

De novo design of buttressed loops for sculpting protein functions

Received: 18 February 2023

Accepted: 29 April 2024

Published online: 30 May 2024

 Check for updates

Hanlun Jiang^{1,2,8}, Kevin M. Jude^{3,4,8}, Kejia Wu^{1,2,5,8}, Jorge Fallas^{1,2}, George Ueda^{1,2}, T. J. Brunette^{1,2}, Derrick R. Hicks^{1,2}, Harley Pyles^{1,2}, Aerin Yang⁴, Lauren Carter^{1,2}, Mila Lamb^{1,2}, Xinting Li^{1,2}, Paul M. Levine^{1,2}, Lance Stewart^{1,2}, K. Christopher Garcia^{3,4,6}✉ & David Baker^{1,2,7}✉

In natural proteins, structured loops have central roles in molecular recognition, signal transduction and enzyme catalysis. However, because of the intrinsic flexibility and irregularity of loop regions, organizing multiple structured loops at protein functional sites has been very difficult to achieve by de novo protein design. Here we describe a solution to this problem that designs tandem repeat proteins with structured loops (9–14 residues) buttressed by extensive hydrogen bonding interactions. Experimental characterization shows that the designs are monodisperse, highly soluble, folded and thermally stable. Crystal structures are in close agreement with the design models, with the loops structured and buttressed as designed. We demonstrate the functionality afforded by loop buttressing by designing and characterizing binders for extended peptides in which the loops form one side of an extended binding pocket. The ability to design multiple structured loops should contribute generally to efforts to design new protein functions.

While antibodies still have central roles in protein therapeutics, progress has been made in drug development using nonantibody-binding proteins that show superior properties in thermal/pH stability, binding affinities, tissue delivery and industrial-scale manufacture^{1–3}. The two main approaches are random library selection methods and computational protein design. Perhaps the most successful scaffold for random library selection has been the ankyrin repeat^{4,5}; libraries of designed ankyrin repeat proteins (DARPs) have been used to identify high-affinity binding proteins via high-throughput screening methods, which have had multiple successes in preclinical studies^{3,5,6}. Ankyrin repeat proteins have a repeating architecture with structured, hairpin-shaped loops extending from the helices to an extended binding groove that is geometrically compatible with many globular protein targets. Despite these successes, the global shape diversity of DARPs

is limited by the use of a single base scaffold. Computational design of binding proteins does not have this limitation as a wide range of scaffolds can be used, with shapes more optimal to bind the target protein of interest. However, this advantage thus far has come with a different limitation—because of the inherent flexibility and lack of extensive backbone hydrogen bonding of long loop regions, protein binder design has focused on scaffolds and binding sites primarily composed of α helical⁷ or β strand⁸ secondary structure, which has limited the achievable local shape diversity.

Results

Design approach

Here we set out to overcome the challenges in de novo design of long loops on the one hand, and the limitations of ankyrin scaffolds in global

¹Department of Biochemistry, University of Washington, Seattle, WA, USA. ²Institute for Protein Design, University of Washington, Seattle, WA, USA.

³Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA, USA. ⁴Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, CA, USA. ⁵Biological Physics, Structure and Design Graduate Program, University of Washington, Seattle, WA, USA. ⁶Department of Structural Biology, Stanford University School of Medicine, Stanford, CA, USA. ⁷Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA. ⁸These authors contributed equally: Hanlun Jiang, Kevin M. Jude, Kejia Wu. ✉ e-mail: kcgarcia@stanford.edu; dabaker@uw.edu

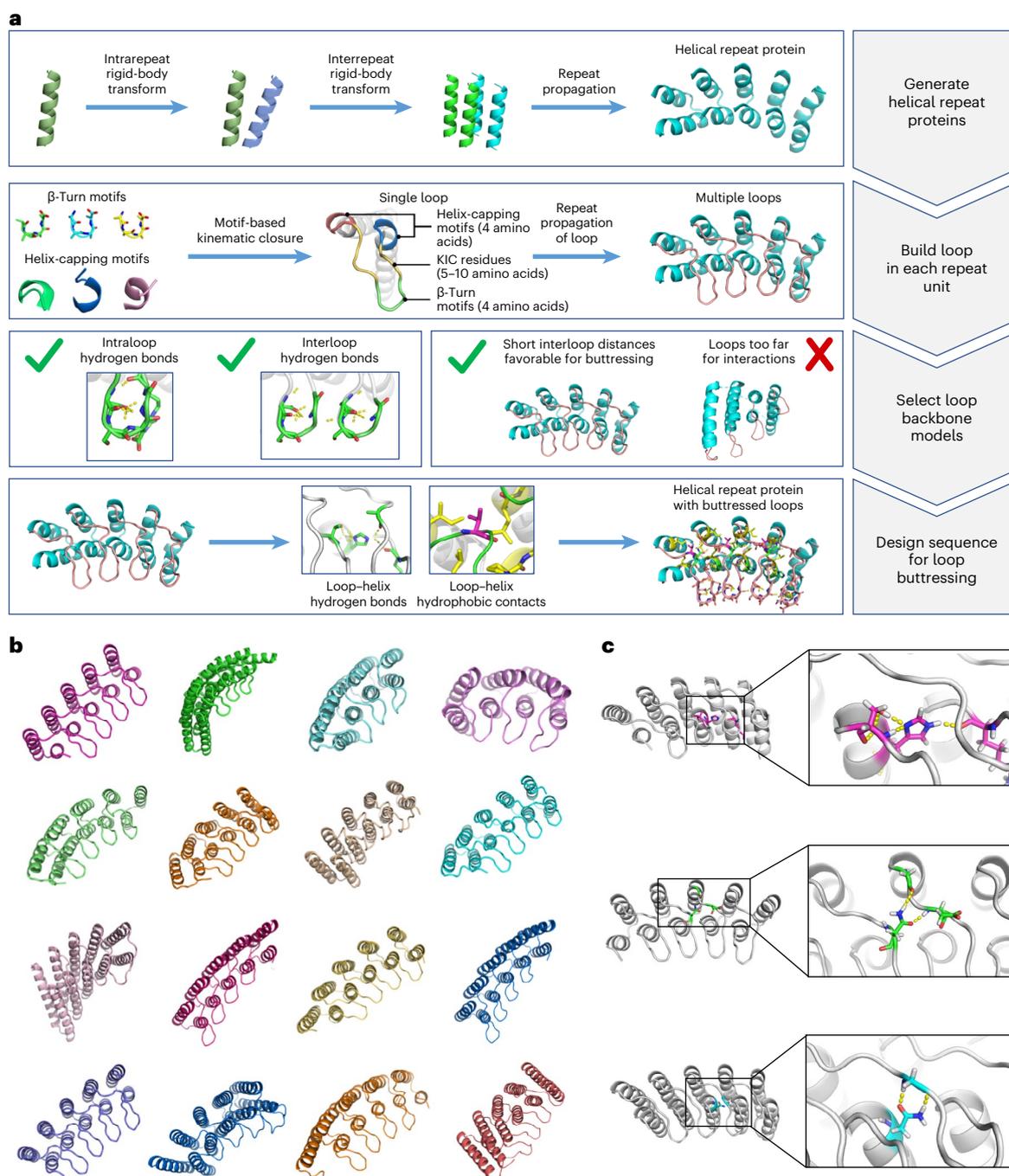


Fig. 1 | Computational design of RBLs. a, Design strategy for generating and stabilizing multiple loops in helical repeat proteins. **b**, A gallery of diverse designed proteins that pass the *in silico* design filters. **c**, Loop buttressing hydrogen bonds in the designed proteins.

shape diversity on the other, by computationally designing repeat proteins with multiple long loops buttressed by loop–loop interactions (Fig. 1a). To achieve this goal, we divided the problem into the following two subproblems: first, the generation of repeating scaffold backbone conformations compatible with loop buttressing, and second, the generation of loop backbone conformations compatible with a dense network of hydrogen bonds and hydrophobic interactions between pairs of loops and between the loops and the underlying scaffold.

We developed a computational method for generating a wide range of repeat protein backbones that are geometrically compatible with the insertion of long loops (Fig. 1a, top row). Previous approaches to designing helical repeat proteins have used fragment assembly methods to assemble repeat units with short loops connecting them⁹.

While these methods can generate considerable diversity, we found that they did not provide sufficient control over backbone positions for designing loop buttressing. Instead, we developed a parametric repeat protein generation method that enables precise control over backbone placement. We generated diverse repeat units consisting of two idealized helices by systematically sampling the lengths of the helices (from 12 to 28 residues) and the six rigid-body degrees of freedom between the two helices. We next sampled the six rigid-body degrees of freedom between repeat units and applied the same transform repeatedly to generate a disconnected repeat protein model. Finally, we connected pairs of sequence-adjacent helices using a native-protein-based loop lookup protocol that grafts on loops (from three to six residues) that best fit onto the termini of the helices¹⁰. In extensive model-building

experiments, we found that to enable the installation of long loops onto these parametrically generated models, the termini of the helices had to be less than 18 Å apart, and we removed backbone models where the distance between termini was greater than this value. We also eliminated poorly packed models with fewer than 28% of the residues in a buried core. The resulting repeat protein models have well-defined core regions and are slightly curved with little or no twisting between the repeat units.

We next sought to develop a general method for building multiple long loops that buttress one another onto protein scaffolds (Fig. 1a, second and third rows). Surveying the structured, long loops in natural proteins, we observed that they frequently contain β turns with strand-like hydrogen bonds flanking the turn residues, which contribute to the stabilization of the specific loop conformation. We also observed that natural proteins often use helix-capping interactions between the sidechain or backbone on the loop residue and the backbone of the helix from which it emanates; this feature helps specify the orientation of the loop as it leaves the helix. Based on these observations, we constructed and curated libraries of β -turn motifs and helix-capping motifs by clustering four-residue native-protein fragments, and we selected the clusters that fulfilled the requirements of hydrogen bonds as described in Methods. During the loop sampling, these motifs were randomly selected and incorporated into a single loop growing from the C-terminus of a helix. Using generalized kinematic closure¹¹, we then connected the C-terminus of the loop to the N-terminus of the next helix in the backbone model. The resulting loop was then propagated to each repeat unit to generate a complete repeat protein model with multiple long loops. To specifically favor loops that could be buttressed with hydrogen bond networks, we required that models have at least two intraloop backbone-to-backbone hydrogen bonds within each repeat unit and at least one interloop backbone-to-backbone hydrogen bond between the repeat unit neighbors. To favor interactions between the long loops and the helices, we further filtered the models by requiring at least five residues within 8 Å of the closest helical residues. The remaining backbone models following these filtering steps contain long loops arranged in sheet-like structures ready for the installation of additional sidechain-based buttressing interactions.

We designed sequences onto these backbones, focusing on further loop stabilization through buttressing (Fig. 1a, bottom row). We began by scanning each position on the long loops for Asn, Asp, His or Gln placements that form backbone-sidechain bidentate hydrogen bonds between loops or between a loop and a helix, and for Val, Leu, Ile, Met or Phe placements that form loop-helix hydrophobic contacts; amino acids meeting these criteria were kept fixed in subsequent design steps. We then performed four rounds of full combinatorial Rosetta protein sequence design with slowly ramped-up fa_{rep} weight to promote core packing. A slight compositional bias toward proline was used in the long loop to increase rigidity. The design models were filtered in Rosetta by the number of buried unsatisfied heavy atoms (≤ 3), core residue hole score (≤ -0.015), total score per residue (≤ -2), packstat (≥ 0.5) and average hydrogen bond energy per residue (≤ -1) in the buttressed long loops. The rigidity of the design models was evaluated using molecular dynamics simulations and the extent to which the designed sequence encodes the structure by AlphaFold^{12,13}. The in silico validated designs span a diverse range of shapes with different repeat protein curvatures and loop geometries (Fig. 1b) and adopt multiple loop buttressing strategies using loop-helix hydrogen bond networks and loop-loop bidentate hydrogen bonds (Fig. 1c). These designed buttressed loops have significantly more diverse structures than the long, hairpin loops in the native ankyrins (Extended Data Fig. 1) and contain more backbone hydrogen bonds (Extended Data Fig. 2).

Experimental characterization

We expressed 102 selected designs (which we call repeat proteins with buttressed loops (RBLs)) in *Escherichia coli* and purified them

by His tag-immobilized metal affinity chromatography. In total, 77 of the purified proteins were soluble (representative models shown in Fig. 2a), 52 were monodisperse and 46 were monomeric, as indicated by multi-angle light scattering coupled with size-exclusion chromatography (SEC-MALS; Fig. 2b). Forty-four of these proteins showed the expected α -helical circular dichroism (CD) spectrum at 25 °C, remained at least partially folded at 95 °C and recovered nearly all the CD signal when cooled down to 25 °C (Fig. 2c). Fourteen designs were further validated by small-angle X-ray scattering (SAXS; Fig. 2d and Extended Data Fig. 3). The experimental scattering curves agreed with profiles computed from the design models.

We determined the crystal structure of design RBL4 at 1.8 Å resolution (Fig. 3a–e). RBL4 contains four helix-long-loop-helix repeat units that are sandwiched by two terminal capping helices. Each long loop is anchored on top of the neighboring helices and stabilized by inter-loop Asn-mediated bidentate hydrogen bond networks as designed, and the design model is in good agreement with the crystal structure with a C α root-mean-square deviation (RMSD) of 1.7 Å (Fig. 3a). The primary discrepancy between the crystal structure and design model is in the inter-repeat transformation—the design model is slightly curved (smaller superhelical radius), while the crystal structure is nearly flat (larger superhelical radius). Within individual repeat units, there is very close agreement between the crystal and design model, with repeat unit C α RMSDs for different repeat units ranging from 0.48 to 0.61 Å (Fig. 3b). The designed loop buttressing interactions—the bidentate interloop hydrogen bonds (Fig. 3c) and loop-helix salt bridges (Fig. 3d)—were accurately recapitulated in the crystal structure. B-factors for the loop residues are elevated compared to the helix residues (Extended Data Fig. 4a), but the fit to the electron density shows that the loops are well ordered (Extended Data Fig. 4b).

Design RBL7 has a similar overall geometry as RBL4, but with a smaller superhelical radius. This design was highly stable and monomeric, with an overall fold validated by SAXS (Fig. 2, second row). We obtained crystals that diffracted poorly with the highest resolution at 4.2 Å. As previous studies suggested that synthetic oligomerization can sometimes assist crystallization¹⁴, we sought to generate a dimeric form of RBL7 by introducing a hydrophobic dimer interface. The redesigned protein, RBL7_C2_3, was soluble and dimeric, and we were able to solve the crystal structure at 3 Å resolution. The crystal structure closely matches the design model, with a C α RMSD over the dimer of 2.9 Å (Fig. 3f) and over the monomer of 1.6 Å (Fig. 3g). The main discrepancies between the crystal and designed structures were in the terminal helices. Similar to design RBL4, the crystal structure confirmed the accuracy of the designed loop buttressing interactions in RBL7_C2_3 (Fig. 3h–j). All of the designed interloop hydrogen bonds at the β turns of long loops were recapitulated in the crystal structure (Fig. 3h). These hydrogen bonds are likely crucial for positioning the long loops and contribute to the close matching between the loops in the design model and those in the crystal structure. Again, the B-factor values of the buttressed loops are slightly elevated in the loops compared to the helices (Extended Data Fig. 4c), but with a good fit to the electron density (Extended Data Fig. 4d).

Design of peptide-binding RBLs

An exciting application of our designed RBLs is to use them as starting points for the computational design of high-affinity binding proteins. This could enable the design of DARPIn-like binders for a wide range of targets without the need for large-scale library selection methods. The ability to design a wide diversity of repeating scaffolds with buttressed loops could considerably expand the space of targets. As a first step toward investigating the design of RBL-based binders, we redesigned the extended groove bordered by the buttressed loops to bind extended peptides. To take advantage of the repeating nature of RBLs, we chose to focus on peptides with a repeating sequence motif—in this case, once a repeat unit is designed to bind a particular short peptide, repeat

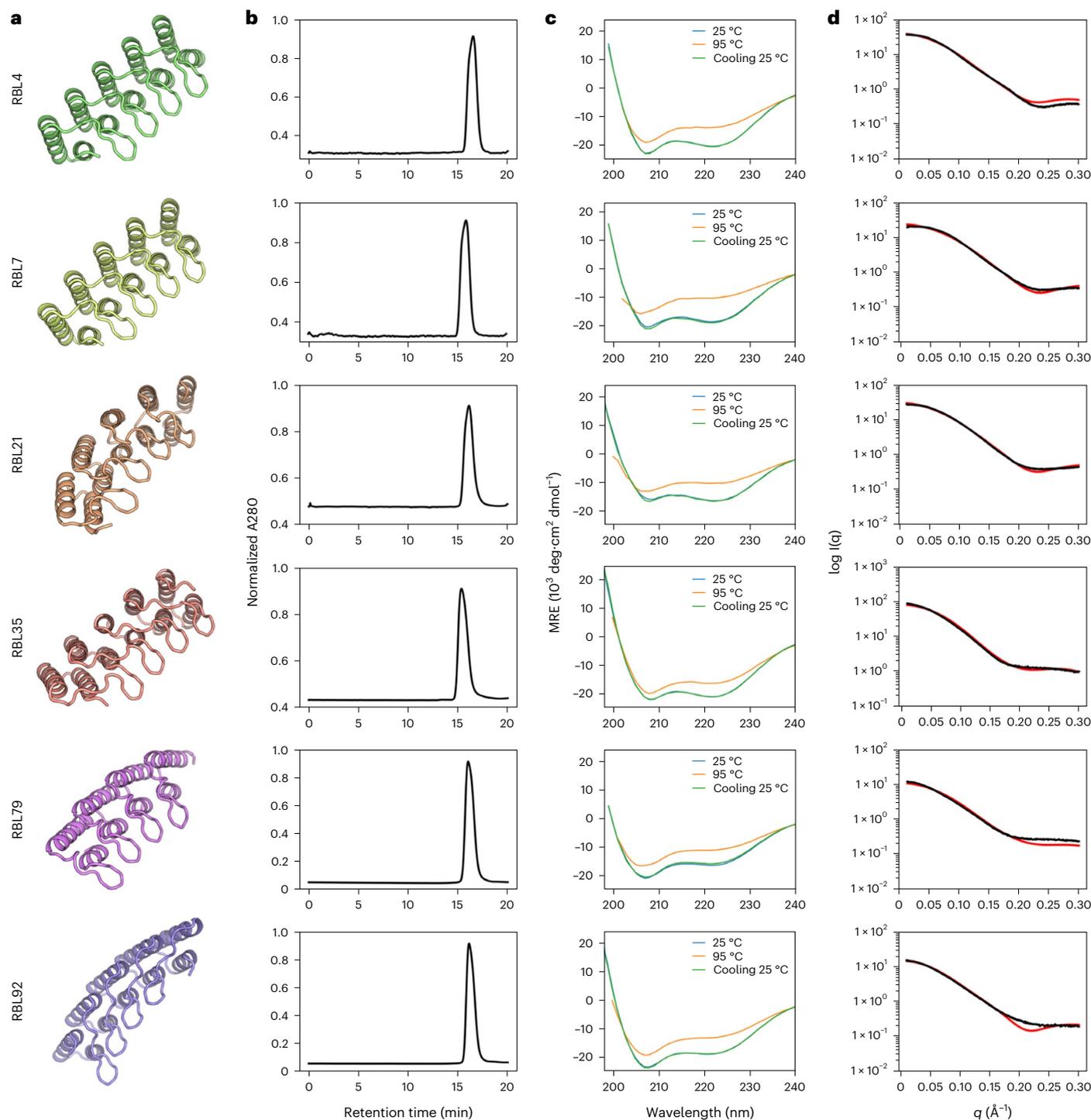


Fig. 2 | Biophysical characterization of designed helical RBLs. a, Design models of six representative designs. **b**, SEC traces monitoring absorbance at 280 nm. **c**, CD spectra collected at 25 °C (blue), 95 °C (orange) and 25 °C after cooling from 95 °C (green). **d**, Overlay of experimental (black) and theoretical (red) SAXS profiles.

proteins containing multiple copies of this unit should bind peptides with multiple copies of the motif, provided the register between the repeat protein and the peptide can be maintained. Generalizing from the observation that some ankyrin family proteins can bind peptides with a PxLPxI/L (x can be any amino acid) sequence motif⁴⁵, we sought to design binders for peptide sequences of the form $(XYZ)_n$, where n is the number of repeats, X is a polar residue interacting with residues in the buttressed loop β turns and Y and Z are hydrophobic residues interacting with the helices and the helix–loopjoint of RBLs (see Fig. 4b

for an example of one peptide repeat unit interacting with an RBL-based peptide binder).

To design binders of $(XYZ)_n$ peptides, we first docked tripeptide repeats in the polyproline II helix conformation to the binding grooves of RBLs guided by the interactions in peptide-binding ankyrin family proteins in the Protein Data Bank (PDB; Methods), and carried out rigid-body perturbations to diversify the docked poses. For each resulting pose, we used the Rosetta sequence design to generate sequences of both RBL and peptide for optimized binding. We

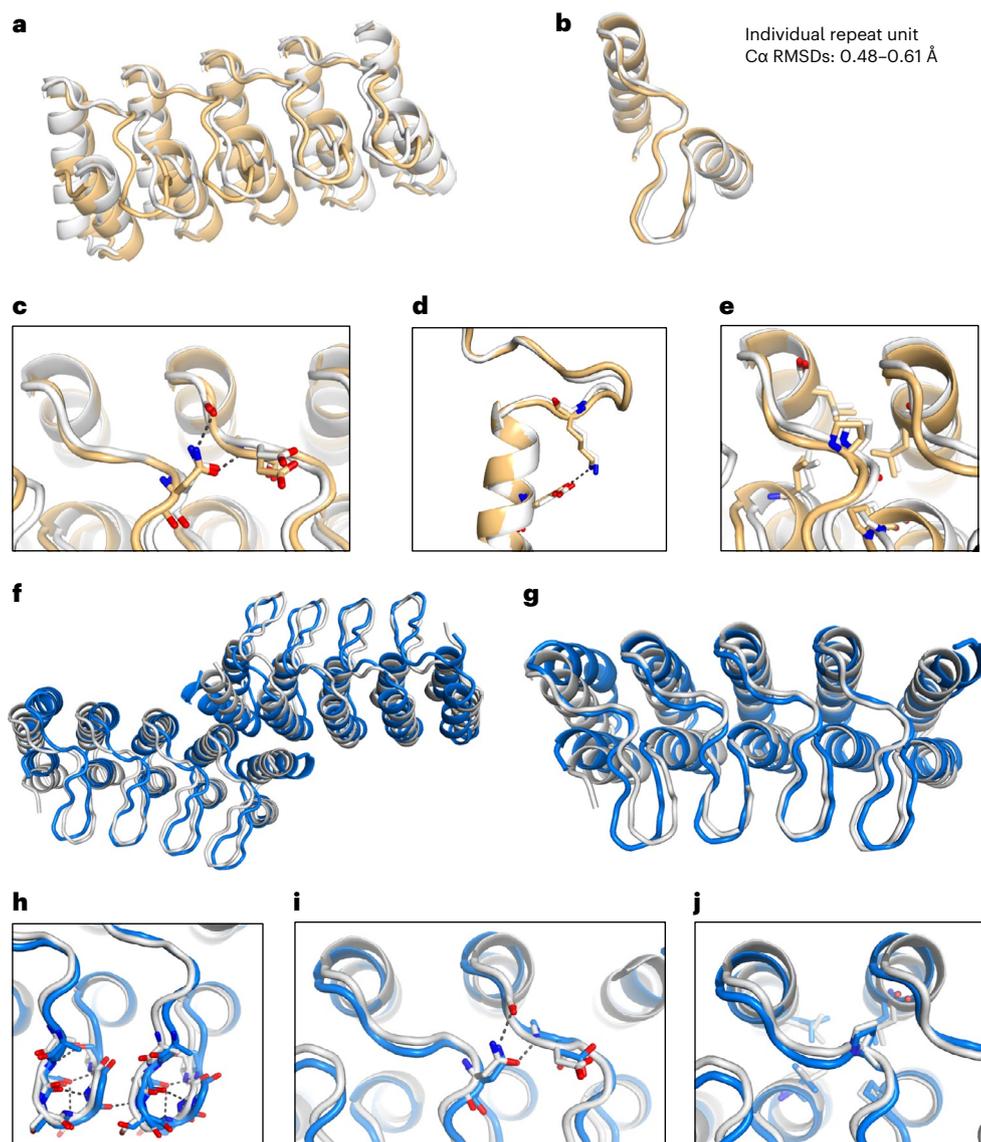


Fig. 3 | Structural characterization by X-ray crystallography. a, Superimposition of crystal structure (yellow) onto the design model of RBL4 (gray). **b**, Alignment of individual repeat units. **c–e**, Accurately designed loop buttressing interactions: bidentate interloop hydrogen bonds (**c**), loop–helix salt bridge (**d**) and loop–helix hydrophobic contacts (**e**). **f**, Superimposition of crystal

structure (blue) onto the design model of RBL7_C2_3 (gray). **g**, Superimposition of a monomer unit in the crystal structure onto the design model. **h–j**, Accurately designed loop buttressing interactions: intraloop and interloop hydrogen bonds (**h**), bidentate interloop hydrogen bonds (**i**) and loop–helix hydrophobic contacts (**j**).

designed 34 proteins to bind six-repeat peptides ((DLP)₆, (KLP)₆ or (DLS)₆), screened them *in silico* based on protein–protein interaction design filters including AlphaFold^{12,13} structure recapitulation, obtained synthetic genes encoding the designs and purified the proteins from *E. coli* expression. In the initial binding screens using split-luciferase assay, seven designs showed clear binding signals. From these designs, we selected the strongest binders for each peptide target and performed fluorescence polarization measurements, which showed the protein–peptide interactions are orthogonal and have high affinities (Fig. 4). All the selected binders (Fig. 4a,d,g) were based on RBL4 with the peptides in nearly identical binding modes. Unlike the natural peptide-binding ankyrins, the designed peptide binders interact specifically with every peptide residue side chain, with the Asp/Lys at the X position forming salt bridges with charged residues from the β-turn tip of RBL, the Leu at the Y position fully buried in the hydrophobic interface and the Pro/Ser interacting with the residues on the bottom of helices (Fig. 4b,e,h). Both the (DLP)₆ binder

and the (KLP)₆ binder bound their target peptides with high affinities ($K_d = 1.2$ nM and <0.3 nM, respectively) and high specificity (Fig. 4c,f). Neither binder strongly bound (DLS)₆, suggesting Pro in the peptide was crucial in the protein–peptide interactions. We sought to rescue the (DLS)₆ binding by installing Gln-mediated bidentate hydrogen bonds (Fig. 4h). The resulting design bound (DLS)₆ with high affinity ($K_d = 2.9$ nM) but retained affinity for (DLP)₆ (Fig. 4i).

Discussion

There are two primary routes forward for engineering new functions using our designed RBLs. First, by analogy with the many DARPins obtained starting from stabilized consensus ankyrin repeat proteins, it should be readily possible to create binders from RBLs by random library generation in conjunction with yeast display and other selection methods for binding to targets of interest. Second, as demonstrated by our design of peptide-binding proteins, computational design methods can be used to generate binders to a wide variety of targets, taking

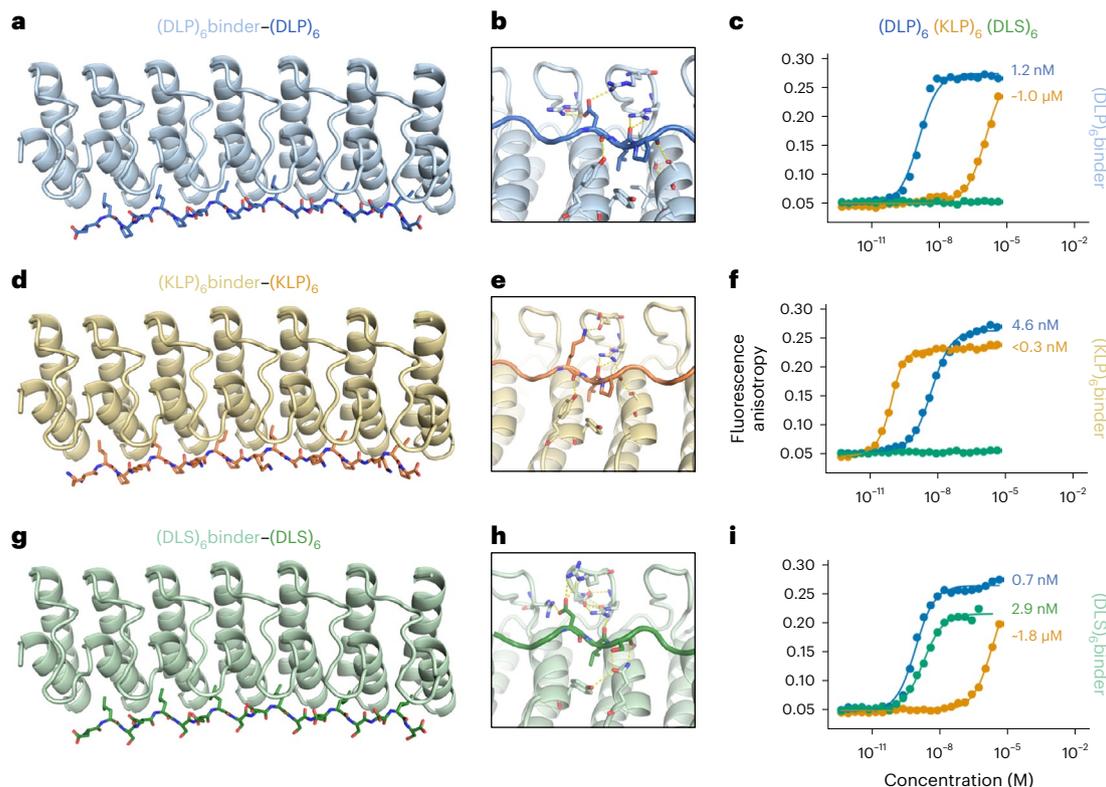


Fig. 4 | Designed repeat peptide-binding RBLs. a,d,g, Design models of peptide-binding proteins in complex with (DLP)₆ (a), (KLP)₆ (d) and (DLS)₆ (g). **b,e,h,** Sequence-specific interactions in the binder–peptide complexes: (DLP)₆binder–(DLP)₆ (b), (KLP)₆binder–(KLP)₆ (e) and (DLS)₆binder–(DLS)₆ (h).

c,f,i, Fluorescence polarization measurement of binding between (DLP)₆binder–(DLP)₆ (c), (KLP)₆binder–(KLP)₆ (f) and (DLS)₆binder–(DLS)₆ (i). For each binder, a titration curve is plotted for the binding of each peptide (blue, (DLP)₆; orange, (KLP)₆; and green, (DLS)₆).

advantage of the diverse geometries that can be achieved with different buttressed loops on different repeat protein scaffolds.

From a fundamental design perspective, the crystal structures presented here show that computational protein design has advanced to the point that proteins with multiple ordered long loops can now be designed. Key to this success was the design of dense networks of hydrogen bonding and nonpolar interactions within and between the loops and between the loops and the underlying secondary structural elements. Our approach, alone or integrated with additional recent progress in loop design¹⁶ and recently developed deep learning approaches for protein design^{17–21} (which do not directly address the challenge of designing structured long loops), should enable the design of structured loops for binding functions and beyond in a wide variety of scaffolds. For example, for enzyme design, multiple loops emanating from designed TIM barrels^{22–24} could be built to stabilize each other and, together with the residues emerging from the top of the β strands and helices in the TIM barrel structure, form an extensively buttressed catalytic site and associated substrate/transition state binding site.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41589-024-01632-2>.

References

- Yu, X., Yang, Y. P., Dikici, E., Deo, S. K. & Daunert, S. Beyond antibodies as binding partners: the role of antibody mimetics in bioanalysis. *Annu. Rev. Anal. Chem.* **10**, 293–320 (2017).
- Simeon, R. & Chen, Z. In vitro-engineered non-antibody protein therapeutics. *Protein Cell* **9**, 3–14 (2018).
- Stumpp, M. T., Dawson, K. M. & Binz, H. K. Beyond antibodies: the DARPin(R) drug platform. *BioDrugs* **34**, 423–433 (2020).
- Mosavi, L. K., Minor, D. L. & Peng, Z. Y. Consensus-derived structural determinants of the ankyrin repeat motif. *Proc. Natl Acad. Sci. USA* **99**, 16029–16034 (2002).
- Pluckthun, A. Designed ankyrin repeat proteins (DARPins): binding proteins for research, diagnostics, and therapy. *Annu. Rev. Pharmacol. Toxicol.* **55**, 489–511 (2015).
- Binz, H. K. et al. High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotechnol.* **22**, 575–582 (2004).
- Cao, L. et al. Design of protein binding proteins from target structure alone. *Nature* **605**, 551–560 (2022).
- Sahtoe, D. D. et al. Transferrin receptor targeting by de novo sheet extension. *Proc. Natl Acad. Sci. USA* **118**, e2021569118 (2021).
- Brunette, T. J. et al. Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580–584 (2015).
- Brunette, T. J. et al. Modular repeat protein sculpting using rigid helical junctions. *Proc. Natl Acad. Sci. USA* **117**, 8870–8875 (2020).
- Bhardwaj, G. et al. Accurate de novo design of hyperstable constrained peptides. *Nature* **538**, 329–335 (2016).
- Jumper, J. & Hassabis, D. Protein structure predictions to atomic accuracy with AlphaFold. *Nat. Methods* **19**, 11–12 (2022).
- Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
- Banatao, D. R. et al. An approach to crystallizing proteins by synthetic symmetrization. *Proc. Natl Acad. Sci. USA* **103**, 16230–16235 (2006).

15. Xu, C. et al. Sequence-specific recognition of a PxLPxI/L motif by an ankyrin repeat tumbler lock. *Sci. Signal* **5**, ra39 (2012).
16. Kundert, K. & Kortemme, T. Computational design of structured loops for new protein functions. *Biol. Chem.* **400**, 275–288 (2019).
17. Watson, J. L. et al. De novo design of protein structure and function with RFDiffusion. *Nature* **620**, 1089–1100 (2023).
18. Dauparas, J. et al. Robust deep learning-based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022).
19. Lee, J. S., Kim, J. & Kim, P. M. Score-based generative modeling for de novo protein design. *Nat. Comput. Sci.* **3**, 382–392 (2023).
20. Ferruz, N. & Höcker, B. Controllable protein design with language models. *Nat. Mach. Intell.* **4**, 521–532 (2022).
21. Madani, A. et al. Large language models generate functional protein sequences across diverse families. *Nat. Biotechnol.* **41**, 1099–1106 (2023).
22. Huang, P. S. et al. De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nat. Chem. Biol.* **12**, 29–34 (2016).
23. Caldwell, S. J. et al. Tight and specific lanthanide binding in a de novo TIM barrel with a large internal cavity designed by symmetric domain fusion. *Proc. Natl Acad. Sci. USA* **117**, 30362–30369 (2020).
24. Chu, A. E., Fernandez, D., Liu, J., Eguchi, R. R. & Huang, P.-S. De novo design of a highly stable ovoid TIM barrel: unlocking pocket shape towards functional design. *Biodes. Res.* **2022**, 9842315 (2022).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

Methods

Computational design method

We developed our computational protein design protocols using Rosetta^{25,26} (2019.01) and PyRosetta4 (release 2019.22)²⁷. Our protocol of parametric repeat protein generation started by building an ideal helix H1 (with a length of 12–28 residues) with the MakeBundleHelix mover in Rosetta^{25,26} and placing it away from the z axis with a given radius and an angle corresponding to its orientation. A second helix, H2 (with a length of 12–28 residues), was then modeled and placed according to the specification of the six rigid-body degrees of freedom for geometry transformation from H1 to H2. By combining H1 and H2 into one pose, we built the first repeat unit R1. Subsequently, we used user-specified six rigid-body degrees of freedom between repeat units to perform a geometric transformation to obtain the second unit R2. We propagated the repeat units based on the number of repeats desired to generate the helical repeat protein backbones. We then connected pairs of sequence-adjacent helices with loops of three to six residues using ConnectChainMover¹⁰. To filter the generated repeat protein backbones, we required a maximum distance of 18 Å between the termini of the helices to be connected by buttressed long loops. We also removed the low-quality backbone models with fewer than 28% of the residues in a buried core.

To design buttressed loops, we developed a hybrid method that assembles native structural motifs via kinematic loop closure. To guide the sampling toward the hairpin-shaped conformations, we constructed a motif library that consists of native β turns. A β -turn motif is defined by having a backbone-to-backbone hydrogen bond between the carbonyl group of residue i and the amine group of residue $i + 3$ (refs. 28,29). In this work, we searched for native β -turn fragments by mining a set of selected PDBs based on 90% maximum sequence identity and a 1.6 Å resolution cutoff from PISCES³⁰. The collected β turns were further clustered by the K -centers algorithm³¹ at a maximum cluster distance of 0.63 Å, resulting in 180 motif clusters. Using the same approach, we compiled a library of native helical capping motifs to guide the sampling of loops connecting helices in the repeat proteins.

We used GeneralizedKIC¹¹ for loop closure. An extended loop fragment was first constructed by stitching native helical capping motifs (four amino acids), β -turn motifs (four amino acids) and KIC residues (five to ten amino acids) with randomized backbone torsion angles. We chose these lengths because we found limited structural diversity for loops with lengths less than nine amino acids. When the loop length exceeded 14 amino acids, it became significantly more difficult to design buttressing interactions to stabilize the entire loop. The torsion angles of β turns were set according to the motifs sampled from the β -turn library, and the Φ/Ψ torsion angles of nonpivot KIC residues were sampled from the Ramachandran distribution, with omega torsion angles fixed at 180°. All the bond lengths were kept fixed at the ideal lengths. The position of the β -turn was randomly sampled in the loop. In each step of GeneralizedKIC, kinematic loop closure was performed to connect the loop to the intended insertion site. Loop conformations were filtered by backbone steric clashes. We further filtered the models by selecting loops with at least two intraloop backbone-to-backbone hydrogen bonds. To avoid helical conformations, we removed the models predicted to have more than five consecutive helical residues by DSSP³². This ensured the extended β -hairpin shape, which contributed to the loop stability and compatibility for buttressing.

To install the loops of the same conformation in each unit of repeat proteins, we used the RepeatPropagationMover in Rosetta^{25,26}. After filtering out the loops with steric clashes, we computed three metrics to help select the best loop conformations for buttressing—number of interloop backbone-to-backbone hydrogen bonds, loop motif score and direction score. We required at least one interloop backbone-to-backbone hydrogen bond between each pair of neighboring loops to enhance the sequence-independent loop buttressing. To select loops with loop–helix hydrophobic contacts, the motif scores

were computed by matching the selected pairs of residues to the known contacting native hydrophobic residue pairs (Val, Leu, Ile, Met and Phe) in PDB³³. The scores for the matched residue pairs in the loop regions were then summed to one total score. Only the loops with a negative total motif score were selected. The direction score described the relative orientation of the loops from the rest of the input repeat proteins. Specifically, we defined the following two vectors: vector **a** started from the center of mass of the two loop terminal residues and pointed to the farthest C α atom of the loop; vector **b** started from the same point as **a** but pointed toward the center of mass of the repeat unit. The direction score was derived by computing the angle between the two vectors.

$$\text{Direction score} = \cos^{-1} \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}| |\mathbf{b}|}$$

The accepted angles ranged from 45° to 135°. We also required at least five residues within 8 Å of the closest helical residues.

Next, we performed a fast sequence design task to identify loop conformations compatible with interloop bidentate hydrogen bond networks. From each propagated set of loops, the loop on the second repeat unit was selected for sequence design. One packing step using PackRotamersMover^{25,26} was conducted separately for each residue on this loop using amino acids that are compatible with forming sidechain-to-backbone bidentate hydrogen bonds—Asn, Asp, Gln or His. We excluded amino acids with longer side chains (Arg and Lys), as their high entropic cost might diminish the free energy contribution of buttressing. After each packing step, bidentate hydrogen bonds between the packed residue and its neighboring residues were counted. A bidentate hydrogen bond was defined as two separate hydrogen bonds forming between atoms in the functional group of the sidechain from a residue on the loop and the backbone of a neighboring repeat unit. The selected amino acid was kept only if it formed interloop bidentate hydrogen bonds; otherwise, the original amino acid (by default, Ala) was kept. In the case where the one-step packing approach failed to generate any interloop bidentate hydrogen bonds, we used an alternative three-stage scheme to maximize the sampling efficiency of bidentate hydrogen bonds—identifying pseudo-bidentate hydrogen bonds, performing constrained minimization for building hydrogen bonds and evaluating the resulting bidentate hydrogen bonds. We defined that a pseudo hydrogen bond has a donor–acceptor distance <3 Å and a hydrogen bond angle >120°. After propagating the designed residue to all the repeat units, we imposed a harmonic distance constraint between each donor and acceptor atoms with a target distance of 2 Å and a s.d. of 0.5 Å. At the minimization stage, we performed symmetric minimization of the loops to improve the interactions of potential hydrogen bonds. Finally, we used the Rosetta score function to examine if the bidentate hydrogen bonds formed in the minimized loop conformations.

To guide the sequence design, we used LayerSelector to define the core, the boundary and the surface layers and specified the allowed amino acids for each layer. We added residue type constraints to fix the identity of the residues participating in loop buttressing bidentate hydrogen bonds, so the stabilizing interactions obtained during loop sampling would be maintained throughout sequence design. Next, we performed four rounds of sequence design using the FastDesign mover under the repeat-symmetric constraints to ensure the repeat units had the same structures and sequences. To improve the solubility and folding of the designs, we subsequently performed one round of FastDesign to remove the solvent-exposed hydrophobic residues on the terminal repeat units. Only polar residues such as Glu, Gln, Lys and Arg were allowed for this round of design. The designed structures were then refined by minimization in Cartesian space and subsequently filtered by the number of buried unsatisfied heavy atoms (≤ 3), hole score normalized by total number of core residues (≤ -0.015), total score normalized by total number of residues (< -2), packstat (≥ 0.5) and hydrogen

bonding energy of each loop residue (≤ -1). Top 10% scoring structures were further tested by in silico validation methods such as molecular dynamics simulations (C α RMSD < 3 Å), AlphaFold^{12,13} (PLDDT > 80, C α RMSD < 3 Å) or RoseTTAFold³⁴ (PLDDT > 80, C α RMSD < 3 Å). Structural similarity between native ankyrin loops and the designed RBL loops was computed by TM-align³⁵.

We performed molecular dynamics simulations using GROMACS (2018.4)³⁶ with the Amber99SB-ILDN force field³⁷. The design models were solvated in dodecahedron boxes of the explicit TIP3P³⁸ waters with the net charge neutralized. We treated long-range electrostatic interactions with the Particle-Mesh Ewald method³⁹. Both short-range electrostatic interactions and van der Waals interactions used a cutoff of 10 Å. Energy minimization was performed using the steepest descent algorithm. A 1-ns equilibration under the NPT ensemble was subsequently performed with position restraints on the heavy atoms. We used Parrinello–Rahman barostat⁴⁰ and velocity-rescaling thermostat⁴¹ for pressure coupling (1 atm) and temperature coupling (310 K), respectively. For the production runs, we launched three 20-ns trajectories under the NPT ensemble for each design model. The C α atom RMSD against the design model was computed for analysis.

Protein expression and characterization

Genes encoding the in silico validated designs were synthesized (IDT) and cloned into pET-29b expression vectors. The plasmids were transformed into Lemo21 (DE3) expression *E. coli* strain (NEB). Protein expression was performed using the auto-induction protocol⁴² at 37 °C for 24 h in 50 ml or 100 ml culture. During the purification, cells were pelleted at 4,000g for 10 min and resuspended in 25 ml lysis buffer (25 mM Tris–HCl (pH = 8), 150 mM NaCl, 30 mM imidazole, 1 mM DNase and 10 mM lysozyme with Pierce Protease Inhibitor Tablets (Thermo Fisher Scientific)). Sonication was subsequently performed for 2.5 min (10 s on and 10 s off per cycle). The lysate was then centrifuged at 16,000g for 30 min. The supernatant was applied to a gravity flow column packed with Ni-NTA resin (Qiagen), followed by 20 ml wash buffer (25 mM Tris–HCl (pH = 8), 150 mM NaCl and 30 mM imidazole) and 5 ml elution buffer (25 mM Tris–HCl (pH = 8), 150 mM NaCl and 400 mM imidazole). The eluted protein was then concentrated and injected into an Akta Pure FPLC device with a flow rate of 0.75 ml min⁻¹ in the running buffer (25 mM Tris–HCl (pH = 8) and 150 mM NaCl). The typical yield of a monodisperse and thermally stable designed RBL is 1–6 g l⁻¹. To perform SEC–MALS, we prepared the purified protein at ~2 mg ml⁻¹ and injected 100 μ l of sample into a Superdex 200 10/300GL column and measured the light scattering signals using a miniDAWN TREOS device (Wyatt Technology). To measure the CD signals, we first prepared the sample at ~0.2 mg ml⁻¹ in 25 mM phosphate buffer in a 1 mm cuvette. A Jasco J-1500 CD spectrometer was used for all CD measurements. We set the range of wavelength from 190 nm to 260 nm and scanned over a three-temperature (25 °C, 95 °C and cooling back to 25 °C) set for each sample. We submitted all samples for SAXS^{43,44} to Advanced Light Source, LBNL for data collection at the SIBYLS 12.3.1 beamline.

Design and characterization of repeat peptide-binding proteins

We used the recently developed protein interface design method⁷ for in silico binder docking and design experiments. Docking of repeat peptides to the binder scaffold was guided by the geometric transformation between native ankyrins and their peptide targets in the crystal structures from PDB¹⁵. Symmetric sequence design was performed for each docked peptide–protein pair following the same protocol used for designing RBLs. All the designed complexes were computationally tested by AlphaFold with a cutoff of PAE_{interaction} ≤ 15 before experimental characterization.

Split-luciferase assay was performed using the Nano-Glo Luciferase Assay System (Promega). The coding sequence of small-BiT was fused to the gene of peptide binders, and the coding sequence

of large-BiT was fused to the coding sequence of the target peptide (GenScript). The BiT-fused proteins and peptides were expressed and purified with the same protocol for RBLs. The purified peptide binders and target peptides were titrated in the presence of Nano-Glo substrate in 96-well plates, and the luminescence was measured on a Synergy Neo2 plate reader (Agilent Technologies). To conduct the fluorescence polarization binding assays, we synthesized the repeat peptide fragments with N-terminal tetramethylrhodamine labels. Fluorescence polarization measurements were performed at 25 °C in a Synergy Neo2 plate reader (Agilent Technologies) with a 530/590 nm filter. A series of twofold dilutions of binder–peptide 80- μ l mixture were performed in 25 mM Tris–HCl (pH = 8), 150 mM NaCl and 0.05% (vol/vol) Tween 20 in 96-well assay plates. The protein concentrations ranged from 4 μ M to 0.47 pM, and the concentration of N-terminal tetramethylrhodamine-labeled peptide was kept at 0.3 nM. The samples were incubated for 3 h before measurement.

Structural characterization by X-ray crystallography

RBL4 was concentrated to 150 mg ml⁻¹ and crystallized by vapor diffusion. Initial crystals formed in the MCSG-2 crystallization screen (Anatrace) and optimized crystals were grown in 100 mM sodium acetate, pH 4.4, and 2% polyethylene glycol 4000. The crystal was cryoprotected with 30% ethylene glycol and flash-cooled in liquid nitrogen. Diffraction was measured at the Advanced Photon Source beamline 23 ID-B. Reflections were indexed, integrated and scaled with autoPROC (1.0.5)⁴⁵. The structure was solved by molecular replacement in Phaser (2.8.3)⁴⁶. Initial attempts using the predicted model were unsuccessful due to clashes. A subsequent search for eight copies of a single helix–loop–helix repeat (76–118 residues) identified two copies of the protein in the asymmetric unit. The model was rebuilt using Phenix AutoBuild (1.18.2_3874)⁴⁷ and completed by iterative rounds of interactive refinement in Coot (0.9.5)⁴⁸ and reciprocal space refinement in Phenix (1.19.1_4122)^{49–52}. The final refinement strategy included reciprocal space refinement, individual atomic displacement parameters, Translation/Libration/Screw refinement using parameters determined with TLSMD (13 June 2012)⁵³ and occupancy refinement of alternate conformations. Model geometry was assessed with MolProbity (implemented in Phenix 1.19.1_4122)⁵⁴. The final model included 99.5% of residues in the favored region of the Ramachandran plot with no outliers.

RBL7_C2_3 was concentrated to 119 mg ml⁻¹ and crystallized by vapor diffusion in 2.4 M sodium malonate, pH 7.0, using the MCSG-1 crystallization screen (Anatrace). The crystal was cryoprotected by the addition of ten volumes of 3.4 M sodium malonate, pH 7.0, and flash-cooled in liquid nitrogen. Reflections were indexed, integrated and scaled with XDS (5 February 2021)⁵⁵. To solve the structure by molecular replacement, an ensemble of monomer structures was generated by AlphaFold and used as a search ensemble in Phaser (2.8.3). The solution contained eight molecules that formed four homodimers. The model was rebuilt with Phenix AutoBuild (1.19.2_4158) with morphing and completed by iterative rounds of interactive refinement in Coot (0.9.8.6) and reciprocal space refinement in Buster (2.10.4)⁵⁶ or Phenix (1.20.1_4487). The final refinement strategy in Phenix included reciprocal space refinement, individual atomic displacement parameters, noncrystallographic symmetry restraints and Translation/Libration/Screw refinement using one group per chain. Model geometry was assessed with MolProbity (implemented in Phenix 1.20.1_4487)⁵⁴. The final model had 98.22% of residues in the favored regions of the Ramachandran plot with no outliers. Composite omit maps were generated in Phenix by sequentially omitting 5% of the final structure model and performing simulated annealing from 5,000 K. Crystallographic software was installed and maintained using SBGrid⁵⁷.

Data analysis and visualization were performed using Python (3.7)⁵⁸, seaborn (0.11.2)⁵⁹, Matplotlib (3.1.3)⁶⁰, Pandas (0.24.2)^{61,62} and PyMOL (2.4.1)⁶³.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All the design models, protein sequences and DNA sequences are available at: https://files.ipd.uw.edu/pub/2023_buttressed_loops/data.tar.gz and Zenodo⁶⁴. Crystal structures and reflection data have been deposited in the RCSB Protein Data Bank with accession IDs **8FRE** (RBL4) and **8FRF** (RBL7_C2_3). X-ray diffraction images have been deposited in the SBGrid Data Bank (**8FRE** and **8FRF**).

Code availability

The design scripts for parametric repeat protein generation and buttressed loop designs are available at https://github.com/hanlunj/buttressed_loops.git and Zenodo⁶⁴.

References

- Leaver-Fay, A. et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).
- Leman, J. K. et al. Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat. Methods* **17**, 665–680 (2020).
- Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26**, 689–691 (2010).
- Marcelino, A. M. & Gierasch, L. M. Roles of beta-turns in protein folding: from peptide models to protein engineering. *Biopolymers* **89**, 380–391 (2008).
- Venkatachalam, C. M. Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. *Biopolymers* **6**, 1425–1436 (1968).
- Wang, G. & Dunbrack, R. L. PISCES: a protein sequence culling server. *Bioinformatics* **19**, 1589–1591 (2003).
- Gonzalez, T. F. Clustering to minimize the maximum intercluster distance. *Theor. Comput. Sci.* **38**, 293–306 (1985).
- Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).
- Fallas, J. A. et al. Computational design of self-assembling cyclic protein homo-oligomers. *Nat. Chem.* **9**, 353–360 (2017).
- Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
- Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).
- Abraham, M. J. et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
- Lindorff-Larsen, K. et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950–1958 (2010).
- Neria, E., Fischer, S. & Karplus, M. Simulation of activation free energies in molecular systems. *J. Chem. Phys.* **105**, 1902–1921 (1996).
- Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
- Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
- Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
- Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
- Hura, G. L. et al. Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat. Methods* **6**, 606–612 (2009).
- Hura, G. L. et al. Comprehensive macromolecular conformations mapped by quantitative SAXS analyses. *Nat. Methods* **10**, 453–454 (2013).
- Vonrhein, C. et al. Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D* **67**, 293–302 (2011).
- McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
- Terwilliger, T. C. et al. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D* **64**, 61–69 (2008).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
- Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352–367 (2012).
- Echols, N. et al. Graphical tools for macromolecular crystallography in PHENIX. *J. Appl. Cryst.* **45**, 581–586 (2012).
- Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D* **75**, 861–877 (2019).
- Headd, J. J. et al. Flexible torsion-angle noncrystallographic symmetry restraints for improved macromolecular structure refinement. *Acta Crystallogr. D* **70**, 1346–1356 (2014).
- Painter, J. & Merritt, E. A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D* **62**, 439–450 (2006).
- Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
- Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
- Bricogne, G. et al. *BUSTER Version X.Y.Z.* (Global Phasing, 2017).
- Morin, A. et al. Cutting edge: collaboration gets the most out of software. *eLife* **2**, e01456 (2013).
- Van Rossum, G. & Drake, F. L. *Python 3 Reference Manual* (CreateSpace, 2009).
- Waskom, M. L. seaborn: statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).
- Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
- The Pandas Development Team. pandas-dev/pandas: pandas. Zenodo <https://doi.org/10.5281/zenodo.3509134> (2024).
- McKinney, W. Data structures for statistical computing in Python. In *Proc. 9th Python in Science Conference* (eds Van der Walt, S. & Millman, J.) 56–61 (SciPy, 2010).
- Schrödinger, L. L. C. The PyMOL molecular graphics system, version 1.8. *CiNii* <https://cir.nii.ac.jp/crid/1370294643858081026> (2015).
- Jiang, H. et al. Data for de novo design of buttressed loops for sculpting protein functions. Zenodo <https://doi.org/10.5281/zenodo.10999147> (2024).

Acknowledgements

We thank F. Praetorius, P. Leung and S. Vazquez for advice on fluorescence polarization assay; B. Wicky and I. Lutz for advice on split-luciferase assay; the Wysocki group at Ohio State University for the support with native mass spectrometry; ALS SIBYLS beamline at Lawrence Berkeley National Laboratory for SAXS data collection; K. VanWormer for wet lab support; L. Goldschmidt for computing support; and D. Silva, C. Xu, H. Bai, C. Norn, P. Salveson, D. Sahtoe, R. Kibler, B. Weitzner, F. DiMaio, P. Bradley, B. Stoddard, K. Lee and F.

Pardo-Avila for helpful discussions. Funding for this work is provided by the Open Philanthropy Project Improving Protein Design Fund (GF129460 to H.J. and D.B.) and the Audacious Project at the Institute for Protein Design (to K.W. and D.B.). Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under contract (DE-AC02-76SF00515 to K.M.J. and K.C.G.). The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health (NIH) and the National Institute of General Medical Sciences (P30GM133894 to K.M.J. and K.C.G.). GM/CA@APS has been funded by the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006 and P30GM138396 to K.M.J. and K.C.G.). This research used resources from the Advanced Photon Source, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract (DE-AC02-06CH11357 to K.M.J. and K.C.G.). The Eiger 16M detector at GM/CA-XSD was funded by an NIH grant (S10 OD012289 to K.M.J. and K.C.G.).

Author contributions

H.J., K.C.G. and D.B. designed the research. H.J. developed the computational loop design protocol and the parametric repeat protein design method based on the work by T.J.B., D.R.H. and H.P. H.J., J.F. and G.U. performed protein expression and purification. H.J. performed CD experiments and analyzed the results of CD and SAXS data.

A.Y. performed yeast protein display. L.C. and M.L. performed and analyzed SEC-MALS. X.L. performed and analyzed mass spectrometry analysis. K.M.J. performed X-ray crystallography and determined the structures of RBL4 and RBL7_C2_3. H.J. and K.W. designed and experimentally characterized repeat peptide-binding proteins. X.L. and P.M.L. synthesized target peptides. L.S. and D.B. supervised the research. H.J., K.M.J. and D.B. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

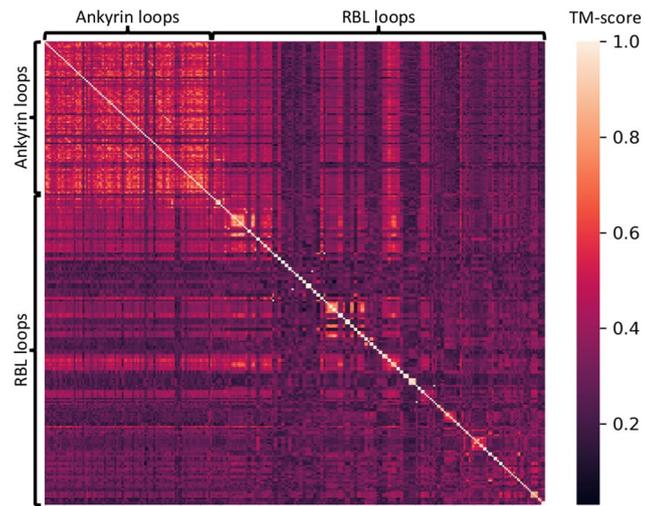
Extended data is available for this paper at <https://doi.org/10.1038/s41589-024-01632-2>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41589-024-01632-2>.

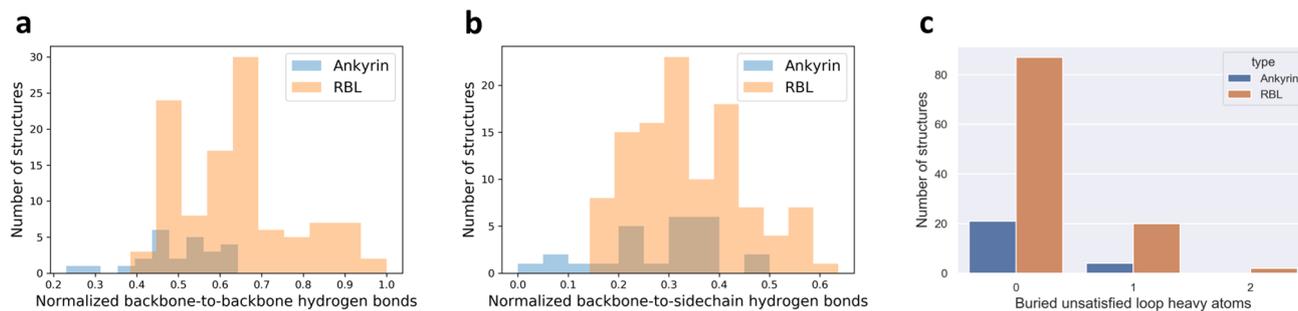
Correspondence and requests for materials should be addressed to K. Christopher Garcia or David Baker.

Peer review information *Nature Chemical Biology* thanks Kale Kundert for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

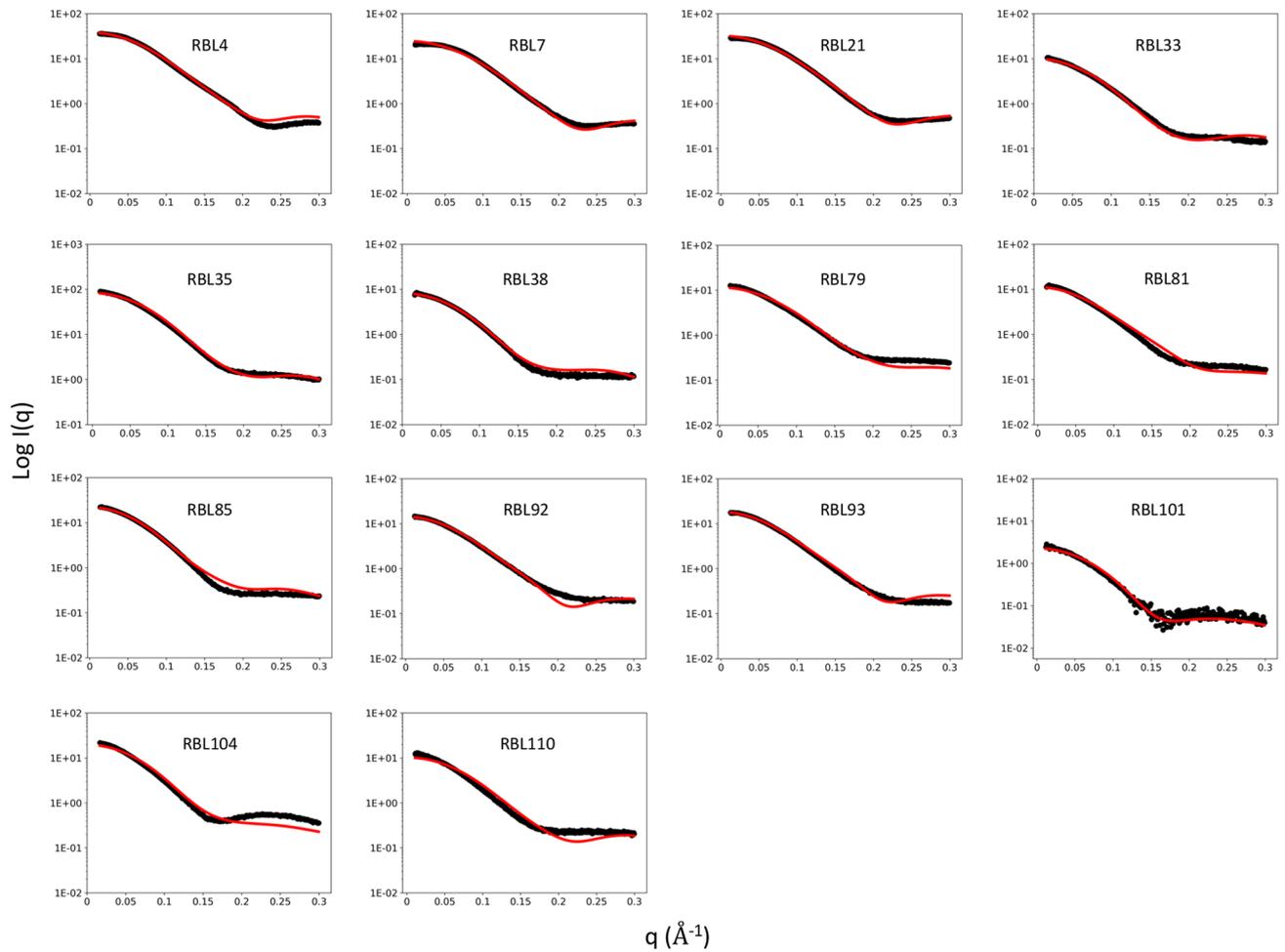


Extended Data Fig. 1 | TM scores between native ankyrin loops and the designed loops from RBLs. Higher TM scores indicate higher structural similarity.

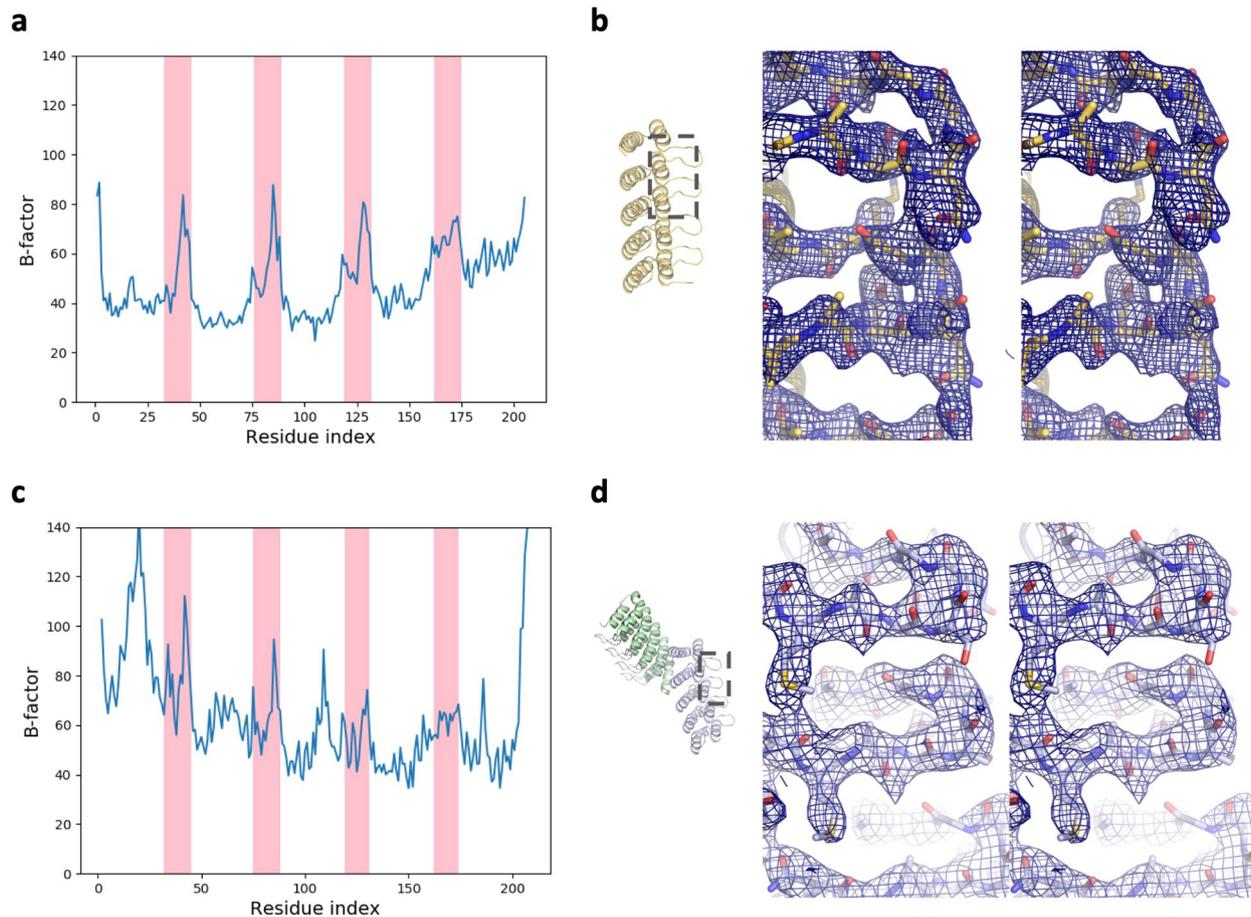


Extended Data Fig. 2 | Comparison of hydrogen bonds and buried unsatisfied loop heavy atoms between ankyrin loops and the designed loops in RBLs. (a) Distribution of the number of backbone-to-backbone hydrogen bonds involving one long loop in each structure normalized by the loop length. **(b)** Distribution of

the number of backbone-to-side chain hydrogen bonds involving one long loop in each structure normalized by the loop length. **(c)** Number of buried unsatisfied loop heavy atoms in one long loop in each structure.



Extended Data Fig. 3 | Characterization of RBLs by small-angle X-ray scattering. The experimental profiles are shown in black, and the theoretical profiles are shown in red.



Extended Data Fig. 4 | Characterization of the buttressed loops by X-ray crystallography. (a,c) Residue-wise B-factor values of the crystal structures of RBL4 (a) and RBL7_C2_3 (c). The regions corresponding to the buttressed loops are highlighted in pink. (b,d) Simulated annealing composite omit maps of RBL4

(b) and RBL7_C2_3 (d). Details of the boxed area showing cross-eyed stereo views of 2mFo-DFc electron density maps are contoured at 1σ over the designed loops. Grid spacing of the maps is $0.25\times$ the resolution of the structure.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Rosetta software suite (2019.01) (<https://www.rosettacommons.org/software>)
PyRosetta-4 (release 2019.22) (<https://www.pyrosetta.org/downloads>)
Source code and design scripts are available at: https://github.com/hanlunj/buttressed_loops.git and <https://doi.org/10.5281/zenodo.10999147>

Data analysis

autoPROC (1.0.5); Phaser (2.8.3); Coot (0.9.5 and 0.9.8.6); XDS (Feb 5, 2021); BUSTER (2.10.4); Phenix (1.19.1_4122 and 1.20.1_4487); Phenix AutoBuild (1.18.2_3874 and 1.19.2_4158); TLSMD (13 June 2012); Molprobitry (implemented in Phenix 1.19.1_4122 and 1.20.1_4487); Python (3.7); seaborn (0.11.2); matplotlib (3.1.3); pandas (0.24.2); PyMOL (2.4.1)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All the design models, protein sequences and DNA sequences are available at: https://files.ipd.uw.edu/pub/2023_buttressed_loops/data.tar.gz and Zenodo (<https://doi.org/10.5281/zenodo.10999147>). Crystal structures and reflection data have been deposited in the RCSB Protein Data Bank with accession IDs 8FRE (RBL4) and 8FRF (RBL7_C2_3). X-ray diffraction images have been deposited in the SGrid Data Bank (8FRE and 8FRF).

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We characterized in total 102 designed RBLs and 34 peptide-binding RBLs that passed our in silico filters which were used to predict the protein folding and protein-peptide interactions. These designs were of significant structural and sequence diversity.
Data exclusions	No data were excluded from analysis.
Replication	Designed proteins were expressed and purified in multiple batches. The samples used for CD, SAXS and X-ray crystallography were independently prepared.
Randomization	We did not employ randomization, as all the samples were characterized by the same procedure and there was no treatment group in our study.
Blinding	Blinding was not relevant to our study, as there was no treatment group in our study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- | n/a | Included in the study |
|-------------------------------------|--|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |

Methods

- | n/a | Included in the study |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |