FEBS Letters

FEBS PRESS science publishing by scientists

RESEARCH LETTER

Molecular handcraft of a well-folded protein chimera

Saacnicteh Toledo-Patiño[1,2], Sara Kathrin Goetz[1,*], Sooruban Shanmugaratnam[1,3], Birte Höcker[1,3] and José Arcadio Farías-Rico[1,4]

1 Max Planck Institute for Developmental Biology, Tübingen, Germany
2 Okinawa Institute of Science and Technology Graduate University, Japan
3 Department of Biochemistry, University of Bayreuth, Germany
4 Synthetic Biology Program, Center for Genome Sciences, National Autonomous University of Mexico, Cuernavaca, Mexico

Correspondence
B. Höcker, Department of Biochemistry, University of Bayreuth, Universitätsstrasse 30, 95447 Bayreuth, Germany
Tel: +49 921 557831
E-mail: birte.hoecker@uni-bayreuth.de
J. A. Farías-Rico, Center for Genome Sciences UNAM, Av. Universidad s/n, Universidad Autónoma del Estado de Morelos, 62209 Cuernavaca, Morelos, México
Tel: +52 7773291777; ext: 38479
E-mail: jafarias@ccg.unam.mx

Present address
*The Boston Consulting Group, Cologne, Germany

(Received 11 August 2023, revised 11 February 2024, accepted 12 February 2024, available online 20 March 2024)

doi:10.1002/1873-3468.14856

Edited by Alfonso Valencia

Modular assembly is a compelling pathway to create new proteins, a concept supported by protein engineering and millennia of evolution. Natural evolution provided a repository of building blocks, known as domains, which trace back to even shorter segments that underwent numerous 'copy-paste' processes culminating in the scaffolds we see today. Utilizing the subdomain-database Fuzzle, we constructed a fold-chimera by integrating a flavodoxin-like fragment into a periplasmic binding protein. This chimera is well-folded and a crystal structure reveals stable interfaces between the fragments. These findings demonstrate the adaptability of α/β-proteins and offer a stepping stone for optimization. By emphasizing the practicality of fragment databases, our work pioneers new pathways in protein engineering. Ultimately, the results substantiate the conjecture that periplasmic binding proteins originated from a flavodoxin-like ancestor.

Keywords: chimeric proteins; homology; protein engineering; protein fold evolution; sequence-based analysis; sub-domain

Proteins play an essential role in the biological processes that support life. Therefore, their origin and evolution have been a matter of in-depth analysis [1–4]. To make sense of the vast structural diversity that we find today, structured proteins are classified into folds; proteins that are classified as the same fold have the same secondary structure elements arranged with the same topological connections. One of the most widely used protein databases is the structural classification of proteins (SCOP) [5]. This database is a hierarchical organization system that groups the protein structures deposited in the protein data bank (PDB) [6] into families, superfamilies, folds, and domains based on their three-dimensional arrangements. In SCOP, proteins belonging to the same family or superfamily are considered homologous groups, while proteins belonging to different folds were not traditionally considered homologous.

The boundaries of homology have recently been expanded due to advancements in homology detection tools. Nowadays we recognize local homologies among proteins belonging to different folds [7,8]. Exploration of the natural protein universe using hidden Markov models has provided strong evidence for the existence of an ancestral set of fragments, which can be regarded as

Abbreviations
CD, circular dichroism; HMM, hidden Markov model; InDel, insertion/deletion; LBP, leucine-binding protein; PBP, periplasmic binding protein; PDB, Protein Data Bank; SCOP, structural classification of proteins; SEC, size exclusion chromatography.

fossils of the first folded proteins [9]. Similarly, other groups have defined short protein sequences as bridging themes connecting regions from different folds [10–13]. The application of such powerful tools for homology detection based on sequence comparisons in combination with efficient structural comparison algorithms [14] has been used to build a database of homologous fragments shared by different folded proteins [13]. These units hold the potential for mimicking the copy-and-paste mechanism of protein evolution to create custom-made protein chimeras [15,16]. For instance, the $(\beta\alpha)_8$-barrel and flavodoxin-like folds, both ancient and widespread in nature [17], were successfully recombined to form a chimeric barrel [18]. In this regard, it has been established that protein engineering studies profit from the analysis of how protein folds originate and evolve [19].

From an evolutionary perspective, the Fuzzle database has been used to shed some light on the origin and evolution of the sugar-binding superfamily of the periplasmic binding protein (PBP) fold [20]. The PBP-like fold has been suggested to have evolved from a flavodoxin-like precursor *via* gene duplication [21]. By determining the structures of two permuted halves of the ribose binding protein, evidence could be provided for the origin of this protein from a precursor half its size, which again adopts a flavodoxin-like fold [22]. In fact, the modern ribose binding protein can be deconstructed into a precursor half its size that can form a functional heterodimer *in vitro* and *in vivo* that binds the ligand ribose with a similar affinity to the modern ribose binding protein [23]. The study of PBPs extends beyond basic biology, as PBPs serve as a model for functional dynamics and have been widely used as scaffolds to design biosensors. Upon ligand binding, they undergo a major conformational change that captures the ligand between the two protein lobes, often termed the Venus flytrap mechanism. This dynamic feature can be leveraged to generate a robust signal, for instance, *via* the insertion of a fluorescent protein [24,25].

Another fold that mirrors the conformational changes observed in PBP, characterized by the opening and closing of its lobes [26], is the HemD-like fold, suggesting its potential for biosensor applications. The HemD-fold encompasses a single superfamily of enzymes that are found in all kingdoms of life. These enzymes catalyze the synthesis of uroporphyrinogen III, which plays a crucial role in the biosynthesis of essential cofactors such as heme, chlorophyll and cobalamin. Sequence-based profile–profile alignments of HemD-like and flavodoxin-like proteins revealed similarities, proposing that the HemD-like fold might have originated from the flavodoxin-like fold involving a short six-residue insertion thereby facilitating a segment-swapped association followed by a gene duplication and fusion event resulting in the formation of a bi-lobular architecture. This could be demonstrated by experimentally reconstructing the most likely evolutionary path to the canonical flavodoxin-like architecture [27]. These combined findings point towards the flavodoxin-like fold as an ancestral template, which can be found in more contemporary folds, an evolutionary link that is experimentally supported by the use of homologous regions to build well-folded chimeras.

In the present work, we explore the modularity and plasticity of the PBPs by performing a molecular handcraft with a shared fragment between a leucine binding protein and a chemotaxis response regulator, which belong to the PBP and the flavodoxin-like folds, respectively. In a process of 'molecular copy-paste', we introduce a homologous fragment from the flavodoxin-like fold into one of the lobes of the PBP scaffold to create a well-folded protein without performing any computational optimization. The resulting protein hybrid showed biophysical properties similar to the PBP-like parent, and an X-ray crystal structure revealed structural integrity but also unexpected structural differences from the intended design. Overall, our work demonstrates how sequence information can be utilized to detect pieces of proteins that can then be recombined to create folded protein chimeras. The experiment highlights the plasticity of proteins and the potential of such 'hopeful monsters' [28] to undergo modifications and give rise to new domains.

Materials and methods

Bioinformatic analysis

To generate sequence-based Hidden Markov Models, multiple sequence alignments were built for SCOP [29] folds c.23.1.1 (CheY response regulator, flavodoxin-like fold) and c.93.1 (leucine binding protein, PBP-like fold type I) contained in the astral database release 1.07 [30]. The alignments were generated employing PSI-BLAST [31], this program is included in the build.pl, the protocol described by Söding [32]. This protocol was employed to generate profile–profile comparisons with default parameters. The secondary structure prediction was turned off (ssm = 0) in order to perform a strictly sequence-based search. The matched homologous regions were aligned with TM-align [14] and manual selections with PDBeFold [33]. The final alignment to perform the chimera building was obtained by filtering according to sequence identity, the highest sequence identity (12%) for the longest aligned fragments with the smallest RMSD (2.6 Å) using default parameters. In summary, we performed a customized version of the protocol used to build the Fuzzle database [13,34,35].

Cloning of wildtype proteins and chimeric variants

The CheY fragment was amplified from a pET21b-CheY construct [18] while the LBP protein was amplified from *Escherichia coli* genomic DNA. The chimeric genes were assembled by PCR employing two fragments from the LBP gene, encoding amino acid 1–140 and 251–346, and one fragment from the CheY gene, encoding residue 2–104. All three fragments contained overlapping regions to facilitate annealing. The generated DNA pieces were purified by gel extraction and mixed in equimolar amounts for assembly *via* PCR. The amplification was performed using primers bearing restriction sites for *Xho*I and *Nde*I. The obtained gene was digested and cloned into a pET21b expression vector, yielding the construct pET21b-LBP-CheY. The derived gene LBPCheY-Δββ was amplified with appropriate primers to remove the C-terminal ββ-hairpin from the pET21a-LBP-CheY construct and cloned into pET21b for expression, as described earlier, yielding pET21b-LBPCheY-Δββ. All constructs were validated *via* DNA sequencing.

Protein expression

BL21 (DE3) and Artic Express (DE3) *E. coli* cells were transformed with the constructs. Cultures were grown in LB medium with 100 $\mu g \cdot mL^{-1}$ ampicillin at 30 °C and shaken until an OD_{600} of ~ 0.6 was reached. Protein expression was induced with 1 mM IPTG for 4 h at 20 and 11 °C, respectively. Cells were harvested by centrifugation and stored at −80 °C. The pellets were resuspended in buffer A (50 mM KP, pH 7.5) supplemented with EDTA-free protease inhibitors. Cells were lysed by sonication at 30% amplitude for 1 min (0.5 s pulse, 0.5 s pause). The resulting lysates were loaded onto an anion exchange column HiLoad26/10 QHP (GE Healthcare, Chicago, IL, USA) previously equilibrated with buffer A. After loading, the column was washed, and then the target proteins were eluted with a salt gradient by mixing with buffer B (50 mM KP, 250 mM NaCl, pH 7.5). Fractions containing the target protein were collected and loaded onto a gel filtration column HiLoad 26/600 Superdex 75 pg (GE Healthcare) previously equilibrated with buffer B. All proteins eluted in a single peak corresponding to monomeric species. Because only a small fraction of the expressed protein was found in the soluble fraction, all proteins were refolded to make use of the high yields of expressed protein in the inclusion bodies. For this, the pellet after sonication was resuspended in 10 mL 6 M guanidine hydrochloride (GdHCl) and incubated for 60 min at 4 °C. Next, 10 mL 1 M GdHCl was added incubating for another 60 min at 4 °C after which the insoluble fraction was removed by centrifugation (26 672 ***g***, 60 min, 4 °C). The soluble protein in the supernatant was diluted with 2 M GdHCl to a final volume of 50 mL, which was finally dialyzed in 3 × 5 L of buffer so that further purification could be continued.

To validate that the refolded proteins corresponded to the native proteins, the obtained refolded proteins were compared *via* circular dichroism (CD) and fluorescence to the parental proteins to corroborate native folding. The pure protein fractions were loaded onto an analytical gel filtration column Superdex 200 10/30 GL (GE Healthcare). All protein variants migrated as monomers and eluted in a single peak.

Biophysical characterization

Analytical gel filtration was performed on a calibrated analytical Superdex S75 10/30 GL (GE Healthcare) column with a flow rate of 0.5 $mL \cdot min^{-1}$ in 50 mM KP, 250 mM NaCl, pH 7.5. CD spectra were recorded with a JASCO model J-810 spectropolarimeter in 50 mM KP, pH 7.5 in a 1 mm cuvette at room temperature. Fluorescence measurements were carried out in the same buffer with a JASCO FP-6500 spectrofluorometer exciting at 280 nm. Temperature-induced unfolding was analyzed by following the far-UV CD signal at 222 nm at slowly increasing temperatures (1 $°C \cdot min^{-1}$). The protein concentrations used for spectroscopic measurements were about 0.25 $mg \cdot mL^{-1}$. Differential scanning calorimetry (DSC) was measured with a Malvern Panalytical Ltd (Malvern, UK) automated PEAQ-DSC. Endotherms for LBPCheY-Δββ were collected in a temperature range from 15 to 80 °C with a heating rate of 1 $°C \cdot min^{-1}$ and a sample concentration of 2 $mg \cdot mL^{-1}$ in 50 mM KP pH 7.5. Data was analyzed after buffer subtraction using the PEAQ-DSC software. Best data interpretation was achieved by applying a 'non-two state fit' with three transitions.

Crystallization and X-ray structure determination of LBPCheY-Δββ

Well-shaped single crystals were obtained in hanging drops following a micro-seeding protocol in 0.3 M Li sulfate, 0.1 M TrisHCl pH 8.5, and 30% PEG 4000 with protein in 50 mM Tris pH 7.5, 250 mM NaCl at a concentration of 6 $mg \cdot mL^{-1}$. Crystals were mounted and flash-cooled in liquid nitrogen. A dataset was collected from a single crystal. Diffraction data were collected at 100 K on a Pilatus 6 M at the beamline X10SA (PX II, Swiss Light Source, PSI) and processed using XDS [36]. The protein formed crystals in tetragonal space group P 41 21 2 with a = b = 129.56, c = 43.64 containing one protein molecule per asymmetric unit. Molecular replacement was performed with PHASER [37] using as templates for the model search the respective fragments from LBP (PDB: 1USG) and CheY (PDB: 1TMY). After identifying the phases, we proceeded with auto-build and refinement cycles to obtain a structure with an overall resolution of 2.45 Å. This structure was deposited already a few years ago in the PDB database with the accession code 4QWV. We now reanalyzed the original data to a resolution of 2.15 Å, processed it with

XDSAPP3 [38] and performed molecular replacement with PHASER [37] using 4QWV as a search model. The resulting model was manually rebuilt with COOT [39] and refined with PHENIX.REFINE [40]. Coordinates and structure factors were validated and deposited in the PDB database with the accession code 8Q52.

Results

Sequence-based comparisons of proteins with PBP and flavodoxin-like folds

The periplasmic-binding protein-like I fold has been described as a potential duplication of the flavodoxin-like fold in SCOP [29]. Similarly, the CATH database [41] and the evolutionary classification of protein domains (ECOD) [42] trim and classify some PBP lobes in the same category as the flavodoxin-like domain, presumably due to their high structural resemblance. However, in the dissection process, the amino acid sequence is split at three sites (Fig. S1). Taking this observation into account, we anticipated that an evolutionary trajectory for the emergence of the PBP fold from the flavodoxin-like fold must have involved more events than just gene duplication and fusion.

To study the likely evolutionary scenario for the emergence of PBP from flavodoxin-like, we performed Hidden Markov model comparisons using HHSUITE [32] as described in the methods. We found that different flavodoxin-like superfamilies match PBP domains with diverse lengths (Fig. S2). For instance, flavodoxin-like domains match both, the N- and C-terminal halves of the PBP D-ribose-binding protein family [22], while it only matches either the N- or C-termini of the remaining PBP families.

In this work, we focused on a particular match that involves a member of the flavodoxin-like fold, the chemotaxis response regulator CheY from *Thermotoga maritima*, and a member of the PBP-type I fold, the leucine-binding protein (LBP) from *Escherichia coli*. CheY has already been used previously to build flavodoxin-like/TIM-barrel fold-chimeras [18]. It has been selected in this work to test whether the flavodoxin-like fold may have acted as an evolutionary template for the elaboration of a number of distinct protein architectures. The homologous region shared by CheY and LBP spans 107 residues (Gly2 to Ser104 and Asp1 to Gln109, respectively) at 90.65% HHsuite probability, *P*-value of 0.038, and 12% sequence identity. The region involves the secondary structural elements $\beta_1\alpha_1\beta_2\alpha_2\beta_3\alpha_3\beta_4$ in CheY, and $\beta_6\alpha_7\beta_7\alpha_8\beta_8\alpha_9\beta_9\alpha_{10}$ in LBP as depicted in Fig. 1.

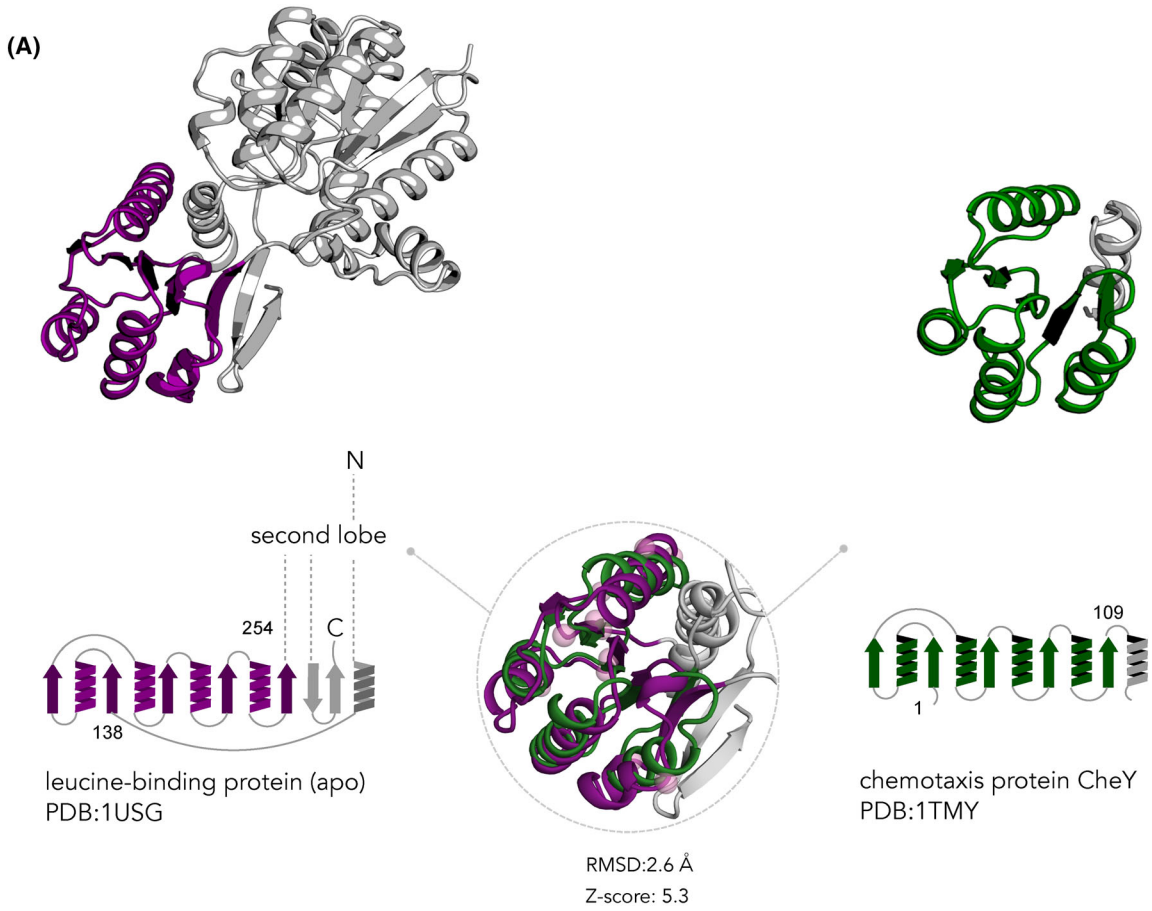
Constructing the LBPCheY chimera

The structural and sequence similarities suggest that at least one lobe of LBP may have originated from the flavodoxin-like fold (Fig. 1). Therefore, we wanted to test the hypothesis by chimeragenesis [15] utilizing a flavodoxin-like fold fragment to recreate the bi-lobular architecture of the PBP-fold and decided to build a chimeric gene using the response regulator CheY as ancestral fragment to be pasted into the LBP protein as acceptor scaffold (Fig. 2).

Two intertwined lobes compose the LBP protein (Fig. 1). The N-terminal lobe crosses the hinge region to the C-terminal lobe at the end of β_5. The β-sheet from the C-terminal lobe starts in β_6 and goes back to the N-terminal lobe in β_{10}, to finally cross one more time after β_{10} to complete the extended β-sheet from the lobe with a C-terminal β-hairpin (composed of the last two β-strands). The pairwise alignment that was obtained through profile-profile comparisons covers the CheY protein sequence from residue 2 in β1 until residue 108 at the beginning of α5. In LBP, the alignment reaches from residue 138 in β6 to residue 253 in β10.

The web server PDBeFold [43] allows users to perform range-restricted structural alignments; we

Fig. 1. Sequence and structural comparison of the Leucine-binding protein and the CheY response regulator. (A) The crystal structures of LBP from *E. coli* (PDB: 1USG, left) and CheY from *T. maritima* (PDB: 1TMY, right) are depicted as cartoons. Gray regions represent segments of the structures where no homology was detected at the sequence level. Purple and green regions represent segments aligned by HHsuite. Topological diagrams are presented below each structure. The central panel illustrates the superimposed homologous regions (RMSD of 2.6 Å over 95 Cα, Z-score of 5.3) with identical residue positions highlighted as pink spheres. (B) The alignment using CheY as query 'Q' and LBP as target 'T' spans 103 residues, with a 90.65% HHsuite probability and 12% sequence identity. Identity refers to the percentage of aligned residue pairs of the query and template master sequences that are identical (highlighted in pink). The line in the middle shows the column score between the query and template amino acid distributions providing a vallable indication of the alignment quality. Similarities are highlighted as gray spheres in the alignment. Symbols such as '=', '-', '.', '+', and '|' are used to represent column scores, indicating the degree of similarity or dissimilarity between aligned residues as follows: '=' column score below −1.5; −: column score between −1.5 and −0.5; '.' column score between −0.5 and +0.5; '+' column score between +0.5 and +1.5, and '|' column score above +1.5. The alignment also displays the consensus sequences of the aligned profiles as well as the predicted secondary structure for query and hit, where 'H', 'E', and 'L' refer to helix, strand and loop, respectively.



(B)

```
HHsuite probability: 90.65
E-value: 0.038
Aligned residues: 107
Sequence identity: 12%

Q ss_pred              eEEEEECCC----HHHHHHHHHHHHHCCCEEEEE------ECCHHHHHHHHHHHCCCEEEEeCCCCCCCHHHHHHHHHH
Q 1TMY               1 GKRVLIVDDA----AFMRMMLKDIITKAGYEVAGE------ATNGREAVEKYKELKPDIVTMDITMPEMNGIDAIKEIMK   70 (118)
Q Consensus          1 G~rILIVDD~--------r~~l~~~L~~~g~~Vv~~------A~~G~eal~~~~~~~PDlvllDi~MP~mdGle~~~~i~~   70 (118)
                       ++||.|+-++    ....+.++..+++.|++|+++      ..|-...+..+++.+||.|+  +..+.-+...+++++++
T Consensus        138 ~k~v~ii~~~~~~g~~~~~~~~~~~~~g~~vv~~~~~~~~~d~~~~i~~l~~~~d~i~~~~~~~~~~~~~~~~~  215 (346)
T 1USG         138 PQRIAIIHDKQQYGEGLARSVQDGLKAANANVVFFDGITAGEKDFSALIARLKKENIDFVY--YGGYYPEMGQMLRQARS  215 (346)
T ss_pred              cceEEEeCCCCCCchHHHHHHHHHHhcCceeeeccccccccHHHHHHHHHHhcCceeE--eccCChhHHHHHHHhH


Q ss_pred              CCCCEEEEEECCCCHHHHHHHHHHCCCCEEEECCCCHHH
Q 1TMY             71 IDPNAKIIVCSAMGQQAMVIEAIKAGAKDFIVKPFQPSR  109 (118)
Q Consensus        71 ~~p~~~iI~~s~~~~~~~~~~a~~~Ga~~yl~KP~~~e~  109 (118)
                     ...+.+++..++..+.+....+-....-.-|+.-|...+.
T Consensus       216 ~g~~~~i~~~~~~~~~~~~~~~~g~~~~~~~~~  254 (346)
T 1USG         216 VGLKTQFMGPEGVGNASLSNIAGDAAEGMLVTMPKRYDQ 254 (346)
T ss_pred              CCceeeecCccccchhHHHHHhhccCCcEEecCCCcCC
```

therefore, only superimposed the regions of the folds that were aligned by sequence. Similar to the sequence alignment, the structural superposition includes the CheY residues from 2 to 104 and in LBP the residues 138 to 248 (Fig. 2A). The structural scores are as follows: an RMSD of 2.667 Å for 95 Cα atoms with a Z-score of 5.35. Interestingly, a Z-score between 3 and 8 is defined as the twilight zone for structural homology [44]. However, if we consider sequence (*P*-value 4.3E-05) and structure-based (Z-score 5.35) scorings

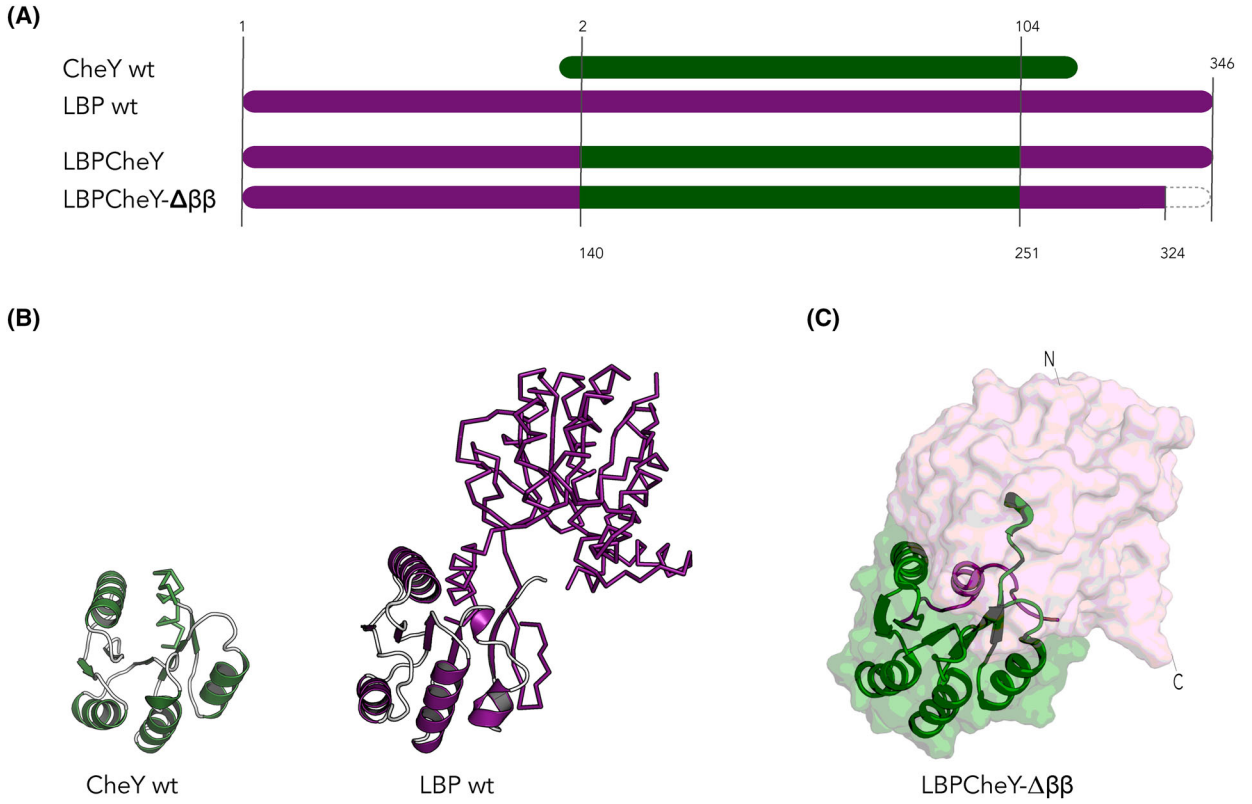


Fig. 2. Construction of the LBPCheY chimera. (A) Cartoon representation of the alignment of LBP from *E. coli* (purple) and CheY from *T. maritima* (green). Hybrid genes (purple-green) depict the inserted regions from the parental proteins. The dashed region corresponds to the ββ element that was removed from the LBPCheY chimera (LBPCheY-Δββ) for crystallization purposes. (B) Crystal structure of LBP (PDB: 1USG, purple) and CheY (PDB: 1TMY, green) displaying homologous regions as detected with HHSUITE as cartoon and non-homologous regions as ribbon. (C) Cartoon representation of the C-terminal lobe of the chimeric protein LBPCheY-Δββ superposed on a surface model. Parts that originate from LBP are in purple, and parts from CheY are in green.

and the significantly long alignment, we could imply a common evolutionary origin (at least locally) of both proteins. Based on the length of the alignment, the high HHsuite probability of 90.65, and the low RMSD upon superposition, it is highly unlikely for the similarity to have evolved through convergent evolution.

To build a homology model of the chimeric design we used Modeler with a structure-based alignment of CheY and LBP as input [45]. From the visual evaluation of the model, we expected that β13 from LBP would establish hydrogen bonds with β5 of CheY in order to generate a chimeric β-sheet. Finally, α6 from LBP must establish hydrophobic interactions with the core of CheY (Fig. 2C). We superimposed the model onto the target scaffold (LPB wild type) and obtained an RMSD of 1.3 Å over 322 Cα (95% residues aligned). The model deviates only somewhat from the scaffold in loop areas, as CheY shows different residue composition in the loops compared to LBP. Next, we continued to build the synthetic gene as described in the methods.

Gene assembly and biophysical characterization of wild-type LBP and LBPCheY chimeric variants

Two chimeric designs were generated, LBPCheY and LBPCheY-Δββ. The latter was constructed by removing a ββ-hairpin from the C-terminal end of LBP on the one hand to test its effect on the structural characteristics as well as to potentially facilitate crystallization as it might be rather flexible (Fig. 2). Similar residues align between β5 in CheY (FIV) and β10 in LBP (MLV) that could allow the formation of a mixed β-sheet in the chimera. On the contrary, residue W336 in β13 that in LBP completes a hydrophobic cluster might be difficult to accommodate in the LBPCheY, which is why both constructs were tested. Analytical gel filtration revealed that both, LBPCheY and LBPCheY-Δββ eluted with similar profiles to the LBP wild-type protein (Fig. 3A),

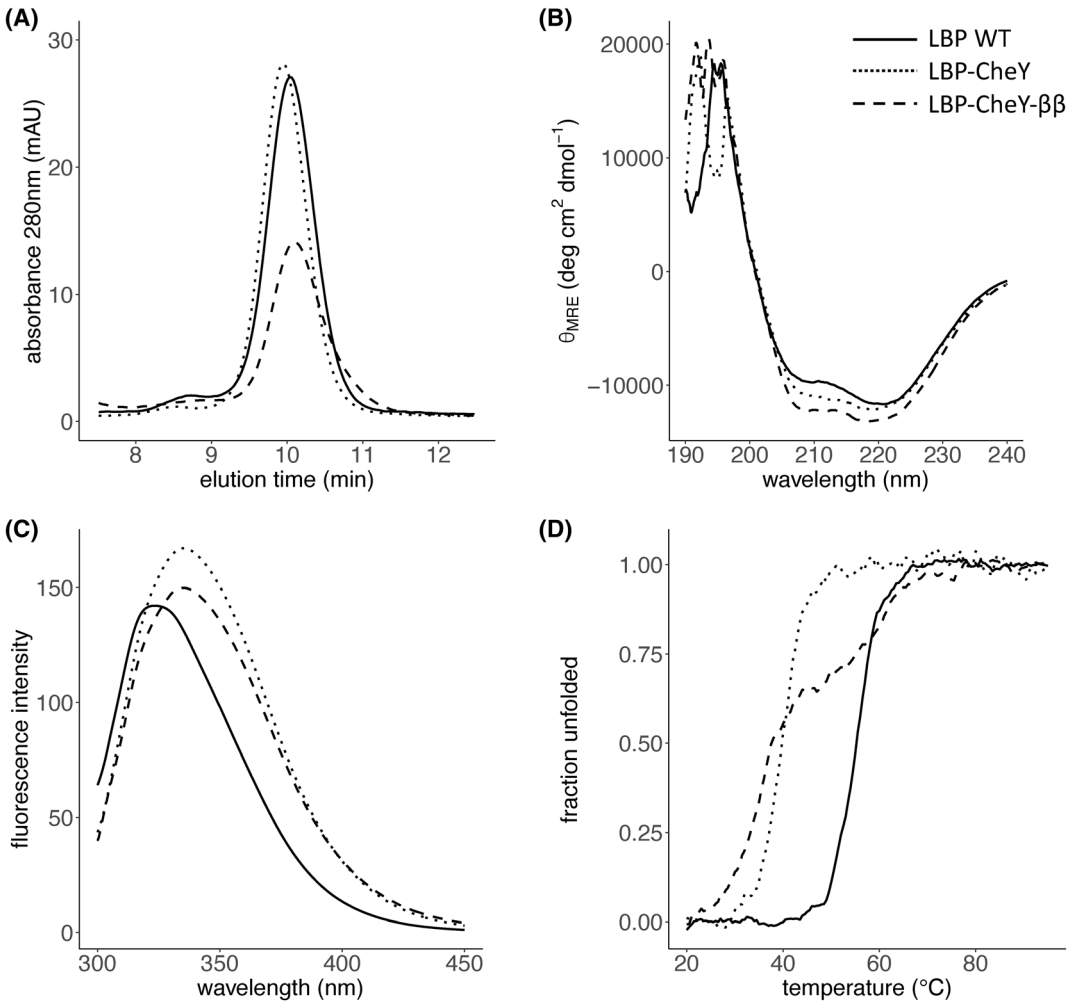


Fig. 3. Biophysical characterization of LBP, LBPCheY and LBPCheY-Δββ. (A) Analytical size exclusion chromatography displays the retention volumes for the corresponding variants. All proteins elute in a single peak. As expected, the larger LBPCheY chimera elutes first, followed by the truncated version LBPCheY-Δββ and LBP. (B) Circular dichroism showing typical spectra of α/β proteins with two strong minima at 208 and 222 nm. (C) Intrinsic fluorescence of the variants excited at 280 nm showing shielding of aromatic residues indicative of globular packing. (D) Thermal denaturation showed a higher stability for the wild type protein at ~ 55 °C and a loss in thermal stability for the engineered variants that unfold at ~ 35–40 °C. All proteins show mostly cooperative unfolding. However, the truncated variant LBPCheY-Δββ presents a two-step denaturation, compared to the wild type and the LBPCheY chimera.

with apparent molecular masses of 36.5, 34.5 and 37.9 kDa, respectively. These values correspond to the calculated molecular masses of the monomeric proteins with 41.0, 37.6 and 39.4 kDa, respectively. As expected, LBPCheY eluted earlier than LBPCheY-Δββ due to the absence of the C-terminal ββ-hairpin.

To evaluate the secondary structural content of the chimeras, circular dichroism (CD) spectroscopy was employed. The CD spectra for all three proteins showed the typical signal expected for an α/β protein, with two negative bands at around 208 and 220 nm (Fig. 3B). Spectroscopic analysis following tryptophan fluorescence confirmed that both chimeras contained aromatic residues that were shielded from solvent thereby indicating tertiary structure formation (Fig. 4C). We then conducted thermal denaturation experiments using the CD signal at 220 nm to assess the stability of the chimeric proteins in comparison to wild-type LBP (Fig. 3D). While LBP exhibited thermal stability with an apparent melting temperature of ~ 62 °C, both chimeras, LBPCheY and LBPCheY-Δββ, showed lower thermal stability with melting temperatures of ~ 40 °C. Despite the differences in denaturation profiles, both chimeras retained some of the cooperative denaturations observed for LBP. Interestingly, LBPCheY-Δββ displayed two transitions at different temperatures which could hint at

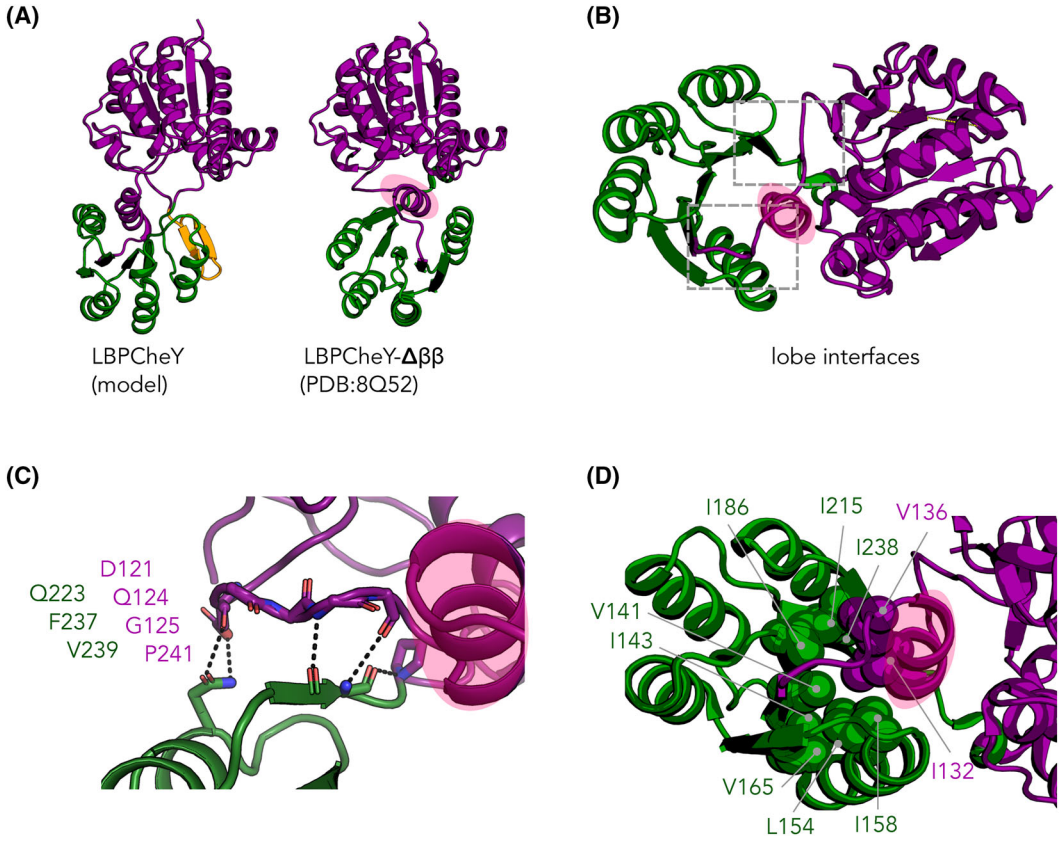


Fig. 4. Structural analysis of the LBPCheY-Δββ crystal structure. (A) A side-by-side comparison of the LBPCheY-Δββ model (left) and the X-ray structure (PDB: 8Q52) illustrates the different relative lobe orientations. The ββ-element in the protein model is highlighted in yellow and the helix interacting with the CheY fragment is highlighted in pink shade. (B) The X-ray structure was rotated 45 degrees to show details of the interface between the fused regions. Boxes indicate the position of the detailed views shown in the next panels. (C) Detail on a loop from LBP that establishes hydrogen bond interactions in β-sheet like fashion with the fifth β-strand from the flavodoxin-like fold fragment. Interactions include Q223-D121 and F237-Q124 whose backbone is depicted as sticks. Furthermore, the interaction involving the backbone of V239 with G125 and P241 is highlighted. (D) The distal part of the interaction shows how the LBP fragment establishes mainly hydrophobic interactions to shield the flavodoxin-like fragment from solvent: I132 and V136 (in purple) with V141, I143, L154, I158, V165, I186, I215, I238 (in green), see text for details.

the uncoupled unfolding of the two lobes without the ββ-hairpin that fixes the conformation of the lobes to each other. This is further supported by the endotherm obtained using DSC, where we observe a major unfolding transition at 37.1 °C, followed by two transitions at 49.9 and 59.9 °C (Fig. S3).

In summary, the biophysical characterization shows that the chimeras are homogeneous and well-folded and thus were successfully designed by combining homologous elements from two different folds identified through sequence information alone.

X-ray structure determination and analysis of LBPCheY-Δββ

We set out to crystallize the chimera LBPCheY to understand how the different fragments interact in the new structural context. The crystals we produced diffracted unfortunately only to 7 Å resolution for the LPBCheY variant. Considering the possibility that β13 of LBP may not establish productive contacts with β5 from CheY in the chimera we decided to remove the C-terminal ββ-hairpin generating LBPCheY-Δββ (Fig. 2).

LBPCheY-Δββ crystallized in space group P41 21 2. Initially, we performed molecular replacement using for the model search the respective fragments from LBP (PDB: 1USG) and CheY (PDB: 1TMY). After identifying the phases, we proceeded with auto-build and refinement cycles to obtain a final structure with R_{work} = 0.20, R_{free} = 0.23 and an overall resolution of 2.45 Å, which was deposited in the PDB as 4QWV. We now reanalyzed the data and obtained a structure to a resolution of 2.15 Å (Table 1). One molecule is

Table 1. Data refinement statistics of the LPBCheY-Δββ structure (PDB: 8Q52). Statistics for the highest-resolution shell are shown in parentheses.

Data collection	
Wavelength (Å)	1.0
Resolution range (Å)	41.36–2.15 (2.23–2.15)
Space group	P 4_12_12
Unit cell	
a, b, c (Å)	129.56, 129.56, 43.64
α, β, γ (°)	90, 90, 90
Total reflections	537 458 (53 588)
Unique reflections	20 824 (2040)
Multiplicity	25.8 (26.3)
Completeness (%)	99.9 (98.9)
Mean I/sigma (I)	14.41 (0.62)
Wilson B-factor	43.7
R_{merge}	0.313 (6.080)
R_{meas}	0.319 (6.198)
R_{pim}	0.062 (1.197)
$CC_{1/2}$	0.999 (0.320)
CC*	1.000 (0.696)
Refinement	
Reflections used in refinement	20 823 (2019)
Reflections used for R_{free}	1041 (102)
R_{work}	0.200 (0.327)
R_{free}	0.252 (0.394)
CC_{work}	0.973 (0.616)
CC_{free}	0.946 (0.553)
Number of non-hydrogen atoms	2599
Macromolecules	2416
Solvent	168
Protein residues	320
RMS (bonds) (Å)	0.002
RMS (angles) (°)	0.420
Ramachandran favored (%)	97.80
Ramachandran allowed (%)	1.89
Ramachandran outliers (%)	0.31
Rotamer outliers (%)	1.17
Clashscore	2.25
Average B-factor	46.8
Macromolecules	46.4
Solvent	48.6

found in the asymmetric unit and all parts of the chimeric protein are well resolved.

In the crystal structure, we observe a large surface of the LBP part interacting with the CheY part in a twisted conformation (Fig. 4). Essentially helix α5 from LBP is shielding the hydrophobic core of CheY, which was expected, however, with somewhat different interactions. The helix packs at in a different angle against the β-sheet of CheY (Fig. S4A). Its plasticity is noteworthy. Compared to LBP it unwinds from its N-terminal end and adopts a length that is similar to the natural helix found in CheY, but much shorter compared to the helix found in wildtype LBP (Fig. S4B). For example, I132 and L133 form hydrophobic interactions and pack against L162 in α7 of LBP. In the chimera I132, L133 and V136 pack against Y161 in the equivalent helix from CheY. However, the preceding segment, which in LBP extends the helix, adopts a completely different conformation in the chimera where it mimics a distorted β-strand that completes the β-sheet from the flavodoxin-like fragment. The β5 from the flavodoxin-like fold (residues Asp-236 to Lys-240) establishes interactions with the loop after α5 in LBP (residues Ser-123 to Pro-126) adopting a more extended conformation. Notably, Q124 and G125 are positioned to establish hydrogen bonds to F237 and V239 (Fig. 4C). This unusual conformation underscores the protein's flexibility in overcoming structural challenges. With respect to the missing ββ-hairpin and respective loss of interactions in LBPCheY-Δββ, this loop completed the β-sheet pattern instead. The interface between both protein fragments, in particular the helix from LBP that shields the core of CheY, is mainly composed of hydrophobic interactions. Hydrophobic cluster analysis with the ProteinTools web server [46] shows that the biggest cluster of residues belonging to the two different folds is comprised of 10 amino acids with Ile 133 and Val 136 from the LBP parent and the other eight residues belonging to the CheY parent (Fig. 4D). The cluster displayed 31 contacts and a total area of 40.4 Å of area/residue.

Discussion

The study of protein evolution has advanced the field of protein design tremendously. Mimicking the way proteins evolve has been and continues to be a powerful source of inspiration for the engineering of protein architectures. In this work, we explored an evolutionary relationship between the PBP-like I fold and the ancient flavodoxin-like superfold. By using sensitive tools for homology detection, we discovered subdomain-sized regions that display sequence and structural conservation. Next, we used the fragment from the more ancestral fold to replace the corresponding sub-domain part in the more derived scaffold and constructed a protein chimera. Two insertions/deletions (InDels) in the alignment are due to longer loops in LBP compared to CheY, presumably required for its binding activity. InDels within flavodoxin-like domains have been implicated to induce segment-swapping of this domain thereby enabling to assemble, for example, the bilobular architecture of the HemD-like fold [27]. Thus, the flavodoxin-like fold can be seen as a protein template for the emergence of widely represented protein architectures such as the TIM-barrel, HemD-like and

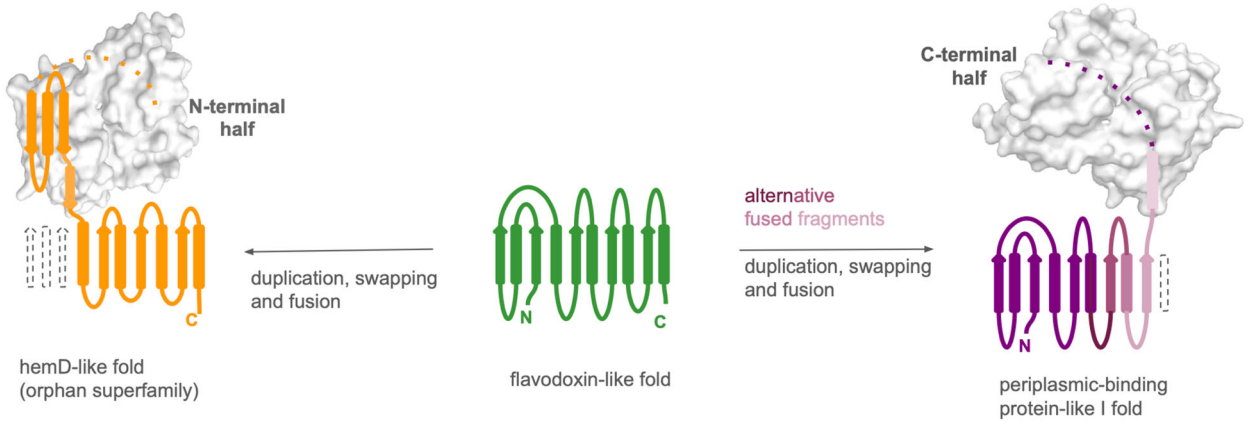


Fig. 5. Model for the evolution of the bi-lobular HemD and PBP-like proteins from a flavodoxin-like precursor template. The flavodoxin-like domain (green) gave rise to the HemD-like fold (orange) and the PBP-like I fold (purple) through different segment swapping, duplication and fusion events.

PBP-like I folds. While both HemD and PBPs are bi-lobular, their similarities to the flavodoxin-like-fold are clearly distinct and thus are expected to have evolved separately from each other though from similar precursor fragments (Fig. 5).

The biophysical characterization of the hybrid protein revealed similar structural features when compared to the wild-type protein LBP. Although attempts to determine the crystal structure of the full-length chimera were not successful, it was possible to determine the structure of a truncated version that lacks the C-terminal ββ-hairpin. While this structure displays a different conformation of the two lobes towards each other than the intended target design, it illustrates the incredible plasticity of α/β-proteins, and Rossman-type proteins in particular [47]. For example, in our design, hydrophobic interactions incorporating a new helix on the one side and a mixed β-sheet on the other side stabilized this 'copy-paste' protein making it available for further adaptations.

It is important to note that in this approach we did not include any step of computational design or mutagenesis. Simply the sequence-based information was sufficient to identify a suitable fragment for exchange and gain a well-folded protein that might be optimized in subsequent steps of protein engineering to generate control of the bilobular orientation and a more native-like LBP structure. This approach could be powerful enough to put together chimeric binding sites that are complementary and might in theory produce bifunctional enzymes. While the crystal structure of the LBPCheY-Δββ chimera displayed new interactions that were not intended in the design, it is remarkable that the independent lobes have the ability to reorder and form a single unit. These observations highlight the plasticity of the bi-lobular rossmannoid architectures, which is particularly evident in the structural variations among the PBP proteins associated with structural dynamics in this fold [48].

This 'hopeful monster' not only illustrates how evolution may have created new folds by recycling and modifying existing gene segments [19] but also demonstrates how evolutionary sequence information can be used to identify combinable units for protein design.

Acknowledgements

We acknowledge assistance from the beamline staff at the SLS. This work was supported by Deutsche Forschungsgemeinschaft grant HO 4022/1-2, the European Research Council (ERC Consolidator Grant 647548 'Protein Lego') and the VolkswagenStiftung (grant 94747) to BH. JAF-R acknowledges funding provided by UNAM-PAPIIT project IA204622, CONAHCYT fund FOP16-2021-01 project number 319320 and Deutscher Akademischer Austauschdienst (DAAD) funding program: Research Stays for University Academics and Scientists 2023 Personal ref. no.: 91877472. Open Access funding enabled and organized by Projekt DEAL.

Author contributions

ST-P, JAF-R and BH conceived and designed the project. ST-P, SKG, SS, JAF-R and BH acquired the data. ST-P, SKG, SS, JAF-R and BH analyzed and interpreted the data. ST-P, JAF-R and BH wrote the article.

Data accessibility

X-ray coordinates have been deposited to the protein data bank (PDB) with accession code 8Q52, succeeding 4QWV.

References

1 Söding J and Lupas AN (2003) More than the sum of their parts: on the evolution of proteins from peptides. *Bioessays* **25**, 837–846.
2 Alvarez-Carreno C, Gupta RJ, Petrov AS and Williams LD (2022) Creative destruction: new protein folds from old. *Proc Natl Acad Sci USA* **119**, e2207897119.
3 Farias-Rico JA and Mourra-Diaz CM (2022) A short tale of the origin of proteins and ribosome evolution. *Microorganisms* **10**, 2115.
4 Höcker B, Schmidt S and Sterner R (2002) A common evolutionary origin of two elementary enzyme folds. *FEBS Lett* **510**, 133–135.
5 Andreeva A, Kulesha E, Gough J and Murzin AG (2020) The SCOP database in 2020: expanded classification of representative family and superfamily domains of known protein structures. *Nucleic Acids Res* **48** (D1), D376–D382.
6 Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN and Bourne PE (2000) The protein data Bank. *Nucleic Acids Res* **28**, 235–242.
7 Romero-Romero S, Costas M, Silva Manzano DA, Kordes S, Rojas-Ortega E, Tapia C, Guerra Y, Shanmugaratnam S, Rodríguez-Romero A, Baker D *et al.* (2021) The stability landscape of de novo TIM barrels explored by a modular design approach. *J Mol Biol* **433**, 167153.
8 Alva V, Remmert M, Biegert A, Lupas AN and Söding J (2010) A galaxy of folds. *Protein Sci* **19**, 124–130.
9 Alva V, Soding J and Lupas AN (2015) A vocabulary of ancient peptides at the origin of folded proteins. *Elife* **4**, e09410.
10 Kolodny R, Nepomnyachiy S, Tawfik DS and Ben-Tal N (2021) Bridging themes: short protein segments found in different architectures. *Mol Biol Evol* **38**, 2191–2208.
11 Kolodny R (2021) Searching protein space for ancient sub-domain segments. *Curr Opin Struct Biol* **68**, 105–112.
12 Qiu K, Ben-Tal N and Kolodny R (2022) Similar protein segments shared between domains of different evolutionary lineages. *Protein Sci* **31**, e4407.
13 Ferruz N, Lobos F, Lemm D, Toledo-Patino S, Farías-Rico JA, Schmidt S and Höcker B (2020) Identification and analysis of natural building blocks for evolution-guided fragment-based protein design. *J Mol Biol* **432**, 3898–3914.
14 Zhang Y and Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* **33**, 2302–2309.
15 Farias-Rico JA and Höcker B (2013) Design of chimeric proteins by a combination of subdomain-sized fragments. *Methods Enzymol* **523**, 389–405.
16 Toledo-Patino S (2019) On the Emergence of the hemD-Like Fold and its Use for Fold-Chimeragenesis, in Biochemistry. Vol. **144**, University of Tübingen, Tuebingen, Germany.
17 Farias-Rico JA, Schmidt S and Höcker B (2014) Evolutionary relationship of two ancient protein superfolds. *Nat Chem Biol* **10**, 710–715.
18 Bharat TA, Eisenbeis S, Zeth K and Höcker B (2008) A beta alpha-barrel built by the combination of fragments from different folds. *Proc Natl Acad Sci USA* **105**, 9942–9947.
19 Höcker B (2013) Engineering chimaeric proteins from fold fragments: 'hopeful monsters' in protein design. *Biochem Soc Trans* **41**, 1137–1140.
20 Ferruz N, Michel F, Lobos F, Schmidt S and Höcker B (2021) Fuzzle 2.0: ligand binding in natural protein building blocks. *Front Mol Biosci* **8**, 715972.
21 Fukami-Kobayashi K, Tateno Y and Nishikawa K (1999) Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history. *J Mol Biol* **286**, 279–290.
22 Michel F, Shanmugaratnam S, Romero-Romero S and Höcker B (2023) Structures of permuted halves of a modern ribose-binding protein. *Acta Crystallogr D Struct Biol* **79** (Pt 1), 40–49.
23 Michel F, Romero-Romero S and Hocker B (2023) Retracing the evolution of a modern periplasmic binding protein. *Protein Sci* **32**, e4793.
24 Ribeiro LF, Amarelle V, Ribeiro LFC and Guazzaroni M-E (2019) Converting a periplasmic binding protein into a synthetic biosensing switch through domain insertion. *Biomed Res Int* **2019**, 4798793.
25 Marvin JS, Schreiter ER, Echevarría IM and Looger LL (2011) A genetically encoded, high-signal-to-noise maltose sensor. *Proteins* **79**, 3025–3036.
26 Mathews MA, Schubert HL, Whitby FG, Alexander KJ, Schadick K, Bergonia HA, Phillips JD and Hill CP (2001) Crystal structure of human uroporphyrinogen III synthase. *EMBO J* **20**, 5832–5839.
27 Toledo-Patino S, Chaubey M, Coles M and Höcker B (2019) Reconstructing the remote origins of a fold singleton from a flavodoxin-like ancestor. *Biochemistry* **58**, 4790–4793.
28 Eisenbeis S, Proffitt W, Coles M, Truffault V, Shanmugaratnam S, Meiler J and Höcker B (2012) Potential of fragment recombination for rational design of proteins. *J Am Chem Soc* **134**, 4019–4022.
29 Lo Conte L, Ailey B, Hubbard TJ, Brenner SE, Murzin AG and Chothia C (2000) SCOP: a structural

classification of proteins database. *Nucleic Acids Res* **28**, 257–259.
30 Brenner SE, Koehl P and Levitt M (2000) The ASTRAL compendium for protein structure and sequence analysis. *Nucleic Acids Res* **28**, 254–256.
31 Altschul SF and Koonin EV (1998) Iterated profile searches with PSI-BLAST—a tool for discovery in protein databases. *Trends Biochem Sci* **23**, 444–447.
32 Söding J (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**, 951–960.
33 Velankar S, van Ginkel G, Alhroub Y, Battle GM, Berrisford JM, Conroy MJ, Dana JM, Gore SP, Gutmanas A, Haslam P *et al.* (2016) PDBe: improved accessibility of macromolecular structure data from PDB and EMDB. *Nucleic Acids Res* **44** (D1), D385–D395.
34 Farias-Rico JA (2013) Evolutionary relationships beyond fold boundaries. Doctoral Thesis, p. 155.
35 Toledo-Patino S (2019) On the Emergence of the hemD-like Fold and its Use for Fold-Chimeragenesis. Doctoral Thesis.
36 Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* **66** (Pt 2), 125–132.
37 McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC and Read RJ (2007) Phaser crystallographic software. *J Appl Cryst* **40** (Pt 4), 658–674.
38 Karine M, Sparta MK, Heinemann U, Mueller U and Weiss MS (2016) XDSAPP2.0. *J Appl Cryst* **49**, 1085–1092.
39 Emsley P, Lohkamp B, Scott WG and Cowtan K (2010) Features and development of coot. *Acta Crystallogr D Biol Crystallogr* **66** (Pt 4), 486–501.
40 Afonine PV, Poon BK, Read RJ, Sobolev OV, Terwilliger TC, Urzhumtsev A and Adams PD (2018) Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74** (Pt 6), 531–544.
41 Knudsen M and Wiuf C (2010) The CATH database. *Hum Genomics* **4**, 207–212.
42 Schaeffer RD, Liao Y, Cheng H and Grishin NV (2017) ECOD: new developments in the evolutionary classification of domains. *Nucleic Acids Res* **45** (D1), D296–D302.
43 Krissinel E and Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60** (Pt 12 Pt 1), 2256–2268.
44 Holm L and Sander C (1995) Dali: a network tool for protein structure comparison. *Trends Biochem Sci* **20**, 478–480.
45 Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen M-Y, Pieper U and Sali A (2006) Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* **Chapter 5**, Unit-5.6.
46 Ferruz N, Schmidt S and Höcker B (2021) ProteinTools: a toolkit to analyze protein structures. *Nucleic Acids Res* **49** (W1), W559–W566.
47 Toth-Petroczy A and Tawfik DS (2014) The robustness and innovability of protein folds. *Curr Opin Struct Biol* **26**, 131–138.
48 Gouridis G, Muthahari YA, de Boer M, Griffith DA, Tsirigotaki A, Tassis K, Zijlstra N, Xu R, Eleftheriadis N, Sugijo Y *et al.* (2021) Structural dynamics in the evolution of a bilobed protein scaffold. *Proc Natl Acad Sci USA* **118**, e2026165118.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Domain dissection of the periplasmic-binding protein-like I proteins by three major domain classification databases.

Fig. S2. Homologous regions between the periplasmic-binding protein-like I fold (PBP) and eight flavodoxin-like (FL) superfamilies.

Fig. S3. DSC measurement of LBPCheY-Δββ.

Fig. S4. Comparison of the LBPCheY-Δββ structure to its parents.