Hypothesis

A common evolutionary origin of two elementary enzyme folds

Birte Höcker[a], Steffen Schmidt[b], Reinhard Sterner[a,*]

[a] *Universität zu Köln, Institut für Biochemie, Otto-Fischer-Strasse 12–14, D-50674 Cologne, Germany*
[b] *European Molecular Biology Laboratory (EMBL) Heidelberg, Meyerhofstrasse 1, D-69117 Heidelberg, Germany*

Received 2 November 2001; accepted 26 November 2001

First published online 6 December 2001

Edited by Gunnar von Heijne

Abstract The $(\beta\alpha)_8$-barrel is the most frequent and most versatile fold among enzymes [Höcker et al., Curr. Opin. Biotechnol. 12 (2001) 376–381; Wierenga, FEBS Lett. 492 (2001) 193–198]. Structural and functional evidence suggests that $(\beta\alpha)_8$-barrels evolved from an ancestral half-barrel, which consisted of four $(\beta\alpha)$ units stabilized by dimerization [Lang et al., Science 289 (2000) 1546–550; Höcker et al., Nat. Struct. Biol. 8 (2001) 32–36; Gerlt and Babbitt, Nat. Struct. Biol. 8 (2001) 5–7]. Here, by performing a comprehensive database search, we detect a striking and unexpected structural and amino acid sequence similarity between $(\beta\alpha)_4$ half-barrels and members of the $(\beta\alpha)_5$ flavodoxin-like fold. These findings provoke the hypothesis that a large fraction of the modern-day enzymes evolved from a basic structural building block, which can be identified by a combination of sequence and structural analyses.

Key words: Enzyme evolution; $(\beta\alpha)_8$-barrel; Flavodoxin-like fold; Histidine biosynthesis; TIM-barrel

1. Introduction

Proteins that have the same major secondary structure elements in the same spatial arrangement and the same topological connections have the same fold [6]. According to this definition, about 10% of all proteins with known three-dimensional structure contain a $(\beta\alpha)_8$-barrel fold, which is the most common fold among enzymes [1,2,7]. It is of fundamental importance to understand how modern $(\beta\alpha)_8$-barrel enzymes, which are highly specific and efficient biocatalysts, have evolved from less specific and efficient precursors. The canonical $(\beta\alpha)_8$-barrel consists of a closed eight-stranded parallel β-sheet, forming a central barrel, which is surrounded by eight α-helices that constitute the outer layer of the structure. The $(\beta\alpha)_8$-barrel is generally regarded as a single structural domain. However, the X-ray structures of *N*′-[(5′-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide-ribonucleotide isomerase (HisA) and imidazoleglycerol phosphate synthase (HisF) from *Thermotoga maritima* showed that these $(\beta\alpha)_8$-barrel enzymes consist of two superimposable half-domains that display significant amino acid sequence similarities. The half-domains designated HisA-N and HisF-N comprise the four N-terminal $(\beta\alpha)$ units, and the half-domains HisA-C and HisF-C comprise the four C-terminal $(\beta\alpha)$ units [3]. The separately produced HisF-N and HisF-C half-domains are stable proteins with native secondary and tertiary structures, which form a functional heterodimer [4]. These data suggest that HisF and HisA evolved from an ancestral half-barrel, which consisted of four $(\beta\alpha)$ units stabilized by dimerization [5].

2. Results and discussion

The hypothesis that the 'half-barrel' is an independently evolving structural domain would be strongly supported if it could be identified in protein folds other than the $(\beta\alpha)_8$-barrel. Therefore, we searched available databases for proteins with significant sequence similarities to either HisF-N or HisF-C, using the programs HMMer [8] and PSI-Blast [9]. However, only a number of $(\beta\alpha)_8$-barrel enzymes were identified in these searches (data not shown). The fact that two proteins do not show significant sequence identity does not exclude, however, their evolutionary relationship [10,11], because nature selects for structural and functional integrity instead of amino acid conservation [12]. Therefore, in order to identify more distant evolutionary relationships, the coordinates of the HisF-N and HisF-C half-barrels were compared with the coordinates of all protein structures of the Protein Data Bank, using the program DALI [13]. Remarkably, besides HisF, HisA and the $(\beta\alpha)_8$-barrel enzyme phosphoenolpyruvate mutase, enzymes with the flavodoxin-like fold gave the best hits in these searches, yielding *Z*-scores higher than 6 and amino acid sequence identities of up to 22% (Table 1). These values suggest that the half-barrels and the flavodoxin-like fold are evolutionarily related [10,11]. The flavodoxin-like fold is a $(\beta\alpha)_5$-fold, comprising five parallel β-strands and five α-helices, which are connected by loops. This fold is frequently encountered, either as an isolated domain or as part of multi-domain enzyme structures [6]. Fig. 1A presents a comparison of the topologies of the half-barrels of HisA and HisF with the flavodoxin-like domain of methylmalonyl-CoA mutase (MMCoA). It is evident that four of the five $(\beta\alpha)$ modules of MMCoA correspond to the four $(\beta\alpha)$ modules of the half-barrels. The second $(\beta\alpha)$ module of MMCoA corresponds to an extra two-stranded β-sheet of the half-barrels of HisA and HisF that are located in the loops connecting β-strands 1/5 with α-helices 1/5 [3]. Fig. 1B presents the superposition of the backbone structure of HisF-C (PDB entry code: 1thf) with the $(\beta\alpha)$ modules 1, 3, 4 and 5 of MMCoA (PDB entry code:

*Corresponding author. Fax: (49)-221-4706731.
E-mail address: reinhard.sterner@uni-koeln.de (R. Sterner).

0014-5793/02/$22.00
PII: S0014-5793(01)03232-X

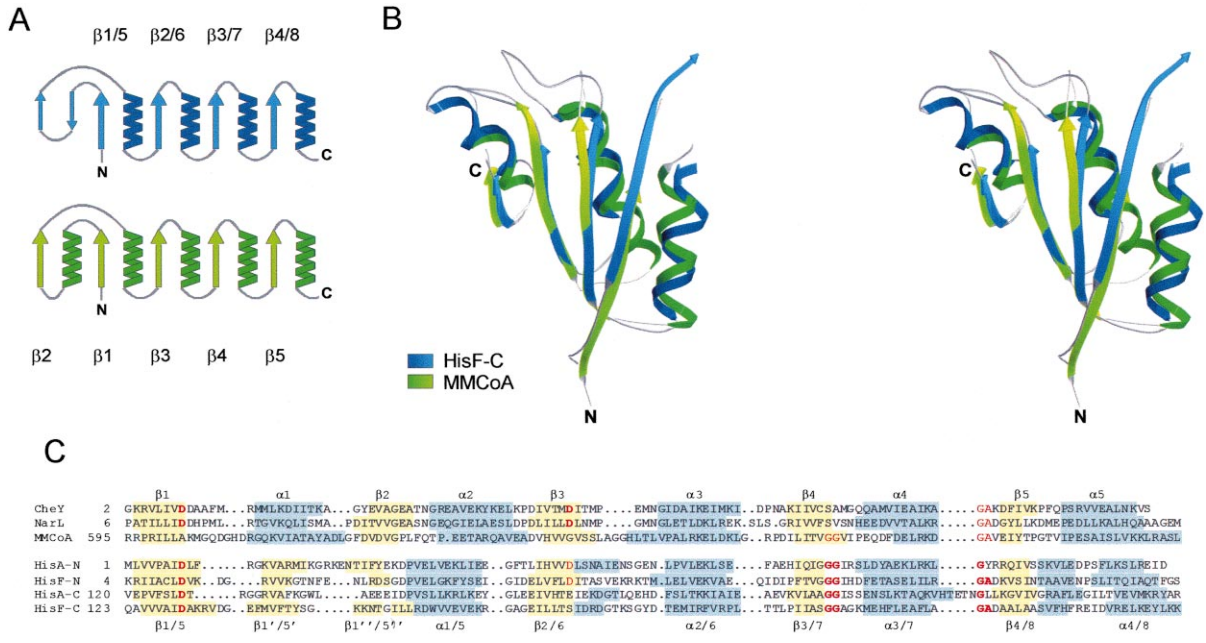

Fig. 1. Conservation of structure and of functionally important amino acid residues between half-barrels and enzymes with the flavodoxin-like fold. A: Topological diagram of N-terminal ($(\beta\alpha)_{1-4}$) and C-terminal ($(\beta\alpha)_{5-8}$) half-barrels (in blue) and the flavodoxin-like $(\beta\alpha)_5$-fold (in green). B: Superposition (in stereo) of ribbon diagrams of the C-terminal half-barrel of HisF from *T. maritima* (HisF-C, in blue; PDB entry code: 1thf) and the flavodoxin-like domain of MMCoA mutase from *Propionibacterium shermanii* (MMCoA, in green; PDB entry code: 1reqA). The superposition was performed with the program STAMP [21]. It yielded an r.m.s. deviation of 1.6 Å for 67 superimposed C_α-atoms. The additional two-stranded β-sheet of HisF-C and the corresponding β-strand 2 and α-helix 2 of MMCoA (cf. A), as well as the C-terminal helices, are omitted for clarity. C: Structure-based amino acid sequence alignment of the flavodoxin-like domains CheY from *T. maritima*, NarL from *E. coli* and MMCoA with the N- and C-terminal half-barrels HisA-N, HisF-N, HisA-C, HisF-C. Residues that are essential for the function of an enzyme are marked in bold red, the corresponding residues in the aligned enzymes are marked in faint red. Amino acids belonging to β-strands and α-helices are marked in yellow and blue (see text for details).

1reqA). The structures are very similar, with an r.m.s. deviation of only 1.6 Å for 67 C_α-atoms. Fig. 1C presents a structure-based amino acid sequence alignment of the half-barrels of HisA and HisF with the flavodoxin-like enzymes MMCoA, CheY from *T. maritima* and NarL (nitrite response regulator protein) from *Escherichia coli*. This alignment supports the equivalence of the individual secondary structure elements of the various enzymes. Moreover, it shows that a number of conserved amino acids that are essential for enzyme function are located at structurally equivalent positions in the half-barrels and the flavodoxin-like enzymes. All known $(\beta\alpha)_8$-barrel enzymes have their active site residues at the C-terminal ends of the β-strands or in the loops that link β-strands with the subsequent α-helices [1]. In HisA and HisF, the two catalytically essential aspartate residues are located at the C-terminal ends of β-strands 1 and 5 [3,14]. These aspartate residues align with a conserved aspartate residue at the C-terminal end of β-strand 1 of CheY and NarL, which is involved in binding of an essential Mg^{2+} ion at the active site [15]. Likewise, the aspartate residue at the C-terminal end of β-strand 3 in CheY and NarL, the phosphorylation of which leads to activation of the protein within the two-component system [16], aligns with an aspartate residue that is highly conserved at the C-terminal end of β-strand 3 in HisA-N and HisF-N. Furthermore, two glycine-rich blocks of amino acids between β-strands 3/7 and 4/8, some of which are important for binding of the phosphate moieties of the bis-phosphorylated substrates of HisA and HisF [17], are partially conserved in CheY, NarL and MMCoA (Fig. 1C).

Table 1
DALI searches [13] with the HisF-N and HisF-C half-barrels identify significant structural similarities to other $(\beta\alpha)_8$-barrels and enzymes with the flavodoxin-like fold

	Hit #	*Z*-score	SeqID (%)
HisF-N vs.			
HisF-N	1	26.5	100
HisF	2	22.1	99
HisA	3	11.3	24
PEP mutase	4	6.4	12
NarL	5	6.4	19
MMCoA	6	6.3	22
HisF-C vs.			
HisF-C	1	28.0	100
HisF	2	20.8	100
HisA	3	9.5	19
MMCoA	4	6.4	8
DHODA	5	5.3	11
MLE	6	5.2	12

Z-score: relative measure for the deviation of the similarity of two structures from the expected mean similarity of unrelated proteins. SeqID (%): percentage of amino acid sequence identity over equivalenced positions. PEP mutase: phosphoenolpyruvate mutase (1pym); *NarL* (1a04); *MMCoA* (1req); DHODA: dihydroorotate dehydrogenase A (2dor); MLE: muconate lactonizing enzyme (1muc).

3. Conclusions

Comprehensive comparisons of sequence, structure and function are required to support the common evolutionary origin of proteins [18]. The results presented in this work show that the structures of the HisF-N and HisF-C half-bar-

rels are strikingly similar to a number of proteins adopting the flavodoxin-like fold (Table 1, Fig. 1A,B). Moreover, there is a local conservation of sequence, which comprises amino acid residues that are essential for catalytic activity (Fig. 1C). These similarities suggest a common evolutionary origin of the $(\beta\alpha)_4$ half-barrels and of enzymes with the flavodoxin-like fold. The absence of a detectable overall amino acid sequence identity between the two folds is not surprising, given the low sequence identities among different $(\beta\alpha)_8$-barrel enzymes, which can often not be detected even by sophisticated methods [7]. For example, the last three enzymes of tryptophan biosynthesis, which are evolutionarily related $(\beta\alpha)_8$-barrels with phosphorylated substrates, display detectable sequence similarities only at their phosphate-binding sites [19]. The most striking difference between the two folds is the presence of an additional $(\beta\alpha)$ module in the flavodoxin-like fold, which corresponds to a two-stranded β-sheet in the half-barrels of HisA and HisF [3] (Fig. 1A). It was recently shown that a single amino acid exchange is sufficient to convert a β-strand into an α-helix at the monomer–monomer interface of the dimeric Arc repressor protein [20]. An analogous event may have led to the observed difference between the $(\beta\alpha)_4$ half-barrels and the flavodoxin-like fold. As a consequence, shielding of the hydrophobic interior of half-barrels from solvent was accomplished either by dimerization, as observed for the HisF halves [4], or by the fifth $(\beta\alpha)$ element, as observed for the flavodoxin-like proteins. The resulting stable folds are found in a broad variety of the contemporary single- and multi-domain enzymes.

Acknowledgements: We thank P. Bork, H. Decker, K. Kirschner, D. Schomburg, T. Schweikardt and M. Wilmanns for comments on the manuscript, and the Deutsche Forschungsgemeinschaft for financial support (STE 891/3-1).

References

[1] Höcker, B., Jürgens, C., Wilmanns, M. and Sterner, R. (2001) Curr. Opin. Biotechnol. 12, 376–381.
[2] Wierenga, R.K. (2001) FEBS Lett. 492, 193–198.
[3] Lang, D., Thoma, R., Henn-Sax, M., Sterner, R. and Wilmanns, M. (2000) Science 289, 1546–1550.
[4] Höcker, B., Beismann-Driemeyer, S., Hettwer, S., Lustig, A. and Sterner, R. (2001) Nat. Struct. Biol. 8, 32–36.
[5] Gerlt, J.A. and Babbitt, P.C. (2001) Nat. Struct. Biol. 8, 5–7.
[6] Murzin, A.G., Brenner, S.E., Hubbard, T. and Chothia, C. (1995) J. Mol. Biol. 247, 536–540.
[7] Copley, R.R. and Bork, P. (2000) J. Mol. Biol. 303, 627–640.
[8] Eddy, S.R. (1998) Bioinformatics 14, 755–763.
[9] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
[10] Hadley, C. and Jones, D.T. (1999) Struct. Fold. Des. 7, 1099–1112.
[11] Lichtarge, O. (2001) Nat. Struct. Biol. 8, 918–920.
[12] Dietmann, S. and Holm, L. (2001) Nat. Struct. Biol. 8, 953–957.
[13] Holm, L. and Sander, C. (1993) J. Mol. Biol. 233, 123–138.
[14] Beismann-Driemeyer, S. and Sterner, R. (2001) J. Biol. Chem. 276, 20387–20996.
[15] Lukat, G.S., Stock, A.M. and Stock, J.B. (1990) Biochemistry 29, 5436–5442.
[16] Stock, A.M., Robinson, V.L. and Goudreau, P.N. (2000) Annu. Rev. Biochem. 69, 183–215.
[17] Thoma, R., Schwander, M., Liebl, W., Kirschner, K. and Sterner, R. (1998) Extremophiles 2, 379–389.
[18] Grishin, N.V. (2001) J. Struct. Biol. 134, 167–185.
[19] Wilmanns, M., Hyde, C.C., Davies, D.R., Kirschner, K. and Jansonius, J.N. (1991) Biochemistry 30, 9161–9169.
[20] Cordes, M.H.J., Walsh, N.P., McKnight, C.J. and Sauer, R.T. (1999) Science 284, 325–327.
[21] Russell, R.B. and Barton, G.J. (1992) Proteins 14, 309–323.