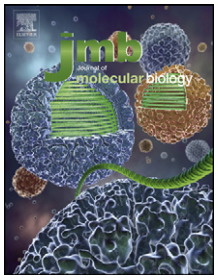


DARPin-Based Crystallization Chaperones Exploit Molecular Geometry as a Screening Dimension in Protein Crystallography

Alexander Batyuk†, Yufan Wu†, Annemarie Honegger, Matthew M. Heberling and Andreas Plückthun

Department of Biochemistry, *University of Zurich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland*

Correspondence to Andreas Plückthun: *plueckthun@bioc.uzh.ch*
http://dx.doi.org/10.1016/j.jmb.2016.03.002
Edited by Amy Keating

Abstract

DARPin libraries, based on a Designed Ankyrin Repeat Protein consensus framework, are a rich source of binding partners for a wide variety of proteins. Their modular structure, stability, ease of *in vitro* selection and high production yields make DARPins an ideal starting point for further engineering. The X-ray structures of around 30 different DARPin complexes demonstrate their ability to facilitate crystallization of their target proteins by restricting flexibility and preventing undesired interactions of the target molecule. However, their small size (18 kDa), very hydrophilic surface and repetitive structure can limit the DARPins' ability to provide essential crystal contacts and their usefulness as a search model for addressing the crystallographic phase problem in molecular replacement. To optimize DARPins for their application as crystallization chaperones, rigid domain–domain fusions of the DARPins to larger proteins, proven to yield high-resolution crystal structures, were generated. These fusions were designed in such a way that they affect only one of the terminal capping repeats of the DARPin and do not interfere with residues involved in target binding, allowing to exchange at will the binding specificities of the DARPin in the fusion construct. As a proof of principle, we designed rigid fusions of a stabilized version of *Escherichia coli* TEM-1 β-lactamase to the C-terminal capping repeat of various DARPins in six different relative domain orientations. Five crystal structures representing four different fusion constructs, alone or in complex with the cognate target, show the predicted relative domain orientations and prove the validity of the concept.

Introduction

The generation of diffraction quality crystals represents a major bottleneck in macromolecular X-ray crystallography. Nucleation and growth of crystals depend on the three-dimensional arrangement of energetically favorable intermolecular interactions, leading to the non-covalent association of the molecules in a well-ordered, rigid three-dimensional lattice. Some proteins seem to achieve this easily under a wide variety of conditions (e.g., T4 lysozyme [1]), while others prove very difficult and some even impossible to crystallize.

Different precipitants and additives modify the relative contribution of hydrophobic and electrostatic interactions to the interaction energies between the molecules in the crystal—high salt conditions weaken electrostatic repulsions and strengthen hydrophobic interactions, and organic modifiers moderate hydrophobic interactions and strengthen electrostatic contributions. Various additives can facilitate crystallization by a wide variety of mechanisms—divalent ions can locate between identical charges and convert a repulsive interaction into an attractive one; hydrophobic molecules bind in protein cavities and pockets to rigidify the protein or intercalate at interfaces between two molecules; and specific ligands and inhibitors decrease conformational heterogeneity by selectively stabilizing one conformation of the protein. Commercial crystallization screening kits have facilitated the search for those often elusive sets of conditions, which in a supersaturated protein solution give rise to sparse

nucleation events, leading to a small number of slowly growing, homogeneous single crystals of sufficient size.

Even small changes in the protein sequence such as single-point mutations can have a profound influence on crystallization success and crystal packing geometry [2]. Screening of homologous proteins from various organisms, deletion of disordered terminal sequences, excision of long flexible loops and production of individual domains out of a larger protein are some of the standard techniques used in the search for more readily crystallizable variants of a protein [3–5]. Additional protein modification options include mutagenesis to remove posttranslational modifications, which can decrease sample heterogeneity, or to decrease surface entropy [6], or to introduce binding sites for heavy atoms to assist in experimental phasing.

To facilitate expression and purification of troublesome proteins, fusing these to well-folded protein affinity domains may be an option [7]. Such tags can increase expression levels, enhance solubility, protect from proteolysis, improve folding yield and facilitate protein purification *via* affinity chromatography. Attached to the N or C terminus of the protein *via* a flexible linker, they are usually removed by proteolytic cleavage prior to crystallization. In some instances, such tags, connected by a very short linker to the protein of interest, have been successfully retained throughout crystallization and have provided crucial crystal contacts [8,9].

For the crystallization of membrane proteins, fusion of protein domains has become a prevalent strategy. Membrane proteins require detergents to mask the large hydrophobic surfaces that are embedded in the membrane within the cellular environment. Particularly for membrane proteins that do not protrude far from the membrane, this represents a critical problem, as only a fraction of the molecular surface is accessible to form specific crystal contacts. Crystallization of G-protein coupled receptors beyond rhodopsin has been greatly enhanced by replacing the large and highly variable third intracellular loop by T4 lysozyme or apo-cytochrome B(562)RIL [10,11], or by fusing these domains to the N terminus of the protein [12].

As an alternative to covalent fusions, poorly crystallizing proteins can be complexed with crystallization chaperones (reviewed in Refs. [13,14]). These chaperone proteins bind non-covalently with high affinity and specificity to the protein of interest to provide alternative potential crystal contacts, which increases the surface area not masked by lipids or detergent, decreases conformational flexibility and prevents undesired interactions of the target protein. Traditionally, Fab fragments derived from monoclonal antibodies have been used to facilitate the crystallization of a wide variety of membrane and soluble proteins [15,16] (e.g., the cytochrome b1c complex, various potassium and chloride channels). More recently, binders based on a variety of alternative frameworks have become available, allowing for the exploration of a wider range of potential crystallization chaperones. However, most are significantly smaller than Fab fragments of ~50 kDa: antibody single-chain fragments (scFv), ~25 kDa; lipocalins (anticalins), ~20 kDa; designed ankyrin repeat proteins (DARPins), ~18 kDa (N3C); camelid VHH domains (nanobodies), ~12 kDa; FN3 domain based binders (e.g., monobodies), ~10 kDa; SH3 domains (e.g., fynomers), ~7.5 kDa; protein A domains (affibodies), ~7 kDa; and cystine-knot miniproteins, ~3–4 kDa. Consequently, these new binders have a smaller surface area and do not have sufficient reach to create essential crystal contacts, which prompts the need for more efficient binders [17].

In this paper, we present an approach to optimize the performance of DARPins [18] as crystallization chaperones. By rigidly fusing various protein domains to either the N- or the C-terminal capping repeat of DARPins selected against a target protein, each DARPin can give rise to a diverse series of distinct crystallization chaperones. Each DARPin can have different fusion partners, and each fusion partner, in turn, can have several different fusion geometries, thereby providing an entire new screening dimension in crystallization trials. This diversity of potential crystal interfaces is generated not by trial-and-error engineering of a hard-to-express crystallization target of unknown structure, but by selecting binders against the target from a DARPin library [19] and then replacing the capping repeats of each binder by a generic set of different fusion caps.

Results

Design of rigid domain–domain fusions

The DARPin fold starts and ends with an α-helix; by extending one of these helices into a compatible terminal helix of a partner domain, a rigid fusion of the two domains can be obtained. The Protein Data Bank (PDB)[‡] was screened for potential fusion partners: high-resolution (<1.2 Å) structures of large (>150 amino acids) protein domains that contain an N- and/or a C-terminal α-helix and that can be produced with high yields in the cytoplasm of *Escherichia coli*. TEM-1 β-lactamase (BL, PDB ID: 1M40, 263 aa, 0.85 Å resolution [20]), containing both an N-terminal and a C-terminal α-helix, was identified as the most promising fusion partner.

To identify viable DARPin–β-lactamase (DB) fusion constructs, an ideal poly-alanine helix was superimposed on the C-terminal helix (residues 148–156) of the DARPin off7 (PDB ID: 1SVX, 2.24 Å resolution) which recognizes maltose-binding protein (MBP; MW = 42 kDa) [21]. Residues 1–23 of BL correspond to the signal peptide not present in the mature protein.

The N-terminal helix (residues 26–38) of the TEM-1 β-lactamase was superimposed on the ideal helix in different registers, resulting in different orientations relative to the DARPin. We considered shared helix lengths from 8 to 32 aa (Fig. 1). Alignments that lead to backbone clashes between the fusion partners were rejected, leaving six potentially viable C-terminal fusions: DB01, 04, 08, 12, 15 and 18 (DB stands for DARPin–β-lactamase fusion, the numbers signify alignment shifts relative to the construct with the shortest shared helix). In the shortest variant, DB01, the shared helix has a length of 9 residues. The longest construct, DB18 with a shared helix length of 26 residues, corresponds to an end-to-end fusion of the two native helices.

Models of the fusion constructs were built using the Homology module of program InsightII (Accelrys, San Diego) to align the templates and generate a continuous model. Shared helix residues with side chains pointing into the core of the DARPin (residues 161, 162, 165 and 166 for a DARPin with three internal repeats) were assigned the DARPin sequence, while side chains pointing into the core of BL (residues 27, 28, 30, 31, 34, 35, 38 of BL) were assigned the BL sequence. Residues involved in steric clashes were mutated based on visual inspection of the models, resulting in version 1 constructs, DB01_v1 to DB18_v1. The Discover module of InsightII was used for energy minimization of the final constructs.

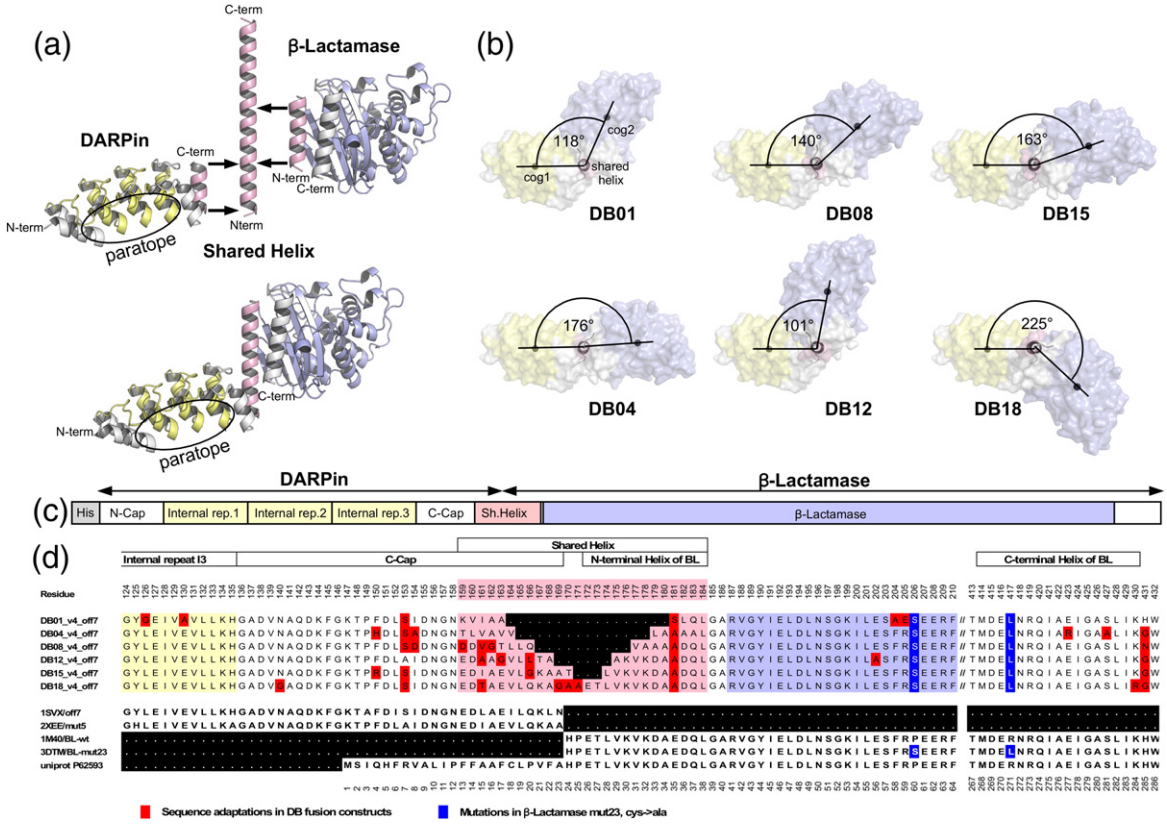
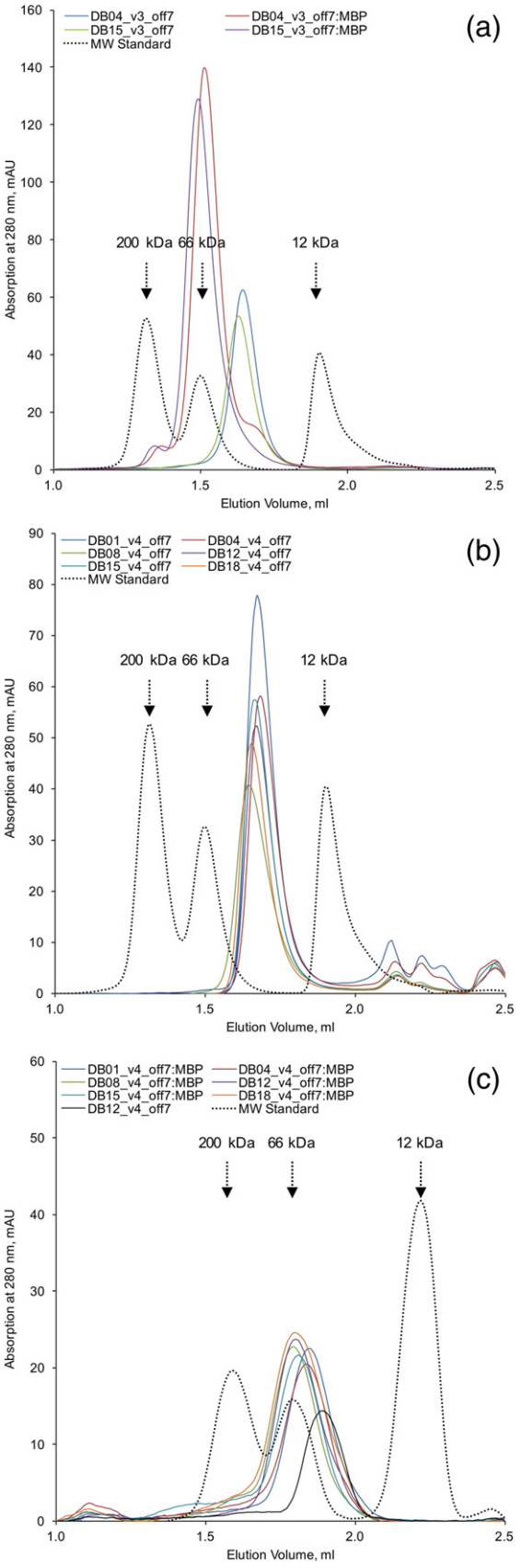


Fig. 1. DARPin–β-lactamase fusion concept. (a) A rigid fusion of a DARPin to a second domain (e.g., β-lactamase) containing an N-terminal helix is achieved through joining the C-terminal helix of a DARPin to the N-terminal helix of the partner domain in such a way that a continuous helix is formed that is embedded in at least one of the two domains along its entire length. Different alignments of the two helices result in different relative orientations of the two domains. (b) The relative orientation of the two domains is defined as the pseudo-torsion angle between the centers of gravity of the Cα positions of DARPin residues 37–135 (the three internal repeats) and of BL residues 41–265 around the axis of the shared helix. Views are along the shared helix. (c) The final protein sequences comprise an N-terminal MRGS-His_6 tag, the N-cap of the DARPin, internal repeats (as needed for binding, here three are shown), part of the C-cap of the DARPin, a shared helix derived from the overlapping C-terminal helix of the DARPin and the N-terminal helix of the β-lactamase, and the remainder of the β-lactamase. (d) Sequences of the shared helix construct with adaptations highlighted that were introduced to ensure folding and stability. Note that positions 1-123 and 211-412 are omitted for space reasons. The bottom five rows show the sequences of the original fusion partners. For complete sequences of all constructs, see Supplementary Fig. S1.



By altering the length of the shared helix, the two domains change relative orientation. We can define a pseudo-torsion angle between DARPin and BL, using the shared helix as the connecting "bond." This pseudo-torsion angle is measured between the centers of gravity of the Cα positions of DARPin residues 37–135 (the three internal repeats) and of BL residues 41–265 around the axis of the shared helix. Thus, the two domains were oriented at an angle of 114° (DB01), 177° (DB04), 138° (DB08), 102° (DB12), 162° (DB15) and 224° (DB18). Expression tests (*E. coli* cytoplasmic expression in standard shake flask culture) showed that constructs DB04, DB15 and DB18 could be expressed in soluble form, yielding about 5 mg/l of purified protein for the variant DB04, and 15 to 18 mg/l of protein for the variants DB15 and DB18 (data not shown). The other constructs were predominantly produced in insoluble form. However, attempts to crystallize the three initial constructs failed.

In a next design iteration, (version 2, DBxx_v2), the sequence of the shared helix and all residues in direct van-der-Waals contact with the shared helix were repacked using Rosetta *fixbb* [22], where we allowed mutations to all amino acids except cysteine, glycine and proline in helical positions and to all amino acids, except cysteine in other positions. However, none of the resulting constructs could be expressed in soluble form.

A third series of constructs (DBxx_v3) retained the same shared helix design as in version 2, but the BL was replaced by a stabilized version. In the first two designs, a BL had been used with two mutations (M180T and V82I compared to UniProt ID: P62593). In version 3, a BL was used (termed "variant 23" [23]) containing nine additional point mutations (P60S, V78I, E145G, A182V, L199P, I206M, A222V, I244V, R271L) reported to increase the melting temperature from 50.8 °C to 69.2 °C. None of the stabilizing residues in the v3 BL are in direct van-der-Waals contact with the shared helix. In addition, Cys 75 and Cys 121 were replaced by Ala, yielding a cysteine-free version of the stabilized β-lactamase to avoid potential heterogeneity of the fusion constructs due

Fig. 2. Analytical SEC result of DB fusion constructs carrying MBP-binding off7 with and without MBP. (a) Two soluble variants from the third design version (DB04_v3_off7 and DB15_v3_off7) are monomers in solution according to their molecular masses, and a clear shift (42 kDa) is observed upon MBP binding. (b) All six variants from the forth design version DBxx_v4_off7 are monomers in solution according to their molecular masses. (c) All designed DBxx_v4_off7 variants bind MBP. A clear shift is observed upon MBP binding, which corresponds to a 42-kDa increase in molecular mass. A mixture of three protein samples (β-amylase, 200 kDa; albumin, 66 kDa; cytochrome *c*, 12 kDa) was used as molecular mass standards, shown as arrows in all figures.

to incomplete disulfide formation in cytoplasmic expression. The Cys → Ala mutations significantly reduced the catalytic activity of the β-lactamase, but this was not relevant for our application. They also decreased the melting temperature by about 2 °C.

Constructs DB04_v3 and DB15_v3 could be expressed in soluble form, with yields of approximately 20 mg/l. While DB15_v3 only produced very poorly diffracting crystals, two variants of DB04_v3 (with different DARPins) yielded diffraction-quality crystals: the original DB04_v3_off7 and DB04_v3_D12, in which the specificity-determining internal repeats had been replaced by those of DARPin 5m3_D12 (abbreviated as D12), recognizing the gp120 V3 loop of the human immunodeficiency virus (HIV) envelope spike [24]. We refer to these fusion proteins as DB (for DARPin–β-lactamase fusion), followed by the shared helix length number (as defined above), design version number and the DARPin variant (e.g., DB04_v3_off7). The complex with its cognate target, MBP, would be called DB04_v3_off7:MBP.

Upon careful re-analysis of the models in comparison to the multiple DARPin and β-lactamase structures, clashes became evident between the N- and C-terminal helices of BL in the fusion construct that warranted a residue mutation to glycine, which was not present in the original BL sequence. In earlier versions, this clash was considered small enough to be rectified by minor main-chain adjustments, since the clash could be relieved by relatively mild energy minimization. New models of the fusion constructs were built based on the structure of C-cap stabilized consensus DARPin mut5 (PDB ID: 2XEE, 2.10 Å resolution, [25]) and stabilized BL variant 23 (3DTM, 2.10 Å resolution, [23]), avoiding all backbone energy minimization except for the residues flanking the splice points between different modeling templates. As in version 1, shared helix residues were assigned the sequence of the domain into which they packed. After Rosetta [22] relaxation with backbone constraints and rescoring of the models, residue scores were analyzed. Residues within or in contact with the shared helix showing repulsive scores >2 were repacked using Rosetta module *fixbb*, now allowing glycine residues within the helices if a clash could not be removed by mutating clashing non-helical positions.

This remodeling (version 4) resulted in minor changes of the relative domain orientation: 114° to 118° (DB01), 177° to 176° (DB04), 138° to 140° (DB08), 102° to 101° (DB12), 162° to 163° (DB15) and 224° to 225° (DB18); the first value refers v1 to v3 and the second value to v4. These changes were mostly due to changes in the take-off angle of the stabilized terminal helix of the DARPin, which had been taken from the C-cap stabilized consensus DARPin mut5 (PDB ID 2XEE) [26] and differed from DARPin off7. As a last modeling step, N-cap and internal repeat sequences of consensus DARPin (PDB ID: 2XEE) were replaced by those of DARPin off7 that recognizes the MBP [21]. The amino acid sequences of all designed models for all 4 versions are listed in Supplementary Fig. S1.

All six v4 constructs could be produced with good yields in *E. coli* and bound the cognate target MBP, which was verified by ELISA (Supplementary Fig. S2) and size exclusion chromatography (Fig. 2). To facilitate the exchange of DARPin binding specificities and shared helix types, a KasI restriction site was introduced between the last internal repeat and the C-cap of the original DARPin (or the C-cap/shared helix module of the fusion constructs) facilitating the exchange of the binding specificity of DARPin off7 with any other DARPin [24,27–29].

Characterization

Eight DB fusions (DB04_v3, DB15_v3 and the six DBxx_v4 constructs) containing the MBP-specific DARPin off7 were expressed as soluble proteins in the cytoplasm of *E. coli* and purified by immobilized metal-ion affinity chromatography, each yielding 20–30 mg of pure protein per liter of shake flask culture. ELISA was used to qualitatively assess binding of the fusion constructs to their target, and all constructs tested were found to retain their specific binding to the target (Supplementary Fig. S2).

Analytical size exclusion chromatography in the absence and presence of the target molecule was used to determine the oligomeric state of the DB fusion constructs and their complexes with the cognate target (MBP, 42 kDa). All eight DB constructs were monomeric at a concentration of 10 μM (Fig. 2a and b) and formed a homogeneous complex with MBP, as indicated by an increase of the apparent molecular weight of the complex by 42 kDa relative to the unliganded DB construct (Fig. 2a and c).

Crystallization

Crystallization was set up using a variety of screens available at the in-house Protein Crystallization Center[§]. Crystals usually appeared within a week and grew to their maximum size within 2–3 weeks (Fig. 3a). Initial screens revealed that some DB fusion variants crystallized readily over a wide pH range. Based on the results of the initial screens, the best hits were further optimized, and the resulting crystals were used in X-ray diffraction experiments. Crystallization conditions are summarized in Table 1.

We were able to determine five structures, representing four out of the six DARPin–β-lactamase fusions, two of them in complex with the cognate target (Fig. 3b): two DB04_v3 constructs containing different DARPin recognition modules (DB04_v3_off7 and DB04_v3_D12 in the unliganded state) and three v4 constructs [DB08_v4_off7 in complex with MBP,

DB12_v4_off7 without ligand and DB15_v4_3G61 complexed with green fluorescent protein (GFP)]. Crystallization failed for the shortest (DB01) and the longest constructs (DB18).

Anti-MBP construct DB04_v3_off7 crystallized in space group $P6_2$ (resolution 2.6 Å) resolution. Two molecules in the asymmetric unit formed the major crystal contact (466.7 $Å^2$) in a back-to-back manner. DB04_v3_D12, which differs only in the DARPin sequence and specificity, crystallized in space group $P2_12_12_1$ with two molecules in the asymmetric unit (resolution 2.1 Å). DB12_v4_off7 crystallized in space group $P2_12_12_1$ with one molecule in the asymmetric unit(resolution 1.62 Å). DB08_v4_off7:MBP and DB15_v4_3G61:GFP were crystallized as complexes with their respective targets MBP and GFP. DB08_v4_off7:MBP crystallized in space group $P2_12_12_1$(resolution 1.86 Å), with two copies in the asymmetric unit. DB15_v4_3G61:GFP crystallized in space group $C121$(resolution 1.37 Å) and one complex in the asymmetric unit. Data collection and refinement statistics for all five structures are shown in Table 1.

Molecular replacement

Initial attempts to determine the structure of the construct DB04_v3_off7 using the designs of the DARPin–β-lactamase fusion as a search model in Phaser [30] produced the correct solution, with two molecules in the asymmetric unit. However, visual inspection of the electron density map, after three rounds of refinement in Phenix [31] (R_{work}/R_{free} = 0.36/0.47), revealed rotational rigid body movement between β-lactamase and DARPin domains centered at the shared helix, leaving the DARPin slightly out of the electron density. To find the precise position of the domains, molecular replacement was performed with the stabilized β-lactamase (PDB ID: 3DTM) and DARPin off7 (PDB ID: 1SVX) as separate search models. Initially, molecular replacement with the new ensemble failed to find all the components, yielding only two β-lactamase molecules. The default selection criteria in the rotation function in Phaser had to be loosened to retain the peaks with a difference of 65% from the mean,

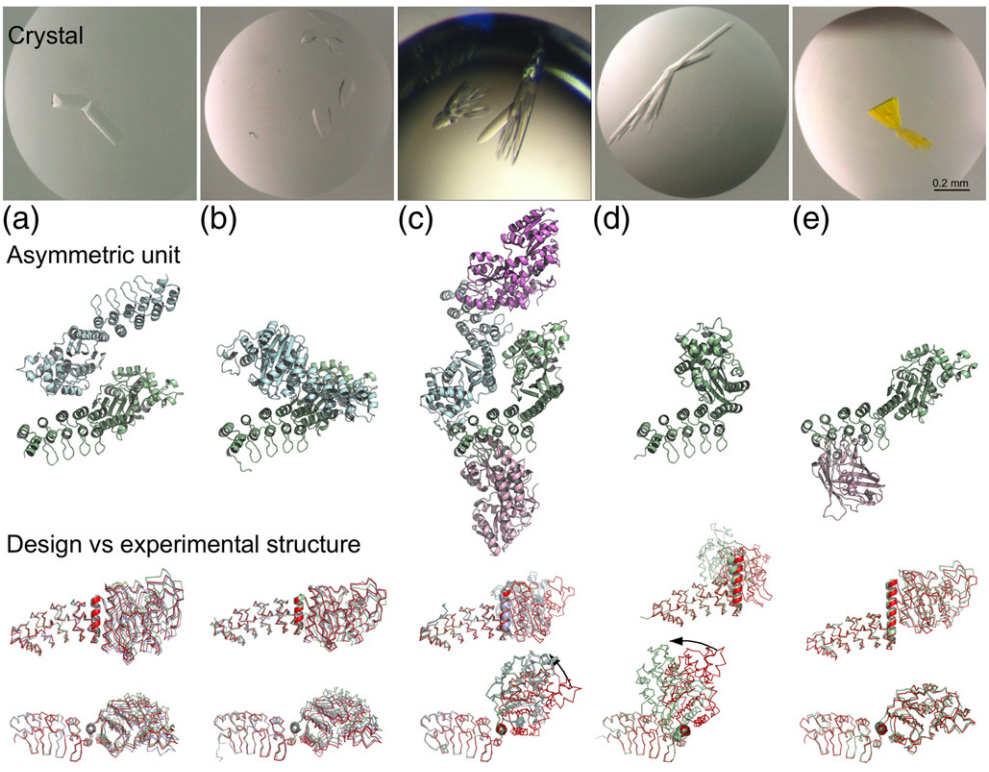


Fig. 3. Experimental structures. Top: crystals of constructs (a) DB04_v3_D12, (b) DB04_v3_off7, (c) DB08_v4_off7:MBP, (d) DB12_v4_off7 and (e) DB15_v4_3G61:GFP. Middle: asymmetric units in these crystals. DARPin fusions are shown in pale green and pale blue; DARPin-bound target proteins MBP (c) and GFP (e) are shown in pink and magenta, respectively. Bottom: comparison of version 4 designs (red) to the experimental structures (pale green and pale blue), views perpendicular to and along the shared helix. Models and structures were superimposed by a least-squares fit of the Cα positions of the DARPin internal repeats (residues 37–135).

Table 1. Data collection and refinement statistics

	DB04_v3_off7	DB04_v3_D12	DB08_v4_off7:MBP	DB12_v4_off7	DB15_v4_3G61:GFP
	PDB ID 5AQA	PDB ID 5AQ7	PDB ID 5AQ9	PDB ID 5AQ8	PDB ID 5AQB
Data collection					
Resolution range (Å)	47.09–2.6 (2.693–2.6)	43.49–2.1 (2.175–2.1)	48.05–1.86 (1.926–1.86)	43.72–1.62 (1.678–1.62)	48.09–1.37 (1.419–1.37)
Space group	$P6_2$	$P2_12_12$	$P2_12_12_1$	$P2_12_12_1$	$C121$
Molecules/AU	2	2	4 (2 complexes)	1	2 (1 complex)
Unit cell parameters					
a, *b*, *c* (Å)	163.12, 163.12, 66.76	79.04, 156.24, 68.47	41.48, 191.12, 219.51	41.07, 76.64, 106.45	90.15, 96.17, 92.78
α, *β*, *γ* (°)	90, 90, 120	90, 90, 90	90, 90, 90	90, 90, 90	90, 118.99, 90
Unique reflections	31,357 (3047)	50,274 (4924)	148,191 (14,557)	43,423 (4268)	143,484 (13,218)
Multiplicity	10.0 (9.4)	12.0 (11.3)	13.1 (13.1)	6.3 (6.0)	6.6 (5.8)
Completeness	1.00 (0.98)	1.00 (0.99)	1.00 (1.00)	1.00 (0.99)	0.99 (0.92)
$\langle I \rangle / \sigma(I)$	10.95 (0.32)	16.06 (1.26)	21.83 (1.81)	11.13 (1.51)	7.00 (0.39)
CC(1/2)	0.997 (0.144)	0.999 (0.547)	1 (0.67)	0.997 (0.705)	0.993 (0.252)
Wilson B-factor	61.93	34.29	29.69	18.66	20.26
Refinement					
R_{work} (%)	0.2293 (0.3825)	0.1899 (0.3205)	0.1783 (0.2698)	0.1862 (0.2661)	0.1598 (0.3922)
R_{free} (%)	0.2789 (0.4032)	0.2300 (0.3435)	0.2086 (0.2986)	0.2352 (0.3686)	0.1789 (0.4189)
Ordered water molecules	65	442	1068	289	707
Protein atoms	6226	6186	12,150	3139	5149
Thiocyanate ion (SCN)	4			3	
Malonate ion (MLI)		2			
Hepes (EPE)				2	
Chromophore (CRO)					1
rmsd of bond lengths	0.003	0.003	0.008	0.005	0.008
rmsd of bond angles	0.84	0.71	1.06	0.84	1.23
Average *B*-factor	78.70	50.70	48.10	26.00	30.80
Ramachandran plot (%)					
Most favored	96.56	98.16	98.10	99.03	98.22
Additionally allowed	3.44	1.84	1.90	0.72	1.78
Generously disallowed	0	0	0	0.24	0
Crystallization conditions	25.0% polyethylene glycol 600 0.15 M potassium thiocyanate 0.1 M Tris (HOAc), pH 8.0	20.0% polyethylene glycol 3350 0.2 M sodium malonate 0.1 M Bis Tris propane, pH 7.5	17.0% polyethylene glycol 6000 0.2 M ammonium chloride 0.05 M Hepes, pH 7.1	17.8% polyethylene glycol 4000 0.27 M potassium thiocyanate 0.15 M ammonium bromide 0.1 M Hepes, pH 7.5	20.0% polyethylene glycol 3350 0.2 M sodium formate 0.1 M Bis Tris propane, pH 8.5

Statistics for the highest-resolution shell are shown in parentheses.

yielding two β-lactamase molecules and one DARPin molecule. Further loosening of the selection criteria to 55% yielded a solution with all four components. The search order did not influence the overall result. The two β-lactamase molecules were always found first, suggesting that the DARPin is not a very potent search model in molecular replacement at this resolution, (3.11 Å with $I/\sigma(I)$ = 2.12) and underlining the value of the fusion protein strategy. Three rounds of refinement of this solution with *phenix.refine* produced a well interpretable electron density map and R_{work}/R_{free} = 0.25/0.35. This electron density map allowed us to build the complete experimental model of the DARPin–β-lactamase fusion variant 04 with Coot [32]. The residues of the shared helix interface were first added as alanines and then mutated to the actual residues. Iterative cycles of model building and refinement were performed until the model was complete with final R_{work}/R_{free} = 0.23/0.28. The structures of DB04_v3_D12 and DB12_v4_off7 were solved by molecular replacement with the same truncated search models as above.

In case of the DB08_v4_off7:MBP complex, initial trials to solve the structure by molecular replacement using the off7:MBP complex (PDB ID: 1SVX) and the stabilized β-lactamase (PDB ID: 3DTM) as search models yielded a correct solution with the final log-likelihood gain (LLG) of 3652 containing only two copies of the β-lactamase [refined translation function *Z*-score (TFZ) of 7.7 and 12.0 for the first and second copies, respectively]. This was surprising since the off7:MBP complex corresponds to 60.1% of the scattering mass in the unit cell versus 30.4% of the scattering mass for the β-lactamase. There was positive electron density in the Fo–Fc difference map, but it was not immediately obvious whether the crystal contained the DB08_v4_off7:MBP complex or only the DB08_v4_off7 fusion protein. The sequential molecular replacement approach was then undertaken to solve this ambiguity. First, the search was performed with β-lactamase, followed by the DARPin. After finding a very convincing solution (final LLG = 5685) with two β-lactamase and two DARPin copies placed, this solution was fixed and an additional search was performed for two copies of MBP. This strategy yielded the correct solution with all the components. The structure of DB15_v4_3G61:GFP complex was then solved by molecular replacement with the sequential search strategy using β-lactamase, DARPin and GFP (PDB ID: 1GFL) as search models.

Since repeat proteins such as DARPins seem not to work very well for phasing by molecular replacement, we compared different strategies, and used the DB15_v4_3G61:GFP dataset for this purpose. The resolution of the dataset (1.37 Å) was cut at 2.5 Å for the calculations and separate searches were performed with β-lactamase followed by the DARPin, β-lactamase alone, DARPin alone, the Rosetta-designed model, and GFP alone for comparison. The results of these calculations (Fig. 4) show that the most successful strategy using the constructs for molecular replacement was to search with β-lactamase followed by the DARPin. The DARPin alone does not perform very well even at this resolution, perhaps because of its repeat nature, leading to register shifts between the repeats. Despite the close agreement between design and experimental structure, the Rosetta-designed model performed poorly due to the movement of the β-lactamase and the DARPin domains relative to each other, which is small but not negligible.

The important role of the β-lactamase domain in crystal contacts

To assess how the β-lactamase fusion partner contributed to successful crystallization of DARPins and DARPin-target complexes, we analyzed all significant crystal contacts (Supplementary Tables ST1–ST5) by calculating the area, and the number of intermolecular H-bonds and salt bridges using PDBePISA[II]. As a typical example, in the unit cell of DB12_v4_off7 (Supplementary Table ST3, Fig. 5a and b), the residues that contribute most to the two major intermolecular crystal contacts are mainly from the β-lactamase domain. The two largest crystal contacts interfaces are formed between DB12_v4_off7 and its symmetry mates ($-x$, $y - 1/2$, $-z + 1/2$) and ($x - 1$, y, z) with areas of 1010.9 Å^2 and 647.1 Å^2, respectively. Seven H-bonds are formed to stabilize the crystal contacts by involving seven key residues (Arg221, Gln226, Lys284, Asn313, Gln344, Asp347 and Asp390) from the β-lactamase domain (Fig. 5c–f). Furthermore, the β-lactamase domain participates in seven out of all eight intermolecular interfaces and provides 93.9% of the total area of all interfaces in the unit cell. It is involved in 16 out of 21 interactions, including all 9 salt bridges.

In the three unliganded DB fusion structures, the β-lactamase domain provides at least 80% of the crystal contact interface area (Supplementary Tables ST1–ST3). In the case of the two complexes, the β-lactamase domain still provides 46.8% (DB08_v4_off7:MBP) and 77.1% (DB15_v4_3G61:GFP) of all crystal packing interface, in spite of providing only 30.3% (26,905/88,727 Da) and 36.2% (26,905/74,376 Da) of the mass of the complex. These results demonstrate that the β-lactamase domain is playing a key role in the formation of crystal contacts by the DB fusion constructs, demonstrating that we achieved the main design goal.

Interactions between the fused DARPins and the cognate target in the complex structures

The interaction between the DARPin off7 and MBP in the context of the DB08_v4_off7:MBP complex was

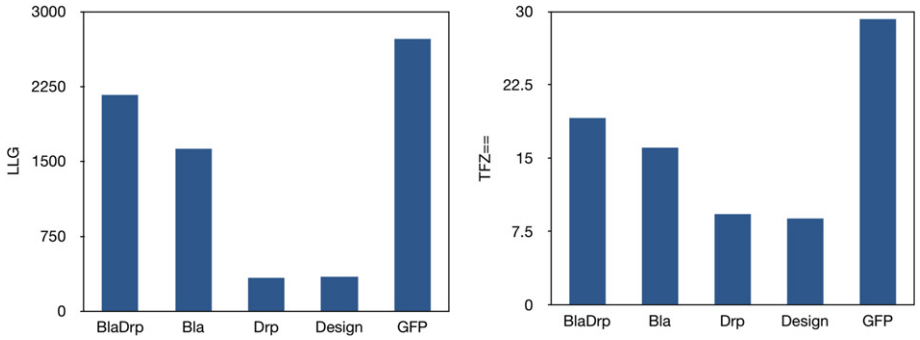


Fig. 4. Performance of search models in molecular replacement. Log-likelihood gain (LLG; left panel) and refined TFZ equivalent (TFZ==; right panel) in various search strategies: β-lactamase followed by the DARPin (BlaDrp), β-lactamase alone (Bla), DARPin alone (Drp), the Rosetta-designed model (Design) and GFP alone (GFP). LLG is an indication of how much "better" the solution is compared to a random solution: the higher the LLG value, the better the solution. The TFZ indicates how many standard deviations (σ) the solution score (X) is above the mean score (μ), $z = (X - \mu)/\sigma$; again, the higher the better.

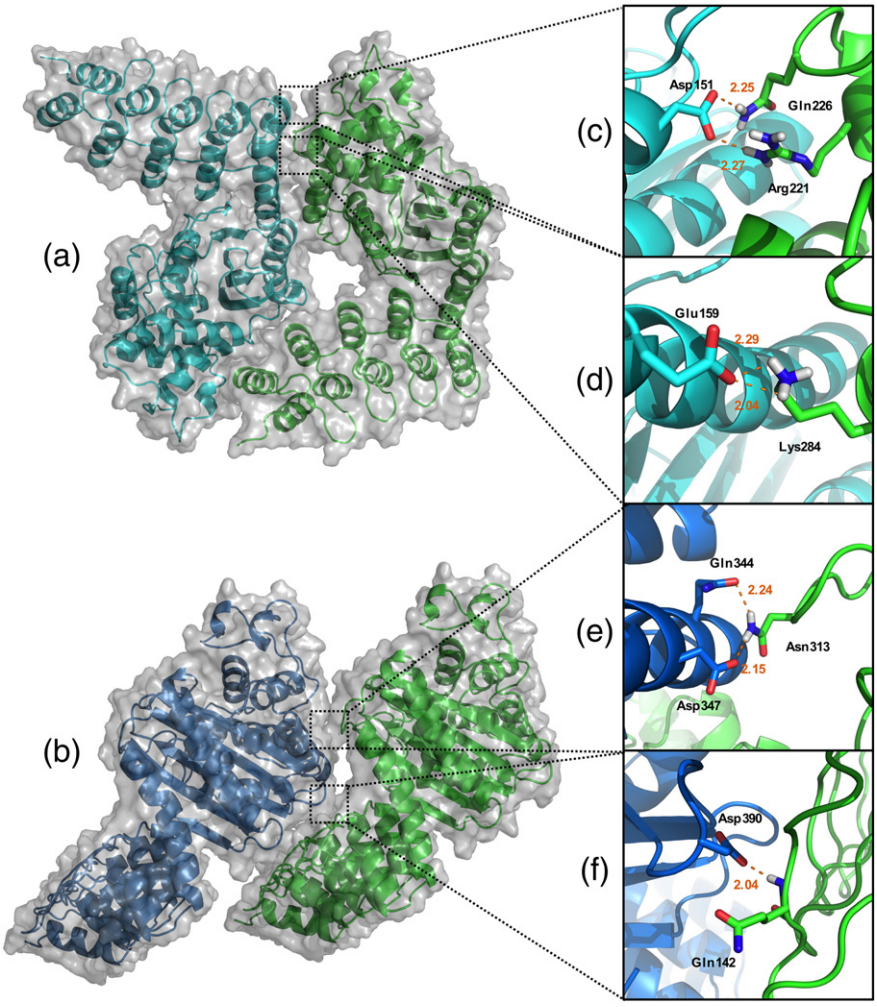


Fig. 5. Major crystal interfaces in DB12_v4_off7. Two major crystal contacts between DB12_v4_off7 and its symmetry mate are shown in cartoon and surface representation (a and b). DB12_v4_off7 is shown in green, its symmetry mates ($-x$, $y - 1/2$, $-z + 1/2$) and ($x - 1$, y, z) in cyan and marine. Seven H-bonds (c–f) are formed to stabilize the crystal contacts predominantly with the β-lactamase domain. The residues involved in forming H-bonds are shown in sticks, H-bonds including the distances are shown in orange.

very similar to the original off7:MBP complex (PDB ID: 1SVX), as expected. DARPin 3G61 is a GFP binder selected by ribosome display [27]. Initial attempts to crystallize the unfused DARPin:GFP complex had failed, while the DB15_v4_3G61:GFP complex yielded crystals that diffracted to 1.37 Å. As expected, the binding interface (Fig. 6a) is formed by the concave randomized surface of the DARPin 3G61 (601 $Å^2$ buried surface) and a slightly larger convex surface on GFP (682 $Å^2$ buried surface), resulting in a total buried surface area of 1283 $Å^2$. The interaction of 3G61 mainly involves residues of the second and third internal repeats of the DARPin. Four intermolecular hydrogen bonds connect 3G61 residues Asp78, Phe88 and Gln110 with GFP residues Lys166, Arg168 and Asn198. The residues contributing to complex formation through H-bonds and hydrophobic interactions are shown in Fig. 6b.

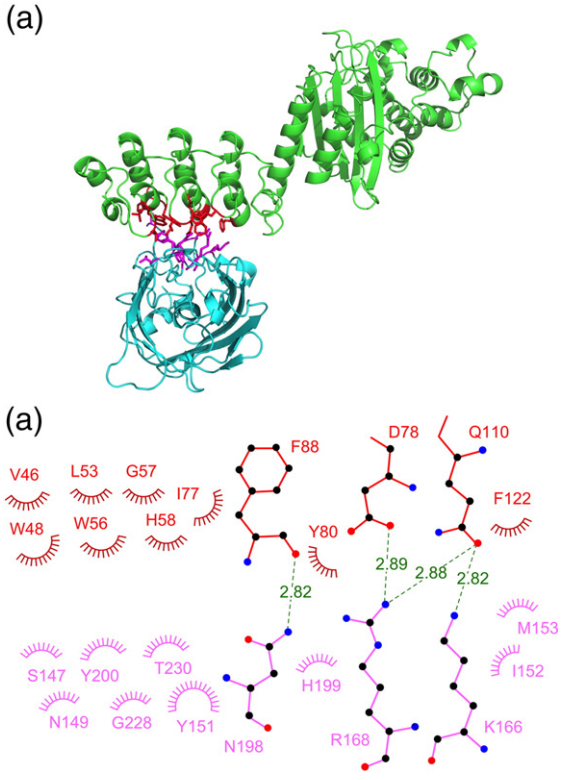


Fig. 6. Crystal structure of the DB15_v4_3G61 in complex with GFP. (a) The overall structure of DB15_v4_3G61 (green) in complex with GFP (cyan). The interacting residues are highlighted in stick mode in red (DB15_v4_3G61) and magenta (GFP). (b) DIMPLOT representation of the interaction between GFP (magenta) and DB15_v4_3G61 (red). H-bonds including the H-bond distances (in green) as well as residues and atoms involved in hydrophobic contacts (indicated by red or magenta rays) are shown.

Self-assembly interaction of DARPin D12

While the structure of DB04_v3_D12 shows two molecules in the asymmetric unit, the largest crystal contact is not formed between these two molecules, but between them and their symmetry mates. The largest crystal contact surface was formed between DB04_v3_D12 and its symmetry mate ($-x-1/2, y-1/2, -z-1$) *via* the randomized binding surface (paratope) of the DARPin D12 (856.3 $Å^2$) (Fig. 7). It should be noted that despite this self-interaction in the crystal, D12- and D12-based constructs are monomers in gel filtration (cf. Ref. [24]).

Comparison of design and experimental structures

The five experimental structures were aligned to the original models by Cα least-squares fits of the overall structures and of the DARPin and β-lactamase domains separately (Fig. 3; Supplementary Table ST6). The N-terminal RGSHis$_6$GS-tag, which is partially visible in some of the structures, was omitted in the comparisons. The two molecules in the asymmetric unit of DB04_v3_off7 aligned to each other with an rmsd of 1.11 Å (406 atoms) and had domain pseudo-torsion angles of 170.6° and 176.4°. The two molecules in DB04_v3_D12 aligned with an rmsd of 1.08 Å (406 atoms) and domain pseudo-torsion angles of 176.6° and 177.4°, compared to 176.5° in the DB04_v4 model. Cross-comparison of the two DB04 structures that crystallized under quite different conditions and in different space groups yielded rmsds between 1.22 Å (DB04_v3_D12 chain A to DB04_v3_off7 chain B) and 2.21 Å (DB04_v3_D12 chain B to DB04_v3_off7 chain A), compared to rmsds between 0.85 Å (DB04_v3_off7 chain B) and 1.43 Å (DB04_v3_off7 chain A) for the comparison to the DB04_v4 model. Thus, the deviation between model and experimental structure is within the range of variation between the different experimental structures (Fig. 3; Supplementary Table ST6). The structure of DB15_v4_3G61:GFP matched the generic DB15_v4 model with an rmsd of 0.801 Å and a domain pseudo-torsion angle of 166.0° compared to the 163.2° predicted by the model.

While a comparison of the two DB molecules in the asymmetric unit of DB08_v4_off7:MBP gives a low rmsd of 0.59 Å (410 atoms), the deviation of the two molecules in the asymmetric unit from the v4 model was 3.48 Å and 3.78 Å, respectively. These larger rmsds are due mainly to deviation of the domain pseudo-torsion angle, which was predicted to be 140.4°, but found to be 119.5° and 119.0° in the experimental structure. With an rmsd of 4.58 Å, DB12_v4_off7 deviated most strongly from its model, DB12_v4, due to a marked kink in the shared

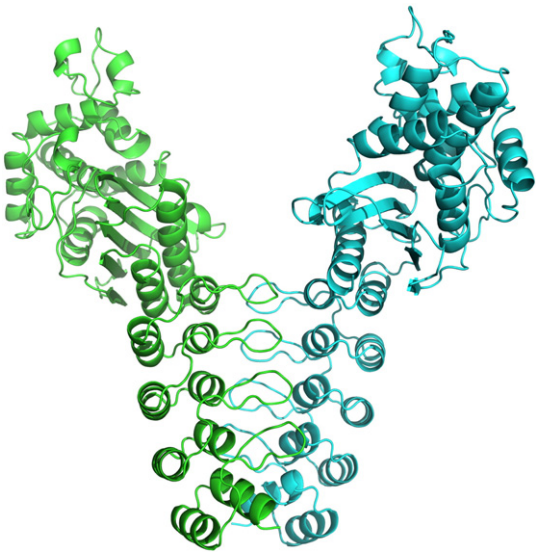


Fig. 7. DARPin D12 forms "self-assembled" crystal contacts in DB04_v3_D12 structure. Major crystal contacts formed through the randomized binding surface of the DARPin D12 domain. In this structure, chain A (green) forms major crystal contacts to the chain B (cyan) of the symmetry-related molecule ($-x - 1/2, y - 1/2, -z - 1$).

helix. At the same time, the internal structures of DARPins in these structures superimposed with rmsds of 0.30 to 0.49 Å, which is comparable to typical variations between DARPins. Moreover, the BL domains superimposed with rmsds of 0.43 to 0.45 Å, confirming the shared helix as the main source of variation. All comparisons are listed in Supplementary Table ST6.

Discussion

To connect two folded domains by a flexible linker into a single protein chain has become a standard technique in protein engineering. To rigidly connect two domains in a predetermined relative orientation is more difficult: naturally evolved rigid domain interactions seem to usually derive from well-packed binding interfaces that are hard to design from scratch and, more importantly, not easily adjustable to a variety of angles of the domains to be connected. In an alternative approach, we relied on a shared secondary structure element to provide a rigid link between two folded proteins. The DARPin structure starts and ends with an α-helix—extending one of these helices into the terminal helix of a partner domain could result in a rigid fusion of the two domains. End-to-end fusions of helices frequently kink or break at the point where the shared helix is exposed to the solvent [9]. Therefore, the shared helix would have to be embedded in at least one of the two domains along its entire length to achieve a rigid fusion.

The PDB[‡] was screened for high-resolution structures of monomeric proteins with a length of >150 aa that contain a (near) N- and/or C-terminal α-helix. Non-redundant hits were analyzed for their suitability as fusion partners by superimposing either the C-terminal helix of the fusion partners and the N-terminal helix of the DARPin or the C-terminal helix of the DARPin and the N-terminal helix of the fusion partner onto an ideal α-helix in different registers. Lengths of the shared helix were considered that spanned between the length of the shorter helix minus 3 aa to the combined length of the two helices plus 3 aa. For each shift of 1 aa in the overlap between the helices, the relative orientation of the two domains changes by a 100° rotation around the axis of the shared helix and a 1.5 Å translation along the axis. Models that lead to backbone clashes between the two proteins were eliminated, as were models that lead to clashes between the fusion partner and a large antigen bound to the DARPin (in this case, MBP). Due to the positioning of the terminal helices relative to the randomized loops of the paratope, fusions to the N-terminal helix of the DARPin are more likely to obstruct its binding site than fusions to the C-terminal helix, even though some viable fusion partners for N-terminal fusions were found (data not shown).

Based on this analysis, we decided on TEM-1 β-lactamase (BL) as a fusion partner for a first test of our strategy. Although wild-type BL contains a disulfide bond and is exported to the periplasm, the protein can also be produced with high yield by cytoplasmic expression in *E. coli*. Two cysteines, normally involved in a disulfide bond that might form only partially in the cytoplasm, were removed to avoid structural heterogeneity. BL contains both an N-terminal and a C-terminal helix, but only fusions to its N-terminal helix were deemed feasible, since the C-terminal amino acid is a tryptophan whose side chain is embedded in the core of the domain. Forcing that residue into a continuous helix would perturb its positioning. For a fusion of the C-terminal helix of a DARPin to the N-terminal helix of BL, six different alignments gave rise to non-clashing conformations. The shortest construct, DB01, has a shared helix length of 9 amino acids, two less than the C-terminal helix of a DARPin (11 aa), and six less than the length of the N-terminal helix of BL (15 aa); a deletion thus predicted to lead to some destabilization of BL. The shared helix of the longest construct tested, DB18, spans 26 aa, corresponding to an end-to-end fusion of the two proteins, restoring α-helical conformation to three DARPin and two β-lactamase terminal residues unresolved in the crystal structure or deviating from α-helical conformation in the experimental structures. The four intermediate constructs contain shared helices that are well embedded along their entire length.

In repacking the overlap between the two fused domains, we at first did not allow any glycine

residues in the shared helix, as these might allow the shared helix to kink. We expected that slightly too close contacts between the elongated N-terminal and the C-terminal helix of BL could be alleviated by minor shifts in the helix positions [1]. However, folding and/or stability of the β-lactamase proved very sensitive to perturbations of its two terminal helices, and none of the Rosetta [22]-repacked constructs yielded properly folded protein upon cytoplasmic expression in *E. coli*. However, replacing the wild-type BL by a previously stabilized mutant (BL "variant 23" [23]) rescued some of the fusion constructs. This indicated that destabilization of the BL portion of the fusion construct was responsible for the folding problems. Once the shared helix constructs were redesigned allowing glycine residues in both helices, in addition to using the stabilized BL, all six constructs were produced as soluble proteins with good yields (20–30 mg purified protein per liter *E. coli* culture).

The shortest (DB01) and the longest (DB18) construct failed to crystallize. In DB01, the very short shared helix represents a 6-aa N-terminal truncation of the lactamase and is not well embedded in the BL and may dissociate to form a flexible linkage. DB18, on the other hand, represents an end-to-end fusion of DARPin and BL, with a glycine residue at the point where the two fused proteins meet, further destabilizing the shared helix at its weakest point and allowing the helix to break. Of the five structures that were obtained, three (DB04_v3_off7, DB04_v3_D12 and DB15_v4_3G61:GFP) conform remarkably well with the design; while in the two others (DB08_v4_off7:MBP and DB12_v4_off7), deviations from the ideal geometry of the shared helix allow the molecules to adapt to a crystal packing that would not be possible for the original design. In the successful v4 designs, computational repacking of the domain interface was restricted to residues with high repulsive scores, leaving the sequence as close as possible to the original sequences of the two proteins. With more extensive repacking of the interface, further stabilization and rigidification of the interface might be possible—whether this would improve the performance of the construct as crystallization chaperones remains to be tested, as there might actually be an advantage to a crystallization chaperone that is able to slightly adapt its conformation to form optimal crystal contacts.

In most structures of the fusions proteins, the major crystal contacts involve the β-lactamase. In contrast, DARPin D12 makes a strong contribution (DB04_v3_D12) through the interaction of its randomized surface, as it was used in the absence of its target. Of course, the target protein may also contribute significantly to crystal contacts in the crystallization of the complex. Nonetheless, a crystallization chaperone is usually employed because the target protein or its complex with a cognate DARPin failed to crystallize. In these cases, rigid fusions of the DARPins to crystallization-competent partners such as the TEM-1 β-lactamase may greatly enhance the chances of successful crystallization.

Conclusions

To extend the utility of DARPins as crystallization chaperones, we have rigidly fused them to another rigid monomeric protein to increase their size and ultimately the surface area to form crucial crystal contacts. A shared helix was established as a general connector, allowing both angle and distance between the domains to be varied. This concept was tested successfully with a stabilized, disulfide-free version of β-lactamase as a fusion partner, leading to well-expressing and well-behaved fusion proteins. Since the shared helix is within the (constant) capping repeat of the DARPin, these designs can be applied to any DARPin of any specificity. This generic applicability extends the search space for crystal formation by the geometric dimension of the crystallization chaperone. A strategy to sequentially use β-lactamase and the DARPins as search models in molecular replacement was also established. While not every complex could be crystallized, this strategy expands the toolbox for structure determination.

Materials and methods

DARPin–β-lactamase fusion design

Potential fusion partners downloaded from the PDB[‡] were screened with Jmol[¶] for terminal helices suitable for fusions to a DARPin. Atomic models of potential fusion constructs were constructed as detailed in the result section using the Homology module of InsightII (Accelrys, San Diego). The Discover module of InsightII was used to minimize the splice points between different templates. To optimize the sequences of the shared helix and contacting residues, Rosetta [22] modules *fixbb*, *score* and *relax* were used. PyMOL[a] was used for model analysis and to generate figures.

Protein expression

Expression of DARPin–β-lactamase fusion proteins was performed with the plasmid pQE30ss (vector pQE30 with double stop codon) in the *E. coli* strain XL1-blue. The fusion proteins contain an N-terminal MRGS-His_6 tag. An overnight culture of each variant was prepared before large-scale expression as follows: 50 ml 2xYT medium (100 μg/l ampicillin, 1% wt/vol glucose) was inoculated with a single colony cultured on LB plates (100 μg/l ampicillin, 1% wt/vol glucose, 1.5% agar). The overnight culture (225 rpm, 37 °C) typically reached an OD_{600} ~ 4.5. The seed cultures were then diluted to a final OD_{600} = 0.1 into 1 l of 2xYT medium

(containing 100 mg/l ampicillin, 1% wt/vol glucose) in a 5-l baffled flask. Expression cultures were placed in a shaker (25 mm radius, 108 rpm, 37 °C) until induction with IPTG (0.5 mM final concentration) at OD_{600} ~ 0.6. Induced cultures were then allowed to shake for 18 h at 30 °C. Overnight expression cultures were harvested by centrifugation (5000 *g*, 4 °C, 15 min). Pellets were frozen in liquid nitrogen and stored at −80 °C.

Protein purification

Frozen cell pellets were thawed at room temperature and re-suspended in 30 ml TBS_{400} buffer [50 mM Tris–HCl (pH 7.4), 400 mM NaCl] containing one EDTA-free protease inhibitor tablet (Roche), DNAse (10 µg) and $MgCl_2$ (10 mM final concentration). The thawed cell pellets were homogenized with a Digitana YellowLine mixer and ruptured by one passage through a TS 1.1 Constant Cell Disrupter System at 30,000 psi. A subsequent sonication step (duty cycle 50%, intensity 5, 1–2 × 30 s pulses) was applied to the crude lysate for additional cell rupture and shearing of DNA. The lysate was clarified by centrifugation in SS34 tubes (28,000 *g*, 4 °C, 30 min) to give a final volume of ~30 ml. The supernatant was then filtered (0.22 µm Millex GP; Millipore) with a sterile syringe before adjusting the pH to ~8.0 with 0.5 M NaOH.

Immobilized metal-ion affinity chromatography on Ni^{2+}-nitrilotriacetic acid (Ni-NTA) resin was used for protein purification at room temperature. Four milliliters of 50% slurry of Ni-NTA Superflow (Qiagen) was applied to a 15-ml fitted chromatography column (Bio-Rad). Before application of the clarified supernatant, the column was equilibrated with 10 column volumes (CVs) of TBS-W buffer [50 mM Tris–HCl (pH 7.4), 400 mM NaCl, 20 mM imidazole, 10% glycerol]. After applying the clarified lysate, a washing step with TBS-W (10 CV) was performed. Protein samples were eluted with 3 CV of TBS-E [50 mM Tris–HCl (pH 7.4), 400 mM NaCl, 250 mM imidazole, 10% glycerol]. The protein was aliquoted to 1 ml aliquots in Eppendorf tubes, frozen in liquid nitrogen and stored at −80 °C.

Size exclusion chromatography was performed for the further purification for the complexes of DB fusion constructs with their targets, MBP or GFP. The DB fusion construct was mixed with its target at 1:1 molar ratio in SEC buffer [10 mM Hepes–Na (pH 7.4) and 150 mM NaCl], and then applied to a preparative Superdex 200 10/300 GL column (GE Healthcare) on an ÄKTAprime system (GE Healthcare). For every protein mixture, the peak fraction with the smallest molecular weight containing both DB fusion and its target in equimolar amounts, as determined by SDS-PAGE, was collected and used for crystallization. Analytical size exclusion chromatography was performed using a Superdex 200 PC 3.2/30 (GE Healthcare) column equilibrated with the SEC buffer on an ÄKTAmicro system (GE Healthcare), and at a protein concentration of 10 µM.

ELISA

To qualitatively assess whether DARPin–β-lactamase fusion constructs retain binding to their target, eight DB fusion constructs from design versions 3 and 4, expressed in soluble form and purified, were tested for binding to their target by ELISA. In this setup, 100 µl of 100 nM biotinylated target protein (MBP) was immobilized on a Maxisorp plate pre-coated with 66 nM Neutravidin and the plate was incubated with 100 µl of 100 nM DB fusion constructs, carrying an N-terminal $RGSHis_6$ tag. Binding was detected with anti-RGS-His-antibody-HRP conjugate (QIAGEN) by measuring the absorbance at OD_{450} with a microplate reader (TECAN Infinite M1000).

Crystallization

Eight commercially available grid screens of the in-house Protein Crystallization Center[§] were used for initial crystallization. The protein was thawed on ice and the buffer was exchanged to HBS_{150} (10 mM Hepes–Na, pH 7.4, 150 mM NaCl) on PD-10 desalting columns (GE Healthcare) according to the manufacturer's instructions. The protein was concentrated to 14–25 mg/ml using an Amicon Ultra Centrifugal Filter Device (Millipore, USA) with a molecular mass cutoff of 10,000 Da. CrystalQuick crystallization plates (Greiner Bio-One) were used for the sitting drop method. The protein was mixed with the reservoir solution in a volume ratio of 1:1, 1:2 and 2:1 for each single condition. Crystal growth took place at 20 °C.

Manual crystallization setups were performed with sitting-drop and hanging-drop crystallization plates from Hampton Research at 20 °C. Reservoir solution of 500 µl was applied and the drop was mixed with the reservoir solution in a volume ratio of 1:1, 1:2 and 2:1 for each single condition.

Data collection and processing

Data were collected from single, cryo-cooled crystals at beamlines PX and PXIII (Swiss Light Source, Villigen, Switzerland) with PILATUS 6M, PILATUS 2M or MARCCD high-resolution diffractometers. Data were processed and scaled with XDS [33].

To find the precise position of the domains, molecular replacement was performed with the stabilized β-lactamase (PDB ID: 3DTM), DARPin off7 (extracted from PDB ID: 1SVX), MBP (extracted from PDB ID: 1SVX) and GFP (PDB ID: 1GFL) as separate search models. The β-lactamase search model was prepared by removing all solvent molecules, as well as the first nine residues corresponding to the N-terminal helix (HPETLVKVK). DARPin off7 was extracted from the PDB file 1SVX, the solvent molecules and the last 11 residues corresponding to the C-terminal helix (NEDLAEILQKL) were removed. The residues were trimmed from the model since the conformation of the shared helix might deviate from the terminal helices in the isolated domain search models. In building the residues of the shared helix interface, they were first added as alanines and then mutated to the actual residues.

Analysis of the structures

All alignments of experimental structures to designed models and the calculation of rmsd were done by PyMOL[a]. The crystal-packing interfaces and intermolecular interactions including H-bonds and salt bridges were analyzed with the PDBePISA server [34]. The calculations of H-bonds and hydrophobic interactions of the DB_15v4_3G61:GFP complex were done by using DIMPLOT [35].

Accession numbers

The atomic coordinates of the described DARPin–β-lactamase fusions and complexes with their cognate targets have been deposited in the PDB[‡] (PDB ID: 5AQ7, 5AQ8, 5AQ9, 5AQA and 5AQB).

Acknowledgments

We thank Beat Blattmann and Céline Stutz-Ducommun for their help in protein crystallization and the staff of beamline PX and PXIII at the Swiss Light Source for support during data collection. We are grateful to Dr. Peer Mittl, Dr. Jing Dong and the members of the Plückthun laboratory for valuable discussions. Yufan Wu was supported by a predoctoral fellowship of the Forschungskredit from the University of Zurich. This work was supported by a Schweizerischer Nationalfonds grant (31003A_146278 to AP).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2016.03.002.

Received 19 January 2016;
Received in revised form 28 February 2016;
Accepted 2 March 2016
Available online 11 March 2016

Keywords:
designed ankyrin repeat proteins;
rigid domain fusions;
X-ray crystallography;
protein design;
protein engineering

Present address: M.M. Heberling, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Nijenborg 4, 9747 AG Groningen, the Netherlands.

† A. Batyuk and Y. Wu contributed equally to this work.

‡ http://www.rcsb.org
§ www.bioc.uzh.ch/research/core-facilities/protein-crystallization-center/experimental-setup/initial-xtal-screens/
|| www.ebi.ac.uk/msd-srv/prot_int/
¶ http://www.jmol.org
a www.pymol.org

Abbreviations used:
BL, *E. coli* TEM-1 β-lactamase; DARPins, designed ankyrin repeat proteins; DBxx, DARPin–β-lactamase fusion construct xx; GFP, green fluorescent protein; LLG, log-likelihood gain; MBP, maltose-binding protein; PDB, Protein Data Bank; TFZ, translation function *Z*-score.

References

[1] W.A. Baase, L. Liu, D.E. Tronrud, B.W. Matthews, Lessons from the lysozyme of phage T4, Protein Sci. 19 (2010) 631–641.
[2] A. Honegger, S. Spinelli, C. Cambillau, A. Plückthun, A mutation designed to alter crystal packing permits structural analysis of a tight-binding fluorescein–scFv complex, Protein Sci. 14 (2005) 2537–2549.
[3] G.E. Dale, C. Oefner, A. D'Arcy, The protein as a variable in protein crystallization, J. Struct. Biol. 142 (2003) 88–97.
[4] A. Ruggiero, G. Smaldone, F. Squeglia, R. Berisio, Enhanced crystallizability by protein engineering approaches: a general overview, Protein Pept. Lett. 19 (2012) 732–742.
[5] K.L. Longenecker, S.M. Garrard, P.J. Sheffield, Z.S. Derewenda, Protein crystallization by rational mutagenesis of surface residues: Lys to Ala mutations promote crystallization of RhoGDI, Acta Crystallogr. D Biol. Crystallogr. 57 (2001) 679–688.
[6] D.R. Cooper, T. Boczek, K. Grelewska, M. Pinkowska, M. Sikorska, M. Zawadzki, et al., Protein crystallization by surface entropy reduction: optimization of the SER strategy, Acta Crystallogr. D Biol. Crystallogr. 63 (2007) 636–645.
[7] M.R. Bell, M.J. Engleka, A. Malik, J.E. Strickler, To fuse or not to fuse: what is your purpose? Protein Sci. 22 (2013) 1466–1477.
[8] D.R. Smyth, M.K. Mrozkiewicz, W.J. McGrath, P. Listwan, B. Kobe, Crystal structures of fusion proteins with large-affinity tags, Protein Sci. 12 (2003) 1313–1322.
[9] M.C. Clifton, D.M. Dranow, A. Leed, B. Fulroth, J.W. Fairman, J. Abendroth, et al., A maltose-binding protein fusion construct yields a robust crystallography platform for MCL1, PLoS ONE 10 (2015), e0125010.
[10] V. Cherezov, D.M. Rosenbaum, M.A. Hanson, S.G. Rasmussen, F.S. Thian, T.S. Kobilka, et al., High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor, Science 318 (2007) 1258–1265.
[11] D.M. Rosenbaum, V. Cherezov, M.A. Hanson, S.G. Rasmussen, F.S. Thian, T.S. Kobilka, et al., GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function, Science 318 (2007) 1266–1273.
[12] G. Fenalti, P.M. Giguere, V. Katritch, X.P. Huang, A.A. Thompson, V. Cherezov, et al., Molecular control of delta-opioid receptor signalling, Nature 506 (2014) 191–196.
[13] G. Sennhauser, M.G. Grütter, Chaperone-assisted crystallography with DARPins, Structure 16 (2008) 1443–1453.
[14] S. Koide, Engineering of recombinant crystallization chaperones, Curr. Opin. Struct. Biol. 19 (2009) 449–457.

[15] L.C. Kovari, C. Momany, M.G. Rossmann, The use of antibody fragments for crystallization and structure determinations, Structure 3 (1995) 1291–1293.
[16] C. Hunte, H. Michel, Crystallisation of membrane proteins mediated by antibody fragments, Curr. Opin. Struct. Biol. 12 (2002) 503–508.
[17] M.A. Bukowska, M.G. Grütter, New concepts and aids to facilitate crystallization, Curr. Opin. Struct. Biol. 23 (2013) 409–416.
[18] H.K. Binz, M.T. Stumpp, P. Forrer, P. Amstutz, A. Plückthun, Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins, J. Mol. Biol. 332 (2003) 489–503.
[19] A. Plückthun, Designed ankyrin repeat proteins (DARPins): binding proteins for research, diagnostics, and therapy, Annu. Rev. Pharmacol. Toxicol. 55 (2015) 489–511.
[20] G. Minasov, X. Wang, B.K. Shoichet, An ultrahigh resolution structure of TEM-1 beta-lactamase suggests a role for Glu166 as the general base in acylation, J. Am. Chem. Soc. 124 (2002) 5333–5340.
[21] H.K. Binz, P. Amstutz, A. Kohl, M.T. Stumpp, C. Briand, P. Forrer, et al., High-affinity binders selected from Designed Ankyrin Repeat Protein libraries, Nat. Biotechnol. 22 (2004) 575–582.
[22] A. Leaver-Fay, M. Tyka, S.M. Lewis, Lange OF, J. Thompson, R. Jacak, et al., ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules, Methods Enzymol. 487 (2011) 545–574.
[23] I. Kather, R.P. Jakob, H. Dobbek, F.X. Schmid, Increased folding stability of TEM-1 beta-lactamase by *in vitro* selection, J. Mol. Biol. 383 (2008) 238–251.
[24] A. Mann, N. Friedrich, A. Krarup, J. Weber, E. Stiegeler, B. Dreier, et al., Conformation-dependent recognition of HIV gp120 by designed ankyrin repeat proteins provides access to novel HIV entry inhibitors, J. Virol. 87 (2013) 5868–5881.
[25] M.A. Kramer, S.K. Wetzel, A. Plückthun, P.R. Mittl, M.G. Grütter, Structural determinants for improved stability of designed ankyrin repeat proteins with a redesigned C-capping module, J. Mol. Biol. 404 (2010) 381–391.
[26] G. Interlandi, S.K. Wetzel, G. Settanni, A. Plückthun, A. Caflisch, Characterization and further stabilization of designed ankyrin repeat proteins by combining molecular dynamics simulations and experiments, J. Mol. Biol. 375 (2008) 837–854.
[27] M. Brauchle, S. Hansen, E. Caussinus, A. Lenard, A. Ochoa-Espinosa, O. Scholz, et al., Protein interference applications in cellular and developmental biology using DARPins that recognize GFP and mCherry, Biol. Open. 3 (2014) 1252–1261.
[28] T. Huber, D. Steiner, D. Röthlisberger, A. Plückthun, *In vitro* selection and characterization of DARPins and Fab fragments for the co-crystallization of membrane proteins: the Na(+)-citrate symporter CitS as an example, J. Struct. Biol. 159 (2007) 206–221.
[29] S.K. Wetzel, C. Ewald, G. Settanni, S. Jurt, A. Plückthun, O. Zerbe, Residue-resolved stability of full-consensus ankyrin repeat proteins probed by NMR, J. Mol. Biol. 402 (2010) 241–258.
[30] A.J. McCoy, R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, R.J. Read, Phaser crystallographic software, J. Appl. Crystallogr. 40 (2007) 658–674.
[31] P.D. Adams, P.V. Afonine, G. Bunkoczi, V.B. Chen, I.W. Davis, N. Echols, et al., PHENIX: a comprehensive python-based system for macromolecular structure solution, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 213–221.
[32] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 486–501.
[33] W. Kabsch, Xds, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 125–132.
[34] E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state, J. Mol. Biol. 372 (2007) 774–797.
[35] A.C. Wallace, R.A. Laskowski, J.M. Thornton, LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions, Protein Eng. 8 (1995) 127–134.