Rapid Reports

High-Resolution Crystal Structure of an Artificial $(\beta\alpha)_8$-Barrel Protein Designed from Identical Half-Barrels[†,‡]

Birte Höcker,[*,§,||] Adriane Lochner,[||,⊥] Tobias Seitz,[⊥] Jörg Claren,[⊥] and Reinhard Sterner[*,⊥]

Max Planck Institute for Developmental Biology, Spemannstrasse 35, D-72076 Tübingen, Germany, and Institute of Biophysics and Physical Biochemistry, University of Regensburg, Universitätsstrasse 31, D-93053 Regensburg, Germany

Received November 18, 2008; Revised Manuscript Received December 17, 2008

ABSTRACT: Ample evidence suggests that the ubiquitous $(\beta\alpha)_8$-barrel enzyme fold has evolved by the duplication and fusion of an ancestral $(\beta\alpha)_4$-half-barrel. To reconstruct this process in the laboratory with a model protein, we earlier fused two copies of the C-terminal half-barrel HisF-C of imidazole glycerol phosphate synthase (HisF) and stepwise stabilized the resulting HisF-CC construct. We now further increased its stability and solubility by introducing two additional amino acid exchanges, which allowed us to crystallize the resulting artificial $(\beta\alpha)_8$-barrel protein HisF-C***C. The analysis of its X-ray structure at 2.1 Å resolution reveals a striking similarity to wild-type HisF, helps us to understand its improved stability, and provides further insights into the evolution of $(\beta\alpha)_8$-barrel proteins.

The molecular evolution of new protein structures and functions is largely based on the duplication and fusion of existing genes and gene fragments (*1, 2*). This becomes particularly obvious in the case of protein folds with internal sequence and structural symmetries (*3*). A prominent example is the $(\beta\alpha)_8$-barrel, which is the most frequent and most versatile enzyme fold (*4, 5*). The canonical $(\beta\alpha)_8$-barrel consists of a central eight-stranded β-sheet, the barrel, which is surrounded by the eight α-helices. The active site is generally located at the C-terminal end of the barrel and in the $\beta\alpha$-loops ("catalytic face"), whereas the N-terminal end of the barrel and the $\beta\alpha$-loops are important for the conformational stability of the fold ("stability face") (*6*). The striking 2-fold sequence and structural symmetry of the $(\beta\alpha)_8$-barrel enzymes *N*′-[(5′-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (ProFAR) isomerase (HisA) and imidazole glycerol phosphate synthase (HisF) suggests that these enzymes have evolved from a common $(\beta\alpha)_4$-half-barrel ancestor by a series of gene duplication and fusion events (*7*). We have reconstructed this process in the laboratory by fusing two identical copies of HisF-C, which is the C-terminal half-barrel of HisF, and by stabilizing the resulting construct HisF-CC stepwise to HisF-C*C and HisF-C**C, using rational design and library selection (*8, 9*). We have now further stabilized HisF-C**C by the introduction of two additional residue exchanges, thus generating HisF-C***C.

The N-terminal half of HisF-C*C contains the amino acid exchanges $A124_NR$ and $A220_NK$, which were introduced into HisF-CC to reconstitute a salt bridge cluster at the stability face of the central β-barrel of the wild-type HisF enzyme (*8*). The N-terminal half of HisF-C**C additionally contains a shortened loop connecting the two half-barrels and the two stabilizing exchanges $Y143_NH$ and $V234_NM$, which were identified by random mutagenesis and selection in vivo (*9*). The high internal 2-fold symmetry suggested to us that the introduction of the corresponding exchanges $Y143_CH$ and $V234_CM$ into the C-terminal half of HisF-C**C might further improve the properties of the resulting HisF-C***C protein (Figure 1a).

Heterologous expression in *Escherichia coli* showed that the fraction of recombinant HisF-C***C that was produced in soluble form (84%) was much larger than the soluble fraction of HisF-C**C (51%) and almost as large as the soluble fraction of the wild-type HisF enzyme (90%) (Figure 1b). The purified HisF-C***C protein unfolds in guanidinium chloride with cooperativity comparable to that of but at denaturant concentrations slightly higher than that of HisF-C**C (Figure 1c). Moreover, HisF-C***C unfolds with cooperativity similar to that of albeit at denaturant concentrations significantly lower than that of wild-type HisF, which must be extraordinarily stable to avoid denaturation in its hyperthermophilic host, *Thermotoga maritima*.

The HisF-C***C protein was crystallized, and its X-ray structure was determined at 2.1 Å resolution by molecular replacement with the coordinates of the C-terminal half of wild-type HisF [residues 123−253, Protein Data Bank (PDB) entry 1thf] (Table S1) (*7*). HisF-C***C forms a $(\beta\alpha)_8$-barrel structure with essentially the same topology and secondary structure content as HisF (Figure 2 and Figure S1). The areas around helices $\alpha8_N$ and $\alpha8_C$ are not well defined and could

[†] This work was supported by a grant from the Deutsche Forschungsgemeinschaft (STE 891/4-3) and institutional funds from the Max-Planck Society.

[‡] The atomic coordinates of HisF-C***C have been deposited in the Protein Data Bank as entry 2w6r.

* To whom correspondence should be addressed. B.H.: telephone, +49-7071-601-322; fax, +49-7071-601-305; e-mail, birte.hoecker@tuebingen.mpg.de. R.S.: telephone, +49-941-943-3015; fax, +49-941-943-2813; e-mail, Reinhard.Sterner@biologie.uni-regensburg.de.

[§] Max Planck Institute for Developmental Biology.

[||] These authors contributed equally to the experiments.

[⊥] University of Regensburg.

10.1021/bi802125b CCC: $40.75
Published on Web 01/23/2009

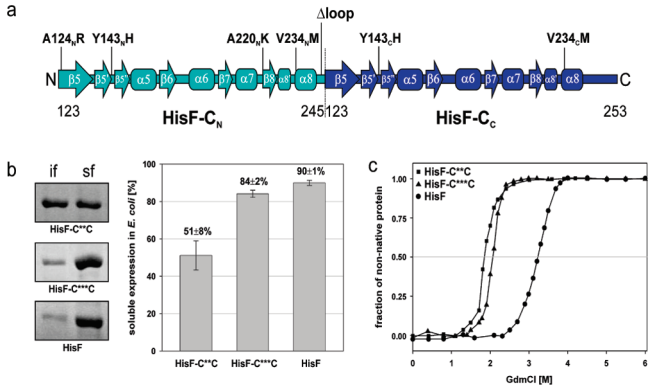

FIGURE 1: Design of the $(\beta\alpha)_8$-barrel protein HisF-C***C from two identical and fused HisF-C half-barrels. (a) Secondary structure elements and stabilizing amino acid exchanges. The exchanges in HisF-C$_N$ and the shortening of the loop connecting the two half-barrels were introduced in earlier work to stabilize HisF-CC via HisF-C*C to HisF-C**C (*8*, *9*). The exchanges Y143$_C$H and V234$_C$M were introduced in this work to further stabilize HisF-C**C to HisF-C***C. (b) Solubility analysis of HisF-C**C, HisF-C***C, and HisF. In the left panel, insoluble (if) and soluble (sf) fractions of *E. coli* host cell extracts analyzed by SDS−PAGE (12.5% acrylamide) and staining with Coomassie blue. In the right panel, the percentage of the proteins produced in soluble form (as deduced from the analysis of three different clones of each variant). (c) Stability of HisF-C**C, HisF-C***C, and HisF in 50 mM potassium phosphate (pH 7.5) at 25 °C. Chemical unfolding was induced by incubation in guanidinium chloride (GdmCl) and monitored by the decrease in fluorescence intensity at 320 nm following excitation at 280 nm. The signals were normalized, and the following $D_{1/2}$ values (in molar GdmCl) were determined: HisF-C**C, 1.90 ± 0.03 M; HisF-C***C, 2.10 ± 0.02 M; HisF, 3.32 ± 0.07 M.

not be built satisfactorily into the density (Figures S1 and S2). All other parts of the structure though have a high resolution and can be compared to the parent HisF structure and an earlier calculated structural model of HisF-C**C (*9*).

The structure of HisF-C***C superimposes nicely with that of HisF. The overall root-mean-square deviation (rmsd) of the two proteins is 1.7 Å, and the rmsd values for the superimposed N-terminal and C-terminal halves are 1.6 and 0.6 Å, respectively. HisF-C***C shows an extremely high 2-fold symmetry, with a rmsd of only 0.6 Å for its two superimposed HisF-C$_N$ and HisF-C$_C$ halves, compared to a rmsd of 1.6 Å for the superimposed HisF-N and HisF-C halves of wild-type HisF (Figure 2). The two aspartate residues of HisF-C***C, D130$_N$ and D130$_C$, which are

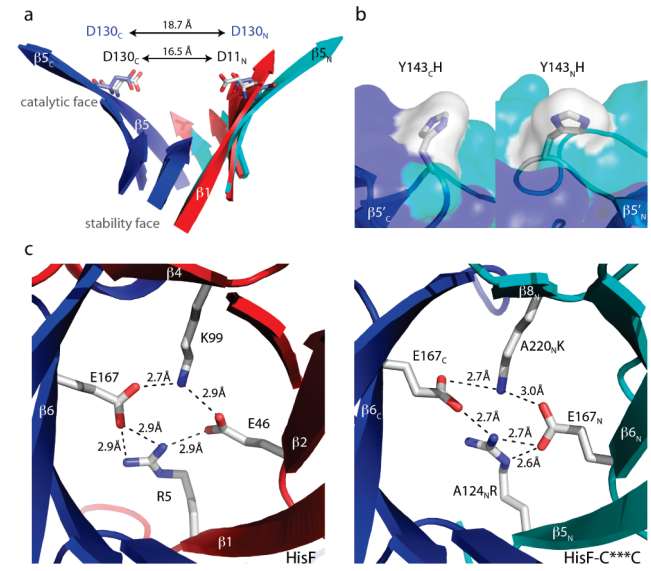

FIGURE 3: Important structural features of HisF-C***C in comparison to those of HisF. (a) The β-barrel of HisF-C***C opens up wider at the catalytic face than HisF, resulting in a distance increase in the functional aspartate residues of ∼2 Å. (b) The exchanges Y143$_N$H and Y143$_C$H, which contribute to the solubility and stability of HisF-C***C, are solvent-exposed in the crystal structure. (c) The stabilizing salt bridge cluster of HisF is formed in a similar way in HisF-C***C through the exchanges A124$_N$R and A220$_N$K. Hydrogen bonds between the involved side chain residues, which are shown as sticks, are indicated by dashed lines up to a distance cutoff of 3.0 Å. Colors as in Figure 2.

located at the C-terminal ends of strands $\beta5_N$ and $\beta5_C$, respectively, correspond to catalytic aspartate residues D11 and D130 of HisF, which are located at the C-terminal ends of strands $\beta1$ and $\beta5$, respectively (Figure 2 and Figure S1).

We therefore tested whether HisF-C***C has HisF activity by incubating high concentrations of the protein (50 μM) with the substrates ammonia and *N*′-[(5′-phosphoribulosyl)-formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (PRFAR). However, no production of ImGP could be detected in the spectroscopic HisF assay (*10*). This lack of measurable activity could be due to the increased distance of catalytic aspartates D130$_N$ and D130$_C$ in comparison to D11 and D130 (Figure 3a). Moreover, although the addition of orthophosphate stabilizes HisF-C***C, we did not find any evidence for the binding of the bisphosphorylated substrate analogue ProFAR (*10*). As the N-terminal phosphate binding site of HisF-C***C is slightly shifted with

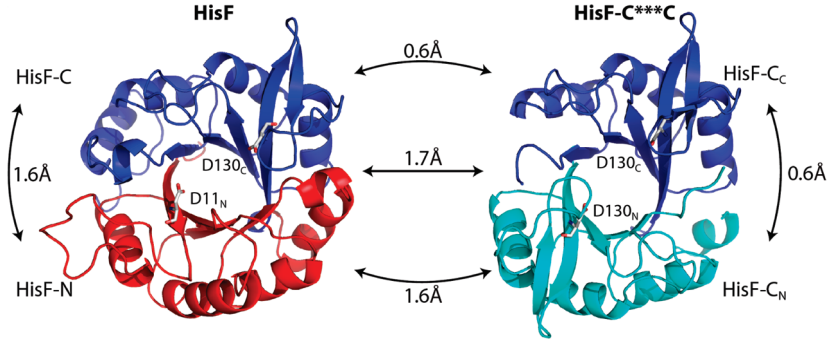

FIGURE 2: Ribbon diagrams of the crystal structures of HisF (PDB entry 1thf) and HisF-C***C. Top view onto the catalytic face (cp. Figure 3a). The functional aspartate residues D11$_N$ and D130$_C$ of HisF and the corresponding D130$_N$ and D130$_C$ residues of HisF-C***C are shown as sticks. The indicated root-mean-square deviations were calculated for the following pairs: HisF−HisF-C***C, 1.7 Å (184 superimposed Cα atoms); N-terminal halves of the two proteins, 1.6 Å (89 Cα atoms); C-terminal halves of the two proteins, 0.6 Å (103 Cα atoms); the halves within HisF, 1.6 Å (89 Cα atoms); the halves within HisF-C***C, 0.6 Å (103 Cα atoms).

respect to that of HisF (not shown), the proper anchoring of the corresponding phosphate group of ProFAR and PRFAR in the designed active site might be hampered, resulting in a low binding affinity.

The X-ray structure further allowed us to analyze the stabilizing effects of the amino acid exchanges that were introduced to generate HisF-C***C from HisF-CC. The histidine residues introduced by library selection [$Y143_N$H (*9*)] and site-directed mutagenesis [$Y143_C$H (this work)] point into the solvent (Figure 3b), which allows them to engage in favorable interactions with water, explaining why these substitutions increase both solubility and stability. Crystal contacts (not shown) appear to stabilize this conformation compared to an alternative one suggested by molecular dynamics simulation of HisF-C**C, in which the imidazole group of $H143_N$ forms a hydrogen bond network with the carboxylate group of $D176_C$ and the phosphate ion bound to the C-terminal half-barrel (*9*). The methionine residues introduced at position 234 are not resolved in the crystal structure, neither in the N-terminal nor in the C-terminal half. Their contribution to increased stability or solubility thus cannot be evaluated on the basis of this structure. A structural model of HisF-C**C, however, suggests that the long methionine side chain protrudes into a hydrophobic pocket of the adjacent half-barrel to form favorable van der Waals interactions (*9*).

Inspection of the structure shows that exchanges $A124_N$R and $A220_N$K at the N-terminal face of the central barrel result in the formation of a salt bridge cluster involving $E167_N$ and $E167_C$, as expected (Figure 3c). This rationally designed salt bridge cluster strengthens the interaction between the N- and C-terminal half-barrel, explaining why the introduced exchanges are beneficial both for the compactness and for the stability of the protein (*8*).

Following the fusion of two identical HisF-C half-barrels, the shortening of the connecting loop and as little as six amino acid exchanges were sufficient to generate the highly stable protein HisF-C***C. The artificial $(\beta\alpha)_8$-barrel shows a high degree of structural similarity to the wild-type HisF enzyme. This finding is remarkable, given that the N-terminal half-barrels HisF-C_N and HisF-N share a degree of sequence identity of only ~25%, and demonstrates that $(\beta\alpha)_8$-barrels can be readily evolved in the laboratory from a half-barrel. Our results make it likely that similar gene duplication and fusion events have led to the generation of an ancestral $(\beta\alpha)_8$-barrel during natural evolution (*6*). Although the sequences of HisF-C and HisF-C***C certainly differ from those of the ancestral proteins due to genetic drift, they are good models for these proteins and the putative evolutionary events connecting them. It will now be interesting to test whether and which catalytic activities can be established on the HisF-C***C scaffold and how many mutations will be necessary to do so. The structure presented here shows how the barrel of HisF-C***C opens up wider than the barrel of HisF, resulting in a larger binding pocket (Figure 3a). This observation and the absence of detectable HisF activity illustrate that a fine-tuning of the active site will be required, which might be hard to achieve by rational approaches and probably will require directed evolution techniques, i.e., random mutagenesis followed by selection or screening.

The regular order of the $(\beta\alpha)$-units has prompted experiments aimed at designing artificial $(\beta\alpha)_8$-barrel enzymes early on. However, the first attempts yielded molten globules with defined secondary structures at best (*11–13*). More recently, a $(\beta\alpha)_8$-barrel that consists of an idealized artificial backbone containing an eight-stranded β-sheet with 4-fold symmetry surrounded by eight α-helices was designed. An optimal amino acid sequence fitting the idealized backbone was then chosen with an automated selection algorithm (*14*). Although the designed protein is stable and appears to adopt a well-defined tertiary structure in solution, no high-resolution X-ray or NMR structure has been published until now. HisF-C***C is therefore the first artificial $(\beta\alpha)_8$-barrel for which a high-resolution structure is available, complementing structural analysis of an artificial nine-stranded barrel that was constructed by fusing HisF-C to parts of a flavodoxin-like $(\beta\alpha)_5$-protein (*15*).

SUPPORTING INFORMATION AVAILABLE

A description of the experimental procedures with supporting references, data collection and refinement statistics from the X-ray structure determination of HisF-C***C (Table S1), structure-based sequence alignment of HisF and HisF-C***C (Figure S1), and electron density maps around helices $\alpha8_N$ and $\alpha8_C$ of HisF-C***C (Figure S2). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Bashton, M., and Chothia, C. (2007) *Structure 15*, 85–99.
2. Chothia, C., Gough, J., Vogel, C., and Teichmann, S. A. (2003) *Science 300*, 1701–1703.
3. Yadid, I., and Tawfik, D. S. (2007) *J. Mol. Biol. 365*, 10–17.
4. Gerlt, J. A., and Raushel, F. M. (2003) *Curr. Opin. Chem. Biol. 7*, 252–264.
5. Wierenga, R. K. (2001) *FEBS Lett. 492*, 193–198.
6. Sterner, R., and Höcker, B. (2005) *Chem. Rev. 105*, 4038–4055.
7. Lang, D., Thoma, R., Henn-Sax, M., Sterner, R., and Wilmanns, M. (2000) *Science 289*, 1546–1550.
8. Höcker, B., Claren, J., and Sterner, R. (2004) *Proc. Natl. Acad. Sci. U.S.A. 101*, 16448–16453.
9. Seitz, T., Bocola, M., Claren, J., and Sterner, R. (2007) *J. Mol. Biol. 372*, 114–129.
10. Beismann-Driemeyer, S., and Sterner, R. (2001) *J. Biol. Chem. 276*, 20387–20396.
11. Beauregard, M., Goraj, K., Goffin, V., Heremans, K., Goormaghtigh, E., Ruysschaert, J. M., and Martial, J. A. (1991) *Protein Eng. 4*, 745–749.
12. Tanaka, T., Kuroda, Y., Kimura, H., Kidokoro, S., and Nakamura, H. (1994) *Protein Eng. 7*, 969–976.
13. Wodak, S. J., Lasters, I., Pio, F., and Claessens, M. (1990) *Biochem. Soc. Symp. 57*, 99–121.
14. Offredi, F., Dubail, F., Kischel, P., Sarinski, K., Stern, A. S., Van de Weerdt, C., Hoch, J. C., Prosperi, C., Francois, J. M., Mayo, S. L., and Martial, J. A. (2003) *J. Mol. Biol. 325*, 163–174.
15. Bharat, T. A., Eisenbeis, S., Zeth, K., and Höcker, B. (2008) *Proc. Natl. Acad. Sci. U.S.A. 105*, 9942–9947.

BI802125B