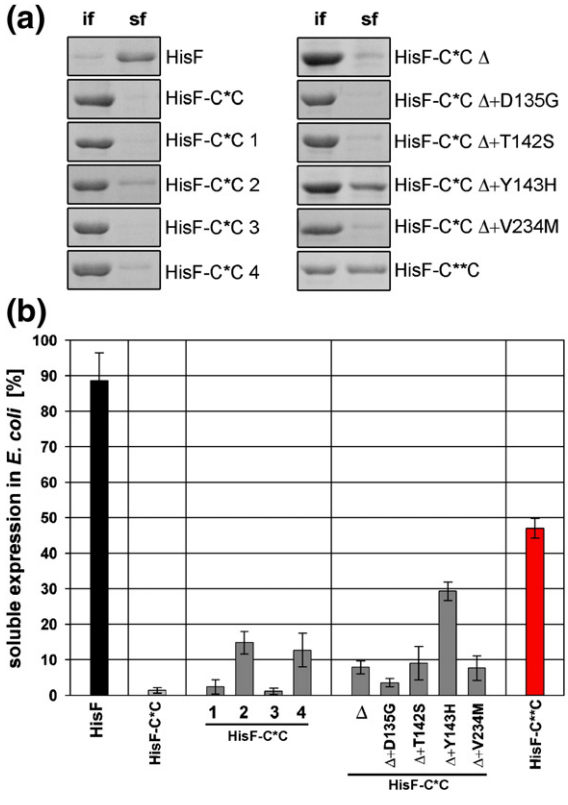

Figure 4. Solubility analysis of selected and designed HisF-C*C variants. (a) Insoluble (if) and soluble (sf) fractions of the different variants as found in cell extracts of the *E. coli* expression strain BL 21(DE3) (SDS–PAGE (12.5% acrylamide) stained with Commassie blue). (b) Percentage of the proteins produced in soluble form (as deduced from the analysis of three different clones of each variant). The protein bands were scanned, and their intensities were determined using Optiquant (version 2.5).

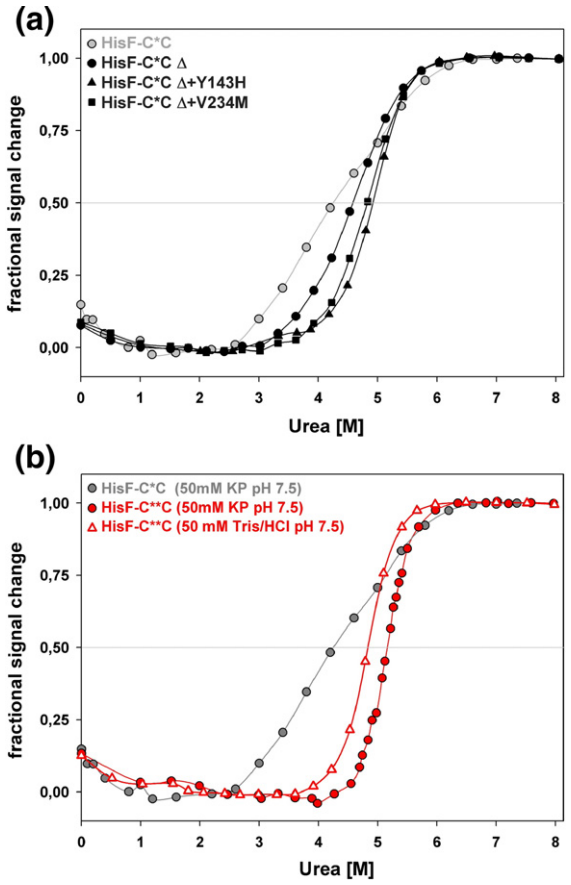

Figure 5. Stability of designed HisF-C*C variants. (a) The unfolding traces of the variants HisF-C*C Δ, HisF-C*C Δ+Y143H and HisF-C*C Δ+V243M are plotted in comparison to HisF-C*C. Proteins at a concentration of 5 μM were incubated with the given concentrations of urea in 50 mM potassium phosphate (KP) (pH 7.5), at 25 °C. (b) To test the stabilising effect of bound phosphate ions, HisF-C**C was unfolded in 50 mM KP (pH 7.5) and in 50 mM Tris–HCl (pH 7.5). Reversible unfolding was followed by the decrease of the fluorescence emission at 320 nm after excitation at 280 nm, and the signals obtained after equilibration were normalized. The $D_{1/2}$ (urea) and apparent *m*-values of the variants are listed in Table 1.

Generation and characterisation of HisF-C**C

In order to produce an even more beneficial HisF-C*C variant, the Y143H and V234M substitutions were combined on the HisF-C*C Δ scaffold. The resulting variant HisF-C**C was produced to 50% as a soluble protein (Figure 4) that unfolds at higher concentrations of urea than any of the originally isolated HisF-C*C variants (Table 1). Its stability is elevated in phosphate compared to Tris buffer (Figure 5(b)), indicating that it possesses two phosphate binding sites as found in wild-type HisF (Figure 1). In order to compare the stability of HisF-C**C with that of HisF, which cannot be completely denatured in urea, unfolding was induced by the stronger denaturant guanidinium chloride. The measurements showed that HisF-C**C is less stable than wild-type HisF but unfolds with higher cooperativity (Figure 6(a)), indicating that the newly generated variant is as compact as the native protein. In accordance with this finding, HisF-C**C and HisF elute from the gel filtration column as equally homogeneous monomers, whose peaks are much sharper than that of HisF-C*C (Figure 6(b)).

To compare the tertiary structures of HisF-C*C and HisF-C**C with that of wild-type HisF, fluorescence emission and near-UV circular dichroism spectra were recorded. The fluorescence spectra are dominated by the single tryptophan residue 156 in helix α5 of HisF, which exists as two copies in HisF-C*C and HisF-C**C. The structural environment of its indole side-chain is hydrophobic in wild-type HisF (Ala106, Val107, and Ile113 from helix α4 of the adjacent HisF-N half barrel) as well as in HisF-C*C and HisF-C**C (Phe 227, His228, Val234/Met234 from helix α8 of the adjacent HisF-C half barrels), and should therefore be shielded from solvent in an intact $(\beta\alpha)_8$-barrel structure. The fluorescence emission maxima of wild-type HisF and HisF-C**C lie at 322 and 323 nm, respectively, whereas the emission maximum of HisF-C*C lies at 327 nm (Figure 6(c)). These results demonstrate that the two tryptophan 156 residues in HisF-C**C are equally well shielded from solvent as the single tryptophan 156 in wild-type HisF. In contrast, the red-shifted emission maximum of HisF-C*C shows that at least one of its tryptophan residues must be surrounded by a more polar environment. Remarkably, the gel filtration profiles and fluorescence emission spectra of HisF-C*C Δ, HisF-C*C Δ+Y143H and HisF-C*C Δ+V234M are identical to those of HisF-C*C (data not shown), demonstrating that the combination of the beneficial substitutions is required to completely shield the tryptophan residues. The pronounced near-UV CD spectra of HisF-C*C and HisF-C**C have a similar shape as the spectrum of wild-type HisF (Figure 6(d)), showing that the aromatic residues are immobilised in an asymmetric environment in both halves. This result proves that HisF-C*C and HisF-C**C do not form "molten globules" but adopt a well-defined tertiary structure.

The unfolding of a great number of proteins with known X-ray structure has shown that the amount of protein surface exposed to solvent (ΔASA) correlates with the denaturation *m*-value.[30] Moreover, ΔASA also correlates with the number of residues of a globular protein. As a consequence, the measured apparent *m*-values of the various HisF-C*C constructs allowed us to estimate the corresponding ΔASA values. The HisF-C*C and HisF-C*C Δ scaffolds contain 274 and 266 residues, corresponding to calculated ΔASA values of 24,575 and 23,831 Å^2, respectively. The apparent *m*-values measured for most HisF-C*C variants indicate that their ΔASA values are smaller by a factor of 2–4, suggesting that the folded states of these proteins do not completely shield their hydrophobic interior from solvent. This result indicates, together with the gel filtration and fluorescence data (Figure 6(b) and (c)), that monomeric HisF-C*C is in equilibrium with

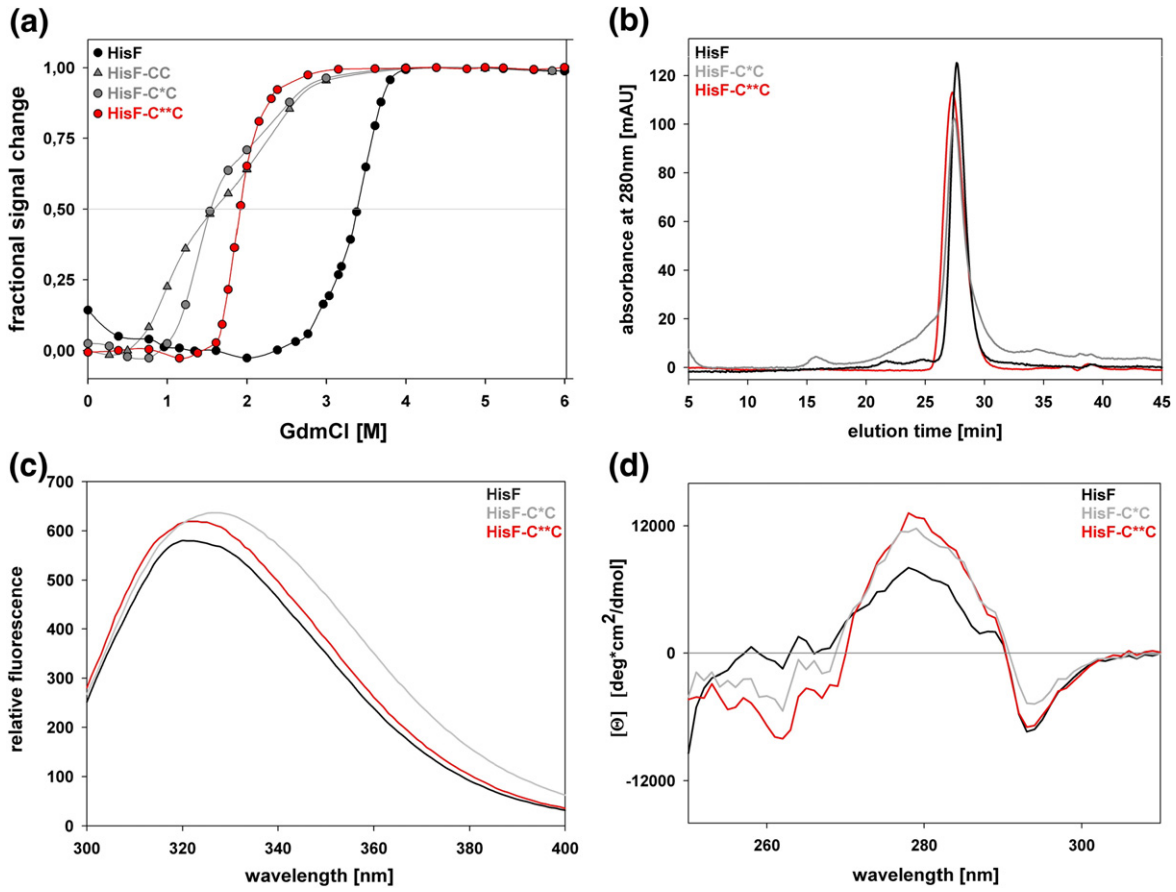

Figure 6. Results of the stepwise optimisation of fused HisF-C half barrels in comparison to wild-type HisF. (a) Traces of reversible chemical unfolding. Proteins at a concentration of 5 μM were incubated with the given concentrations of GdmCl in 50 mM potassium phosphate (pH 7.5), at 25 °C. Unfolding was followed by the decrease of the fluorescence emission at 320 nm after excitation at 280 nm, and the signals obtained after equilibration were normalised. The transitions midpoints of both HisF-CC and HisF-C*C lie at 1.5 M GdmCl; the multi-phasic unfolding traces did not allow to determine apparent *m*-values. The transition midpoints of HisF-C**C and wild-type HisF lie at 1.9 M and 3.4 M GdmCl, respectively; the unfolding of these variants is highly cooperative with apparent *m*-values of 17.3 kJ mol^{-1} M^{-1} for HisF-C**C and 14.2 kJ mol^{-1} M^{-1} for wild-type HisF. (b) Elution profiles of analytical gel filtration. Each protein (0.15–0.2 mg) was applied on a Superdex 75 (Amersham) column equilibrated with 50 mM potassium phosphate (pH 7.5), containing 300 potassium chloride, at 25 °C. The elution times of the main peaks correspond to the monomers, and the shape of the flanking region of HisF-C*C indicates the population of ill-defined higher oligomers. (c) Fluorescence emission spectra. The spectra shown were monitored after excitation at 280 nm of 5 μM protein in 50 mM potassium phosphate (pH 7.5), at 25 °C. The measured emission intensity of HisF, which is mainly caused by its single Trp156, was multiplied by a factor of 2 in order to compensate for the presence of two Trp156 residues in HisF-C*C and HisF-C**C. The emission maxima are as follows: HisF: 322(±2) nm; HisF-C**C: 323(±2) nm; HisF-C*C: 327(±1) nm. The values given are the mean and standard deviation of three independent measurements. (d) Near-UV CD spectra. The spectra shown are the mean of ten individual spectra, which were recorded at 25 °C in 50 mM potassium phosphate (pH 7.5).

a small fraction of a partially unfolded and aggregation-prone conformation. In contrast, the deduced ΔASA value of HisF-C**C of 21,782 $Å^2$ indicates that this variant is a compactly folded protein with a buried hydrophobic core, again in accordance with the results from analytical gel filtration and fluorescence emission spectroscopy (Figure 6(b) and (c)).

Molecular dynamics simulations

In order to gain insight into the mechanism of stabilisation caused by the various amino acid substitutions, the crystal structure of HisF (PDB code 1thf)[25] was used as the basis to construct models of duplicated and fused half barrels. To start with, two HisF-C units were covalently linked, and the resulting HisF-CC variant was then relaxed by 2 ns of molecular dynamics (MD) simulation at 300K using the MAB-force field as implemented in MOLOC‡.[31] Remarkably, the presence of two bound phosphate ions in the simulations, one in each half barrel as observed in the crystal structure of HisF (Figure 1), stabilised HisF-CC as well as the other variants. Simulations performed without bound phosphate

‡ http://www.moloc.ch

led to the collapse of the phosphate binding site and a partially distorted protein backbone. Starting from HisF-CC, structures of the variants HisF-C*C, HisF-C*C Δ, and HisF-C**C were stepwise calculated and relaxed by MD, which resulted in the model displayed in Figure 7(a). The gain in stability caused by the exchanges introduced at the half barrel interface ("interaction energy") was analysed using a pseudo-thermodynamic cycle. To this end, the energies of the two isolated, solvated $(\beta\alpha)_4$-half barrels were subtracted from the energy of the intact $(\beta\alpha)_8$-barrel structure.[32] The calculated interaction energy of HisF-CC (440(±10) kJ/mol) is smaller than that of HisF-C*C (565(±10) kJ/mol), confirming the stabilizing effect of the reconstituted wild-type salt bridge cluster in layer 4 of the central β-barrel of HisF-C*C[28] (Figure 7(b)). Moreover, the deduced interaction energy of 125 kJ/mol for this cluster is reasonable for a network of four charged residues.[33] The calculated interaction energy of HisF-C*C Δ (557(±10) kJ/mol) is identical within error to that of HisF-C*C, indicating that the shortening of the connecting loop between the two half barrels does not influence their interaction. In contrast, the presence of the two exchanges Y143H and V234M at the half barrel interface of HisF-C**C leads to an

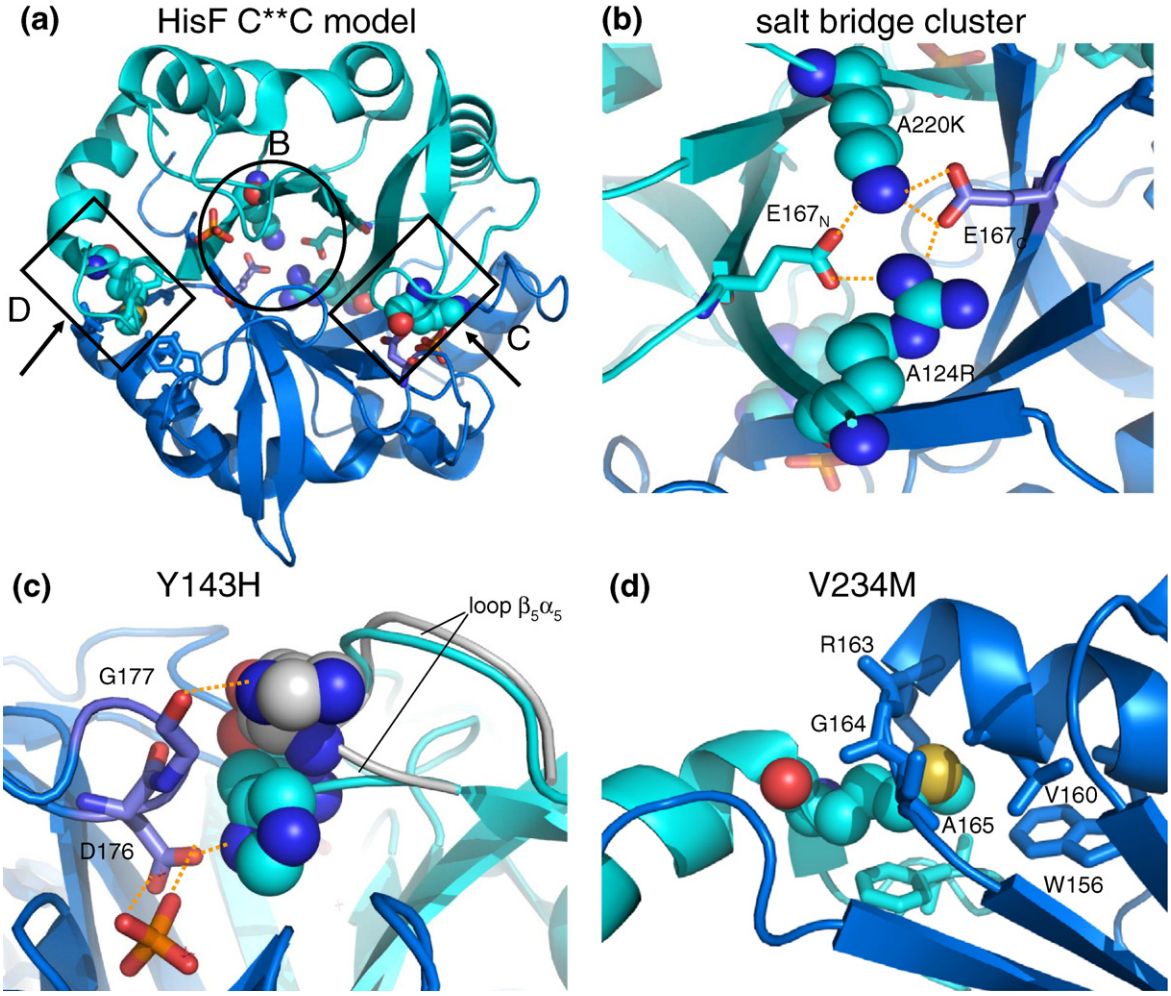

Figure 7. Structural model of HisF-C**C based on MD simulations. (a) A top view onto the C-terminal face of the central β-barrel is shown as a ribbon diagram. The introduced stabilizing residues are highlighted as van der Waals-spheres, and other important residues are shown as sticks. The two phosphate ions bound to the N- and C-terminal half barrels are depicted as tetrahedrons. The colour code is the same as for Figure 1(c); the mutated N-terminal half barrel is shown in cyan, and the wild-type C-terminal half barrel is shown in blue. The arrows indicate the direction of the side views shown in (c) and (d). (b) The reconstituted salt bridge cluster forming layer 4 of the central β-barrel is composed of N- and C-terminal glutamate residues ($E167_N$, $E167_C$) as well as arginine and lysine residues introduced into the N-terminal half (A124R and A220K). H-bonds are represented by dotted orange lines. (c) The histidine residue introduced into the N-terminal half barrel (Y143H) forms a H-bond network with Asp176 and the phosphate ligand of the C-terminal half barrel. In a possible alternative conformation, His143 forms an H-bond with Gly177 and is solvent exposed. The corresponding alternative closed and open conformations of loop β5α5 (residues 141–149) and the carbon atoms of the imidazole side-chain of His143 in the two conformations are shown in cyan and grey, respectively. (d) The buried methionine residue introduced into the N-terminal half barrel (V234M) is protruding into a hydrophobic pocket of the C-terminal half barrel, where it comes in van der Waals contact with Trp156, Val160, Arg163, Gly164 and Ala165.

increase by 42 kJ/mol in the calculated interaction energy to 599(±10) kJ/mol. The histidine residue at position 143 is located in loop β5α5 near the phosphate-binding site of the C-terminal half barrel. During the molecular dynamics relaxation a hydrogen bond network between His143, the bound phosphate, the conserved crystal water molecules next to them and Asp176 was established, which provides about 30 kJ/mol of interaction energy. However, the simulations suggest that an alternative conformation exists in which the imidazole side-chain of His143 is solvent-exposed and forms an H-bond with the carbonyl oxygen of Gly177 (Figure 7(c)). The introduced methionine at position 234 is in hydrophobic contact with several residues of the C-terminal half barrel (Val160, Arg163, Gly164, Ala165; Figure 7(d)). The energy decomposition suggests mainly hydrophobic interactions of about 12 kJ/mol contributed by the van der Waals energy of the MAB force field.[34] The calculated interaction energy between HisF-N and HisF-C within wild-type HisF is 638(±10) kJ/mol. Taken together, the results of the MD simulations qualitatively reflect the experimentally determined stability increase from HisF-CC *via* HisF-C*C and HisF-C**C to HisF, indicating that the beneficial amino acid substitutions can be interpreted as an improved interaction between the fused half barrels.

Discussion

Structurally defined and stable domains with a given function constitute the basic building blocks of proteins.[35,36] It follows that the duplication, fusion and recombination of domain-encoding genes are the evolutionary mechanisms used by nature to generate complex multi-domain proteins.[37] When analysing the evolution of the versatile and ubiquitous $(\beta\alpha)_8$-barrel proteins, we noticed that some representatives of this highly symmetrical fold consist of sub-domains comprising $(\beta\alpha)_4$-half barrels.[9] In an attempt to reconstruct the evolution of the $(\beta\alpha)_8$-barrel HisF from the $(\beta\alpha)_4$-half barrel HisF-C, we earlier generated and characterised the HisF-CC construct which consists of two identical half barrels linked by a Gly-Ser-Gly linker. Using rational design, HisF-CC was optimised to HisF-C*C by reconstituting the native salt bridge cluster which forms layer 4 at the N-terminal face of the central β-barrel of HisF.[28] Since model building did not point to other beneficial amino acid exchanges to be introduced by rational design, we used a combination of random mutagenesis and selection *in vivo* to further stabilise HisF-C*C.

Pitfalls of the selection system

Various directed evolution methods for increasing protein solubility and stability have been described.[38–42] We chose an approach in which a mixture of *his*F-C*C genes with a randomised N-terminal half was fused to the *cat* gene, and selected for HisF-C*C-CAT fusion proteins which allowed their host cells to grow in the presence of high concentrations of chloramphenicol.[29,43,44] Although the original vector pCFN1[29] was modified such that without ligation of the randomised N-terminal *his*F-C* fragment a frameshift was generated, 100% of the variants selected in the presence of 400 μg/ml and about 95% of the variants selected in the presence of 350 μg/ml and 300 μg/ml chloramphenicol were false-positives. These clones carried plasmids with extensively truncated *his*F-C*C mutants in frame with the *cat* gene, generating CAT proteins with short N-terminal extensions that did not compromise its high solubility. We speculate that these deletions have mainly occurred during the cloning procedure of the *his*F-C* library because cells transformed with the empty pCFN-*his*FC vector (Figure 2) were highly sensitive to chloramphenicol. Probably the false-positives were generated by unspecific digestions of the used restriction enzymes XhoI and BamHI, and were strongly enriched as a result of the high selection pressure. However, we cannot exclude that some of the observed deletions were caused by recombination events in *E. coli*. The sequencing of a number of false-positive clones showed that the inserts were diverse, which precluded the identification and subsequent removal of unwanted restriction "hot spots" by site directed mutagenesis. Moreover, from the 11 selected unique variants with a complete *his*F-C*C insert, in the absence of CAT seven proteins did not show a higher stability or solubility than their parent HisF-C*C. The reliability of screening methods using fused reporter proteins frequently suffers from unexpected interactions between reporter and target, which can influence the outcome of the assay. In the case of CAT, which is an obligate trimer, the linkage to a target protein might lead in certain cases to higher-order complexes with unpredictable effects on the solubility of the fusion constructs.[40,42]

Structural basis of stabilisation

From the true-positive variants, HisF-C*C 1 and 3 were slightly and HisF-C*C 2 and 4 were considerably more stable and soluble than HisF-C*C (Table 1; Figures 3 and 4). These proteins altogether contained eight different amino acid exchanges of which three (G245R, V248M, L250Q) were located in the long αβ loop connecting the two HisF-C units (Figure 1(b)). The shortening of this loop from 13 to five residues in HisF-C*C Δ increased the stability and the solubility of the protein significantly (Table 1; Figures 4 and 5). This result is in line with the observation that αβ loops of $(\beta\alpha)_8$-barrels are generally short and contribute to the conformational stability of the fold.[11] Along the same lines, a number of proteins from hyperthermophiles carry shorter loops compared to their homologues from mesophiles, which causes a decrease of the entropy difference between the unfolded and the folded state that results in an increase of the stabilisation free energy.[45,46] Remarkably, the deletion of another two

residues resulting in a loop containing only the introduced Gly-Ser-Gly linker decreased stability and solubility (data not shown), possibly due to the introduction of conformational strain.

The tyrosine residue at position 143 was replaced in the two best variants, with cysteine in HisF-C*C 4 and with histidine in HisF-C*C 2 (Table 1). The substitution Y143C in HisF-C*C 2 leads to the formation of an intermolecular disulfide bridge (Figure 3(b)). However, the cytoplasm of *E. coli* is reducing and HisF-C*C 2 is more stable than HisF-C*C also in the presence of DTT (Figure 3(a)). It is therefore unclear whether the cystine was formed already in the host cell during the *in vivo* selection or only upon contact of the protein with air oxygen during its purification. However, the X-ray structures of a number of cytoplasmic enzymes from hyperthermophilic Archaea suggest that cystine residues can contribute to extreme protein thermostability of intracellular proteins,[47–50] an hypothesis which is supported by the computational analysis of microbial genomes.[51,52] The introduction of new S–S bonds[53] is a widely used protein stabilisation strategy, and the beneficial effect of S–S bonds is ascribed to the decrease of the entropy of the unfolded state.[54–58] In addition, the detected disulfide bond in HisF-C*C 2 could also stabilise the protein by increasing its association state from the monomer to the dimer. Along these lines, a number of (βα)$_8$-barrels and other proteins from hyperthermophiles are higher-order oligomers than their counterparts from mesophiles.[59–61] The structural basis of the stabilising effect of the Y143H exchange is less clear. Although the MD simulations suggest that the introduced imidazole ring leads to the formation of a hydrogen bonding network (Figure 7(c)), alternative mechanisms are feasible. Since the imidazole side-chain of histidine is more polar than the aromatic ring of tyrosine, its higher solvatation energy could stabilise the native state. Along these lines, position 143 is located in the long loop connecting strand β5 with helix α5, which was highly flexible in the simulations and might easily become solvent-exposed. Accordingly, we have frequently observed in the simulations geometries in which the imidazole ring is not in contact to Asp176 and the bound phosphate ion but forms an H-bond with the carbonyl group of Gly177 at the surface of the protein (Figure 7(c)). In accordance with this finding, the Y143H exchange leads to a strong improvement of the solubility of HisF-C*C (Figure 4). The substitution V234M was found in three of the four improved HisF-C*C variants and is the single exchange present in HisF-C*C 3 (Table 1). The MD simulations suggest that the long and linear hydrophobic methionine side-chain protrudes into a patch of hydrophobic residues of the adjacent half barrel (Figure 7(d)). This newly formed hydrophobic cluster is well shielded from solvent, which explains why HisF-C*C 3 and HisF-C*C Δ + V234M are no more soluble than HisF-C*C and HisF-C*C Δ, respectively (Figure 4). However, HisF-C**C is more soluble than HisF-C*C Δ + Y143 H, suggesting that the effect of the V234M exchange can be influenced by the protein scaffold in which it is introduced. Along the same lines, the increase of the apparent *m*-value caused by the V234M substitution is somewhat more pronounced in the background of HisF-C*C Δ + Y143H than in the background of HisF-C*C Δ and much stronger than in the background of HisF-C*C (Table 1). The basis of this context dependence is difficult to elucidate in the absence of highly resolved X-ray structures of the different variants. One can speculate that the presence of the Y143H exchange and the shortening of the loop induce a better defined local structure in which methionine 234 can unfurl its stabilizing effect more efficiently.

Implications for the evolution of (βα)$_8$-barrels

The structural superposition of the wild-type HisF half barrels HisF-N and HisF-C reveals that about 100 of the approximately 120 aligned residues are different.[25] It is therefore remarkable that four amino acid substitutions in the N-terminal half of HisF-CC were sufficient to generate the HisF-C**C protein, whose tertiary structure must be highly similar to that of HisF. Moreover, neither the Y143H nor the V243M substitution in the N-terminal half of HisF-C**C reconstituted residues found at the equivalent positions in HisF-N. These findings strengthen the hypothesis that half barrels are flexible and independently evolving structural entities, which can be mixed and matched by nature to generate the highly symmetrical and catalytically versatile and ubiquitous TIM-barrel fold.[10,11]

Earlier attempts to design (βα)$_8$-barrels from individual (βα) units mainly generated molten globules that contained a considerable amount of secondary structure but did not fold into well-defined tertiary structures.[11] However, recently an idealised (βα)$_8$-barrel with native-like properties was generated on the basis of a two-step computational approach.[62] Since the central eight-stranded β-sheet was constructed on the basis of a 4-fold symmetry, the successful design supports the hypothesis that (βα)$_2$ units constitute the smallest independently evolving units within (βα)$_8$-barrels.[10,15] However, it has not been reported whether this protein is a monomer with a tertiary structure that is equally well defined and compact as that of HisF-C**C.

In HisF, the catalytically essential residues aspartate 11 in β-strand 1 and aspartate 130 in β-strand 5 are related by a 2-fold symmetry axis along the central barrel axis and perfectly superimpose when HisF-N and HisF-C are structurally aligned.[9,25,63] HisF-C**C carries two structurally equivalent aspartate residues in its two β-strands 5, and both chemical unfolding and MD simulations suggest that HisF-C**C possesses intact sites to bind the two terminal phosphate moieties of the HisF substrate *N*′-[(5′-phosphoribulosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (PRFAR) (Figures 5(b) and 7(a)). Nevertheless, HisF-C**C did not show measurable catalytic

activity, even at a protein concentration of 50 μM (data not shown). Probably, minor perturbations of the active site geometry of HisF-C**C prevent the productive binding and turnover of PRFAR, which shall now be established by further rounds of random mutagenesis and selection *in vivo* supported by molecular dynamics studies *in silico*.

Materials and Methods

Construction of a *his*F-C*C gene library with randomised N-terminal half

The chloramphenicol-acetyltransferase (CAT) fusion vector pCFN-*his*FC (Figure 2) was derived from pCFN1.[29] This vector includes phage f1 and pBR322 origins of replication, the ampicillin resistance gene (*bla*) and a strong isopropyl-β-thiogalactoside (IPTG)-inducable promotor (P_{trc}). Sequences inserted into the multiple cloning site (MCS) of pCFN1 between the restriction sites BglII and XbaI lead to the expression of proteins fused at their N-terminus with a His_6-tag and the FLAG epitope, and at their C-terminus with CAT. An amber stop-codon is inserted between the XbaI site and the *cat* gene such that an amber suppressor *E. coli* strain is needed for expression of the full-length fusion gene. pCFN-*his*FC was constructed in two steps. First the 5′ BglII restriction site of the MCS was changed into a XhoI site by using the QuikChange™ Site-Directed Mutagenesis Kit (Stratagene) with the complementary oligonucleotides 5′-CTACAAGGACGATGACGACTCGAGCTCACGTCTAGA-3′ and 5′-GCGAGCCTACTCTAGACGTGAGCTCGAGTCGTCATC-3′ as overlapping primers,[64] yielding pCFN1(XhoI) (new restriction site underlined). Next *his*F-C was amplified by PCR using a *his*F-C containing pCFN1 vector (pCFN1-*his*FC) as template and the oligonucleocleotides 5′-ATACTCGAGTTGGGATCCGGTCAGGCCGTTGTC-3′ with XhoI and BamHI sites (underlined) as 5′-primer, and 5′-CTCGCGAGCCTACTCTAGAAA-3′ with a XbaI site (underlined) as 3′-primer. The amplified *his*F-C fragment was cloned into pCFN1(XhoI) using XhoI and XbaI restriction sites, introducing a new BamHI site between XhoI and *his*F-C and yielding pCFN-*his*FC. In order to avoid expression of HisF-C - CAT fusion proteins through religated vectors in the upcoming selection, a frame shift between the coding region of FLAG and the BamHI site was introduced.

The *his*F-C* gene, which is identical to *his*F-C except for the substitutions A124R and A220K,[28] was subjected to error-prone PCR using the GeneMorph® II Random Mutagenesis Kit (Stratagene) under conditions which should introduce about two nucleotide exchanges per gene. To this end, the oligonucleotides 5′-AAGAAGGAGATATACTCGAGCCAGCGC-3′ and 5′-TTGTCGACGGAGCTCGAATTCGGATCC-3′ were used as 5′ and 3′ primers, respectively. These contained XhoI and BamHI restriction sites (underlined) and reconstituted the correct reading frame for the translation of the *his*F-C*C-cat fusion construct after cloning of the randomised *his*F-C* fragments into pCFN-*his*FC.

The plasmids obtained were used to transform *E. coli* XL1-BlueMrf′cells (Stratagene) by electroporation, which were plated on medium containing ampicillin to select for plasmid uptake. As estimated from the number of grown colonies and the ligation efficiency tested by colony PCR, the *his*F-C*C-*cat* gene repertoire contained approximately 7.4×10^4 different variants. The grown colonies were scratched off the plates and resuspended, followed by the isolation of the pCFN-*his*F C*C plasmid library. Sequencing of several clones grown under such non-selective conditions showed that Mutazyme® II DNA Polymerase had inserted two mutations per gene on average, as intended, with a range from zero to seven exchanges. The distribution of mutations showed a bias towards GC over AT base-pair exchanges (70% *versus* 30%), and of transitions over transversions (75% *versus* 25%).

In vivo selection for soluble HisF-C*C variants

The amber-suppressor *E. coli* strain JM101[65] was transformed by electroporation with the pCFN-*his*F C*C gene library, as well as with the controls pCFN-*his*FC containing "wild-type" *his*F-C* and pCFN-*his*FC containing no insert. In each case, expression was induced immediately after transformation by adding 200 μg/ml of IPTG. Following shaking of the transformants for 2 h at 37 °C, the cells were streaked on plates containing 200 μg/ml IPTG and different concentrations of chloramphenicol (0–400 μg/ml), and incubated at 37 °C for 16 h. Cells transformed with "wild-type" pCFN-*his*F C*C were able to grow on plates with up to 250 μg/ml chloramphenicol, whereas cells transformed with empty pCFN-*his*FC vector grew only in the absence of the antibiotic, due to the introduced frame shift between the coding region of FLAG and the BamHI site mentioned above. From the pCFN-*his*F C*C gene library several colonies grown in the absence of chloramphenicol, and the >330 colonies grown on plates containing 300, 350 and 400 μg/ml chloramphenicol were picked and subjected to plasmid isolation. Colony PCR showed that 97% of these colonies were false-positives, which carried plasmids with truncated inserts that had lost the major part of the *his*F-C*C sequence and thus contained essentially only the *cat* gene.

Cloning of selected and designed *his*F-C*C variants into pET24a(+)

In order to produce the selected HisF-C*C variants in the absence of CAT as well as the N-terminal His_6- and FLAG-tags, the *his*F-C* inserts from the 11 true positive colonies grown in the presence of high chloramphenicol concentrations were sub-cloned from pCFN-*his*F C*C into pET24a(+) vectors containing the wild-type *his*F-C gene (pET24a(+)-*his*FC_2).[28] To this end, the *his*F-C* inserts were cut out from pCFN-*his*F C*C with the help of the restriction enzymes XhoI and BamHI. The resulting fragment was amplified by PCR using the oligonucleotides 5′-ATACATATGCAGCGCGTTGTCGTGGCGATA-3′ with a NdeI site (underlined) as 5′-primer and 5′-TCGACGGAGTCTAGATTCGGATCCCAACCC-3′ with a BamHI site (underlined) as 3′-primer. The amplification product was ligated with pET24-*his*FC_2, yielding the various pET24a(+)-*his*F C*C variants. In these constructs, the genes of *his*F-C* and *his*F-C are connected by the BamHI site in such a way that in the HisF-C*C proteins both halves are linked by a Gly-Ser-Gly stretch.

For construction of the variant HisF-C*C Δ in which the αβ loop connecting the N- and C-terminal halves was shortened from 13 to 5 residues, *his*F-C* was amplified by PCR using an isolated *his*F-C* fragment as template. The oligonucleotide 5′-ATACATATGCAGCGCGTTGTCGTGGCGATA-3′ with a NdeI site (underlined) was used as 5′-primer and the oligonucleotide 5′-

ATAGACGGATCCGCCGTGTTTTTTGAGGTACTCTTT-CAGTTC-3′ with a BamHI site (underlined) was used as 3′-primer. The amplified fragment *his*F-C* Δ was cloned into pET24a(+)-*his*FC_2, yielding pET24a(+)-*his*FC*CΔ. As a consequence, the halves connecting αβ-loop in HisF-C*C Δ ends with Gly245 (Figure 1(c)), and the Gly-Ser-Gly stretch connecting the two halves was preserved.

To introduce the single amino acid exchanges D135G, T142S, Y143H and V234M into the N-terminal half of HisF-C*C Δ the *his*F-C* sequence was amplified by PCR using an isolated *his*F-C* fragment as template. The oligonucleotides 5′-GATATACATATGCAGCG-CGTTGTCGTGGCGATAGATGCAAAAAGAGTG-**GGT**GGAGAGTTC-3′ (for generation of *his*F-C*-D135G), 5′-GATATACATATGCAGCGCGTTGTCGTGGCGATA-GATGCAAAAAGAGTGGATGGAGAGTTCATGGTCTT-C**AGC**TACTCCGGAAAGAAG-3′ (for generation of *his*F-C*-T142S), 5′-GATATACATATGCAGCG-CGTTGTCGTGGCGATAGATGCAAAAAGAGT-GGATGGAGAGTTCATGGTCTTCACC**CAC**TCCG-GAAAGAAG-3′ (for generation of *his*F-C*-Y143H) and 5′-ATACATATGCAGCGCGTTGTCGTGGCGATA-3′ (for generation of *his*F-C*-V234M) were used as 5′-primers. The oligonucleotides 5′-ATAGACGGATCCGCCGTGT-TTTTTGAGGTACTCTTTCAGTTC-3′ (for generation of *his*F-C*-D135G, *his*F-C*-T142S and *his*F-C*-Y143H) and 5′-ATAGACGGATCCGCCGTGTTTTTTGAGGTACTCTTT-CAGTTCTCT**CAT**GTC-3′ (for generation of *his*F-C*-V234M) were used as 3′-primer. As a consequence, at the 5′ terminus a *Nde*I-site (underlined) and at the 3′ terminus a BamHI site (underlined) were introduced together with the specific codon exchange (bold). The 3′-primers are designed in such a way that the amplified *his*F-C* fragment is shortened like in *his*F-C* Δ. Thus, the N-terminal half of each single variant ends at Gly245, and the two halves are linked by a Gly-Ser-Gly stretch. The cloning of the amplification products into pET24-*his*FC_2 yielded pET24a(+)-*his*F C*C Δ + D135G, pET24a(+)-*his*F C*C Δ + T142S, pET24a(+)-*his*F C*C Δ + Y143H and pET24a(+)-*his*F C*C Δ + V234M.

For construction of *his*F-C**C in pET24a(+), *his*F-C* was amplified by PCR using an isolated *his*F-C* fragment as template and the oligonucleotides mentioned above which introduce the Y143H and the V234M exchanges as well as the loop shortening. The cloning of the amplification fragment into pET24a(+)-*his*FC_2 yielded pET24a(+)-*his*F C**C. All constructs were sequenced entirely to exclude inadvertent mutations.

Heterologous expression of *his*F-C*C variants and protein purification

Heterologous expression of the *his*F-C*C variants was performed in *E. coli* BL21(DE3) (Stratagene) transformed with the various pET24a(+) plasmids containing the encoding sequences. To this end, four liters of Luria Broth (LB) media supplemented with 75 μg/ml of kanamycin (for maintenance of pET24a(+)) were inoculated with a pre-culture and incubated at 37 °C. After an OD_{600} of 0.6 was reached, expression was induced by adding 1 mM IPTG, and growth continued for another 4 h. Cells were harvested by centrifugation (Sorvall/RC5B, GS3 rotor, 15 min, 4000 rpm at 4 °C), washed with 10 mM potassium phosphate (pH 7.5), and centrifuged again. The cells were resuspended in 140 ml of the same buffer, lysed by sonification (Branson Sonifier W-250D; 2×2 min in 15 s intervals, 50% pulse, 0 °C), and centrifuged again (Sorvall/RC5B, SS34 rotor, 30 min, 13,000 rpm at 4 °C) to separate the soluble from the insoluble fraction of the cell extract.

The variants HisF-C*C 2 and 4, as well as HisF-C*C Δ, Δ + D135G, Δ + T142S, Δ + Y143H, Δ + V234M were purified from the soluble cell fraction. To this end, the extract was loaded onto a nickel Sepharose column (HisTrap FF crude 5 ml; GE Healthcare) which had been equilibrated with 50 mM potassium phosphate, 300 mM NaCl, 1 mM imidazole (pH 7.5). The column was washed with the equilibration buffer, and the bound His_6-tagged protein was eluted by applying a linear gradient of 1 mM–300 mM imidazole. Fractions with pure protein were pooled and dialysed extensively against 50 mM potassium phosphate (pH 7.5). In the case of the HisF-C*C 4, the dialysis buffer contained either 1 mM dithiothreitol (DTT) to avoid disulfide-linkage *via* Y143C or no DTT to allow for cystine formation. According to SDS–PAGE (12.5% (w/v) acrylamide), all proteins were pure >95%.

The variants HisF-C*C, HisF-C*C 1 and 3 were purified from the insoluble cell fraction, which was washed with 10 mM potassium phosphate (pH 7.5), resuspended in 80 ml of the same buffer containing 6 M guanidinium chloride (GdmCl) and incubated for 30 min at room temperature. Following the addition of 80 ml of 1 M GdmCl and further incubation for 30 min the suspension was centrifuged (Sorvall/RC5B, SS34 rotor, 30 min, 13,000 rpm at 4 °C). The supernatant was mixed with 280 ml of 2 M GdmCl and then dialysed extensively against 50 mM potassium phosphate (pH 7.5), to induce protein refolding. The solution was centrifuged (Sorvall/RC5B, SS34 rotor, 30 min, 13,000 rpm at 4 °C) to separate protein aggregates. Most refolded HisF-C*C variants were pure to >95%. Otherwise, minor impurities were removed by applying nickel Sepharose chromatography as described above.

Wild-type HisF, HisF-C and HisF-CC were produced in *E. coli* BL21(DE3) as mentioned above and purified as described.[28]

Analysis of protein solubility

HisF and the HisF-C*C variants were produced in 50 ml cultures of *E. coli* BL21(DE3) cells as described above. From each culture a 2 ml aliquot was centrifuged (Eppendorf 5415R, 1 min, 13,000 rpm at 4 °C). The cells were resuspended in 1 ml of 100 mM potassium phosphate, 2 mM EDTA (pH 7.5), and lysed by sonification (Branson Sonifier W-250D, 45 sec, 15% pulse, 0 °C). The soluble and insoluble fractions of 100 μl of the cell extract were separated. The two preparations were supplemented with polyacrylamide gel loading buffer and applied at different dilutions on 12.5% SDS- polyacrylamide gels. The protein bands were subsequently stained with Coomassie blue and their intensities were quantified using OptiQuant (version 02.50). The protein bands shown in Figure 4(a) represent undiluted samples of soluble and insoluble cell fractions. To avoid quantification errors caused by a non-linear relationship between the staining signal and the amount of protein, protein bands of the soluble and insoluble fractions with similar staining intensity but different degrees of dilution were compared, followed by back calculation to the undiluted solutions. The data shown in Figure 4(b) are the result of three independent experiments, which allowed us to determine mean values and standard deviations.

Analytical methods

Analytical gel filtration was performed by using a calibrated Superdex 75 column (Amersham). In each case 0.15–0.2 mg of the protein was applied on the column and

eluted with a flow rate of 0.5 ml/min in 50 mM potassium phosphate, 300 mM potassium chloride (pH 7.5), at 25 °C. HisF-C*C 4 was analysed either in the presence or absence of 1 mM DTT. Fluorescence emission spectra of HisF, HisF-C*C and HisF-C**C were measured at 25 °C with a Cary Eclipse spectrophotometer (Varian) using 5 μM protein dissolved in 50 mM potassium phosphate (pH 7.5), and an excitation wavelength of 280 nm. The generated emission is mainly caused by the single tryptophan 156 of HisF which exists as two copies in HisF-C*C and HisF-C**C, due to its location in the C-terminal half of HisF. Near-UV CD spectra were recorded at 25 °C with an Aviv CD spectrometer 62-DS (d = 1 cm) using 68 μM HisF, 42 μM HisF-C*C and 54 μM HisF-C**C dissolved in 50 mM potassium phosphate (pH 7.5). The measured intensities were normalized to molar ellipticities.

Protein unfolding induced by urea or GdmCl was followed in 50 mM potassium phosphate (pH 7.5), by the decrease of the fluorescence intensity at 320 nm after excitation at 280 nm. The signal was measured after different time intervals until no further change was observed to ensure that the equilibrium was reached. The obtained intensities were normalised and plotted as fractional change of the native signal. This allowed us to determine the midpoint of unfolding $D_{1/2}$ [M], which represents the concentration of urea where 50% of the protein had non-native structure and served as an operational measure for conformational stability. Moreover, the unfolding traces were analysed according to the two-state model[66] to determine an apparent m-value, which served as an operational measure for the cooperativity of the unfolding process. All proteins could be refolded after complete unfolding in 8 M urea or 6 M GdmCl by removing the denaturant *via* dilution or dialysis against 50 mM potassium phosphate (pH 7.5). To investigate the stabilizing effect of phosphate on HisF-C**C the unfolding of this protein was also measured in 50 mM Tris–HCl (pH 7.5). To completely remove phosphate ions, purified HisF-C**C was unfolded by 8 M urea and refolded by extensive dialysis against 50 mM Tris–HCl (pH 7.5).

The ammonia dependent HisF activity was measured as described.[9] ΔASA (Å^2) values, which denote the difference of solvent exposed surface between the folded and unfolded states were calculated as described.[30]

Molecular dynamics simulations

The structural and energetic properties of the various proteins were analysed by molecular dynamics calculations using the molecular modelling program MOLOC‡[31] and the force field MAB[34] with implicit solvation. The constructs were built on the basis of the published X-ray structure of HisF (PDB code 1thf)[25] by a 180° rotation around the barrel axis of the C-terminal half barrel and its subsequent fitting onto the N-terminal half barrel using DeepView 3.7§. Subsequently, first the loops were relaxed and later the whole structure by minimization and 2 ns of molecular dynamics at 300 K. Starting from HisF-CC, models for HisF-C*C, HisF-C*C Δ, and HisF-C**C were constructed and relaxed by MD. For lack of a reliable method to calculate the free energy of the unfolded state, "interaction energies" between the linked half barrels were calculated using a pseudo-thermodynamic circle.[32] The energetic properties were analysed after an equilibration phase of 500 ps.

§ http://www.expasy.org/spdbv/

Acknowledgements

We thank Adriane Lochner for experimental help, and Dr. Birte Höcker for discussion and comments on the manuscript. This work was supported by grants of the Deutsche Forschungsgemeinschaft to M.B. (SPP 1170:BO 3066/1-1) and R.S. (STE 891/4-3, 4-4).

References

1. Lupas, A. N., Ponting, C. P. & Russell, R. B. (2001). On the evolution of protein folds: are similar motifs in different protein folds the result of convergence, insertion, or relics of an ancient peptide world? *J. Struct. Biol.* **134**, 191–203.
2. Riechmann, L. & Winter, G. (2006). Early protein evolution: building domains from ligand-binding polypeptide segments. *J. Mol. Biol.* **363**, 460–468.
3. Marcotte, E. M., Pellegrini, M., Thompson, M. J., Yeates, T. O. & Eisenberg, D. (1999). A combined algorithm for genome-wide prediction of protein function. *Nature*, **402**, 83–86.
4. Marcotte, E. M., Pellegrini, M., Ng, H. L., Rice, D. W., Yeates, T. O. & Eisenberg, D. (1999). Detecting protein function and protein-protein interactions from genome sequences. *Science*, **285**, 751–753.
5. Yadid, I. & Tawfik, D. S. (2007). Reconstruction of functional β-propeller lectins via homo-oligomeric assembly of shorter fragments. *J. Mol. Biol.* **365**, 10–17.
6. Gerlt, J. A. & Raushel, F. M. (2003). Evolution of function in $(\beta/\alpha)_8$-barrel enzymes. *Curr. Opin. Chem. Biol.* **7**, 252–264.
7. Glasner, M. E., Gerlt, J. A. & Babbitt, P. C. (2007). Mechanisms of protein evolution and their application to protein engineering. *Adv. Enzymol. Relat. Areas Mol. Biol.* **75**, 193–239.
8. Henn-Sax, M., Höcker, B., Wilmanns, M. & Sterner, R. (2001). Divergent evolution of $(\beta\alpha)_8$-barrel enzymes. *Biol. Chem.* **382**, 1315–1320.
9. Höcker, B., Beismann-Driemeyer, S., Hettwer, S., Lustig, A. & Sterner, R. (2001). Dissection of a $(\beta\alpha)_8$-barrel enzyme into two folded halves. *Nature Struct. Biol.* **8**, 32–36.
10. Nagano, N., Orengo, C. A. & 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.
11. Sterner, R. & Höcker, B. (2005). Catalytic versatility, stability, and evolution of the $(\beta\alpha)_8$-barrel enzyme fold. *Chem. Rev.* **105**, 4038–4055.
12. Pujadas, G. & Palau, J. (1999). TIM barrel fold: structural, functional and evolutionary characteristics in natural and designed molecules. *Biologica (Bratislava)*, **54**, 231–253.
13. Vega, M. C., Lorentzen, E., Linden, A. & Wilmanns, M. (2003). Evolutionary markers in the $(\beta/\alpha)_8$-barrel fold. *Curr. Opin. Chem. Biol.* **7**, 694–701.
14. Wierenga, R. K. (2001). The TIM-barrel fold: a versatile framework for efficient enzymes. *FEBS Letters*, **492**, 193–198.

15. Nicolet, Y. & Drennan, C. L. (2004). AdoMet radical proteins - from structure to evolution - alignment of divergent protein sequences reveals strong secondary structure element conservation. *Nucl. Acids Res.* **32**, 4015–4025.
16. Zitzewitz, J. A., Gualfetti, P. J., Perkons, I. A., Wasta, S. A. & Matthews, C. R. (1999). Identifying the structural boundaries of independent folding domains in the α subunit of tryptophan synthase, a β/α barrel protein. *Protein Sci.* **8**, 1200–1209.
17. Zitzewitz, J. A. & Matthews, C. R. (1999). Molecular dissection of the folding mechanism of the α subunit of tryptophan synthase: an amino-terminal autonomous folding unit controls several rate-limiting steps in the folding of a single domain protein. *Biochemistry*, **38**, 10205–10214.
18. Eder, J. & Kirschner, K. (1992). Stable substructures of eightfold $\beta\alpha$-barrel proteins: fragment complementation of phosphoribosylanthranilate isomerase. *Biochemistry*, **31**, 3617–3625.
19. Soberon, X., Fuentes-Gallego, P. & Saab-Rincon, G. (2004). *In vivo* fragment complementation of a $(\beta/\alpha)_8$ barrel protein: generation of variability by recombination. *FEBS Letters*, **560**, 167–172.
20. Bertolaet, B. L. & Knowles, J. R. (1995). Complementation of fragments of triosephosphate isomerase defined by exon boundaries. *Biochemistry*, **34**, 5736–5743.
21. Pan, H., Raza, A. S. & Smith, D. L. (2004). Equilibrium and kinetic folding of rabbit muscle triosephosphate isomerase by hydrogen exchange mass spectrometry. *J. Mol. Biol.* **336**, 1251–1263.
22. Silverman, J. A. & Harbury, P. B. (2002). The equilibrium unfolding pathway of a $(\beta/\alpha)_8$ barrel. *J. Mol. Biol.* **324**, 1031–1040.
23. Gu, Z., Zitzewitz, J. A. & 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.
24. Thoma, R., Schwander, M., Liebl, W., Kirschner, K. & Sterner, R. (1998). A histidine gene cluster of the hyperthermophile *Thermotoga maritima*: sequence analysis and evolutionary significance. *Extremophiles*, **2**, 379–389.
25. 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.
26. Höcker, B., Schmidt, S. & Sterner, R. (2002). A common evolutionary origin of two elementary enzyme folds. *FEBS Letters*, **510**, 133–135.
27. Heinz, D. W., Essen, L. O. & Williams, R. L. (1998). Structural and mechanistic comparison of prokaryotic and eukaryotic phosphoinositide-specific phospholipases C. *J. Mol. Biol.* **275**, 635–650.
28. 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.
29. Maxwell, K. L., Mittermaier, A. K., Forman-Kay, J. D. & Davidson, A. R. (1999). A simple *in vivo* assay for increased protein solubility. *Protein Sci.* **8**, 1908–1911.
30. Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138–2148.
31. Gerber, P. R. & Müller, K. (1995). MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry. *J. Comput. Aided Mol. Des.* **9**, 251–268.
32. van Gunsteren, W. F. & Berendsen, H. J. (1987). Thermodynamic cycle integration by computer simulation as a tool for obtaining free energy differences in molecular chemistry. *J. Comput. Aided Mol. Des.* **1**, 171–176.
33. Halgren, T. A. (1996). Merck Molecular Force Field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comp. Chem.* **17**, 520–552.
34. Gerber, P. R. (1998). Charge distribution from a simple molecular orbital type calculation and non-bonding interaction terms in the force field MAB. *J. Comput. Aided Mol. Des.* **12**, 37–51.
35. Vogel, C., Bashton, M., Kerrison, N. D., Chothia, C. & Teichmann, S. A. (2004). Structure, function and evolution of multidomain proteins. *Curr. Opin. Struct. Biol.* **14**, 208–216.
36. Orengo, C. A. & Thornton, J. M. (2005). Protein families and their evolution - a structural perspective. *Annu. Rev. Biochem.* **74**, 867–900.
37. Vogel, C., Teichmann, S. A. & Pereira-Leal, J. (2005). The relationship between domain duplication and recombination. *J. Mol. Biol.* **346**, 355–365.
38. Arnold, F. H. & Georgiou, G. (2003). Directed enzyme evolution - screening and selection methods. In *Methods in Molecular Biology* (Walker, J. M., ed), p. 230, Humana Press, Totowa, NJ.
39. Eijsink, V. G., Gaseidnes, S., Borchert, T. V. & van den Burg, B. (2005). Directed evolution of enzyme stability. *Biomol. Eng.* **22**, 21–30.
40. Hart, D. J. & Tarendeau, F. (2006). Combinatorial library approaches for improving soluble protein expression in *Escherichia coli*. *Acta Crystallog. sect. D*, **62**, 19–26.
41. Roodveldt, C., Aharoni, A. & Tawfik, D. S. (2005). Directed evolution of proteins for heterologous expression and stability. *Curr. Opin. Struct. Biol.* **15**, 50–56.
42. Waldo, G. S. (2003). Genetic screens and directed evolution for protein solubility. *Curr. Opin. Chem. Biol.* **7**, 33–38.
43. Sieber, V., Martinez, C. A. & Arnold, F. H. (2001). Libraries of hybrid proteins from distantly related sequences. *Nature Biotechnol.* **19**, 456–460.
44. Sieber, V. (2003). Selection for soluble proteins via fusion with chloramphenicol acetyltransferase. *Methods Mol. Biol.* **230**, 45–55.
45. Yano, J. K. & Poulos, T. L. (2003). New understandings of thermostable and peizostable enzymes. *Curr. Opin. Biotechnol.* **14**, 360–365.
46. Miyazono, K., Sawano, Y. & Tanokura, M. (2005). Crystal structure and structural stability of acylphosphatase from hyperthermophilic archaeon *Pyrococcus horikoshii OT3*. *Proteins: Struct. Funct. Genet.* **61**, 196–205.
47. Toth, E. A., Worby, C., Dixon, J. E., Goedken, E. R., Marqusee, S. & Yeates, T. O. (2000). The crystal structure of adenylosuccinate lyase from *Pyrobaculum aerophilum* reveals an intracellular protein with three disulfide bonds. *J. Mol. Biol.* **301**, 433–450.
48. Chiu, H. J., Johnson, E., Schröder, I. & Rees, D. C. (2001). Crystal structures of a novel ferric reductase from the hyperthermophilic archaeon *Archaeoglobus fulgidus* and its complex with $NADP^+$. *Structure*, **9**, 311–319.
49. Karlström, M., Stokke, R., Steen, I. H., Birkeland, N. K. & Ladenstein, R. (2005). Isocitrate dehydrogenase from the hyperthermophile *Aeropyrum pernix*:

X-ray structure analysis of a ternary enzyme-substrate complex and thermal stability. *J. Mol. Biol.* **345**, 559–577.
50. Arai, R., Kukimoto-Niino, M., Kuroishi, C., Bessho, Y., Shirouzu, M. & Yokoyama, S. (2006). Crystal structure of the probable haloacid dehalogenase PH0459 from *Pyrococcus horikoshii OT3*. *Protein Sci.* **15**, 373–377.
51. Mallick, P., Boutz, D. R., Eisenberg, D. & Yeates, T. O. (2002). Genomic evidence that the intracellular proteins of archaeal microbes contain disulfide bonds. *Proc. Natl Acad. Sci. USA*, **99**, 9679–9684.
52. Beeby, M., O'Connor, B. D., Ryttersgaard, C., Boutz, D. R., Perry, L. J. & Yeates, T. O. (2005). The genomics of disulfide bonding and protein stabilization in thermophiles. *PLoS Biol.* **3**, e309.
53. Dombkowski, A. A. (2003). Disulfide by design: a computational method for the rational design of disulfide bonds in proteins. *Bioinformatics*, **19**, 1852–1853.
54. Matthews, B. W., Nicholson, H. & Becktel, W. J. (1987). Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc. Natl Acad. Sci. USA*, **84**, 6663–6667.
55. Matsumura, M., Signor, G. & Matthews, B. W. (1989). Substantial increase of protein stability by multiple disulphide bonds. *Nature*, **342**, 291–293.
56. Betz, S. F. (1993). Disulfide bonds and the stability of globular proteins. *Protein Sci.* **2**, 1551–1558.
57. Mansfeld, J., Vriend, G., Dijkstra, B. W., Veltman, O. R., Van den Burg, B., Venema, G. *et al.* (1997). Extreme stabilization of a thermolysin-like protease by an engineered disulfide bond. *J. Biol. Chem.* **272**, 11152–11156.
58. Ivens, A., Mayans, O., Szadkowski, H., Jürgens, C., Wilmanns, M. & Kirschner, K. (2002). Stabilization of a $(\beta\alpha)_8$-barrel protein by an engineered disulfide bridge. *Eur. J. Biochem.* **269**, 1145–1153.
59. Sterner, R. & Liebl, W. (2001). Thermophilic adaptation of proteins. *Crit. Rev. Biochem. Mol. Biol.* **36**, 39–106.
60. Vieille, C. & Zeikus, G. J. (2001). Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* **65**, 1–43.
61. Walden, H., Bell, G. S., Russell, R. J., Siebers, B., Hensel, R. & Taylor, G. L. (2001). Tiny TIM: a small, tetrameric, hyperthermostable triosephosphate isomerase. *J. Mol. Biol.* **306**, 745–757.
62. Offredi, F., Dubail, F., Kischel, P., Sarinski, K., Stern, A. S., Van de Weerdt, C. *et al.* (2003). *De novo* backbone and sequence design of an idealized α/β-barrel protein: evidence of stable tertiary structure. *J. Mol. Biol.* **325**, 163–174.
63. 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.
64. Höcker, B. (2002). *Experimentelle Rekonstruktion der Evolution von Proteinen am Beispiel zweier (β α)8-Barrel Enzyme aus der Histidin-Biosynthese*. PhD thesis, University of Cologne, Germany.
65. Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*, **33**, 103–119.
66. Santoro, M. M. & Bolen, D. W. (1988). Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry*, **27**, 8063–8068.

Edited by F. Schmid

(Received 7 May 2007; received in revised form 5 June 2007; accepted 13 June 2007)
Available online 19 June 2007