Biochemistry
Communication
Cite This: Biochemistry 2019, 58, 4790−4793
pubs.acs.org/biochemistry

Reconstructing the Remote Origins of a Fold Singleton from a Flavodoxin-Like Ancestor

Saacnicteh Toledo-Patiño,[†,‡] Manish Chaubey,[‡] Murray Coles,[‡] and Birte Höcker*,[†,‡]

[†]Department of Biochemistry, University of Bayreuth, 95447 Bayreuth, Germany
[‡]Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany

Supporting Information

ABSTRACT: Evolutionary processes that led to the emergence of structured protein domains left footprints in the sequences of modern proteins. We searched for such hints employing state-of-the-art sequence analysis and found evidence that the HemD-like fold emerged from the flavodoxin-like fold through segment swap and gene duplication. To verify this hypothesis, we reverted these evolutionary steps experimentally, constructing a HemD-half that resulted in a protein with the canonical flavodoxin-like architecture. These results of fold reconstruction from the sequence of a different fold strongly support our hypothesis of common ancestry. It further illustrates the plasticity of modern proteins to form new folded proteins.

Proteins are the nanomachines of life. They are built from only 20 different amino acids, yet their enormous versatility covers nearly all fundamental requirements to sustain life. In the past 60 years, more than 40 thousand unique protein structures have been solved, revealing that in spite of their multiple functions and shapes, all proteins result from the combination of a limited number of independently folding units, the domains. To date, little is known about domain emergence. Previous structural and sequence analysis strongly suggest that they arose through a combination of a recurrent set of subdomain-sized fragments, rather than *de novo*.[1,2] Since similar structures can be the result of convergent evolution,[3] sequence-based evidence is required to confirm whether two proteins share a common origin. This represents a challenge when sequence identities are low, as it is the case for remote homologues. To overcome this, sensitive sequence analysis methods that employ profile-profile comparison of Hidden-Markov-Models (HMMs) have been developed and applied.[4,5] The current work makes use of this robust tool to explore the evolutionary events that mediated the emergence of the HemD-like fold.

The HemD-like fold is a fold singleton, meaning it is adopted by a single homonymous superfamily, which encloses only variants of one known enzyme, the uroporphyrinogen III synthase (U3S). This enzyme is present in all kingdoms of life and catalyzes the synthesis of uroporphyrinogen III, an intermediate molecule in the biosynthesis of essential cofactors such as heme, chlorophyll, F430, and cobalamin. Malfunction of this essential enzyme leads to congenital erytroporia porphyria, a rare disease that causes severe skin photosensitivity.[6] At a structural level, U3S displays a symmetrical bilobular architecture (Figure 1), whose hinge region provides space for ligand binding and catalysis. Previous structural studies on U3S led to the proposal of two possible evolutionary scenarios for its emergence: (1) duplication and circular permutation or (2) duplication, swapping, and fusion.[7]

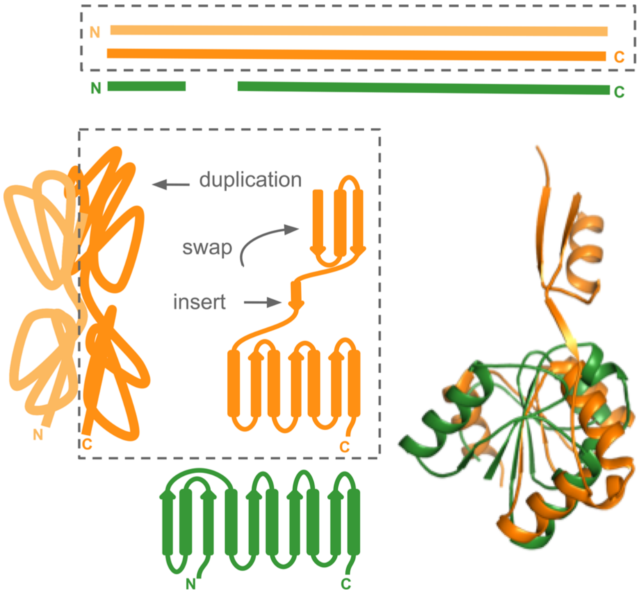

Figure 1. Sequence-based profile alignments compared to their structural superpositions. HHpred profile comparisons showed N- and C-terminal sequences of hemD-like halves (shades of orange) to align with each other and with flavodoxin-like proteins (green). At a structural level, hemD-like halves correspond to each other but not to the flavodoxin-like domain due to an $\alpha\beta\alpha$-swap as shown schematically and in a superposition of *P. aeruginosa* U3S (PDB ID 4es6) and *T. thermophilus* LitR (PDB ID 3whp).

However, insufficient sequence evidence has been available to distinguish them. In their work, Szilágyi et al.[7] also state that a swapping event would be difficult to revert, since the duplicated halves, though originally identical, have diverged substantially and would no longer stabilize the “unswapped” domains through residue-residue contacts.

Received: October 9, 2019
Revised: November 12, 2019
Published: November 14, 2019

Here we provide sequence-based evidence that supports an alternative emergence path for the HemD-like fold, consisting of an insert-assisted segment swap, gene duplication, and fusion originating from the flavodoxin-like fold. To test this hypothesis experimentally, we further set out to revert these events, one by one, starting from U3S and yielding a well-folded protein, which in fact adopts the canonical flavodoxin-like architecture. Our results not only offer a plausible path of emergence for HemD-like proteins but also provide another example of common ancestry across folds.

■ SEQUENCE AND STRUCTURE COMPARISONS

Profile-based comparison methods provide powerful tools for detecting remote homology. The efficiency of this method relies on the recruitment of information from multiple sequences rather than single ones. This information is stored as a scoring matrix (profile) that contains the likelihood of mutations, deletions, or extensions to occur at each position of a multiple sequence alignment. We generated such a profile for all U3S proteins of known structure using HHpred[8] and compared it against pdb70 and scop70. The gathered alignments showed that N-terminal halves aligned with their C-terminal counterparts, suggesting a gene duplication event (Table S1). In addition, U3S halves also align well with proteins adopting a flavodoxin-like fold, suggesting that the duplication originated from this scaffold. We proceeded to manually truncate N- and C-termini of the previously employed U3S queries to generate a second profile database and assessed a new search. This yielded alignments that recruited more flavodoxin-like proteins. The best-scoring alignment corresponded to the B12-binding domain of LitR from *Thermus thermophilus* (PDB 3whp, residues 164−271), a protein from the cobalamin-binding superfamily, which aligns with 30% identity over 105 residues with the C-terminal half of U3S from *Pseudomonas aeruginosa* (Figure S1). This value is remarkably high and statistically relevant according to a sequence homology benchmark by Sander & Schneider,[9] which estimates that 24.8% sequence identity is required in order to validate homology of sequence alignments ≥80 amino acids in length. The sequence identity between LitR and the U3S C-terminal half thus strongly supports common ancestry between these folds.

At a structural level, N- and C-terminal halves of HemD-like proteins align over regions that extend beyond the lobes. These regions even align with RMSDs of up to 2.1 Å over >100 residues (Cα), whereas none of the U3S lobes can be superimposed over more than 75 residues (Figure S3). Looking at the HemD-like topology, it becomes apparent that flavodoxin-like proteins do not superimpose well with the N- and C-terminal halves of U3S due to a difference in architecture, an $\alpha\beta\alpha$-segment swap that is caused by an insertion of about six amino acids only present in the HemD-like profiles (Figure 1 and Figure S1). This suggests that the insert may have mediated the swap rather than circular permutation, as previously suggested.

To investigate the effect of a circular permutation event, we permutated our database of halves manually, generating their sequence profiles to assess in a third search against pdb70 and scop70. The alignments showed that none of the permuted halves identify flavodoxin-like proteins with higher sequence identities than the original halves. Moreover, none of the permutated sequences showed an identity toward flavodoxin-like above the homology threshold, speaking in favor of insertion-mediated segment swapping.

■ EXPERIMENTAL RECONSTRUCTION

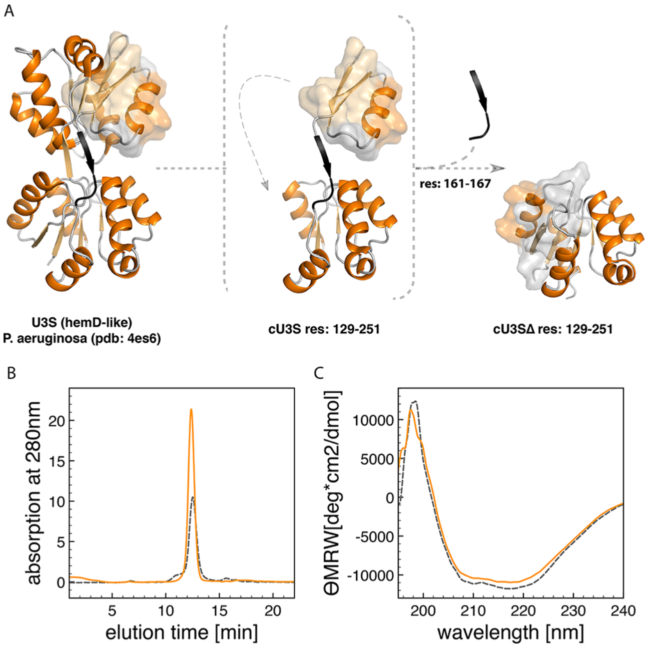

Figure 2. Steps to reconstruct a precursor of hemD-like proteins. The C-terminal half of uroporphyrinogen III synthase (U3S) from *P. aeruginosa* was truncated (cU3S) and its insert removed (cU3SΔ), revealing the canonical flavodoxin-like architecture as determined by NMR spectroscopy (A). Biophysical characterization of the truncated halves cU3S (gray, dashed) and cU3SΔ (orange, solid line) via size exclusion chromatography (B) and CD spectroscopy (C).

To reconstruct these events experimentally, we created two constructs. First, we truncated the C-terminal half of U3S from *P. aeruginosa* (cU3S) and further removed six residues that form one strand of the β-bridge between the two lobes (cU3SΔ) (Figure 2a). Both proteins were soluble upon heterologous expression in *E. coli*. However, a large part of cU3S was also found in inclusion bodies after expression. In addition, cU3S was prone to aggregation, precipitating during the purification process, particularly at concentrations above 1 mg/mL. It should be noted, however, that temporary misfolding and formation of aggregation-prone intermediates are a typical feature of proteins with a flavodoxin-like architecture.[10] In contrast, cU3SΔ remained stable at concentrations above 20 mg/mL. Size exclusion chromatography showed that both proteins elute as monomers at the concentrations tested (Figure 2b), except after refolding, in which case cU3S forms multimers, illustrating the tendency of cU3S to associate with itself. Both monomeric proteins showed similar CD spectra (Figure 2c) corresponding to well-folded proteins with comparable $\alpha\beta$-content. Further, thermal denaturation revealed a higher thermostability and cooperativity of unfolding for cU3SΔ (~60 °C) than cU3S (~40 °C) (Figure S2).

We proceeded to elucidate the structure of cU3SΔ by NMR spectroscopy (PDB 6TH8), which revealed the canonical flavodoxin-like topology and architecture (Figure 2a and Figure S4, Table S3). Structural searches against pdb90, employing DALI,[11] identified high similarities both to the U3S lobes and flavodoxin-like proteins with significant Z scores of up to 14 and 11, respectively (Table S2). Even though the aligned structures present a different orientation of some secondary structural elements, these discrepancies are also observed for structures within a protein family (Figure S5). The different domain assignment by different databases is noteworthy. Whereas CATH[12] considers the HemD-like lobes as independent domains, ignoring the connecting β-strands that are assigned only to the lower lobe, SCOP[13] classifies the scaffold as a single domain due to the discontinuity of the polypeptide chain along the two lobes and the fact that both lobes are required for function (functional domain).

IMPLICATIONS FOR PROTEIN FOLD EVOLUTION

Our results strongly support the emergence of HemD-like from flavodoxin-like proteins. The precursor may have acquired a short insert, which facilitated its dimerization via segment swap, thereby creating a hinge region for binding and catalysis. Interestingly, segment swaps have been reported as likely artifacts in crystallization for other flavodoxin-like proteins (PDB 4q37[5] and PDB 3c85), illustrating the plasticity of this particular fold. Overall, segment swaps appear fairly frequently; e.g., about 13% of multidomain proteins in the PDB are reported to present segment swaps.[14] An interesting example is glyoxalase I, whose active site is located at the interface of two identical, noncovalently linked segment-swapped domains.[15] U3S may have functioned in a similar manner. After insertion, the halves may have overcome a selective pressure due to better domain association. Finally, duplication provided additional constraints to the proteins' flexibility, decreasing further the entropic costs for ligand binding and catalysis. We consider the emergence of flavodoxin-like proteins via HemD-like truncation less likely than the described scenario as the flavodoxin-like fold is estimated to be one of the oldest structural domains.[16] Thus, it can be seen as a primary rather than as a derived scaffold.

Taken together, this work illustrates how evolution can be reconstructed experimentally. It further highlights the importance of duplication and fusion in the development of complex protein domains similar to the evolution of $(\beta\alpha)_8$-barrel proteins from half-barrels.[17,18] Here, however, the evolutionary process also includes a segment swap, and the reconstruction led to a robust canonical flavodoxin-like protein, thereby highlighting the plasticity of modern proteins to form new forms, which can be harnessed in protein design from natural protein fragments.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.9b00900.

Experimental methods and solution structure statistics (PDF)

Accession Codes

The Uniprot ID of hemD is P48246, and the NCBI accession is 4ES6_A. The PDB ID of the NMR structure is 6TH8 and the BMRB ID 34452.

AUTHOR INFORMATION

Corresponding Author

*E-mail: Birte.Hoecker@uni-bayreuth.de.

ORCID

Murray Coles: 0000-0001-6716-6150
Birte Höcker: 0000-0002-8250-9462

Funding

This project was funded by the ERC (Consolidator Grant 647548 "Protein Lego" to B.H.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Prof. Gunter Schneider and Dr. Robert Schnell for kindly providing us with the U3Spa construct and Sooruban Shanmugaratnam for excellent technical assistance.

REFERENCES

(1) Söding, J., and Lupas, A. N. (2003) More than the Sum of Their Parts: On the Evolution of Proteins from Peptides. *BioEssays 25* (9), 837−846.

(2) Murzin, A.G. P (1998) How Far Divergent Evolution Goes in Proteins. *Curr. Opin. Struct. Biol. 8* (3), 380−387.

(3) Bork, P., Sander, C., and Valencia, A. (1993) Convergent Evolution of Similar Enzymatic Function on Different Protein Folds: The Hexokinase, Ribokinase, and Galactokinase Families of Sugar Kinases. *Protein Sci. 2* (1), 31−40.

(4) Söding, J., Biegert, A., and Lupas, A. N. (2005) The HHpred Interactive Server for Protein Homology Detection and Structure Prediction. *Nucleic Acids Res. 33*, W244−248.

(5) Farias-Rico, J. A., Schmidt, S., and Höcker, B. (2014) Evolutionary Relationship of two ancient protein superfolds. *Nat. Chem. Biol. 10* (9), 710−715.

(6) Fortian, A., Castano, D., Ortega, G., Lain, A., Pons, M., and Millet, O. (2009) Uroporhyrinogen III Synthase Mutations Related to Congenital Erythropoietic Porphyria Indentify Key Helix for Protein Stability. *Biochemistry 48* (2), 454−461.

(7) Szilagyi, A., Györffy, D., and Zavodszky, P. (2017) Segment Swapping Aided the Evolution of Enzyme Function: The Case of Uroporphyrinogen III Synthase. *Proteins: Struct., Funct., Genet. 85* (1), 46−53.

(8) Alva, V., Nam, S. Z., Söding, J., and Lupas, A. N. (2016) The MPI Bioinformatics Toolkit as an Integrative Platform for Advanced Protein Sequence and Structure Analysis. *Nucleic Acids Res. 44* (W1), W410−W415.

(9) Sander, C., and Schneider, R. (1991) Database of Homology-Derived protein structures and the Structural Meaning of Sequence Alignment. *Proteins: Struct., Funct., Genet. 9* (1), 56−68.

(10) Houwman, J. A., and van Mierlo, C. P. M. (2017) Folding of proteins with a flavodoxin-like architecture. *FEBS J. 284*, 3145−3167.

(11) Holm, L., and Laakso, L. M. (2016) DALI Server Update. *Nucleic Acids Res. 44* (W1), W351−W355.

(12) Dawson, N. L., Lewis, T. E., Das, S., Lees, J. G., Lee, D., Ashford, P., Orengo, C. A., and Sillitoe, I. (2017) CATH: an Expanded Resource to Predict Protein Function through Structure and Sequence. *Nucleic Acids Res. 45* (D1), D289−D295.

(13) Chandonia, J. M., Fox, N. K., and Brenner, S. E. (2017) SCOPe:Manual Curation and Artifact Removal in the Structural Classification of Proteins - extended Database. *J. Mol. Biol. 429* (3), 348−355.

(14) Szilagyi, A., Zhang, Y., and Zavodszky, P. (2012) Intra-Chain 3D Segment Swapping Spawns the Evolution of New Multidomain Protein Architectures. *J. Mol. Biol. 415* (1), 221−235.

(15) Cameron, A. D., Olin, B., Riderström, M., Mannervik, B., and Jones, T. A. (1997) Crystal Structure of Human Glyoxalase I - Evidence for Gene Duplication and 3D Domain Swapping. *EMBO. J. 16*, 3386−3395.

(16) Caetano-Anolles, G., Kim, H. S., and Mittenthal, J. E. (2007) The Origin of Modern Metabolic Networks Inferred from

Phylogenomic Analysis of Protein Architecture. *Proc. Natl. Acad. Sci. U. S. A. 104* (22), 9358−9363.

(17) Höcker, B., Beismann-Driemeyer, S., Hettwer, S., Lustig, A., and Sterner, R. (2001) Dissection of a ($\beta\alpha$)8-barrel enzyme into two folded halves. *Nat. Struct. Biol. 8* (1), 32−36.

(18) Höcker, B., Claren, J., and Sterner, R. (2004) Mimicking enzyme evolution by generating new ($\beta\alpha$)8-barrels from ($\beta\alpha$)4-half-barrels. *Proc. Natl. Acad. Sci. U. S. A. 101* (47), 16448−164453.