Synthetic Biology Cite This: ACS Synth. Biol. 2018, 7, 2226–2235

An Interface-Driven Design Strategy Yields a Novel, Corrugated **Protein Architecture**

Mohammad ElGamacy,[†] Murray Coles,[†][©] Patrick Ernst,[‡] Hongbo Zhu,[†] Marcus D. Hartmann,[†][©] Andreas Plückthun,[‡] and Andrei N. Lupas*^{,†}

[†]Department of Protein Evolution, Max-Planck-Institute for Developmental Biology, 72076 Tübingen, Germany ^{*}Department of Biochemistry, University of Zurich, 8057 Zurich, Switzerland

S Supporting Information

ABSTRACT: Designing proteins with novel folds remains a major challenge, as the biophysical properties of the target fold are not known a priori and no sequence profile exists to describe its features. Therefore, most computational design efforts so far have been directed toward creating proteins that recapitulate existing folds. Here we present a strategy centered upon the design of novel intramolecular interfaces that enables the construction of a target fold from a set of starting fragments. This strategy effectively reduces the amount of computational sampling necessary to achieve an optimal sequence, without compromising the level of topological control. The solenoid architecture has been a target of



extensive protein design efforts, as it provides a highly modular platform of low topological complexity. However, none of the previous efforts have attempted to depart from the natural form, which is characterized by a uniformly handed superhelical architecture. Here we aimed to design a more complex platform, abolishing the superhelicity by introducing internally alternating handedness, resulting in a novel, corrugated architecture. We employed our interface-driven strategy, designing three proteins and confirming the design by solving the structure of two examples.

KEYWORDS: alternating handedness, novel fold, protein design, protein structure, repeat protein

omputational design has thus far been very successful in diversifying the geometries and sequences of existing folds. This has been largely assisted by the presence of one or more starting structures for redesigning a particular fold, and the associated data that underpin its sequence determinants. In contrast, designing novel folds with a predetermined backbone blueprint, while offering a vast new range of designable folds, renders the gross sequence and rotamer sampling problem intractable. A fragment-based approach can greatly reduce this search space, as starting building blocks already carry intrinsic folding information. In addition to maintaining control over the level of adherence to a target fold, this approach also offers the possibility of coarse-graining the assembly problem by choice of the building block sizes. The latter may range from secondary structural elements to large subdomain or domainsized fragments.¹ This effectively decomposes the problem into searching for optimal interfragment interfaces and loops. This promises to focus the available computing resources on accurately and exhaustively exploring restricted spaces, instead of sparsely exploring much larger ones. Here we demonstrate the capacity of this interface-driven approach as an efficient means for novel fold design.

For many years repeat proteins—in particular solenoids have been a central topic of protein design. Unlike globular proteins, their low contact order and compositional uniformity have made them excellent platforms for investigating sequencestructure relationships and dissecting the energetics of protein folding.² They have also served a wide range of applications as antibody-like tailored synthetic binding proteins selected from libraries, and some have even progressed to late-stage clinical trials.³ Because of their favorable biophysical properties, they have also been developed into crystallization chaperones. Initially, design efforts on solenoids were aimed at generating more robust variants through sequence idealization.⁵ More recently, the vast potential of solenoid proteins as tunable scaffolds has motivated computational design aimed at expanding the available repertoire of solenoid configurations with atomic accuracy.⁶ These controlled geometries have included previously unobserved forms. However, despite this considerable success, to date the general solenoid architecture has not been altered. Here we aim to move beyond the solenoid, exploiting an incremental increase in the topological complexity to create a corrugated arrangement so far not observed in nature.

Solenoid proteins are characterized by a uniform connectivity between repeat units and thus wind into a continuous

Received: May 28, 2018 Published: August 27, 2018

ACS Synthetic Biology



Figure 1. Solenoid repeat, corrugated repeat and design strategy. (A) Natural solenoids have repeats with a single junction type (blue circle) and a uniform handedness; they can thus be described by a sawtooth wave. The wave represents the torsion angle along the superhelical axis between the first and the *n*th residue. (B) A corrugated fold, represented by a triangle wave, would entail bihanded repeats, and thus require two junction types; Figure S9 shows actual values for idealized templates. (C) A two-stage strategy of interface design and loop construction; Geo: geometric filtering calculations, PMF: Potential of mean force energy calculations, RFD: Rotational force dissipation simulations. The target fold is built from a four-helix-bundle (orange) and its translational symmetry image (purple); top panel. The interface was then spanned by a grafted loop (purple); bottom panel.

superhelix.⁷ This implies a single interunit junction type (defined as an interface and a connecting loop), where the units are bound to possess the same handedness. A waveform description of this periodic fold can be made by plotting the change in dihedral angle around the superhelical axis versus sequence position, which takes the form of a sawtooth wave (Figure 1A and Figure S9). The next step in complexity involves the alternating use of two junction types with opposite handedness; under the waveform description this topology adopts the form of a triangle wave (Figure 1B and Figure S9). Such a bihanded topology can be obtained by doubling the size of the building block and introducing a new junction of the opposite handedness to that of the starting block. In contrast to solenoids, the alternating handedness eliminates supercoiling. Here we have taken this approach to construct the corrugated target architecture, using an interface-driven design strategy that builds upon existing, simpler structural blocks and minimizes the amount of sampling required to achieve a target fold.

RESULTS AND DISCUSSION

Design Strategy. To construct the target topology, two unique helical hairpins, two unique interfaces, and two unique loops are required. The use of an up-down four-helix-bundle as a starting point provides two hairpins, a single interface and a single loop. The design of a second interface with the translated image of the bundle and of a second connecting loop is then sufficient to complete the target fold (Figure 1C). For this purpose, we employ a two-stage strategy: The first aimed at designing an intramolecular interface between two arbitrarily posed building blocks. This arbitrary docking step is only constrained by N- to C-terminal distance between the two blocks, a distance that can be defined by the allowed loop length. The second stage is aimed at contructing a loop across this interface.

We began by compiling a set of four-helix-bundles from the Protein Data Bank (PDB) that satisfied a set of geometric criteria defining regularity, bundle height range, and internal hairpin similarity. Initial poses were built between the bundle backbones and their translation images. The relative orientations were made to minimize the twist and curvature at the connecting interface along the central axis. This step was followed by the main sampling routine, where a combination of sequence sampling, side chain rotamer sampling, backbone refinement, and rigid-body docking were performed with an initially softened steric repulsion term. For efficient sampling, different iterations of the main Monte Carlo sampling loops were interlaced with a geometric filter. The latter being aimed at eliminating solutions with poor residue packing quality, before further rounds of design are resumed. The sequence sampling was restricted to the interface positions, while conformational refinement was performed globally (Materials and Methods, Figure S1). With the goal of further filtering the generated decoys by estimating the interaction free energy of the designed interfaces, more expensive potential-of-meanforce calculations were conducted through variable-velocity, variable-force steered molecular dynamics (SMD) simulations. The interaction free energy between the building blocks was calculated from the convolution of the velocity and force functions, and was used to rank the candidates accepted for the



Figure 2. All three designs were folded. The first column shows the designed models as cartoon representation. The second column shows the respective CD spectra of the designs. The third column shows the melting curves of the designs, where BRIC1 and BRIC3 exhibit monophasic unfolding, while BRIC2 does not thermally unfold below 100 $^{\circ}$ C.

loop design stage. The described simulation setup applies a more adaptive pulling scheme of a previous constant-velocity setup that we have previously benchmarked against a subset of a protein–protein affinity data set⁸ (Materials and Methods). This accelerated form of free energy estimation method has been shown to be particularly suited for protomers that do not undergo major conformational changes upon unbinding,⁹ which was assumed here given the nature of our building blocks.

The next stage was to construct the loop that connects the newly designed interface. For this we searched the PDB for loop configurations that could serve as initial templates. The search routine scanned structures with a gapped sliding window, based on a generic description of the geometry defined by the ending and starting segments of adjacent repeat units (Figure S2). This description was obtained using the dihedrals profile, the axial vectors of the relevant segments and their orientations (Materials and Methods). The grafted loops were then subjected to combined sequence and conformer sampling, and all resulting loop compositions were evaluated using an accelerated molecular dynamics scheme. The routine applies a linear ramp of rotational force across the peptide bond at the center of the loop in a crankshaft fashion. This linear force titration, results in a nonlinear rotational response. the resulting nonlinear rotational kinetic energy is evaluated across the simulation time. The loop compositions that required the highest force magnitudes to induce rotational motion were selected for experimental evaluation (Materials and Methods, Figure S3).

Choice of Building Blocks. Three starting template bundles were adopted from three different natural proteins, to evaluate the generality of the approach and the choice of purely geometric criteria for template inclusion. The first design, BRIC1 (for Bihanded Repeat with Internal Corrugation), was constructed from a template bundle from the CheA histidine phosphotransfer domain (PDB: 115N);¹⁰ the second, BRIC2, from the DRNN four-helix-bundle, which had previously undergone a total computational redesign of its hydrophobic core (PDB: 2LCH);¹¹ and the third, BRIC3, from a focal adhesion targeting domain (PDB: 3B71).¹² While one 2LCH is a monomeric solution structure, 1I5N and 3B71 do not possess any crystallographic arrangement similar to that proposed in Figure 1C. In the design process, BRIC1 underwent 12 mutations on the N-terminal face of the designed interface and 13 on the C-terminal face. BRIC2 underwent 12 mutations on the N-terminal face and 12 mutations on the C-terminal face; in addition to 2 mutations in the core of each bundle. BRIC3 underwent 20 mutations on



Figure 3. Experimental structures of BRIC1 and BRIC2 confirm the design. (A) The crystal structure of the 3D domain-swapped dimer, with individual protomers (colored by sequence position from cyan to blue) superimposed on the design (yellow). (B) The low-resolution NMR model of the BRIC1 monomer superimposed on the design (colors as in panel A). (C) The crystal structure of BRIC2 (blue) superimposed on the design (orange).

the N-terminal face and 17 mutations on the C-terminal face. Structure-based sequence alignments of the designs to their respective starting templates are shown in Table S1. The interface for BRIC1 was bridged by a 5-residue-loop, for BRIC2 by a 7-residue-loop, and for BRIC3 by a 6-residue-loop containing a disulfide bridge. For each of the designs, we explored experimentally the minimal form consisting of two repeat units. For BRIC1, we retained the native C-terminal helix of the phosphotransfer domain as a C-terminal capping helix.

Biophysical Characterization. We expressed the proteins in Escherichia coli and purified them using immobilized metal ion affinity chromatography (Materials and Methods). All three were primarily monomeric by analytical size-exclusion chromatography, although they all showed pH-dependent oligomerization. Their well-dispersed 1D NMR spectra were consistent with folded proteins (Figures S6 and S7) and their circular dichroism (CD) spectra with predominantly helical secondary structure (Figure 2). In thermal unfolding experiments, BRIC1 and BRIC3 showed single-phase equilibrium unfolding at 86 and 67 °C, respectively, while BRIC2 did not exhibit any melting transition. The monophasic melting transition of BRIC1 corresponds to that of a single, compact domain. This emphasizes the success of our interface design, as it implies that the enthalpy of the designed junction matches that of the native one. The three constructs underwent crystallization screening and only BRIC1 readily yielded diffracting crystals, BRIC2 was fused to a crystallization chaperone, while BRIC3 did not express in sufficient yield in M9 minimal medium or in fusion with the crystallization chaperone

Crystal Structure of BRIC1. Crystallization screens yielded well-diffracting BRIC1 crystals, for which we obtained data to 2.5 Å resolution in space group C2. The crystals contained one BRIC1 monomer in the asymmetric unit, which was unambiguously located in a molecular replacement trial searching with the full design model, in the first attempt and with high contrast. However, after initial refinement, it became apparent that the connectivity between the two four-helical halves of the protein differed from the design. Clear electron

density showed the linker in an extended conformation, resulting from a domain-swapped dimeric assembly. In this assembly, two elongated BRIC1 protomers, related by the crystallographic 2-fold symmetry, are associated in an antiparallel fashion, such that the N-terminal four-helix bundle of one protomer interfaces with the C-terminal bundle of the other protomer, and vice versa (Figure 3 and Table S2). Given that BRIC1 also shows a minor dimeric form in solution (Figure S4), it appears that this form was selectively crystallized. As a result, the inter-repeat interface has entirely retained the designed interface features. This had a swapped backbone RMSD to the design of 1.82 Å (all-atom RMSD was 2.1 Å) across the entire structure, excluding the loop.

NMR Structure of BRIC1. To address the nature of the monomeric form of BRIC1, we prepared isotope labeled samples for solution NMR. Diffusion coefficients measured on freshly prepared samples were consistent with the designed monomer (Figure S5). However, dimeric and higher oligomeric forms accumulated over time, impacting on the quality of spectra. This feature, combined with the ambiguity intrinsic to repeat sequences, precluded full resonance assignment and thus high-resolution structure determination. We therefore adopted a strategy aimed at creating a lowresolution model, using a sample selectively ¹³C-labeled on methionine methyl groups to define interhelical contacts (Materials and Methods). An initial observation was the similarity of chemical shifts between the repeats, indicating that both adopt very similar structures. Interhelical contacts then defined intra- and inter-repeat junctions very similar to those observed in the crystal, with the C-terminal repeat identified by contacts to the unambiguously assigned C-terminal capping helix. The compiled data were sufficient to define the monomer structure, using the domain-swapped crystallographic protomer as a starting point (Figure 3B and Table S3). The calculated monomer ensemble agrees well with the design, with an average backbone RMSD of 1.8 Å (all-atom RMSD ranged from 2.5 to 2.9 Å, excluding the capping helix).

Crystal Structure of BRIC2. In contrast to BRIC1, BRIC2 did not yield well-diffracting crystals in the first attempt. For this reason, a rigid shared helix fusion to DARPin D12



Figure 4. Architectural uniqueness and interface accuracy of the BRIC designs. (A) A comparison between the idealized BRIC architecture and the closest architectures through structural similarity searches. The structures are colored by chain path from blue to yellow. (B) Polar and Cartesian disparity between the designs and experimental structures of BRIC1 and BRIC2. The top panel shows the angular deviation from design values for the tilt (θ), bend (β) and curvature (κ) across the designed interface in green, teal and cyan, respectively. Each dot represents either an NMR model or one of the asymmetric unit chains (the single chain of the BRIC1 crystal structure is represented by orange crosses). The bottom panel shows the CAPRI evaluation criteria (L_{rms} , I_{rms} and f_{nat}) for the designed interfaces (defined in the Materials and Methods). The red, green and blue dashed horizontal lines mark the high, medium and acceptable ranks, respectively. Error bars represent the standard error across asymmetric unit chains of NMR models (some error bars are within the dot diameter).

(designed ankyrin repeat protein D12) was constructed. DARPin D12 had been previously identified as wellcrystallizing under many different conditions and thus serving as a crystallization chaperone when rigidly fused to other repeat proteins.^{4a} An N-terminal fusion of the DARPin was built in silico and both the shared helix and residues within 5 Å proximity were sequence-optimized using Rosetta fixed backbone design, as previously described.^{4a} Crystals appeared after 25 days, diffracted to 3.0 Å resolution and the data were integrated in space group P1. For the molecular replacement, a model of the DARPin was used as a search model and the design models of BRIC2 were manually fitted into the density (refinement statistics are provided in Table S4). Out of the four molecules of the asymmetric unit, chain B and D looked as designed and a slight bend of the shared helix was observed for chains A and C, due to crystal forces (Figure S8). Clear electron density was visible for the whole BRIC2 domain and the designed loop, connecting the two repeats, could be built. In comparison to BRIC1, BRIC2 was monomeric and no domain swap was observed in all of the four chains, proving the successful interface design between the two helical bundles (Figure 3C). The overall backbone RMSD ranged from 2.27 to 3.0 Å (all-atom RMSD ranged from 2.8 to 3.4 Å), and

confirmed both the design and the potential of DARPin D12 as a crystallization chaperone.

Architectural Uniqueness and Interface Design Precision. To contrast the BRIC architecture to the nearest existing folds, we conducted structure searches against the entire PDB using PDBeFOLD¹³ and DALI,¹⁴ and the ECOD database¹⁵ using TM-align.¹⁶ No folds were found that structurally align along the full length of our designed structures. Any similarity detected was largely localized to a four-helix-bundle substructure. PDBeFOLD did not recover any significantly related hits, while the best TM-align hits had TM-scores <0.55 and did not share significant similarity with our BRICs. For the DALI searches we selected three structures based on their alignment lengths and secondary structures arrangements. Figure 4A shows the structures and idealized topologies of these hits contrasted against the BRIC topology. Two of these hits (3D19 and 4AKK) were topologically similar to each other, but with opposite chain paths. These were composed of two uniformly handed four-helix-bundles with Nand C-termini abutting each other at the connecting interface; a close-ended configuration that results from the parallel orientation of the helical hairpins to the main axis. The third hit (3AY5) consisted of two dissociated antiparallel helical domains, with one being a right-handed, side-connected

bundle, and the other a left-handed, diagonally connected bundle.

To evaluate the interface design precision, we measured the polar and Cartesian error across the interface (Figure 4B). The polar deviations of the interface between designs and experimental structures were defined in terms of tilt (θ) , bend (β) and curvature (κ). Polar deviations in BRIC1 were lower in the solution structure than in the crystal structure, with the highest deviation in the κ dimension. BRIC2, however, exhibited a larger range of deviations between the asymmetric unit protomers, particularly, along the β and κ dimensions. As these large interprotomer deviations would potentially build-up in a multirepeat scenario, we carried out molecular dynamics simulations for a five-bundle-repeat model of BRIC2. The deviations in the solvated simulations were of smaller magnitude, averaging below 8.0°, 6.6° and 3.0° for $|\theta|$, $|\beta|$ and $|\kappa|$, respectively, at the four designed interfaces. We therefore expect most of these deviations to stem from the crystal packing. For estimating the Cartesian precision at the interface, we calculated the evaluation criteria used in the CAPRI interaction prediction competition: $L_{\rm rms}$, $I_{\rm rms}$ and $f_{\rm nat}$.¹ The three structures ranked *medium* on the $L_{\rm rms}$ score. The BRIC1 solution structure ranked medium on the I_{rms} score, while the two crystal structures ranked acceptable. All three structures ranked high on the f_{nat} score. In spite of the asymmetric nature of the two-sided interface design, the intramolecular four-helix-bundle backbone RMSD was minor; 0.8 Å within design and 1.1 Å within structure for BRIC1, and 0.9 Å within design and 1.3 Å within structure for BRIC2. The design vs structure values were 0.6 Å for both respective bundles of BRIC1, and 1.2 and 1.3 Å for the first and second bundles of BRIC2, which affirms the rigid incorporation of the building blocks.

CONCLUSIONS

At the frontier of protein design is the aim to provide new scaffolds for functionalization, this potential has made repeat architectures attractive design targets. Internal cross-alignments in our experimental structures show that minimal structural perturbation has been introduced to the starting building blocks, leading to the possibility of constructing longer repeats. With this architecture, the large-sized building blocks can harbor functional features from selected parent blocks, or afford more extensive engineering owing to their expanded substructural diversity, as compared to repeats with smaller building blocks.

Protein design efforts have so far been biased toward assembling idealized secondary structure elements. As such they do not reflect natural proteins, where structural deformities are common and often associated with functional motifs. The difficulty of sampling such deformities is an inherent barrier to the designability of these motifs.¹⁸ A successful design strategy that sidesteps this problem, and even creates new topologies, has been to combine natural substructures by directly using structurally similar fragments as overlapping connectors.¹⁹ This, however, does not allow strict control of target topologies, since it is contingent upon the existence of overlaps that yield viable intramolecular interfaces. In contrast to this connectivity-driven approach, we introduce an interface-driven approach that is capable of delivering novel topologies from an arbitrary arrangement of building blocks. Moreover, this strategy employs sequence and conformational sampling focused only on the junctions

between building blocks, and separates interface optimization from loop design, thus adding to the overall efficiency.

MATERIALS AND METHODS

Computational Design. The interface sequence design stage was performed in multiple consecutive rounds filtering the top 10-20 candidates from each round and feeding them as input to the next. Each round was performed using a RosettaScripts²⁰ protocol comprising two generic Monte Carlo loops separated by packstat²¹ and total energy (talaris2013 scoring function²²) filters. Each loop executed a protocol comprising soft-repulsion sequence sampling, backbone optimization,²³ docking and conformational refinement. Between the consecutive rounds, under- or overpacking was evaluated by calculating the average deviation from highresolution structures packing density probability. The last round's output was filtered through an accelerated SMD routine that aims at approximately assessing the potential of mean force of unbinding across the designed interface. The free energy of unbinding (W) was evaluated as $W_{t_a \to t_c} = \int_{t_a}^{t_c} \mathbf{v}(t)$ $\mathbf{F}(t) dt$ where $\mathbf{F}(t)$ and $\mathbf{v}(t)$ are the pulling force and velocity vectors at time t, respectively. One partner was fixed and aligned against a reference orientation while the other was pulled along a single dimension through a loose spring to achieve a variable-velocity, variable-force SMD setup that yields the free energy profile along the unbinding path. The protein was modeled using the CHARMM36 force field,²⁴ where the simulations were performed in explicit solvent (TIP3P water model) and 0.15 M sodium chloride as NPT ensembles at 310 K and 1 atm using a Langevin thermostat and a Langevin barostat as implemented in the NAMD engine.²⁵ Particle Mesh Ewald electrostatics grid of 1 Å resolution was used with a long-range cutoff set at 12 Å (switching at 10 Å) and a time step of 2 fs. The reference pulling velocity (\mathbf{v}_{ref}) was calibrated to 2.5 Å/ns with a spring constant (k) of 20 kcalmol⁻¹·Å⁻² where the applied force (**F**(*t*)) was computed as $-\nabla \left| \frac{1}{2} k [t \mathbf{v}_{ref} - (\mathbf{r}_t - \mathbf{r}_o) \cdot \mathbf{n}]^2 \right| (\mathbf{r}_t \text{ being the position vector of})$ the steered atom group and \boldsymbol{n} being the pulling direction vector). The systems underwent 2000 steps of conjugate gradient minimization before random atom velocities initialization and force application on the backbone carbonyl carbon atoms. The calculated work was used to rank designs for the next stage.

The loop design stage begins with a structural search using a gapped sliding window across the whole PDB, where the landing sites are defined by two N-to-C vectors and a single (φ, ψ) array. Given the latter representation, every subject landing site was compared to the query geometry by means of dihedral profiles similarity, landing sites lengths similarity and landing sites relative orientation similarity. Loop lengths of 4 and up to 8 were searched for, with landing sites of lengths ranging from 4 to 8 residues. The best matches according to the previous metrics were then grafted onto the top ranking interface designs and subjected to loop mutagenesis using a Rosetta script that performs sequence sampling, backrub refinement, and side chain refinement in a Monte Carlo looper. The designed loops were evaluated by applying reciprocating crankshaft force across the peptide bond at the center of the loop with a reciprocation frequency of 20 fs⁻¹. A 60 ps span of equilibration was followed by equal torques applied to the peptide bond hydrogen and oxygen atoms around the peptide bond axis, starting by an angular

acceleration of 2 rad·ps⁻². The latter rotational acceleration was incrementally ramped up every 40 fs by a value of 2 rad·ps⁻² using the updated atomic positions every 20 fs so as not to apply any forces against the peptide axis itself. The simulation was performed in triplicates in durations of 300 ps with similar system setup parameters to the SMD described above. The distributions of the loop atoms root-mean-squared-fluctuation and rotational kinetic energy were assessed to choose the designs of the lowest mean and standard deviation of these variables. The top designs at this point were directly taken to the laboratory.

Expression and Purification. The genes were acquired from Synbio Technologies, already cloned into pET-28a(+) using NcoI and NdeI cloning sites and in-frame with an Nterminal hexaHis-tag and a thrombin cleavage site, while harboring a kanamycin resistance gene as a selection marker. The plasmids were used to transform chemically competent E. coli BL21(DE3) by means of heat-shock. The expression procedure entailed growing of the cells in LB medium and inducing with IPTG at OD600 of 0.5-1 with overnight expression at 25 °C. For expression of labeled protein, a preculture in LB medium was grown, cells collected, and resuspended in M9 minimal medium (240 mM Na₂HPO₄, 110 mM KH₂PO₄, 43 mM NaCl), supplemented with 10 μ M FeSO₄, 0.4 µM H₃BO₃, 10 nM CuSO₄, 10 nM ZnSO₄, 80 nM MnCl₂, 30 nM CoCl₂ and 38 μ M kanamycin sulfate, to an OD_{600} of 0.5–1. After 40 min of incubation at 25 °C, 2.0 g ¹⁵N-labeled ammonium chloride (Sigma-Aldrich cat. no. 299251) and 6.25 g ¹³C D-glucose (Cambridge Isotope Laboratories, Inc. cat. no. CLM-1396) were added, or 100 mg methyl-¹³C L-methionine (Sigma-Aldrich cat. no. 299146) in case of selective labeling in a 2.5 L culture. After another 40 min IPTG was added to 1 mM final concentration for overnight expression. Cells were collected by centrifugation at 5000g for 15 min, lysed by a Branson Sonifier S-250 (Fisher Scientific) in hypotonic 50 mM Tris-HCl buffer supplemented with one tablet of the complete protease cocktail (Sigma-Aldrich cat. no. 4693159001) and 3 mg of lyophilized DNase I (5200 U/mg; Applichem cat. no. A3778). The insoluble fraction was pelleted by 25 000g centrifugation for 50 min, and the soluble fraction was filtered (0.45 μ m filter pore size) and directly applied to a Ni-NTA column. A 5 mL HisTrapFF immobilized nickel column (GE Healthcare Life Sciences cat. no. 17-5255-01) was used for this purpose, washed consecutively by 30 mL 150 mM NaCl, 30 mM Tris buffer (pH 8.5) at 0, 30, and 60 mM imidazole. Fractions were collected by a gradient elution at >60 mM imidazole. The eluate was concentrated using 10 kDa MWCO centrifugal filters (Merck Millipore cat. no. UFC901024) and loaded onto an equilibrated Superdex 75 gel filtration column (GE Healthcare Life Sciences cat. no. 17517401). The gel filtration buffer used was always 100 mM sodium phosphate buffer (for NMR and CD transparency) composed to the target pH, where BRIC1 was eluted in pH 8.5, while BRIC2 and BRIC3 at pH 5.5. An AktaFPLC system (GE Healthcare Life Sciences) was used for all chromatography runs.

For the D12-BRIC2 chimera, the shared helix between D12 and BRIC2 was introduced by assembly PCR and the resulting fragment was cloned into a pQE30LIC_3C (Qiagen) based plasmid *via* BamHI and HindIII restriction sites. Chemocompetent BL21DE3 cells were transformed with the plasmid and the protein was expressed in autoinduction medium at 25 °C for 16 h.²⁶ Cells were resuspended in 50 mM Tris/HCl pH 8, 500 mM NaCl, 20 mM imidazole and lysed *via* sonication. Insoluble material was spun down by centrifugation for 30 min at 30 000g and the supernatant was loaded on 5 mL NiNTA resin, equilibrated with resuspension buffer. The column was washed with 25 mL resuspension buffer and protein was eluted with 15 mL resuspension buffer containing 250 mM imidazole. The elution fraction was dialyzed overnight against 50 mM Tris/HCl pH8, 300 mM NaCl and the N-terminal 10xHis-tag was removed by cleavage with 3C-protease (2% w/w). Following a second NiNTA step to remove the protease and the His-tag, the protein solution was concentrated to 5 mL and further purified by gel filtration on an S200 16/600 column (GE healthcare) equilibrated with 10 mM Tris/HCl pH 8, 100 mM NaCl.

Biophysical Characterization. The analytical gel filtration experiments were all done on a Superdex 200 10/300 GL (GE Healthcare Life Sciences cat. no. 17517501), and the collected fractions from the eluate were used for CD or NMR measurements directly after. ¹H NMR spectra were collected on a Bruker AVIII-800. NMR diffusion ordered spectroscopy experiments were performed on a Bruker AVIII-600 using the relevant functionality in the TopSpin software, running the analysis over multiple aliphatic proton peaks. The structurebased prediction of the diffusion coefficient was done using the HYDROpro software,²⁷ setting the corresponding temperature to 310 K and viscosity to 0.007 P. CD spectra were recorded on a Jasco J-810 spectrometer, with a spectral scan window of 200-240 nm, with a sweep delta of 0.1 nm while averaging over 5 scans. Melting curves were measured from 20 to 100 °C, recording the ellipticity at 222 nm every 0.5 °C, while heating at a 1 °C/min rate.

X-ray Crystallography. For BRIC1 crystallization, the protein was concentrated to 13 mg/mL in 25 mM Tris buffer, pH 8.5, 150 mM NaCl. The D12-BRIC2 fusion was concentrated to 40 mg/mL in 10 mM Tris buffer, pH 8, 100 mM NaCl. Sitting-drop vapor diffusion crystallization trials were performed in 96-well format, equilibrating drops containing 300 nL of protein solution and 300 nL of reservoir solution against 50 µL of reservoir solution. For D12-BRIC2, the drop size was 150 nL + 150 nL and the reservoir contained 75 μ L of mother liquor. Best diffracting crystals were obtained with a reservoir solution containing 20% v/v PEG 500 MME, 10% w/v PEG 20 000, 30 mM MgCl₂, 30 mM CaCl₂ and 100 mM Tris-BICINE pH 8.5, loop-mounted, and flash-frozen in liquid nitrogen. For D12-BRIC2, an initial hit was found in 0.2 M (NH₄)₂SO₄, 25% w/v PEG 3350 and 100 mM Bis-Tris pH 5.5. A fine screen with two perpendicular gradients of the PEG concentration and the pH was set up to yield diffracting crystals, which were flash-frozen in mother liquor containing 20% v/v ethylene glycol. Data were collected at beamline X10SA at the Swiss Light Source, at 100 K with an X-ray wavelength of 1 Å and a PILATUS 6M-F detector (Dectris) for BRIC1 or an EIGER 16 M X detector (Dectris) for D12-BRIC2. Data for BRIC1 were indexed, integrated and scaled to a resolution of 2.5 Å in space group C2, using XDS.²⁸ For D12-BRIC2, two crystals were indexed and integrated in space group P1. After merging the two data sets, the data were scaled to 3 Å. According to the unit cell dimensions, one BRIC1 monomer was expected in the asymmetric unit with a solvent content of 50%. Molecular replacement was carried out using MOLREP,²⁹ using the designed coordinates as a search model. A unique solution was found in the first attempt with high contrast. After rigid-body refinement with Refmac5,30

different conformation of the designed connecting loop became apparent and was manually rebuilt in Coot.³¹ The structure was completed and finalized in cycles of manual modeling in Coot and refinement with BUSTER or PHENIX-refine.³² Data processing and refinement statistics are summarized in Tables S2 and S4.

NMR Structure Determination. Spectra were recorded at 310 K on Bruker AVIII-600 and AVIII-800 spectrometers. Backbone sequential assignments were made using standard triple-resonance experiments and by tracing strong NOESY contacts between sequential amide protons in helical segments. Aliphatic side chain assignments were completed with TOCSY-based experiments, while partial aromatic assignments were made by linking aromatic spin systems to unambiguously assigned aliphatic groups in NOESY spectra. The oligomeric purity of samples was checked with diffusion-ordered (DOSY) spectra. These confirmed that fresh samples used in diffusion experiments were predominantly monomeric.

To identify interhelical contacts, we exploited the uneven distribution of methionine residues observed in the dimeric crystal structure. The 16 methionine residues in this structure fall into three broad clusters, one within each repeat and a third at the inter-repeat interface. To assign these we produced a sample selectively ¹³C-labeled on methionine methyl groups on a ¹²C, ¹⁵N-labeled background. Members of each cluster could be identified by contacts between the labeled methyl groups in a 3D CCH-NOESY experiment.³³ Contacts to unambiguously assigned protons in a ¹³C-HSQC-NOESY spectrum then allowed the assignment of all members within each cluster. Thus, assigned, these methyl groups were effective probes of the interhelical interfaces providing 34 long-range distance restraints. These were applied, in simulated annealing calculations, together with other unambiguously assigned contacts and TALOS-based dihedral restraints. A summary of the input data and final structure statistics is given in Supplementary Table S3.

Structures were calculated with XPLOR-NIH (version 2.9.4) using a monomer extracted from the domain-swapped dimer as a starting structure; i.e., an open structure with no interunit interface. Simulated annealing runs were first aimed at closing this interface by treating the four-helix bundles as pseudorigid bodies. The resulting set of 50 structures defined an interface very similar to that observed in the crystal structure. Refinement was performed using atomistic molecular dynamics computations in isothermal-isobaric ensembles to accommodate large conformational changes, where the overall explicit solvent simulations setup was similar to that described above. A total of 135 ns were collected while deploying the NMR-derived dihedral and distance restraints using the harmonic restraint terms $k_{\text{torsion}}(\theta_t - \theta_{\text{ref}})^2$ and $k_{\text{distance}}(x_t - \theta_{\text{ref}})^2$ $(x_{ref})^2$, respectively. Here $k_{torsion}$ and $k_{distance}$ are the dihedral and distance spring constants (set at 1 and 0.1, respectively), θ_t is the φ or ψ angle at time t, x_t is the atom pair distance at time t, while θ_{ref} and x_{ref} are the NMR-derived values. Fifty frames from these runs were picked on the basis of agreement with distance restraints and minimized under restraints in XPLOR-NIH to regularize covalent geometry. The final ensemble consisted of 26 structures chosen on the basis of lowest restraint violations.

Structural Analysis. Searching among existing structures for similar folds was performed using three different methods. The PDBeFOLD¹³ and DALI¹⁴ servers were used to search against the entire PDB for similar existing folds to the

experimental structures of BRIC1 and BRIC2. The resulting hits were sorted by their alignment lengths, and manually inspected the top 100 hits for similar topologies. Additionally, the ECOD database¹⁵ (ECOD40 subset) was searched using TM-align¹⁶ for the same purpose. Only hits with TM-score equal or above 0.5 were manually inspected for potential similarity.

Polar precision at the designed interface was assessed by calculating the deviation between the designs and experimental structures for three quantities; the tilt (θ) , bend (β) and curvature (κ) across the designed interface. The three quantities are supposed to represent the plane-projected angular change between the two helical hairpins across the designed interface, along the three mutually orthogonal planes. The assessment of the designed interface accuracy in Cartesian and qualitative terms was done using the CAPRI interface criteria: $L_{\rm rms}$, $I_{\rm rms}$ and $f_{\rm nat}$.¹⁷ The $L_{\rm rms}$ represents the backbone RMSD of the protein unit downstream of the designed interface, after structurally aligning the pair by their upstream units. The $I_{\rm rms}$ was calculated as the backbone RMSD between the residues at the designed interface (defined by a distance cutoff of 10 Å). The f_{nat} represents the number of contacts common across the designed interface between the design and experimental structure, divided by the total number of contacts in the experimental structure. A contact is defined by the existence of any interatomic distance within 5 Å between two residues across either side of the interface (the designed loop residues were not considered).

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.8b00224.

Figures S1-S9; Tables S1-S3 (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: andrei.lupas@tuebingen.mpg.de.

ORCID [©]

Murray Coles: 0000-0001-6716-6150 Marcus D. Hartmann: 0000-0001-6937-5677 Andreas Plückthun: 0000-0003-4191-5306

Andrei N. Lupas: 0000-0002-1959-4836

Notes

The authors declare no competing financial interest. The crystal structure of the dimeric BRIC1 has been deposited

in the RCSB Protein Data Bank under the code 6FF6. The crystal structure of the D12-BRIC2 fusion has been deposited in the RCSB Protein Data Bank under the code 6FES.

ACKNOWLEDGMENTS

We thank Reinhard Albrecht and Vincent Truffault from the Max Planck Institute for Developmental Biology for technical assistance in structure determination experiments. We also thank Beat Blattmann from the UZH Protein Crystallization Center for setting up the crystallization experiments for the D12-BRIC2 chimera. This work was supported by institutional funds of the Max Planck Society and by grant 310030B_166676 from the Swiss National Science Foundation to AP. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

(1) Woolfson, D. N., Bartlett, G. J., Burton, A. J., Heal, J. W., Niitsu, A., Thomson, A. R., and Wood, C. W. (2015) De novo protein design: how do we expand into the universe of possible protein structures? *Curr. Opin. Struct. Biol.* 33, 16–26.

(2) (a) Kloss, E., Courtemanche, N., and Barrick, D. (2008) Repeatprotein folding: new insights into origins of cooperativity, stability, and topology. Arch. Biochem. Biophys. 469 (1), 83-99. (b) Rowling, P. J. E., Sivertsson, E. M., Perez-Riba, A., Main, E. R. G., and Itzhaki, L. S. (2015) Dissecting and reprogramming the folding and assembly of tandem-repeat proteins. Biochem. Soc. Trans. 43 (5), 881-888. (c) Kajander, T., Cortajarena, A. L., Main, E. R. G., Mochrie, S. G. J., and Regan, L. (2005) A New Folding Paradigm for Repeat Proteins. J. Am. Chem. Soc. 127 (29), 10188-10190. (d) Mello, C. C., and Barrick, D. (2004) An experimentally determined protein folding energy landscape. Proc. Natl. Acad. Sci. U. S. A. 101 (39), 14102-14107. (e) Lowe, A. R., and Itzhaki, L. S. (2007) Rational redesign of the folding pathway of a modular protein. Proc. Natl. Acad. Sci. U. S. A. 104 (8), 2679-2684. (f) Wetzel, S. K., Settanni, G., Kenig, M., Binz, H. K., and Plückthun, A. (2008) Folding and Unfolding Mechanism of Highly Stable Full-Consensus Ankyrin Repeat Proteins. J. Mol. Biol. 376 (1), 241-257. (g) Wetzel, S. K., Ewald, C., Settanni, G., Jurt, S., Plückthun, A., and Zerbe, O. (2010) Residue-Resolved Stability of Full-Consensus Ankyrin Repeat Proteins Probed by NMR. J. Mol. Biol. 402 (1), 241-258.

(3) Plückthun, A. (2015) Designed Ankyrin Repeat Proteins (DARPins): Binding Proteins for Research, Diagnostics, and Therapy. *Annu. Rev. Pharmacol. Toxicol.* 55 (1), 489–511.

(4) (a) Wu, Y., Batyuk, A., Honegger, A., Brandl, F., Mittl, P. R. E., and Plückthun, A. (2017) Rigidly connected multispecific artificial binders with adjustable geometries. *Sci. Rep.* 7 (1), 11217. (b) Batyuk, A., Wu, Y., Honegger, A., Heberling, M. M., and Plückthun, A. (2016) DARPin-Based Crystallization Chaperones Exploit Molecular Geometry as a Screening Dimension in Protein Crystallography. *J. Mol. Biol.* 428 (8), 1574–1588.

(5) (a) Main, E. R. G., Jackson, S. E., and Regan, L. (2003) The folding and design of repeat proteins: reaching a consensus. Curr. Opin. Struct. Biol. 13 (4), 482-489. (b) Parmeggiani, F., Huang, P.-S., Vorobiev, S., Xiao, R., Park, K., Caprari, S., Su, M., Seetharaman, J., Mao, L., Janjua, H., Montelione, G. T., Hunt, J., and Baker, D. (2015) A General Computational Approach for Repeat Protein Design. J. Mol. Biol. 427 (2), 563-575. (c) Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P., and Plückthun, A. (2003) Designing Repeat Proteins: Well-expressed, Soluble and Stable Proteins from Combinatorial Libraries of Consensus Ankyrin Repeat Proteins. J. Mol. Biol. 332 (2), 489-503. (d) Parmeggiani, F., Pellarin, R., Larsen, A. P., Varadamsetty, G., Stumpp, M. T., Zerbe, O., Caflisch, A., and Plückthun, A. (2008) Designed Armadillo Repeat Proteins as General Peptide-Binding Scaffolds: Consensus Design and Computational Optimization of the Hydrophobic Core. J. Mol. Biol. 376 (5), 1282-1304.

(6) (a) Brunette, T. J., Parmeggiani, F., Huang, P.-S., Bhabha, G., Ekiert, D. C., Tsutakawa, S. E., Hura, G. L., Tainer, J. A., and Baker, D. (2015) Exploring the repeat protein universe through computational protein design. *Nature* 528, 580. (b) Parmeggiani, F., and Huang, P.-S. (2017) Designing repeat proteins: a modular approach to protein design. *Curr. Opin. Struct. Biol.* 45, 116–123. (c) Doyle, L., Hallinan, J., Bolduc, J., Parmeggiani, F., Baker, D., Stoddard, B. L., and Bradley, P. (2015) Rational design of α -helical tandem repeat proteins with closed architectures. *Nature* 528, 585. (d) Park, K., Shen, B. W., Parmeggiani, F., Huang, P.-S., Stoddard, B. L., and Baker, D. (2015) Control of repeat protein curvature by computational protein design. *Nat. Struct. Mol. Biol.* 22 (2), 167–174.

(7) Kobe, B., and Kajava, A. V. (2000) When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. *Trends Biochem. Sci.* 25 (10), 509–515.

(8) Kastritis, P. L., Moal, I. H., Hwang, H., Weng, Z., Bates, P. A., Bonvin, A. M. J. J., and Janin, J. (2011) A structure-based benchmark for protein–protein binding affinity. *Protein Sci.* 20 (3), 482–491.

(9) Chen, P.-C., and Kuyucak, S. (2011) Accurate Determination of the Binding Free Energy for KcsA-Charybdotoxin Complex from the Potential of Mean Force Calculations with Restraints. *Biophys. J.* 100 (10), 2466–2474.

(10) Mourey, L., Da Re, S., Pédelacq, J.-D., Tolstykh, T., Faurie, C., Guillet, V., Stock, J. B., and Samama, J.-P. (2001) Crystal Structure of the CheA Histidine Phosphotransfer Domain that Mediates Response Regulator Phosphorylation in Bacterial Chemotaxis. *J. Biol. Chem.* 276 (33), 31074–31082.

(11) Murphy, G. S., Mills, J. L., Miley, M. J., Machius, M., Szyperski, T., and Kuhlman, B. (2012) Increasing Sequence Diversity with Flexible Backbone Protein Design: The Complete Redesign of a Protein Hydrophobic Core. *Structure (Oxford, U. K.)* 20 (6), 1086–1096.

(12) Garron, M.-L., Arthos, J., Guichou, J.-F., McNally, J., Cicala, C., and Arold, S. T. (2008) Structural Basis for the Interaction between Focal Adhesion Kinase and CD4. *J. Mol. Biol.* 375 (5), 1320–1328.

(13) Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 60, 2256–2268.

(14) Holm, L., and Laakso, L. M. (2016) Dali server update. Nucleic Acids Res. 44, W351–W355.

(15) Cheng, H., Schaeffer, R. D., Liao, Y., Kinch, L. N., Pei, J., Shi, S., Kim, B.-H., and Grishin, N. V. (2014) ECOD: An Evolutionary Classification of Protein Domains. *PLoS Comput. Biol.* 10 (12), e1003926.

(16) Zhang, Y., and Skolnick, J. (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* 33 (7), 2302–2309.

(17) Mendez, R., Leplae, R., De Maria, L., and Wodak, S. J. (2003) Assessment of blind predictions of protein–protein interactions: Current status of docking methods. *Proteins: Struct., Funct., Genet.* 52 (1), 51–67.

(18) (a) Kim, D. E., Blum, B., Bradley, P., and Baker, D. (2009) Sampling bottlenecks in de novo protein structure prediction. *J. Mol. Biol.* 393 (1), 249–260. (b) Huang, P.-S., Boyken, S. E., and Baker, D. (2016) The coming of age of de novo protein design. *Nature* 537, 320.

(19) Jacobs, T. M., Williams, B., Williams, T., Xu, X., Eletsky, A., Federizon, J. F., Szyperski, T., and Kuhlman, B. (2016) Design of structurally distinct proteins using strategies inspired by evolution. *Science* 352 (6286), 687.

(20) Fleishman, S. J., Leaver-Fay, A., Corn, J. E., Strauch, E.-M., Khare, S. D., Koga, N., Ashworth, J., Murphy, P., Richter, F., Lemmon, G., Meiler, J., and Baker, D. (2011) RosettaScripts: A Scripting Language Interface to the Rosetta Macromolecular Modeling Suite. *PLoS One 6* (6), e20161.

(21) Sheffler, W., and Baker, D. (2009) RosettaHoles: Rapid assessment of protein core packing for structure prediction, refinement, design, and validation. *Protein Sci.* 18 (1), 229–239.

(22) Leaver-Fay, A., O'Meara, M. J., Tyka, M., Jacak, R., Song, Y., Kellogg, E. H., Thompson, J., Davis, I. W., Pache, R. A., Lyskov, S., Gray, J. J., Kortemme, T., Richardson, J. S., Havranek, J. J., Snoeyink, J., Baker, D., and Kuhlman, B. (2013) Scientific Benchmarks for Guiding Macromolecular Energy Function Improvement. *Methods Enzymol.* 523, 109–143.

(23) Smith, C. A., and Kortemme, T. (2008) Backrub-like backbone simulation recapitulates natural protein conformational variability and improves mutant side-chain prediction. *J. Mol. Biol.* 380 (4), 742–756.

(24) Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M., and MacKerell, A. D. (2012) Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ , ψ and Side-Chain χ 1 and χ 2 Dihedral Angles. J. Chem. Theory Comput. 8 (9), 3257–3273. (25) Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005) Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* 26 (16), 1781–1802.

(26) Studier, F. W. (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expression Purif.* 41 (1), 207–234.

(27) Ortega, A., Amorós, D., and García de la Torre, J. (2011) Prediction of Hydrodynamic and Other Solution Properties of Rigid Proteins from Atomic- and Residue-Level Models. *Biophys. J.* 101 (4), 892–898.

(28) Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* 26 (6), 795–800.

(29) Vagin, A., and Teplyakov, A. (2000) An approach to multi-copy search in molecular replacement. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 56* (12), 1622–1624.

(30) Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S., and Dodson, E. J. (1999) Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 55 (1), 247–255.

(31) Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 60, 2126–2132.

(32) (a) Bricogne, G. B. E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P. S, A., Smart, O. S., Vonrhein, C., and Womack, T. O. (2017) *BUSTER*, version 2.10.3, Global Phasing Ltd., Cambridge, UK. (b) Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr, Sect. D: Biol. Crystallogr.* 66 (2), 213–221.

(33) Diercks, T., Coles, M., and Kessler, H. (1999) An efficient strategy for assignment of cross-peaks in 3D heteronuclear NOESY experiments. J. Biomol. NMR 15 (2), 177–180.