

## RESEARCH ARTICLE SUMMARY

## PROTEIN DESIGN

## Reconfigurable asymmetric protein assemblies through implicit negative design

Danny D. Sahtoe<sup>†</sup>, Florian Praetorius<sup>†</sup>, Alexis Courbet, Yang Hsia, Basile I. M. Wicky, Natasha I. Edman, Lauren M. Miller, Bart J. R. Timmermans, Justin Decarreau, Hana M. Morris, Alex Kang, Asim K. Bera, David Baker<sup>\*</sup>

**INTRODUCTION:** For many current challenges in synthetic biology, it would be desirable to have bio-orthogonal and modular sets of interacting proteins that are folded and soluble when alone but rapidly and specifically associate when mixed. Although pairs of proteins with these properties are found in nature, with the exception of single-helix coiled-coil peptides that are not folded in isolation, it has been very challenging to generate new bio-orthogonal pairs by design. This is because designed, largely nonpolar interfaces that drive association between two different chains can also mediate self-association of individual chains into large oligomers or aggregates that disassociate very slowly. For example, designed heterodimeric

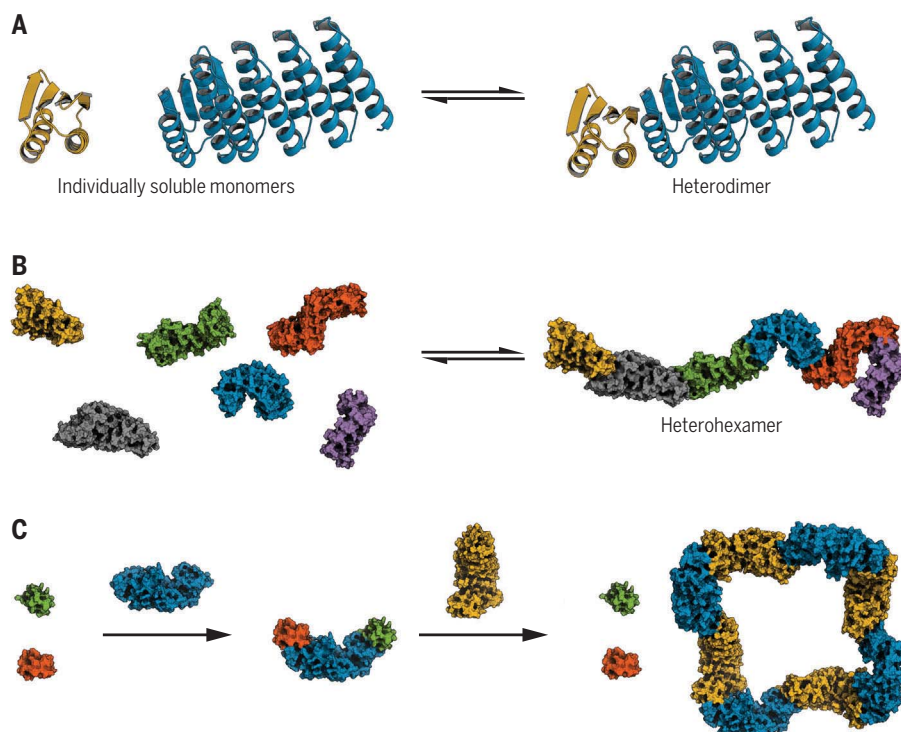
helical bundles with specific and orthogonal interfaces do not readily assemble from individually purified monomers but require coexpression, thermal or chemical denaturation, or incubation for more than a week for assembly from individually prepared components.

**RATIONALE:** We sought to design heterodimers that (i) spontaneously assemble upon mixing of stable and soluble individual components, (ii) allow dynamic exchange of components, and (iii) are amenable to rigid fusion to enable the assembly of higher-order hetero-oligomeric complexes with defined structures. We reasoned that these properties could be achieved with rigid designed protomers with exposed  $\beta$  sheet

edges that associate to form a continuous  $\beta$  sheet in the heterodimer (see the figure): Any off-target interaction that does not allow for strand-pair formation should be highly unfavorable because of the high thermodynamic cost of the burial of backbone polar groups away from water in the absence of compensating hydrogen-bonding interactions, thus reducing the probability of undesired homo-oligomerization.

**RESULTS:** We designed 12  $\alpha$ - $\beta$  heterodimers that readily assemble from individually expressed and purified monomers. By rigidly fusing two different protomers to designed helical repeat proteins, we generated bivalent connector proteins that can bind two different partners. Using one or more of these bivalent connectors, we successfully assembled linear heterotrimers, heterotetramers, heteropentamers, and a heterohexamer with distinct shapes (see the figure). We further assembled branched heterotetramers using trivalent connectors that can bind three different partners in defined orientations. By rigidly fusing our protomers to previously designed homooligomers, we created hubs that can bind three or four copies of the same binding partner. Using symmetry-aware helical fusion, we designed a closed C4-symmetric ring. All of these hetero-oligomeric complexes readily assemble from individually prepared components. The components function as designed in living cells, mediating the assembly of liquid-liquid condensates or more static aggregates depending on the interaction affinities, and designed assemblies can be reconfigured by addition of components providing access to lower free-energy states (see the figure).

**CONCLUSION:** Our reversible heterodimeric assemblies open the door to many exciting new synthetic biology and nanomaterial applications. Bivalent connectors can be used to induce interaction between otherwise monomeric proteins to modulate biological function, and symmetric hubs can present multiple copies of ligands or antigens to cluster cell-surface receptors. Bio-orthogonal signaling systems can be constructed by using the heterodimer components in synthetic receptors and ligands. Reconfigurable higher-order nanomaterials—one-dimensional (1D) fibers, 2D lattices, and 3D nanocages and crystals—can be created using our designed components to drive geometrically precise association between the material components. ■



**Schematic representation of reconfigurable protein assemblies.** (A and B) Formation of a heterodimer (A) or heterohexamer (B) from monomers that are stable and soluble in isolation. (C) Two components A (orange) and C (green) are monomeric and do not interact. Addition of a bivalent connector B (blue) brings them in close proximity. Subsequent addition of component B' (yellow) leads to the formation of a symmetric B<sub>4</sub>B'<sub>4</sub> ring and release of free A and C.

The list of author affiliations is available in the full article online.

<sup>\*</sup>Corresponding author. Email: [dabaker@uw.edu](mailto:dabaker@uw.edu)

<sup>†</sup>These authors contributed equally to this work.

Cite this article as D. D. Sahtoe *et al.*, *Science* **375**, eabj7662 (2022). DOI: [10.1126/science.abj7662](https://doi.org/10.1126/science.abj7662)

**S READ THE FULL ARTICLE AT**  
<https://doi.org/10.1126/science.abj7662>

## RESEARCH ARTICLE

## PROTEIN DESIGN

## Reconfigurable asymmetric protein assemblies through implicit negative design

Danny D. Sahtoe<sup>1,2,3†</sup>, Florian Praetorius<sup>1,2†</sup>, Alexis Courbet<sup>1,2,3</sup>, Yang Hsia<sup>1,2</sup>, Basile I. M. Wicky<sup>1,2</sup>, Natasha I. Edman<sup>1,2,4,5</sup>, Lauren M. Miller<sup>1,2</sup>, Bart J. R. Timmermans<sup>1,2</sup>, Justin Decarreau<sup>1,2</sup>, Hana M. Morris<sup>1,2</sup>, Alex Kang<sup>1,2</sup>, Asim K. Bera<sup>1,2</sup>, David Baker<sup>1,2,3\*</sup>

Asymmetric multiprotein complexes that undergo subunit exchange play central roles in biology but present a challenge for design because the components must not only contain interfaces that enable reversible association but also be stable and well behaved in isolation. We use implicit negative design to generate  $\beta$  sheet-mediated heterodimers that can be assembled into a wide variety of complexes. The designs are stable, folded, and soluble in isolation and rapidly assemble upon mixing, and crystal structures are close to the computational models. We construct linearly arranged hetero-oligomers with up to six different components, branched hetero-oligomers, closed C4-symmetric two-component rings, and hetero-oligomers assembled on a cyclic homo-oligomeric central hub and demonstrate that such complexes can readily reconfigure through subunit exchange. Our approach provides a general route to designing asymmetric reconfigurable protein systems.

**D**ynamic reconfigurable multiprotein complexes play key roles in central biological processes (1). The subunits are generally monomeric in isolation, allowing the assemblies to reconfigure by successive addition or removal of one or more components. Such modulation is essential to their function; for example, subunit loss and addition underlie the molecular mechanisms of protein complexes that drive DNA replication and transcription (2, 3). The ability to de novo design such multicomponent reconfigurable protein assemblies would enable the realization of sophisticated new functions. Previous design efforts have generated cyclic oligomeric and higher-order symmetric nanostructures such as icosahedral nanocages with as many as 120 subunits and two-dimensional (2D)-layers with many thousands of regularly arrayed components (4–8). Essential to this is the symmetry and cooperativity of assembly, which strongly favors just one of a large number of possible states. Once formed, these assemblies are therefore typically quite static and exchange subunits only on long time scales, which is advantageous for applications such as nanoparticle vaccine design and multivalent receptor engagement (9).

The design of reconfigurable asymmetric assemblies is more challenging, because there

is no symmetry “bonus” favoring the target structure (as is attained, for example, in the closing of an icosahedral cage) and because the individual subunits must be stable and soluble in isolation in order to reversibly associate. Reconfigurable asymmetric protein assemblies could, in principle, be constructed using a modular set of protein-protein interaction pairs (heterodimers), provided that first, the individual subunits are stable and monomeric in isolation so that they can be added and removed; second, the interacting pairs are specific; and third, they can be rigidly fused through structured connectors to other components. Rigid fusion, as opposed to fusion by flexible linkers, is important to program the assembly of structurally well-defined complexes; most higher-order natural protein complexes have, despite their reconfigurability, distinct overall shapes that are critical for their function. Although there are designed orthogonal sets of interacting proteins that have one of these properties, designed proteins that have all of these properties are lacking. The components of designed helical-hairpin heterodimers (10, 11) on their own form homodimers or other higher-order homomeric aggregates that disassemble on very long time scales (10, 12), making them unsuitable for use in constructing reconfigurable higher-order assemblies. Heterodimeric coiled coils assemble from peptides that are soluble and monomeric, but the monomers are unfolded before binding their partners (13, 14), complicating their use in structurally defined rigid fusions.

We set out to design sets of interacting protein pairs for constructing reconfigurable assemblies (Fig. 1A). The first challenge is the

systematic design of proteins with interaction surfaces that drive association with cognate partners but not self-association. Hydrophobic interactions drive protein complex assembly, but these same hydrophobic interactions can also promote homomerization. Previously designed heterodimeric helical bundles featured, in addition to hydrophobic interactions, explicit hydrogen-bond networks that contribute to binding specificity and make the interface more polar. However, the individual protomers, either helical hairpins or individual helices, lack a hydrophobic core and are thus flexible and unstable as monomers, allowing a wide range of potential off-target homo-oligomers to form (Fig. 1B). Explicit negative-design methods favor one state by considering the effect of amino acid substitutions on the free energies of both states (15–17). However, such methods cannot be readily applied to disfavor self-association, because there are, in general, a large number of possible self-associated states that cannot be systematically enumerated.

We instead sought to use implicit negative design (18) by introducing three properties that collectively make self-associated states unlikely to have low free energy: First, in contrast to the flexible coiled coils and helical hairpins used in previous designs, we aimed for well-folded individual protomers stabilized by substantial hydrophobic cores; this property limits the formation of slowly exchanging homo-oligomers (Fig. 1B). Second, we constructed interfaces in which each protomer has a mixed  $\alpha$ - $\beta$  topology and contributes one exposed  $\beta$  strand to the interface, giving rise to a continuous  $\beta$  sheet across the heterodimer interface (19–21) (Fig. 1C). The exposed polar backbone atoms of this “edge strand” limit self-association to arrangements that pair the  $\beta$  edge strands; most other homomeric arrangements are unlikely because they result in the energetically unfavorable burial of the polar backbone atoms on the  $\beta$  edge strand (Fig. 1C). Third, taking advantage of the restrictions in possible undesired states resulting from the two properties noted above, we explicitly modeled the limited number of homo-oligomeric states and designed in additional elements that were likely to sterically occlude such states (Fig. 1D).

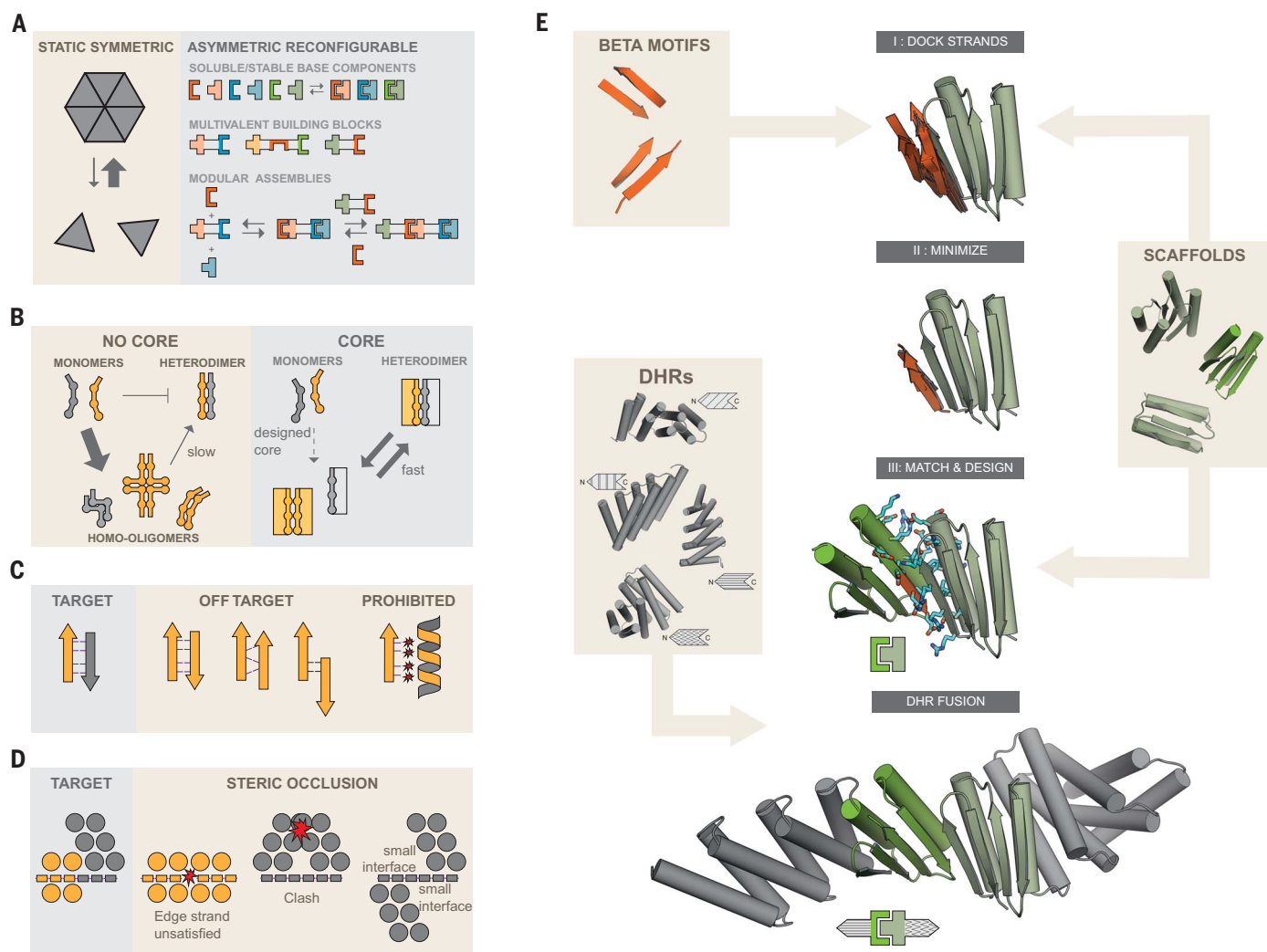
## Results

To implement these properties, we chose to start with a set of mixed  $\alpha$ - $\beta$  scaffolds that were designed by Foldit players (22). The selected designs contain sizable hydrophobic cores, exposed edge strands required for  $\beta$  sheet extension (19), and one terminal helix (either N or C) available for rigid helical fusion (Fig. 1E) (23). Using blueprint-based backbone building (24, 25), we designed additional helices at the other terminus for a subset of the scaffolds to

<sup>1</sup>Department of Biochemistry, University of Washington, Seattle, WA 98195, USA. <sup>2</sup>Institute for Protein Design, University of Washington, Seattle, WA 98195, USA. <sup>3</sup>Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA. <sup>4</sup>Molecular and Cellular Biology Graduate Program, University of Washington, Seattle, WA 98195, USA. <sup>5</sup>Medical Scientist Training Program, University of Washington, Seattle, WA 98195, USA.

\*Corresponding author. Email: dabaker@uw.edu

†These authors contributed equally to this work.



**Fig. 1. Strategies for the design of asymmetric hetero-oligomeric complexes.**

(A) Many design efforts have focused on cooperatively assembling symmetric complexes (left) with little subunit exchange. In this study, we sought to create asymmetric hetero-oligomers from stable heterodimeric building blocks that can modularly exchange subunits (right). Design strategies for preventing subunit self-association are illustrated in the following panels. (B) Protomers that have a substantial hydrophobic core (right rectangles) are less likely to form stable homo-oligomers than protomers of previously designed heterodimers that lack hydrophobic monomer cores. (C) In  $\beta$  sheet extended interfaces, most homodimer states that bury non-hydrogen-bonding polar edge-strand atoms are energetically inaccessible. Potential homodimers are more likely to form by  $\beta$  sheet extension. These are restricted to only two orientations (parallel and antiparallel) and a limited

number of offset registers. Arrows and ribbons represent strands and helices, respectively; thin lines indicate hydrogen bonds; and red stars indicate unsatisfied polar groups. (D) By modeling the limited number of  $\beta$  sheet homodimers across the  $\beta$  edge strand, structural elements may be designed that specifically block homodimer formation or make it unlikely because of small interfaces but still allow heterodimer formation. Circles indicate helices, rectangles indicate  $\beta$  strands, and stars indicate steric clashes. (E) To design reversible heterodimers,  $\beta$  strands are docked to the edge strands of hydrophobic core-containing protein scaffolds [in this paper, from Foldit (22)], a second scaffold is superimposed on the docked strand creating a protein-protein complex, the amino acids at the protein-protein interface are optimized for high affinity and specificity binding, and finally DHRs are fused to the terminal helices.

enable rigid fusion at both the N and C termini (fig. S1). Heterodimers with paired  $\beta$  strands across the interface were generated by superimposing one of the two strands from each of a series of paired  $\beta$  strand templates onto an edge  $\beta$  strand of each scaffold (Fig. 1E, top) and then optimizing the rigid body orientation and the internal geometry of the partner  $\beta$  strand of the template to maximize hydrogen-bonding interactions across the interface

(Fig. 1E, second row). This generates a series of disembodied  $\beta$  strands that form an extended  $\beta$  sheet for each scaffold; for each of these, an edge  $\beta$  strand from a second scaffold was superimposed on the disembodied  $\beta$  strand to form an extended  $\beta$  sheet (Fig. 1E, third row). The interface side chain-side chain interactions in the resulting protein-protein docks were optimized using Rosetta combinatorial sequence design (26). To limit excessive hydrophobic inter-

actions, we generated explicit hydrogen-bond networks across the heterodimer interface (17) or constrained the amino acid composition to favor polar residues while penalizing buried unsatisfied polar groups (27). This resulted in interfaces that, outside of the polar hydrogen bonding of the  $\beta$  strands, contained both hydrophobic interactions and polar networks. To further disfavor unwanted homodimeric interactions (Fig. 1D, right), we rigidly fused



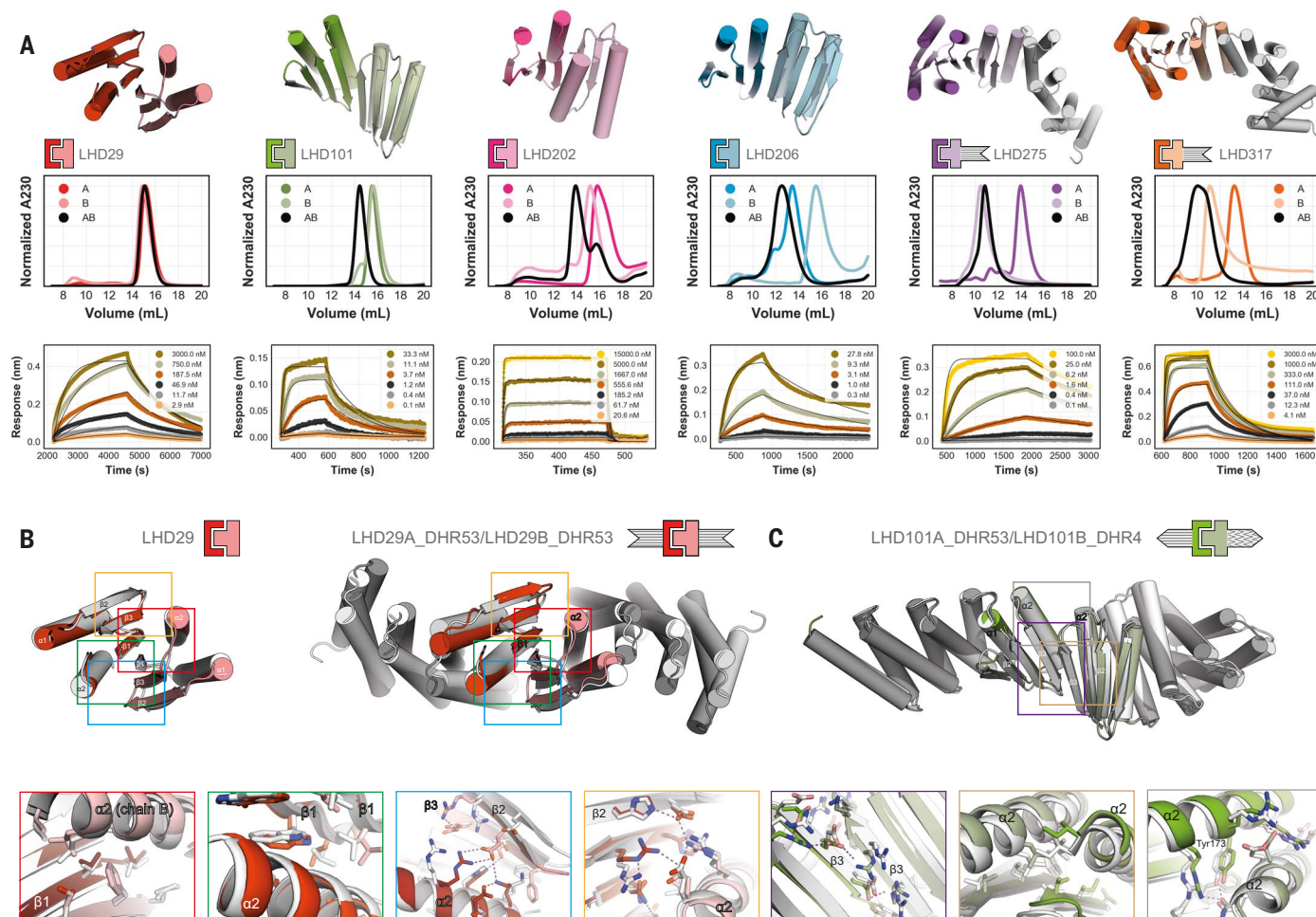
designed helical repeat proteins (DHRs) to terminal helices (23, 28). Because these DHRs have different shapes, they also serve to diversify building-block shapes for subsequent higher-order assembly design. Designed heterodimers were selected for experimental characterization on the basis of binding energy, the number of buried unsatisfied polar groups, buried surface area, and shape complementarity (see Materials and methods).

We coexpressed the selected heterodimers in *Escherichia coli* using a bicistronic expression system that encodes one of the two protomers with a C-terminal polyhistidine tag and the other protomer either with no tag or with a green fluorescent protein (GFP) tag at the N terminus. Complex formation was initially assessed using nickel affinity chromatography; designs for which both protomers were present in SDS-polyacrylamide gel electrophoresis (PAGE) after nickel pulldown were

subjected to size exclusion chromatography (SEC) and liquid chromatography–mass spectrometry (LC-MS). Of the 238 tested designs, 71 passed the bicistronic screen and were selected for individual expression of protomers. Of these, 32 formed heterodimers from individually purified monomers as confirmed by SEC, native MS, or both (Fig. 2A and figs. S2 and S3A). In SEC titration experiments, some protomers were monomeric at all injection concentrations, whereas others self-associated at higher concentrations (fig. S4). Both LHD101 protomers and their fusions were monomeric even at injection concentrations greater than 100  $\mu$ M (fig. S4). LHD275A, LHD278A, LHD317A, and a redesigned version of LHD29 with a more polar interface (LHD274) were also predominantly monomeric (figs. S4 and S5). Designs for which isolated protomers were poorly expressed, polydispersed in SEC, or did not yield stable, soluble, and functional

rigid DHR fusions were discarded along with designs that were very similar to other designs but otherwise stable and soluble. The remaining 11 heterodimers span three main structural classes [Fig. 2A, fig. S2, and data S1 (LHD components)]. In class one, the central extended  $\beta$  sheet is buttressed on opposite sides by helices that contribute additional interface interactions (LHD29 and LHD202 in Fig. 2A); in class two, the helices that provide additional interactions are on the same side of the extended central sheet (LHD101 and LHD206 in Fig. 2A); and in class three, both sides of the central  $\beta$  sheet extension are flanked by helices (LHD275 and LHD317 in Fig. 2A).

We monitored the kinetics of heterodimer formation and dissociation through biolayer interferometry (BLI) (Fig. 2A, fig. S2, and table S1) by immobilizing individual biotinylated protomers onto streptavidin-coated sensors and adding the designed binding partner.



**Fig. 2. Designed heterodimer characterization.** (A) Characterization of six designed heterodimers. Design models are shown in the top row; the color scheme for the different designs is maintained throughout the paper. Normalized SEC traces of individual protomers (A and B) and complexes (AB) are shown in the middle row and kinetic binding traces with global kinetic fits of in vitro biolayer interferometry binding assays in the bottom row. (B and C) Crystal structures (in colors) of the designs LHD29, LHD29A53/B53, and LHD101A53/B4 overlaid on design models (light gray). Colored rectangles in the full models (top row) match the corresponding detailed views (bottom row). Sequences and models for all proteins are provided in data S1.

Unlike previously designed heterodimers, binding reactions equilibrated rapidly, with affinities ranging from micromolar to low nanomolar (fig. S3C and table S1). Association rates were quite fast and ranged from  $10^6 \text{ M}^{-1} \text{ s}^{-1}$  for the fastest heterodimer to  $10^2 \text{ M}^{-1} \text{ s}^{-1}$  for the slowest heterodimer LHD29, which is still an order of magnitude faster than the fastest associating designed helical hairpin heterodimer DHD37 (10) (Fig. 2A, fig. S6A, and tables S1 and S2). For LHD101 and LHD206, we independently determined the equilibrium dissociation constant ( $K_d$ ) with a split luciferase-based binding assay in *E. coli* lysates and obtained very similar values, indicating that heterodimer association is not affected by high concentrations of noncognate proteins (fig. S6, D and E, and table S3).

We determined the crystal structures of two class-one designs, LHD29 (2.2 Å) and LHD29A53/B53 (2.6 Å) in which both protomers are fused to DHR53 (Fig. 2B and table S4). In the central extended  $\beta$  sheet, the LHD29 design closely matches the crystal structure (red and green boxes in Fig. 2B and table S5). Aside from backbone  $\beta$  sheet hydrogen bonds, this part of the interface is supported by primarily hydrophobic packing interactions between the side chains of each interface  $\beta$  edge strand. The two flanking helices on opposite sides of the central  $\beta$  sheet (blue and orange boxes in Fig. 2B) contribute predominantly polar contacts to the interface and are also similar in the crystal structure and design model. Apart from crystal contact-induced subtle backbone rearrangements in strand two of LHD29B that promote the formation of a polar interaction network (blue box in Fig. 2B), most interface side chain-side chain interactions agree with the design model. As for unfused LHD29, the interface of LHD29A53/B53 resembles the designed model; at the fusion junction and repeat protein regions, deviations are slightly larger (table S5).

We also determined the structure of a class-two design, LHD101A53/B4 (2.2 Å), in which protomer A is fused to DHR53 and B to DHR4 (Fig. 2C and tables S4 and S5). The crystal structure agrees well with the design model at both the interface and fusion junctions, as well as the repeat protein regions. In class-two designs, the interface  $\beta$  strand pair is reinforced by flanking helices that, unlike in class-one designs, are in direct contact with both each other and the interface  $\beta$  sheet. The solvent-exposed side of the  $\beta$  interface consists primarily of electrostatic interactions (purple box in Fig. 2C), whereas the buried side consists exclusively of hydrophobic side chains. Together with apolar side chains on the flanking helices of both protomers, these residues form a closely packed core interface (brown box in Fig. 2C) that is further stabilized by solvent-exposed polar interactions between

the flanking helices. Notably, the designed semiburied polar interaction network centered on Tyr<sup>173</sup> is recapitulated in the crystal structure (gray box in Fig. 2C).

As described above, the third of our implicit negative-design principles was to incorporate structural elements incompatible with  $\beta$  sheet extension in homodimeric species (Fig. 1D). To assess the utility of this principle, we took advantage of the limited number of possible off-target edge-strand interactions that can form (Fig. 1C); we docked all protomers against themselves on the edge strand that participates in the heterodimer interface and calculated the Rosetta binding energy after relaxation of the resulting homodimeric dock (fig. S7). Homodimer docks of the protomers that chromatographed as monomers in SEC had unfavorable energies compared with those that showed evidence of self-association in agreement with our initial hypothesis (Fig. 1D), and visual inspection of these docks suggested that homodimerization was likely prevented by the presence of sterically blocking secondary-structure elements (fig. S7).

Twenty-eight additional rigid fusion proteins that were generated using the 11 base heterodimers and LHD274 (Fig. 3A) retained both the oligomeric state and binding activity of the unfused counterparts, indicating that the designed heterodimers are quite robust to fusion (figs. S3D, S6E, and S8). There are 74 different possible heterodimeric complexes that can be assembled from these fusions, each with different shapes. Most of the fusions involve protomers of LHD274 and LHD101; fusions to LHD101 protomers alone enable the formation of 30 distinct heterodimeric complexes (fig. S9).

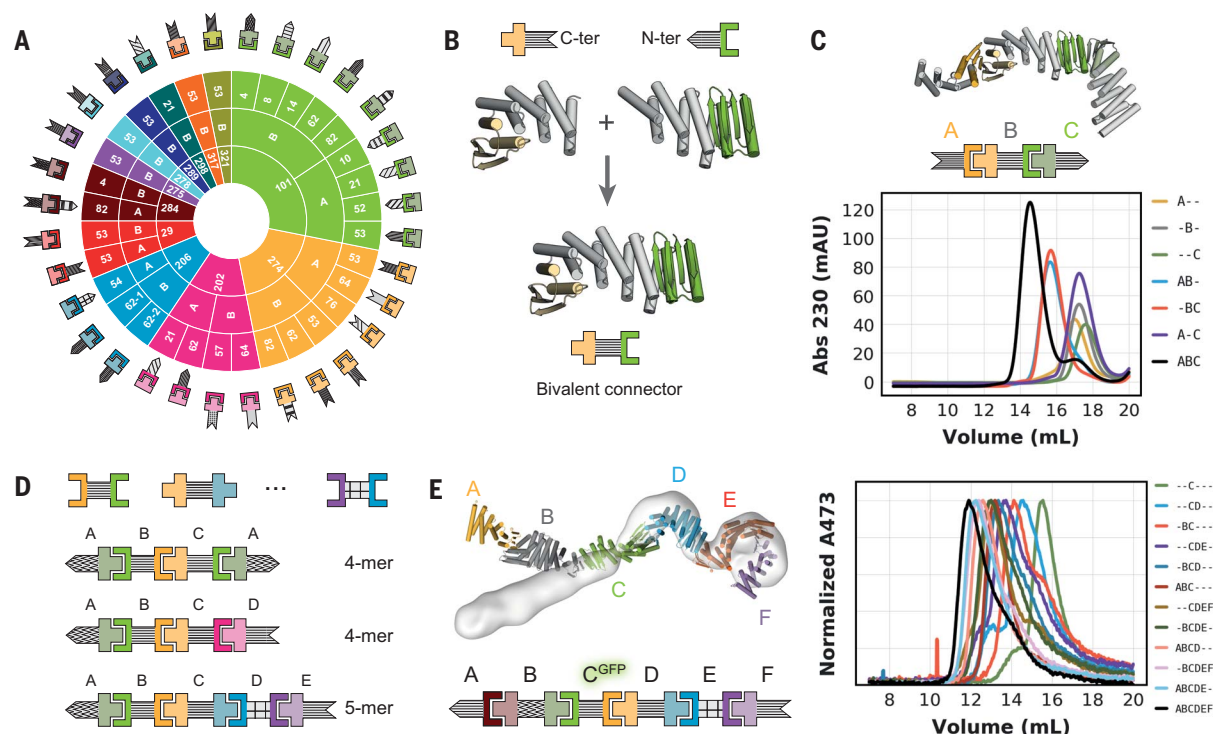
Larger multicomponent hetero-oligomeric protein assemblies require subunits that can interact with more than one binding partner at the same time. To this end, we generated single-chain bivalent connector proteins. Designed protomers that share the same DHR as the fusion partner and have compatible termini can be simply spliced together into a single protein chain on overlapping DHR repeats (Fig. 3B). Mixing a linear connector ("B") with its two cognate binding partners ("A" and "C") yields a linearly arranged heterotrimer ("ABC") in which the two terminal capping components A and C are connected through component B but otherwise are not in direct contact with each other (Fig. 3C). We analyzed the assembly of this heterotrimer and controls by SEC (Fig. 3C) and observed stepwise assembly of the ABC heterotrimer with clear baseline separation from AB and BC heterodimers, as well as from monomeric components (Fig. 3C). Using experimentally validated linear connectors created using the above-described modular splicing approach (Fig. 3D, fig. S10A, and data S1), we assembled 20 heterotrimers in total,

including one verified by negative-stain electron microscopy (nsEM) (figs. S10B and S11). The absence of off-target complexes in these assemblies corroborates the orthogonality of the heterodimer interfaces (fig. S12).

By using more than one connector subunit, larger linear hetero-oligomers can be generated. We constructed and confirmed assembly of ABCA and ABCD heterotetramers, each containing two different linear connectors (B and C) and either one or two terminal caps (two A or A and D), an ABBA heterotetramer using a homodimeric central connector (two B) and one terminal cap (two A), and a nsEM-verified heteropentamer (ABCDE) containing three different linear connectors and two caps (Fig. 3D and figs. S13 and S14). We followed the assembly of an ABCDEF heterohexamer in SEC by GFP-tagging one of the components and monitoring GFP absorbance. The full assembly, as well as subassemblies generated as controls, eluted as monodisperse peaks, with elution volumes agreeing well with expected assembly sizes (Fig. 3E). nsEM reconstruction of the hexamer confirmed that all components were present (Fig. 3E and fig. S15A). Deviation of the experimentally observed shape from the design model likely arises from small deviations from the model in one of the components that cause a lever-arm effect (Fig. 2B).

In total, by combining the bivalent connectors with each other and with monovalent terminal caps, we constructed 36 hetero-oligomers with up to six different chains and confirmed their assembly by SEC and EM [Fig. 3, C and E; figs. S10, S11, S13, and S15; and data S1 (experimentally validated assemblies)]. This number can be readily increased to 489 by including all available components [Fig. 3A, fig. S10A, and data S1 (all\_theoretical\_assemblies)]. Because all fusions have structured helical linkers, the overall molecular shapes of the complexes and the spatial arrangement of individual components are well defined, which should be useful for scaffolding and other applications. Our linear assemblies resemble elongated modular multiprotein complexes found in nature (fig. S15B), like the Cullin RING E3 ligases (29) that mediate ubiquitin transfer by geometrically orienting the target protein and catalytic domain.

We next sought to go beyond linear assemblies and build branched and closed assemblies. Trivalent connectors can be generated from heterodimers in which one protomer has both N- and C-terminal helices (LHD275A, LHD278A, LHD289A, and LHD317A). Such protomers can be fused to two helical repeat proteins and spliced together with different halves of other heterodimer protomers via a common DHR repeat (Figs. 3, A and B, and 4A). The resulting branched trivalent connectors ("A") are capable of binding the three cognate binding partners ("B," "C," and "D")



**Fig. 3. Design of higher-order assemblies.** (A) Schematic overview of experimentally validated heterodimer-DHR fusions. In the colored circle at the center, the inner ring represents the heterodimer, the middle ring the protomer chain that is fused, and the outer ring the DHR (28) fusion partner. In the design model cartoons outside the colored circle, the patterning of the DHRs (in gray) is consistent throughout the paper. (B) Schematic representation of the design-free alignment method used to generate bivalent connectors from heterodimer-DHR fusions. Shown are LHD274B fused to the N terminus of DHR53 (274B53) (top left), LHD101A fused to the C terminus of DHR53 (101A53) (top right), and bivalent connector FDBO (bottom). (C) Representations of a heterotrimer (top) comprising the bivalent connector in (B) (“B”) and two of the rigid fusions shown in (A) (“A” is 274A53 and “C” is 101B62) and SEC traces for all possible

combinations of the trimer components (bottom). Abs 230, absorbance at 230 nm; mAU, milli-arbitrary units. **(D)** Schematic representations of three examples of bivalent connectors (see fig. S10A for a full list) and experimentally validated higher-order assemblies (see figs. S10 and S11). **(E)** Shown on the left is an overlay of the heterohexamer design model (in colors) and nsEM density (light gray). Shown on the right are SEC traces of partial and full mixtures of the hexamer components ("A" is 284A82, "B" is DF284, "C" is DFA-GFP, "D" is DF206, "E" is DF275A, and "F" is 275B). Absorbance was monitored at 473 nm to follow the GFP-tagged component C. Sequences, models, and chain-to-construct mapping are provided in data S1, affinities of individual interactions in tables S1 and S3, and the mapping of schemes to names for individual components in fig. S25.

simultaneously and conceptually resemble Ste5 and related scaffolding proteins that organize mitogen-activated protein (MAP) kinase signal transduction pathways in eukaryotes (30). Through SEC analyses, we verified the assembly of two different tetrameric branched ABCD complexes, each containing one trivalent branched connector bound to three terminal caps (Fig. 4A and fig. S16). For one of these, the complex was confirmed by nsEM class averages and 3D reconstructions, which indicate not only that all binding partners are present but also that the shape closely matches the designed model (Fig. 4A and fig. S16A).

A different type of branched assemblies are “star shaped” oligomers with cyclic symmetries, akin to natural assemblies formed by immunoglobulin M (IgM) and the inflammasome (31, 32). Using the alignment approach described above (Fig. 3B), we fused our building blocks (Fig. 3A) to previously designed homooligomers (23, 33) that terminate in helical

repeat proteins (Fig. 4, B and C). Such fusions yield central homo-oligomeric hubs (“A<sub>*n*</sub>”) that can bind multiple copies of the same binding partner (“*n*\*B”). We generated C3- and C4-symmetric “hubs” that can bind three or four copies of their binding partners, respectively (Fig. 4, B and C). In both cases, the oligomeric hubs are stable and soluble in isolation and readily form the target complexes when mixed with their binding partners, as confirmed by SEC, nsEM class averages, and 3D reconstructions (Fig. 4, B and C, and figs. S17 to S19). For the C4-symmetric hub, in the absence of its binding partner, we observed an additional concentration-dependent peak on SEC (Fig. 4C and figs. S18A and S19A), indicating formation of a higher-order complex. This is likely a dimer of C4 hubs, because the C4 hub contains the redesigned protomer LHD274B that, despite its reduced homodimerization propensity compared with parent design LHD29B, still weakly homodimerizes (fig. S5). Addition of

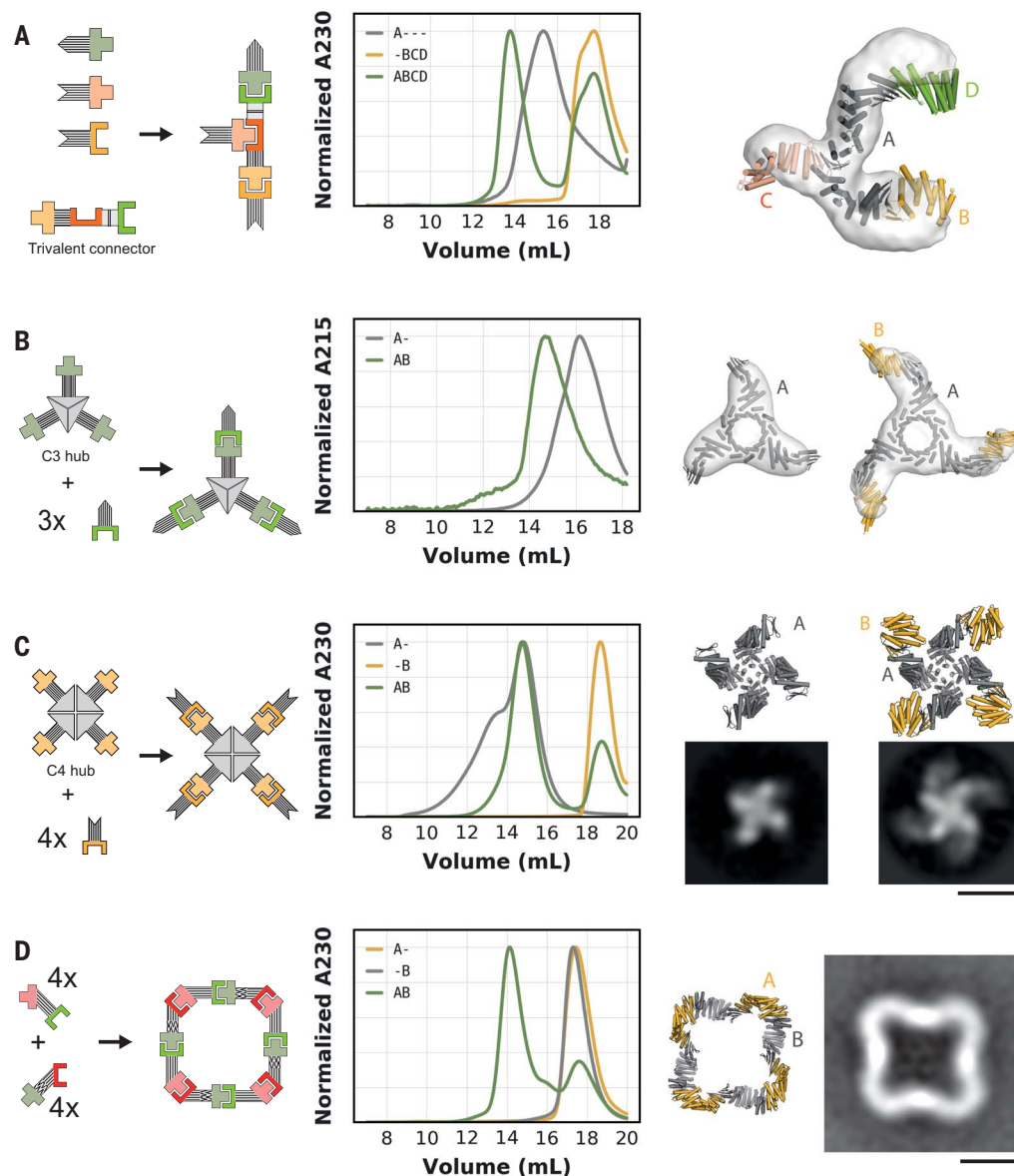
the binding partner drives reconfiguration of this higher-order assembly into the on-target octameric ( $A_4B_4$ ) complex (Fig. 4C).

In addition to linear and branched assemblies, we designed closed symmetric two-component assemblies. Designing these presents a more complex geometric challenge, because the interaction geometry of all pairs of subunits must be compatible with a single closed 3D structure of the entire assembly. We used architecture-aware rigid helical fusion (7, 34) to generate two bivalent connector proteins from the crystal-verified fusions of LHD29 and LD101 (Fig. 2B) that allow assembly of a perfectly closed C4-symmetric hetero-oligomeric two-component ring (Fig. 4D). Individually expressed and purified components are stable and soluble monomers in isolation, as confirmed by SEC, multiangle light scattering (MALS) and native MS (Fig. 4D and fig. S20). Upon mixing, the components form a higher-order complex that, by native MS and MALS, comprises four copies of each component.



**Fig. 4. Design of branched and closed hetero-oligomeric assemblies.**

Schematic depictions of designs are shown in the first column, SEC binding data in the second column, and designed models in colors overlayed on a nsEM reconstruction in the third column. **(A)** A trivalent connector ("A" is TF10) binds three different binding partners ("B" is 274A53, "C" is 317B, and "D" is 101B62). **(B)** The C3-symmetric hub presenting three copies of LHD101B (chain A) binds three copies of its binding partner ("B" is 101A53). **(C)** The C4-symmetric hub presenting four copies of LHD274B binds its cognate binding partner (274A53). Representative nsEM class averages are shown on the right. **(D)** C4-symmetric closed ring comprising two components (A and B) assemblies from constituent components. A representative nsEM class average is shown on the right. Scale bars, 10 nm.

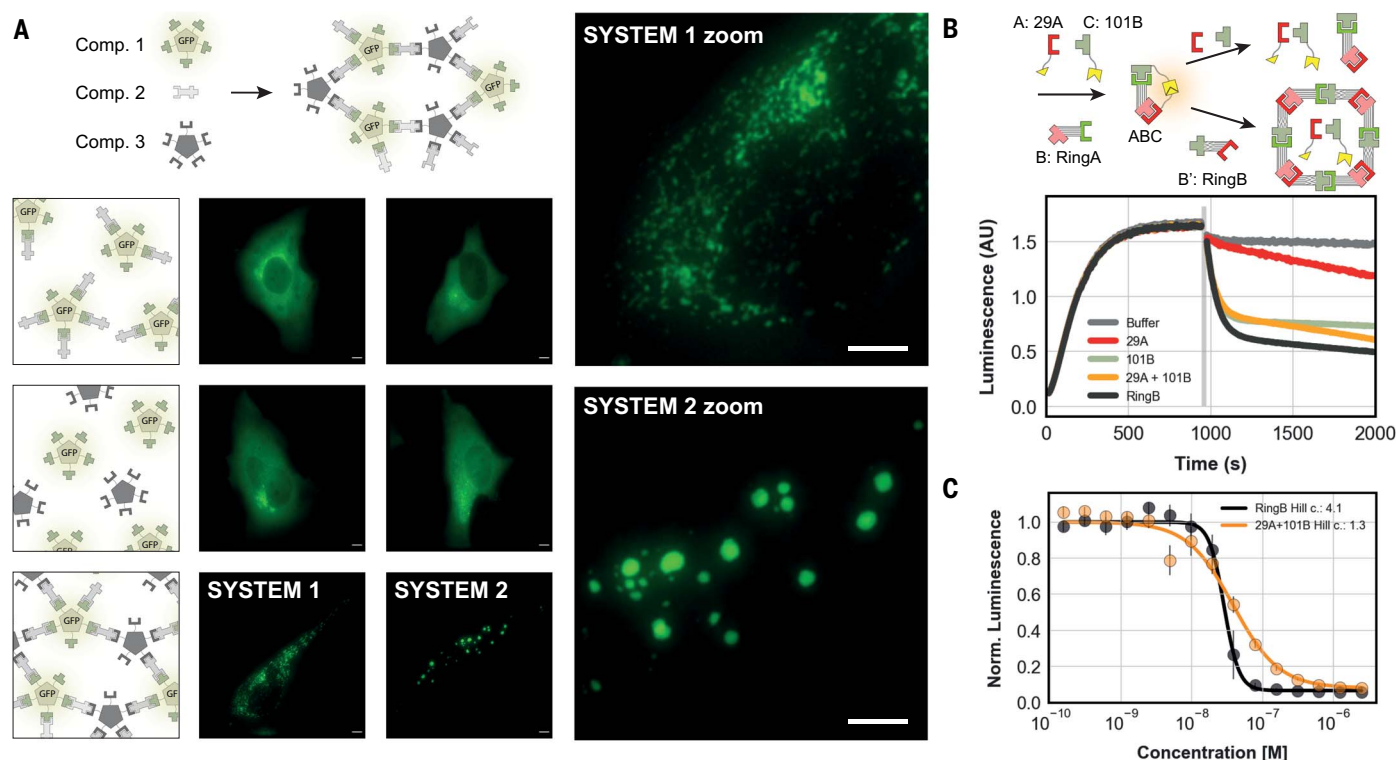


nsEM confirmed that this higher-order complex is similar to the designed C4-symmetric ring (Fig. 4D and fig. S21).

To determine whether our components function as designed in living cells, and to evaluate their use in constructing conditional assemblies, we fused one heterodimer protomer to a previously designed GFP-tagged C5 homo-oligomer (7) and a second protomer of a different heterodimer to an untagged C5 homo-oligomer. Transient expression of the two constructs in HeLa cells led to a distributed and diffuse GFP signal throughout the cell (Fig. 5A and fig. S22), suggesting that the components do not interact with each other or self-associate. However, when a bivalent connector (Fig. 3B) designed to link the two homo-oligomers was also expressed, the GFP signal redistributed into discrete puncta con-

sistent with the expected three-component extended meshwork (Fig. 5A and fig. S22). Notably, changing just one of the two heterodimer interfaces in the assembly from a high- to a low-affinity interface had a notable effect on the morphology of the puncta. When both interfaces had nanomolar affinity (system 1 in Fig. 5A), there were many small puncta, whereas substitution with a micromolar affinity heterodimer with a more rapid dissociation rate led to large droplet-like puncta (system 2 in Fig. 5A). These results show that, as designed, the components of the heterodimers are well behaved in isolation and assemble when combined in cells. The morphology differences further suggest that the ability to modulate dissociation rates and affinities of designed components could be advantageous for probing phase transitions in cells.

Because our designed building blocks are stable in solution and not kinetically trapped in off-target homo-oligomeric states, the assemblies they form can, in principle, reconfigure, as outlined in Fig. 1A and observed for the C4-symmetric hub shown in Fig. 4C. To examine reconfiguration dynamics, we constructed an ABC linear heterotrimer in which the B connector component is one of the two components of the ring shown in Fig. 4D and the A and C capping components are tagged with split luciferase fragments. In the absence of B, components A and C do not interact, and luciferase activity is not reconstituted (Fig. 5B). Upon addition of B, the heterotrimer forms, resulting in luciferase activity (Fig. 5B). Addition of the other ring component (B') to the preformed ABC trimer leads to a rapid decrease in luciferase activity, consistent with



**Fig. 5. Inducible and reconfigurable assemblies.** (A) Cross-linking of homopentamers by bivalent connectors in cells. Schematic representations of the components are at the top and in the first column, and fluorescence microscopy images of cells expressing different combinations of the components are in the second and third columns. High-affinity system 1 (second column) uses LHD101 and LHD275; low-affinity system 2 (third column) uses LHD101 and LHD321. See fig. S22 for additional control images. Scale bars, 5  $\mu$ m. (B) Schematic representation of an ABC heterotrimer (top) with split luciferase activity (yellow

shapes) undergoing subunit exchange through addition of non-luciferase-tagged components. Real-time luminescence measurements (bottom) of samples containing the mixture ABC shown at the top left. The gray bar indicates the addition of either buffer (gray trace), component RingB, or non-luciferase-tagged components LHD29A and LHD101B. AU, arbitrary units. (C) Titration of either component RingB or non-luciferase-tagged components LHD29A and LHD101B to the preformed ABC heterotrimer. Data are fitted to the Hill equation. Error bars represent SD.

disassembly of the trimer and formation of the ring (Fig. 5B and fig. S23, A to C). Because ring formation is cooperative owing to the additional interactions made upon ring closure, we reasoned that the concentration dependence of ABC trimer dissociation would be steeper upon addition of B' than with untagged A and C. To investigate this, we titrated B' and non-luciferase-tagged variants of A and C into the preformed trimer. There was a steep concentration dependence to the loss in luciferase signal upon addition of B' with a Hill coefficient of 4.1 (Fig. 5C and fig. S23D), consistent with the cooperative formation of a symmetrically closed ring ( $B_4B'_4$ ). By contrast, the loss of luciferase signal upon addition of nontagged A and C had a Hill coefficient close to 1, as expected for formation of a non-cooperative linear assembly (Fig. 5C and fig. S23D). In both cases, reconfiguration occurred on the several-minute time scale (fig. S23, B and C). We also observed reconfiguration of heterotrimers using SEC and BLI (fig. S24). This behavior, although common in naturally

evolved protein complexes, has been difficult to achieve by design, because it requires that the individual protomers not self-associate on their own. Our design principles pave the way for the design of functions requiring reconfigurable multiprotein complexes.

### Discussion

Our implicit negative-design principles enable the de novo design of heterodimer pairs for which the individual protomers are stable in solution and readily form their target heterodimeric complexes upon mixing, unlike previously designed assemblies. Rigid fusion of components through structured helical linkers enables the design of higher-order asymmetric multiprotein complexes in which individual subunits have well-defined positions relative to each other. Although rigidly fused building blocks may still exhibit flexibility (molecular breathing), fusion with structured connectors allows more control of subunit orientation than can be achieved by flexible linker fusion and enables fine-tuning of protein complex

geometries. Because of the small sizes of our unfused protomers (between 7 and 15 kDa without DHR or tags), complexes can readily be functionalized through genetic fusion of subunits with proteins of interest. Our bivalent or trivalent connectors can then be used to colocalize and geometrically position two or three such target protein fusions, respectively, and our symmetric hubs can be used to colocalize and position multiple copies of the same target fusion. Because of the modularity of our system, the same set of target fusions can be arranged in multiple different arrangements with adjustable distances, angles, and copy numbers by simply using different components (fig. S25). Because of the solubility and stability of the designs in isolation, complexes can be assembled stepwise (see, for example, Fig. 5A). The asymmetric complexes generated with our components will, in general, have low assembly cooperativity, so the fraction of fully assembled complex will be sensitive to the concentrations of the individual components over a broad range, enabling



subunit exchange and complex reconfiguration in response to signal inputs for synthetic biology and other applications. Because the thermodynamics and kinetics of our designed interfaces are not altered by fusion, the fraction of full assemblies and subassemblies, as well as assembly dynamics, can, in principle, be predicted based on the properties of the individual interfaces (fig. S23A). We expect that the design approach and components presented here will lead to a new generation of reconfigurable protein assemblies for a wide range of applications, including intracellular control for synthetic biology, design of protein logic gates, reprogramming cells from the outside by arraying receptor binding modules with specific geometries, processive multienzyme complexes, and designed molecular machines.

## Materials and methods

### Protein design

#### Docking procedure

As scaffolds for generating edge-strand heterodimers, we used mixed  $\alpha$ - $\beta$  proteins designed by citizen scientists (22) and variants of the Foldit scaffolds that were either expanded with additional helices (see backbone generation methods) and/or fused to designed helical repeat (DHR) proteins (28). Edge-strand docking was performed as described previously (19). Exposed edge strands suitable for docking were identified by calculating the solvent-accessible surface area of the  $\beta$  sheet backbone atoms in all the scaffolds used in the docking procedure. Next, the  $C\alpha$  atoms of each strand of short two-stranded parallel and antiparallel  $\beta$  sheet motifs were aligned to the exposed edge strand, yielding an aligned clashing strand and free docked strand. After removal of the aligned clashing strand, the docked strand was trimmed at the N and/or C terminus to remove potential clashes and subsequently minimized using Rosetta FastRelax (35) to optimize backbone-to-backbone hydrogen bonds. Docks that failed a specified threshold value (typically  $-4$  using ref2015) for the backbone hydrogen-bond score term in Rosetta (hbond\_lr\_bb) were discarded. The minimized docked strands were then geometrically matched to the scaffold library using the MotifGraftMover to create a docked protein-protein complex (36).

#### Interface design

The interface residues of the docked heterodimer complexes were optimized using Rosetta combinatorial sequence (37–40) design using “ref2015,” “beta\_nov16,” or “beta\_genpot” as score functions (41). The interface polarity of the docked heterodimer complexes were fine-tuned in several ways (see supplementary materials for a description of the design xml’s).

First, the HBNetMover (11) was used to install explicit hydrogen-bond networks that contained at least three hydrogen bonds across the interface. Later design rounds consisted of two separate interface sequence-optimization steps. First, interface residues were optimized without compositional constraints, yielding a substantial number of hydrophobic interactions in the interface. The best designs were subsequently selected, and hydrophobic residue pairs with the lowest Rosetta energy interactions across the interface were stored as a seed hydrophobic interaction hotspot (42). In a second round, a polar-interaction network was designed around the fixed hydrophobic hotspot interaction using compositional constraints that favor polar interactions (27). Designs were filtered on interface properties such as binding energy, buried surface area, shape complementarity, degree of packing, and presence of unsatisfied buried polar atoms. A final selection was made by visual inspection of models.

#### Homodimer self-docking

In later design rounds, the propensity for homodimerization was explicitly assessed in silico. Each individual chain of a heterodimer was docked onto itself through edge-strand docking (19) (see also the Docking procedure section). This creates a set of disembodied strands that pair with the scaffold edge strand that also participates in the heterodimeric complex. Homodimer docks were generated by aligning the heterodimerizing edge strand of a second copy of the scaffold back onto the disembodied docked strand (see fig. S7A). Docks with different  $\beta$  register offsets and orientations (parallel and antiparallel) were created. Docks were next converted to polypeptide and clash-checked. Docks where the repulsive Rosetta scoreterm (fa\_rep) was higher than 250 (scorefunction ref2015) were discarded (i.e., no homodimer possible). Surviving docks were converted to full atom models and minimized using FastRelax (35) followed by scoring and assessing of homodimer interface metrics such as binding energy, buried surface area, shape complementarity, degree of packing, and presence of unsatisfied buried polar atoms.

#### Backbone generation and scaffold design

De novo designed protein scaffolds created by Foldit players (22) were expanded with C-terminal polyvaline helices using blueprint-based backbone generation (24, 25). The amino acid identities of the newly built helices and their surrounding region were optimized using Rosetta combinatorial sequence design using a flexible backbone. The resulting models were folded in silico using Rosetta folding simulations, and trajectories that converged to the designed model

structure without off-target minima were selected for rigid fusion and heterodimer design.

#### Design of rigid fusions

To generate rigid fusions of scaffolds or heterodimers to DHRs, we adapted the HFuse pipeline (7, 23): Fusion junctions were designed using the Fastdesign mover to allow backbone movement, and additional filters were included to ensure sufficient contact between the DHR and the fusion partner. When fusing to heterodimers, an additional filter was used to prevent additional contacts between the DHR and the other protomer of the dimer. Bivalent connectors were generated by aligning two proteins that share the same DHR along their shared helical repeats and subsequently splicing together the sequences. To build the C3-symmetric hub, we used a previously published crystal structure of a 12-repeat toroid ring (33). The starting structure was relaxed, its  $z$  axis was aligned, and it was cut into three C3 symmetric chains. Then the HFuse software (7, 23) was used to sample DHR fusions to the exposed helical C termini, and the newly created interfaces were redesigned using RosettaScripts. For the C4-symmetric hub, we used a previously published C4 symmetric homo-oligomer that already contained a N-terminal DHR. Both DHR-containing hubs were fused to LHD protomers in the same way as described above for the bivalent connectors.

#### Design of C4 rings

Using the relaxed crystal structures of LHD29 and LHD101 fused to their respective DHRs, the WORMS software (7, 9, 34) was used to fuse the two heterodimers into cyclic symmetrical rings. Because one construct has exposed N termini and the other has exposed C termini, they were able to be fused head to tail without introduction of further building blocks. Briefly, the first three repeats of each repeat protein were allowed to be sampled as fusion points to ensure that the heterodimer interface was not altered. After fusion into cyclic structures, fixed backbone junction design was applied to the new fusion point using RosettaScripts (39), optimizing for shape complementarity (43). One design from each symmetry—C3, C4, C5, and C6—was selected for experimental testing.

#### Protein expression and purification

Synthetic genes encoding designed proteins and their variants were purchased from Genscript or Integrated DNA technologies (IDT). Bicistronic genes were ordered in pET29b, with the first cistron being either without tag or with an N-terminal sfGFP tag followed by the intercistronic sequence TAAAGAAGGAGATATCA-TATG. The second cistron was tagged with a polyhistidine His6 $\times$  tag at the C terminus.

Plasmids encoding the individual protomers were ordered in pET29b with an N-terminal polyhistidine His6 $\times$  tag followed by a TEV cleavage site, N-terminal polyhistidine His6 $\times$  tag followed by a snac cleavage site or C-terminal polyhistidine His6 $\times$  tag preceded by a snac tag. For enzymatic biotinylation reactions, an Avi-Tag was included at either the N or C terminus (see data S1 for detailed construct information). Proteins were expressed in BL21 LEMO *E. coli* cells by autoinduction using TBII media (Mpbio) supplemented with 50 $\times$ 5052, 20 mM MgSO<sub>4</sub>, and trace metal mix or in almost TB media containing 12 g of peptone and 24 g of yeast extract per liter supplemented with 50 $\times$ 5052, 20 mM MgSO<sub>4</sub>, trace metal mix, and 10 $\times$  phosphate buffer. Proteins were expressed under antibiotic selection at 37°C overnight or at 18°C for 24 hours after initial growth for 6 to 8 hours at 37°C. Cells were harvested by centrifugation at 4000g and lysed by sonication after resuspension of the cells in lysis buffer (100 mM Tris pH 8.0, 200 mM NaCl, 50 mM imidazole pH 8.0) containing protease inhibitors (Thermo Scientific) and Bovine pancreas DNaseI (Sigma-Aldrich). Proteins were purified by immobilized metal affinity chromatography (IMAC). Cleared lysates were incubated with 2 to 4 ml of nickel-nitriloacetic acid (NTA) beads (Qiagen) for 20 to 40 min before washing beads with 5 to 10 column volumes of lysis buffer, 5 to 10 column volumes of high-salt buffer (10 mM Tris pH 8.0, 1 M NaCl), and 5 to 10 column volumes of lysis buffer. Proteins were eluted with 10 ml of elution buffer (20 mM Tris pH 8.0, 100 mM NaCl, 500 mM imidazole pH 8.0).

Designs were finally polished using SEC on either Superdex 200 Increase 10/300GL or Superdex 75 Increase 10/300GL columns (GE Healthcare) using 20 mM Tris pH 8.0, 100 mM NaCl or 20 mM Tris pH 8.0, 300 mM NaCl. Cyclic assemblies of C3 and C4 symmetries were purified using a Superose 6 increase 10/300GL (GE Healthcare). The two component C4 rings were purified by SEC in 25 mM Tris pH 8.0, 300 mM NaCl. Peak fractions were verified by SDS-PAGE and LC/MS and stored at concentrations between 0.5 and 10 mg/ml at 4°C or flash frozen in liquid nitrogen for storage at -80°C. Designs that precipitated at low concentration upon storage at 4°C could, in general, be salvaged by increasing the salt concentration to 300 to 500 mM NaCl.

For structural studies, designs with a polyhistidine tag and TEV recognition site were cleaved using TEV protease (his6-TEV). TEV cleavage was performed in a buffer containing 20 mM Tris pH 8.0, 100 mM NaCl, and 1 mM TCEP using 1% (w/w) his6-TEV and allowed to proceed overnight at room temperature. Uncleaved protein and his6-TEV were separated

from cleaved protein using IMAC followed by SEC. Designs carrying a C-terminal SNAC-polyhistidine tag [GGSHHWGS(...)(HHHHHH)] were cleaved chemically by on-bead nickel-assisted cleavage (44): Nickel-bound designs were washed with 10 column volumes of lysis buffer followed by 5 column volumes of 20 mM Tris pH 8.0, 100 mM NaCl. Proteins were subsequently washed with 5 column volumes of SNAC buffer (100 mM CHES, 100 mM acetone oxime, 100 mM NaCl, pH 8.6). Beads were next incubated with 5 column volumes of SNAC buffer with 2 mM NiCl<sub>2</sub> for more than 12 hours at room temperature on a shaking platform to allow cleavage to take place. Next, the flow-through that contained cleaved protein was collected. The flow-throughs of two additional washes (SNAC buffer/SNAC buffer with 50 mM Imidazole) of 3 to 5 column volumes were also collected to harvest any remaining weakly bound protein. Cleaved proteins were finally purified by SEC.

For mammalian cell expression, synthetic genes encoding designed proteins were purchased from Genscript and cloned into mammalian expression vectors. LHD101B-C5 was cloned into the KpnI-XbaI site of pCDNA3.1+N-eGFP in frame with enhanced GFP (eGFP). Both LHD275B\_53\_0\_LHD101A and LHD321B\_53\_LHD101A were cloned into the NheI-XbaI site of pCDNA3.1+C-HA. LHD275A-C5 and LHD321A-C5 were cloned into the KpnI-XbaI site of pCDNA3.1+N-HA.

#### Cell culture and transient transfections

HeLa cells (ATCC CCL-2) were cultured in Dulbecco's modified Eagle's medium (DMEM) (Gibco) that was supplemented with 1 mM L-glutamine (Gibco), 4.5 g/liter D-glucose (Gibco), 10% fetal bovine serum (FBS), and (1 $\times$ ) nonessential amino acids (Gibco). Cells were cultured at 37°C and 5% CO<sub>2</sub> and passaged twice per week. To passage, cells were dissociated using 0.05% trypsin EDTA (Gibco) and split 1:5 or 1:10 into a new tissue culture (TC)-treated T75 flask (Thermo Scientific ref 156499).

HeLa cells were plated at 20,000 cells per well in Cellview cell culture slides (Greiner Bio-One ref 543079). Twenty-four hours later, cells were transiently transfected at a concentration of 1875 ng total DNA per well and 1  $\mu$ g/ $\mu$ l PEI-MAX (Polysciences) mixed with Opti-MEM medium (Gibco). Transfected cells were incubated at 37°C and 5% CO<sub>2</sub> for 24 to 36 hours before being imaged.

#### Fluorescence microscopy and image processing

3D images were acquired with a commercial OMX-SR system (GE Healthcare). A 488-nm Topica diode laser was used for excitation. Emission was collected on a PCO.edge scientific complementary metal-oxide semiconductor (sCMOS) camera using an Olympus 60 $\times$  1.42NA PlanApochromat oil immersion lens. Images

were acquired in a 1024 by 1024 field of view (pixel size 6.5  $\mu$ m) with no binning. Acquisition was controlled with AcquireSR acquisition control software. Z-stacks were collected with a step size of 500 nm and 15 slices per image. Images were deconvolved with an enhanced ratio using SoftWoRx 7.0.0 (GE Healthcare). Cell images were sum-projected using Fiji v2.1.0. Scale bars equal 5  $\mu$ m.

#### Enzymatic protein biotinylation

Avi-tagged (GLNDIFEAAQKIEWHE; see supplementary materials) proteins were purified as described above. The BirA500 (Avidity, LLC) biotinylation kit was used to biotinylate 840  $\mu$ l of protein from the IMAC elution in a 1200  $\mu$ l (final volume) reaction according to the manufacturer's protocol. Reactions were incubated at 4°C overnight and purified using SEC on a Superdex 200 10/300 Increase GL (GE Healthcare) or S75 10/300 Increase GL (GE Healthcare) in SEC buffer (20 mM Tris pH 8.0, 100 mM NaCl).

#### Biolayer interferometry

Biolayer interferometry experiments were performed on an OctetRED96 BLI system (ForteBio, Menlo Park, CA). Streptavidin-coated biosensors were first equilibrated for at least 10 min in Octet buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.05% surfactant P20) supplemented with 1 mg/ml bovine serum albumin (SigmaAldrich). Enzymatically biotinylated designs were immobilized onto the biosensors by dipping the biosensors into a solution with 10 to 50 nM protein for 30 to 120 s. This was followed by dipping in fresh octet buffer to establish a baseline for 120 s. Titration experiments were performed at 25°C while rotating at 1000 rpm. Association of designs was allowed by dipping biosensors in solutions containing designed protein diluted in octet buffer until equilibrium was approached followed by dissociation by dipping the biosensors into fresh buffer solution to monitor the dissociation kinetics. Steady-state and global kinetic fits were performed using the manufacturer's software (Data Analysis 9.1) assuming a 1:1 binding model.

#### SEC binding assays

Complexes and individual components were diluted in 20 mM Tris pH 8.0, 100 mM NaCl. After overnight equilibration of the mixtures at room temperature or 4°C, 500  $\mu$ l of sample was injected onto a Superdex 200 10/300 increase GL (dimers, linear assemblies) or Superose 6 increase 10/300 GL (symmetric assemblies) (all columns from GE healthcare) using the absorbance at 230 or 473 nm (for GFP-tagged components) as readout. Dimers were mixed at monomer concentrations of 5  $\mu$ M or higher. Trimer and ABCD tetramer mixtures contained 5  $\mu$ M of the bivalent

connector and 7.5  $\mu\text{M}$  of each terminal cap. (Lower absolute concentrations with the same ratios were used for some trimers.) ABCA tetramer mixtures contained 5  $\mu\text{M}$  per bivalent connector and 15  $\mu\text{M}$  terminal cap. The hexamer mixture contained 3  $\mu\text{M}$  of components C and D, 3.6  $\mu\text{M}$  of B and E, and 4.4  $\mu\text{M}$  of A and F. The branched assembly shown in Fig. 4A contained 2.8  $\mu\text{M}$  of the trivalent connector and 4  $\mu\text{M}$  of each cap. For the exchange experiment shown in Fig. S24A, the ABC trimer was preincubated at concentrations of 6  $\mu\text{M}$  B and 9  $\mu\text{M}$  each of A and C. C' was then added to reach a final concentration of 2  $\mu\text{M}$  B, 3  $\mu\text{M}$  each of A and C, and 6  $\mu\text{M}$  C'.

### Native mass spectrometry

Sample purity, integrity, and oligomeric state were analyzed by on-line buffer exchange MS in 200 mM ammonium acetate using a Vanquish ultra-high performance LC system coupled to a Q Exactive ultra-high mass range Orbitrap mass spectrometer (Thermo Fisher Scientific). A self-packed buffer exchange column was used (P6 polyacrylamide gel, BioRad) (45). The recorded mass spectra were deconvolved with UniDec version 4.2+ (46).

### Crystal structure determination

For all structures, starting phases were obtained by molecular replacement using Phaser (47). Diffraction images were integrated using XDS (48) or HKL2000 (49) and merged and scaled using Aimless (50). Structures were refined in Phenix (51) using phenix.autobuild and phenix.refine or Refmac (52). Model building was performed using COOT (53).

Proteins were crystallized using the vapor diffusion method at room temperature. LHD29 crystals grew in 0.2 M sodium iodide, 20% PEG3350; LHD29A53/B53 crystals in 3.2 M ammonium sulfate, 0.1 M citric acid pH 4.0; and LHD101A53/B4 crystals in 2.4 M sodium malonate pH 7.0. Crystals were harvested and cryoprotected using 20% PEG200 for LHD29, 20% PEG400 for LHD29A53/B53, and 20% glycerol for LHD101A53/B4 before data was collected at the Advanced Light Source (Berkeley, USA). The structures were solved by molecular replacement using either computationally designed models of individual chains A or B or the full heterodimer complex as search models.

The root mean square deviation (RMSD), TMscore, and local distance difference test (LDDT) metrics between the designed models and corresponding crystal structures were calculated as described previously (54, 55). Protein structure graphics were prepared using PyMOL (Schrödinger).

### Electron microscopy

SEC peak fractions were concentrated before nsEM screening. Samples were then immediately diluted 5 to 150 times in Tris-buffered

saline (TBS) buffer (25 mM Tris pH 8.0, 25 mM NaCl) depending on sample concentration. A final volume of 5  $\mu\text{l}$  was applied to negatively glow discharged, carbon-coated 400-mesh copper grids (01844-F, TedPella, Inc.) and then washed with Milli-Q Water and stained using 0.75% uranyl formate as previously described (56). Air-dried grids were imaged on a FEI Talos L120C TEM (FEI Thermo Scientific, Hillsboro, OR) equipped with a 4K  $\times$  4K Gatan OneView camera at a magnification of 57,000 $\times$  and pixel size of 2.51. Micrographs were imported into CislTEM software or cryoSPARC software, and a circular blob picker was used to select particles that were then subjected to 2D classification. Ab initio reconstruction and homogeneous refinement in Cn symmetry were used to generate 3D electron density maps (57, 58).

### Constructs for luciferase assays

Split luciferase reporter constructs were ordered as synthetic genes from Genscript. Each design was N-terminally fused to a superfolder GFP (sfGFP) (for protein quantification in lysate) and C-terminally fused to either smBIT or lgBiT of the split luciferase components. A Strep-tag was included at the N terminus for purification, and a glycine-serine (GS)-linker was inserted between the design and the split luciferase component.

### Expression for multiplexed luciferase assay

Plasmids were transformed into Lemo21(DE3) cells (New England Biolabs) and grown in 96-deep-well plates overnight at 37°C in 1 ml of LB containing 50  $\mu\text{g}/\text{ml}$  of kanamycin sulfate. The next day, 100  $\mu\text{l}$  of overnight cultures were used to inoculate 96-deep-well plates containing 900  $\mu\text{l}$  of TBII medium (MP Biomedicals) with 50  $\mu\text{g}/\text{ml}$  of kanamycin sulfate, and the cultures were grown for 2 hours at 37°C before induction with 0.1 mM isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG). Protein expression was carried out at 37°C for 4 hours before the cells were harvested by centrifugation (4000g, 5 min). Cell pellets were resuspended in 100  $\mu\text{l}$  of lysis buffer (10 mM sodium phosphate, 150 mM NaCl, pH 7.4, 1 mg/ml lysozyme, 0.1 mg/ml DNase I, 5 mM  $\text{MgCl}_2$ , 1 tablet per 50 ml of cOmplete protease inhibitor (Roche), 0.05% v/v Tween 20), and cells were lysed by performing three freeze-thaw cycles (1 hour incubations at 37°C followed by freezing at  $-80^\circ\text{C}$ ). The lysate was cleared by centrifugation (4000g, 20 min), and the soluble fraction was transferred to a 96-well assay plate (Corning, cat. no. 3991). Concentrations of the constructs in soluble lysate were determined by sfGFP fluorescence using a calibration curve.

### Lysate production for multiplexed luciferase assay

Neutral lysate for preparing serial dilutions was prepared by transforming Lemo21(DE3)

with the pUC19 plasmid. Transformations were used to inoculate small overnight cultures, which were used to inoculate 0.5-liter TBII cultures (all cultures contained 50  $\mu\text{g}/\text{ml}$  of carbenicillin). Cells were grown for 24 hours at 37°C before being harvested. Pellets were resuspended in the same lysis buffer, followed by sonication. The lysate density was adjusted with lysis buffer to have its OD280 (optical density at 280 nm) match that of pUC19 control wells from the 96-well expression plate.

### Expression and purification of luciferase constructs

Plasmids were transformed into Lemo21 (DE3) cells and used directly to inoculate 50 ml of autoinduction media (TBII supplemented with 0.5% w/v glucose, 0.05% w/v glycerol, 0.2% w/v lactose monohydrate, and 2 mM  $\text{MgSO}_4$ , 50  $\mu\text{g}/\text{ml}$  kanamycin sulfate). The cultures were incubated at 37°C for 20 to 24 hours before harvesting the cells by centrifugation (4000g, 5 min). Cells were resuspended in 10 ml of lysis buffer [100 mM Tris, 150 mM NaCl, pH 8, 0.1 mg/ml lysozyme, 0.01 mg/ml DNase I, 1 mM phenylmethylsulfonyl fluoride (PMSF)] and lysed by sonication. The insoluble fraction was cleared by centrifugation (16,000g, 45 min), and the proteins were purified from the soluble fraction by affinity chromatography using Strep-Tactin XT Superflow High-Capacity resin (IBA Lifesciences). Elutions were performed with 100 mM Tris, 150 mM NaCl, 50 mM biotin, pH 8, and the proteins were further purified by SEC using a Superdex 200 10/300 increase column equilibrated with 20 mM sodium phosphate, 100 mM NaCl, pH 7.4, 0.05% v/v Tween 20.

### Luciferase binding assays

All assays were performed in 20 mM sodium phosphate, 100 mM NaCl, pH 7.4, 0.05% v/v Tween 20. Depending on the source of the protein used in the assay (purified components or lysate), soluble lysate components were also present. Reactions were assembled in 96-well plates (Corning, cat. no. 3686) in the presence of Nano-Glo substrate (Promega, cat. no. N1130) and diluted 100 $\times$  or 500 $\times$  for kinetics and endpoint measurements, respectively, and the luminescence signal was recorded on a Synergy Neo2 plate reader (BioTek).

Kinetic binding assays were performed under pseudo first-order conditions, with the final concentration of one protein at 1 nM and the other at 10 nM. Stock solutions were mixed in a 1:1 volume ratio in the presence of substrate, and the dead-time between mixing and starting the measurement (typically 15 to 30 s) was added during data processing. For long kinetic measurements (Fig. S6A), the proteins were premixed and kept in a sealed tube at room temperature over the course of the experiment. Aliquots were taken at regular intervals, mixed



with substrate, and immediately recorded. All kinetic measurements were fitted to a single exponential decay function:

$$S = A \cdot \exp(-k_{\text{obs}} \cdot t) + B$$

where  $t$  is time (the independent variable) and  $S$  is the observed luminescence signal (the dependent variable) and the fitted parameters are the amplitude  $A$ , the observed rate constant  $k_{\text{obs}}$ , and the endpoint luminescence  $B$ .

Equilibrium binding assays were performed with one component kept constant at 1 nM while titrating the other protein. Serial dilutions curves were prepared over 12 points, with a one-quarter dilution factor between each step. The concentration of protein in the soluble lysate provided the highest concentration point of the curve. To avoid serial dilution of the other lysate components, all stocks were prepared with neutral lysate. The assembled plates were incubated overnight at room temperature before adding substrate and immediately measuring luminescence. The data was fitted to the following equation to obtain  $K_d$  values:

$$S = S_0 + S_1 \cdot f_{AB} + a_2 \cdot B_T \cdot S_2$$

$$f_{AB} = [A_T + B_T + K_d - (A_T + B_T + K_d)^2 - 4A_TB_T]/2A_T$$

where  $A_T$  and  $B_T$  are the total concentrations of each species (the independent variables,  $A_T = 1$  nM,  $B_T$  is the titrated species) and  $S$  is the observed signal (the dependent variable). The fitted parameters are the pre-saturation baseline  $S_0$ , the postsaturation baseline  $S_1$ , and the correction terms  $a_2$  and  $S_2$ .

Ternary complex equilibrium binding experiments were performed with pure protein, using the concentration indicated in fig. S23 for the constant components, and titrating B. After assembly, the plates were incubated overnight before adding substrate and immediately measuring luminescence.

Ternary complex reconfiguration kinetics (Fig. 5B and fig. S23) were measured with pure proteins. Components A (1 nM) and C (100 nM) were briefly preincubated in the presence of substrate (1/500 dilution) before adding component B (50 nM) to start the reaction. Once the association reactions were complete, the assay plate was briefly taken out of the plate reader; out-competing protein(s) (100 nM each in Fig. 5B and fig. S23B and 1000 nM each in fig. S23C) were added to the reactions; and data acquisition was resumed.

Ternary complex thermodynamic out-competitions (Fig. 5C and fig. S23D) were measured with purified proteins. Final concentrations of components A-smBiT, B, and C-IgBiT were 1, 50, and 100 nM final, respec-

tively. The out-competitor(s) (B' or untagged A+C) were titrated from 10 uM down to about 1 pM over 24 points, with a one-half dilution factor between each step. Reactions were incubated at room temperature for 2 to 5 hours before adding substrate (1/500 dilution) and measuring luminescence. The averages of four experiments were fitted to the Hill equation:

$$S = S_0 + (S_1 - S_0)/[1 + (K/L)^n]$$

where  $L$  is the total concentration of the out-competitor(s) (the independent variable) and  $S$  is the observed signal (the dependent variable). The fitted parameters are the presaturation baseline  $S_0$ , the postsaturation baseline  $S_1$ , the transition midpoint  $K$ , and the Hill coefficient  $n$ .

### Simulation of ternary complex

Systems of ordinary differential equations describing the kinetics of interactions between the species involved in the formation of the ternary complex (fig. S23A) were numerically integrated using `scipy.integrate.odeint` as implemented in `Scipy` (version 1.6.3). Steady-state values were used to determine the distribution of species at thermodynamic equilibrium.

The ternary system is composed of the following species: A, B, C, AB, BC, and ABC. The following set of equations was used to describe the system:

$$\begin{aligned} d[A]/dt &= -k_1[A][B] + k_{-1}[AB] - \\ &\quad k_1[A][BC] + k_{-1}[ABC] \end{aligned}$$

$$\begin{aligned} d[B]/dt &= -k_1[A][B] + k_{-1}[AB] - \\ &\quad k_2[B][C] + k_{-2}[BC] \end{aligned}$$

$$\begin{aligned} d[C]/dt &= -k_2[B][C] + k_{-2}[BC] - \\ &\quad k_2[AB][C] + k_{-2}[ABC] \end{aligned}$$

$$\begin{aligned} d[AB]/dt &= k_1[A][B] - k_{-1}[AB] + \\ &\quad k_{-2}[ABC] - k_2[AB][C] \end{aligned}$$

$$\begin{aligned} d[BC]/dt &= k_2[B][C] - k_{-2}[BC] + \\ &\quad k_{-1}[ABC] - k_1[A][BC] \end{aligned}$$

$$\begin{aligned} d[ABC]/dt &= k_1[A][BC] - k_{-1}[ABC] + \\ &\quad k_2[AB][C] - k_{-2}[ABC] \end{aligned}$$

where  $k_i$  describes bimolecular association rate constants and  $k_{-i}$  represents unimolecular dissociation rate constants.  $K_1 = k_{-1}/k_1$ , and  $K_2 = k_{-2}/k_2$  describe the affinity of the A:B and B:C interfaces, respectively.

### REFERENCES AND NOTES

1. S. E. Tusk, N. J. Delalez, R. M. Berry, Subunit exchange in protein complexes. *J. Mol. Biol.* **430**, 4557–4579 (2018). doi: [10.1016/j.jmb.2018.06.039](https://doi.org/10.1016/j.jmb.2018.06.039); pmid: [29959924](https://pubmed.ncbi.nlm.nih.gov/29959924/)

2. C. Engel, S. Neyer, P. Cramer, Distinct mechanisms of transcription initiation by RNA polymerases I and II. *Annu. Rev. Biophys.* **47**, 425–446 (2018). doi: [10.1146/annurev-biophys-070317-033058](https://doi.org/10.1146/annurev-biophys-070317-033058); pmid: [29792819](https://pubmed.ncbi.nlm.nih.gov/29792819/)
3. P. M. J. Burgers, T. A. Kunkel, Eukaryotic DNA replication fork. *Annu. Rev. Biochem.* **86**, 417–438 (2017). doi: [10.1146/annurev-biochem-061516-044709](https://doi.org/10.1146/annurev-biochem-061516-044709); pmid: [28301743](https://pubmed.ncbi.nlm.nih.gov/28301743/)
4. S. Gonen, F. DiMaio, T. Gonen, D. Baker, Design of ordered two-dimensional arrays mediated by noncovalent protein-protein interfaces. *Science* **348**, 1365–1368 (2015). doi: [10.1126/science.aaa9897](https://doi.org/10.1126/science.aaa9897); pmid: [26089516](https://pubmed.ncbi.nlm.nih.gov/26089516/)
5. Y. Hsia et al., Design of a hyperstable 60-subunit protein dodecahedron. [corrected]. *Nature* **535**, 136–139 (2016). doi: [10.1038/nature18010](https://doi.org/10.1038/nature18010); pmid: [27309817](https://pubmed.ncbi.nlm.nih.gov/27309817/)
6. N. P. King et al., Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510**, 103–108 (2014). doi: [10.1038/nature13404](https://doi.org/10.1038/nature13404); pmid: [24870237](https://pubmed.ncbi.nlm.nih.gov/24870237/)
7. Y. Hsia et al., Design of multi-scale protein complexes by hierarchical building block fusion. *Nat. Commun.* **12**, 2294 (2021). doi: [10.1038/s41467-021-22276-z](https://doi.org/10.1038/s41467-021-22276-z); pmid: [33863889](https://pubmed.ncbi.nlm.nih.gov/33863889/)
8. A. J. Ben-Sasson et al., Design of biologically active binary protein 2D materials. *Nature* **589**, 468–473 (2021). doi: [10.1038/s41586-020-03120-8](https://doi.org/10.1038/s41586-020-03120-8); pmid: [33408408](https://pubmed.ncbi.nlm.nih.gov/33408408/)
9. R. Divine et al., Designed proteins assemble antibodies into modular nanocages. *Science* **372**, eabd9994 (2021). doi: [10.1126/science.abd9994](https://doi.org/10.1126/science.abd9994); pmid: [33795432](https://pubmed.ncbi.nlm.nih.gov/33795432/)
10. Z. Chen et al., Programmable design of orthogonal protein heterodimers. *Nature* **565**, 106–111 (2019). doi: [10.1038/s41586-018-0802-y](https://doi.org/10.1038/s41586-018-0802-y); pmid: [30568301](https://pubmed.ncbi.nlm.nih.gov/30568301/)
11. S. E. Boyken et al., De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science* **352**, 680–687 (2016). doi: [10.1126/science.aad8865](https://doi.org/10.1126/science.aad8865); pmid: [27151862](https://pubmed.ncbi.nlm.nih.gov/27151862/)
12. Z. Chen et al., De novo design of protein logic gates. *Science* **368**, 78–84 (2020). doi: [10.1126/science.aay2790](https://doi.org/10.1126/science.aay2790); pmid: [32241946](https://pubmed.ncbi.nlm.nih.gov/32241946/)
13. H. Gradišar, R. Jerala, De novo design of orthogonal peptide pairs forming parallel coiled-coil heterodimers. *J. Pept. Sci.* **17**, 100–106 (2011). doi: [10.1002/psc.1331](https://doi.org/10.1002/psc.1331); pmid: [21234981](https://pubmed.ncbi.nlm.nih.gov/21234981/)
14. C. L. Edgell, A. J. Smith, J. L. Beesley, N. J. Savery, D. N. Woolfson, De novo designed protein-interaction modules for in-cell applications. *ACS Synth. Biol.* **9**, 427–436 (2020). doi: [10.1021/acssynbio.9b00453](https://doi.org/10.1021/acssynbio.9b00453); pmid: [31977192](https://pubmed.ncbi.nlm.nih.gov/31977192/)
15. A. Leaver-Fay, R. Jacak, P. B. Stranges, B. Kuhlman, A generic program for multistate protein design. *PLOS ONE* **6**, e20937 (2011). doi: [10.1371/journal.pone.0020937](https://doi.org/10.1371/journal.pone.0020937); pmid: [21754981](https://pubmed.ncbi.nlm.nih.gov/21754981/)
16. A. Leaver-Fay et al., Computationally designed bispecific antibodies using negative state repertoires. *Structure* **24**, 641–651 (2016). doi: [10.1016/j.str.2016.02.013](https://doi.org/10.1016/j.str.2016.02.013); pmid: [26996964](https://pubmed.ncbi.nlm.nih.gov/26996964/)
17. J. J. Havranek, P. B. Harbury, Automated design of specificity in molecular recognition. *Nat. Struct. Biol.* **10**, 45–52 (2003). doi: [10.1038/nsb877](https://doi.org/10.1038/nsb877); pmid: [12459719](https://pubmed.ncbi.nlm.nih.gov/12459719/)
18. S. J. Fleishman, D. Baker, Role of the biomolecular energy gap in protein design, structure, and evolution. *Cell* **149**, 262–273 (2012). doi: [10.1016/j.cell.2012.03.016](https://doi.org/10.1016/j.cell.2012.03.016); pmid: [22500796](https://pubmed.ncbi.nlm.nih.gov/22500796/)
19. D. D. Sahtoe et al., Transferrin receptor targeting by de novo sheet extension. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2021569118 (2021). doi: [10.1073/pnas.2021569118](https://doi.org/10.1073/pnas.2021569118); pmid: [33879614](https://pubmed.ncbi.nlm.nih.gov/33879614/)
20. P. B. Stranges, M. Machiusi, M. J. Miley, A. Tripathy, B. Kuhlman, Computational design of a symmetric homodimer using  $\beta$ -strand assembly. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20562–20567 (2011). doi: [10.1073/pnas.1115124108](https://doi.org/10.1073/pnas.1115124108); pmid: [22143762](https://pubmed.ncbi.nlm.nih.gov/22143762/)
21. H. Remaut, G. Waksman, Protein-protein interaction through  $\beta$ -strand addition. *Trends Biochem. Sci.* **31**, 436–444 (2006). doi: [10.1016/j.tibs.2006.06.007](https://doi.org/10.1016/j.tibs.2006.06.007); pmid: [16828554](https://pubmed.ncbi.nlm.nih.gov/16828554/)
22. B. Koepnick et al., De novo protein design by citizen scientists. *Nature* **570**, 390–394 (2019). doi: [10.1038/s41586-019-1274-4](https://doi.org/10.1038/s41586-019-1274-4); pmid: [31168091](https://pubmed.ncbi.nlm.nih.gov/31168091/)
23. T. J. Brunette et al., Modular repeat protein sculpting using rigid helical junctions. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 8870–8875 (2020). doi: [10.1073/pnas.1908768117](https://doi.org/10.1073/pnas.1908768117); pmid: [32245816](https://pubmed.ncbi.nlm.nih.gov/32245816/)
24. Y.-R. Lin et al., Control over overall shape and size in de novo designed proteins. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E5478–E5485 (2015). doi: [10.1073/pnas.1509508112](https://doi.org/10.1073/pnas.1509508112); pmid: [26396255](https://pubmed.ncbi.nlm.nih.gov/26396255/)
25. N. Koga et al., Principles for designing ideal protein structures. *Nature* **491**, 222–227 (2012). doi: [10.1038/nature11600](https://doi.org/10.1038/nature11600); pmid: [23135467](https://pubmed.ncbi.nlm.nih.gov/23135467/)
26. J. K. Leman et al., Macromolecular modeling and design in Rosetta: Recent methods and frameworks. *Nat. Methods* **17**,

- 665–680 (2020). doi: [10.1038/s41592-020-0848-2](https://doi.org/10.1038/s41592-020-0848-2); pmid: [32483333](https://pubmed.ncbi.nlm.nih.gov/32483333/)
27. B. Coventry, D. Baker, Protein sequence optimization with a pairwise decomposable penalty for buried unsatisfied hydrogen bonds. *PLOS Comput. Biol.* **17**, e1008061 (2021). doi: [10.1371/journal.pcbi.1008061](https://doi.org/10.1371/journal.pcbi.1008061); pmid: [33684097](https://pubmed.ncbi.nlm.nih.gov/33684097/)
  28. T. J. Brunette *et al.*, Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580–584 (2015). doi: [10.1038/nature16162](https://doi.org/10.1038/nature16162); pmid: [26675729](https://pubmed.ncbi.nlm.nih.gov/26675729/)
  29. J. R. Lydeard, B. A. Schulman, J. W. Harper, Building and remodelling Cullin-RING E3 ubiquitin ligases. *EMBO Rep.* **14**, 1050–1061 (2013). doi: [10.1038/embor.2013.173](https://doi.org/10.1038/embor.2013.173); pmid: [24232186](https://pubmed.ncbi.nlm.nih.gov/24232186/)
  30. L. K. Langeberg, J. D. Scott, Signalling scaffolds and local organization of cellular behaviour. *Nat. Rev. Mol. Cell Biol.* **16**, 232–244 (2015). doi: [10.1038/nrm3966](https://doi.org/10.1038/nrm3966); pmid: [25785716](https://pubmed.ncbi.nlm.nih.gov/25785716/)
  31. H. W. Schroeder Jr., L. Cavacini, Structure and function of immunoglobulins. *J. Allergy Clin. Immunol.* **125**, S41–S52 (2010). doi: [10.1016/j.jaci.2009.09.046](https://doi.org/10.1016/j.jaci.2009.09.046); pmid: [20176268](https://pubmed.ncbi.nlm.nih.gov/20176268/)
  32. P. Broz, V. M. Dixit, Inflammasomes: Mechanism of assembly, regulation and signalling. *Nat. Rev. Immunol.* **16**, 407–420 (2016). doi: [10.1038/nri.2016.58](https://doi.org/10.1038/nri.2016.58); pmid: [27291964](https://pubmed.ncbi.nlm.nih.gov/27291964/)
  33. L. Doyle *et al.*, Rational design of  $\alpha$ -helical tandem repeat proteins with closed architectures. *Nature* **528**, 585–588 (2015). doi: [10.1038/nature16191](https://doi.org/10.1038/nature16191); pmid: [26675735](https://pubmed.ncbi.nlm.nih.gov/26675735/)
  34. I. Vulovic *et al.*, Generation of ordered protein assemblies using rigid three-body fusion. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2015037118 (2021). doi: [10.1073/pnas.2015037118](https://doi.org/10.1073/pnas.2015037118); pmid: [34074752](https://pubmed.ncbi.nlm.nih.gov/34074752/)
  35. M. D. Tyka *et al.*, Alternate states of proteins revealed by detailed energy landscape mapping. *J. Mol. Biol.* **405**, 607–618 (2011). doi: [10.1016/j.jmb.2010.11.008](https://doi.org/10.1016/j.jmb.2010.11.008); pmid: [21073878](https://pubmed.ncbi.nlm.nih.gov/21073878/)
  36. A. Chevalier *et al.*, Massively parallel de novo protein design for targeted therapeutics. *Nature* **550**, 74–79 (2017). doi: [10.1038/nature23912](https://doi.org/10.1038/nature23912); pmid: [28953867](https://pubmed.ncbi.nlm.nih.gov/28953867/)
  37. P. Hosseinzadeh *et al.*, Comprehensive computational design of ordered peptide macrocycles. *Science* **358**, 1461–1466 (2017). doi: [10.1126/science.aap7577](https://doi.org/10.1126/science.aap7577); pmid: [29242347](https://pubmed.ncbi.nlm.nih.gov/29242347/)
  38. B. Dang *et al.*, De novo design of covalently constrained mesosize protein scaffolds with unique tertiary structures. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 10852–10857 (2017). doi: [10.1073/pnas.1710695114](https://doi.org/10.1073/pnas.1710695114); pmid: [28973862](https://pubmed.ncbi.nlm.nih.gov/28973862/)
  39. S. J. Fleishman *et al.*, RosettaScripts: A scripting language interface to the Rosetta macromolecular modeling suite. *PLOS ONE* **6**, e20161 (2011). doi: [10.1371/journal.pone.0020161](https://doi.org/10.1371/journal.pone.0020161); pmid: [21073610](https://pubmed.ncbi.nlm.nih.gov/21073610/)
  40. G. Bhardwaj *et al.*, Accurate de novo design of hyperstable constrained peptides. *Nature* **538**, 329–335 (2016). doi: [10.1038/nature19791](https://doi.org/10.1038/nature19791); pmid: [27626386](https://pubmed.ncbi.nlm.nih.gov/27626386/)
  41. R. F. Alford *et al.*, The Rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017). doi: [10.1021/acs.jctc.7b00125](https://doi.org/10.1021/acs.jctc.7b00125); pmid: [28430426](https://pubmed.ncbi.nlm.nih.gov/28430426/)
  42. S. Chaudhury, S. Lyskov, J. J. Gray, PyRosetta: A script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26**, 689–691 (2010). doi: [10.1006/jmbi.1993.1648](https://doi.org/10.1006/jmbi.1993.1648); pmid: [8263940](https://pubmed.ncbi.nlm.nih.gov/8263940/)
  43. M. C. Lawrence, P. M. Colman, Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950 (1993). doi: [10.1006/jmbi.1993.1648](https://doi.org/10.1006/jmbi.1993.1648); pmid: [8263940](https://pubmed.ncbi.nlm.nih.gov/8263940/)
  44. B. Dang *et al.*, SNAC-tag for sequence-specific chemical protein cleavage. *Nat. Methods* **16**, 319–322 (2019). doi: [10.1038/s41592-019-0357-3](https://doi.org/10.1038/s41592-019-0357-3); pmid: [30923372](https://pubmed.ncbi.nlm.nih.gov/30923372/)
  45. Z. L. VanAernum *et al.*, Rapid online buffer exchange for screening of proteins, protein complexes and cell lysates by native mass spectrometry. *Nat. Protoc.* **15**, 1132–1157 (2020). doi: [10.1038/s41596-019-0281-0](https://doi.org/10.1038/s41596-019-0281-0); pmid: [32005983](https://pubmed.ncbi.nlm.nih.gov/32005983/)
  46. M. T. Marty *et al.*, Bayesian deconvolution of mass and ion mobility spectra: From binary interactions to polydisperse ensembles. *Anal. Chem.* **87**, 4370–4376 (2015). doi: [10.1021/acs.analchem.5b00140](https://doi.org/10.1021/acs.analchem.5b00140); pmid: [25799115](https://pubmed.ncbi.nlm.nih.gov/25799115/)
  47. A. J. McCoy *et al.*, Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007). doi: [10.1107/S0021889807021206](https://doi.org/10.1107/S0021889807021206); pmid: [19461840](https://pubmed.ncbi.nlm.nih.gov/19461840/)
  48. W. Kabsch, XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010). doi: [10.1107/S0907444909047337](https://doi.org/10.1107/S0907444909047337); pmid: [20124692](https://pubmed.ncbi.nlm.nih.gov/20124692/)
  49. Z. Otwinowski, W. Minor, Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997). doi: [10.1016/S0076-6879\(97\)76066-X](https://doi.org/10.1016/S0076-6879(97)76066-X)
  50. M. J. Winn *et al.*, Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011). doi: [10.1107/S0907444910045749](https://doi.org/10.1107/S0907444910045749); pmid: [21460441](https://pubmed.ncbi.nlm.nih.gov/21460441/)
  51. P. D. Adams *et al.*, PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010). doi: [10.1107/S0907444909052925](https://doi.org/10.1107/S0907444909052925); pmid: [20124702](https://pubmed.ncbi.nlm.nih.gov/20124702/)
  52. G. N. Murshudov, A. A. Vagin, E. J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997). doi: [10.1107/S0907444996012255](https://doi.org/10.1107/S0907444996012255); pmid: [15299926](https://pubmed.ncbi.nlm.nih.gov/15299926/)
  53. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004). doi: [10.1107/S0907444904019158](https://doi.org/10.1107/S0907444904019158); pmid: [15572765](https://pubmed.ncbi.nlm.nih.gov/15572765/)
  54. Y. Zhang, J. Skolnick, TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005). doi: [10.1093/nar/gki524](https://doi.org/10.1093/nar/gki524); pmid: [15849316](https://pubmed.ncbi.nlm.nih.gov/15849316/)
  55. V. Mariani, M. Biasini, A. Barbato, T. Schwede, ICDT: A local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* **29**, 2722–2728 (2013). doi: [10.1093/bioinformatics/btt473](https://doi.org/10.1093/bioinformatics/btt473); pmid: [23986568](https://pubmed.ncbi.nlm.nih.gov/23986568/)
  56. B. L. Nannenga, M. G. Iadanza, B. S. Vollmar, T. Gonen, Overview of electron crystallography of membrane proteins: crystallization and screening strategies using negative stain electron microscopy. *Curr. Protoc. Protein Sci.* **72**, 17.15.1–17.15.11 (2013). doi: [10.1093/bioinformatics/btt473](https://doi.org/10.1093/bioinformatics/btt473); pmid: [23986568](https://pubmed.ncbi.nlm.nih.gov/23986568/)
  57. T. Grant, A. Rohou, N. Grigorieff, cisTEM, user-friendly software for single-particle image processing. *eLife* **7**, e35383 (2018). doi: [10.7554/eLife.35383](https://doi.org/10.7554/eLife.35383); pmid: [29513216](https://pubmed.ncbi.nlm.nih.gov/29513216/)
  58. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017). doi: [10.1038/nmeth.4169](https://doi.org/10.1038/nmeth.4169); pmid: [28165473](https://pubmed.ncbi.nlm.nih.gov/28165473/)
  59. D. D. Sahtoe, F. Praetorius, A. Courbet, Y. Hsia, N. I. Edman, D. Baker, Data for: Reconfigurable asymmetric protein assemblies through implicit negative design, Version 1, Zenodo (2021). doi: [10.5281/zenodo.5717329](https://doi.org/10.5281/zenodo.5717329)

## ACKNOWLEDGMENTS

We acknowledge Baker lab members for discussion; B. Sankaran and beamline scientists at the Advanced Light Source for crystallographic data collection and support; F. Busch, A. Norris, and the Wysocki lab for native mass spectrometry measurements; R. Mout for C5-LHD101B construct sequences; L. Carter for SEC-MALS analysis; M. Ahlrichs, C. Ogohara, and M. Murphy for mammalian cell transfections; C. Miller for discussions on mammalian cell assays; and T. Sixma and H. Dietz for critical reading of the manuscript. **Funding:** This work was funded by EMBO long term fellowship ALTF 1295-2015 and ALTF 139-2018 (D.D.S. and B.I.M.W.); Washington Research Foundation Innovation Fellowship (D.D.S.); Human Frontiers Science Program long term fellowship (F.P. and A.C.); DARPA Biostasis HR001118S0034 (Y.H. and D.B.); Open Philanthropy Project Improving Protein Design Fund (D.B., F.P., A.K.B., and A.K.); The Audacious Project at the Institute for Protein Design (L.M.M., H.M.M., B.J.R.T., and N.I.E.); Eric and Wendy Schmidt by recommendation of the Schmidt Futures (D.B. and Y.H.); The Howard Hughes Medical Research Institute (D.B. and A.C.); and NIH Resource for Native Mass Spectrometry Guided Structural Biology P41 GM128577 (V. Wysocki, Ohio State University). **Author contributions:** Conceptualization: D.D.S., F.P., D.B.; Methodology: D.D.S., F.P.; Investigation: D.D.S., F.P., A.C., Y.H., B.I.M.W., N.I.E., L.M.M., H.M.M., B.J.R.T., J.D., A.K.B.; Visualization: D.D.S., F.P.; Funding acquisition: D.D.S., F.P., A.C., B.I.M.W., D.B.; Supervision: D.B.; Writing – original draft: D.D.S., F.P., D.B.; Writing – review and editing: D.D.S., F.P., D.B. D.D.S. and F.P. developed the hetero-oligomer design pipeline, performed design calculations and experiments, and analyzed all data. A.C. designed and characterized the homo-oligomeric C3 hub. Y.H. designed and characterized the two component C4 ring. B.I.M.W. performed and analyzed split luciferase binding assays. N.I.E. designed and characterized the homo-oligomeric C4 hub. A.C., Y.H., and N.I.E. performed nsEM and 3D reconstructions. B.J.R.T. designed scaffolds. L.M.M. and H.M.M. purified designs. D.D.S. and J.D. analyzed mammalian cell-based assays. D.D.S., A.B. and A.K. determined crystal structures. D.B. supervised research. **Competing interests:** D.D.S., F.P., A.C., N.I.E., Y.H., B.J.R.T., and D.B. are inventors on a provisional patent application submitted by the University of Washington for the design, composition, and function of the proteins created in this study. **Data and materials availability:** Crystallographic models have been deposited in the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) (accession codes 6wmk, 7mwq, and 7mwr). All data are available in the main text or the supplementary materials. Design scripts, protein sequences, design models, and models of assemblies are also available through Zenodo (59).

## SUPPLEMENTARY MATERIALS

[science.org/doi/10.1126/science.abj7662](https://science.org/doi/10.1126/science.abj7662)

Figs. S1 to S25

Tables S1 to S5

Reference (60)

MDAR Reproducibility Checklist

Data S1 and S2

[View/request a protocol for this paper from Bio-protocol.](#)

31 May 2021; accepted 13 December 2021

[10.1126/science.abj7662](https://doi.org/10.1126/science.abj7662)

## Reconfigurable asymmetric protein assemblies through implicit negative design

Danny D. SahtoeFlorian PraetoriusAlexis CourbetYang HsiaBasile I. M. WickyNatasha I. EdmanLauren M. MillerBart J. R. TimmermansJustin DecarreauHana M. MorrisAlex KangAsim K. BeraDavid Baker

*Science*, 375 (6578), eabj7662. • DOI: 10.1126/science.abj7662

### Designing dynamic protein complexes

Protein complexes play important roles in biological processes, and many complexes are dynamic, with subunits exchanging to facilitate different functions. It has been challenging to design stable and soluble monomeric proteins that reversibly associate into hetero-oligomers. Sahtoe *et al.* used a strategy called implicit negative design to construct proteins with interaction interfaces that drive association with a selected partner but not self-association. The resulting designs are stably folded in solution and provide the modules for assembly into a wide variety of complexes. They can be functionalized, allowing target proteins to be displayed in defined geometries, and complex subunits can be exchanged by varying the available concentrations of components. —VV

### View the article online

<https://www.science.org/doi/10.1126/science.abj7662>

### Permissions

<https://www.science.org/help/reprints-and-permissions>

Use of think article is subject to the [Terms of service](#)