

# Four-component protein nanocages designed by programmed symmetry breaking

<https://doi.org/10.1038/s41586-024-07814-1>

Received: 16 June 2023

Accepted: 11 July 2024

Published online: 18 December 2024

Open access

 Check for updates

Sangmin Lee<sup>1,2,3,4,6</sup>, Ryan D. Kibler<sup>1,2,6</sup>, Green Ahn<sup>1,2</sup>, Yang Hsia<sup>1,2</sup>, Andrew J. Borst<sup>1,2</sup>, Annika Philomin<sup>2</sup>, Madison A. Kennedy<sup>1,5</sup>, Buwei Huang<sup>1,2</sup>, Barry Stoddard<sup>5</sup> & David Baker<sup>1,2,3</sup>✉

Four, eight or twenty C3 symmetric protein trimers can be arranged with tetrahedral, octahedral or icosahedral point group symmetry to generate closed cage-like structures<sup>1,2</sup>. Viruses access more complex higher triangulation number icosahedral architectures by breaking perfect point group symmetry<sup>3–9</sup>, but nature appears not to have explored similar symmetry breaking for tetrahedral or octahedral symmetries. Here we describe a general design strategy for building higher triangulation number architectures starting from regular polyhedra through pseudosymmetrization of trimeric building blocks. Electron microscopy confirms the structures of T = 4 cages with 48 (tetrahedral), 96 (octahedral) and 240 (icosahedral) subunits, each with 4 distinct chains and 6 different protein–protein interfaces, and diameters of 33 nm, 43 nm and 75 nm, respectively. Higher triangulation number viruses possess very sophisticated functionalities; our general route to higher triangulation number nanocages should similarly enable a next generation of multiple antigen-displaying vaccine candidates<sup>10,11</sup> and targeted delivery vehicles<sup>12,13</sup>.

Natural and designed protein nanocages consist of one or multiple unique components arranged with tetrahedral, octahedral and icosahedral point group symmetry<sup>1,3–7,14–16</sup>. Although there are no other point group symmetries, in the icosahedral case, viruses access larger and more complex higher triangulation (T) number architectures by interspersing varying numbers of hexagons between 12 pentagonal substructures (pentons), which expand the structure without changing the curvature. Viruses achieve this symmetry breaking by placing at symmetrically non-equivalent positions either the same subunit in different conformations (quasisymmetry)<sup>3–8</sup>, or closely related but distinct subunits (pseudosymmetry)<sup>9</sup>. The accessing of higher T number icosahedral structures by such symmetry breaking is critical to the remarkable functionality of viruses, such as the ability to package and deliver large nucleic acid cargos. Similarly, the de novo design of higher T number protein assemblies could enable new approaches to nucleic acid delivery and, as the potency of nanoparticle immunogens can increase with increasing valency of display, could lead to more potent vaccines. However, although protein design has had considerable success in designing symmetric assemblies that assemble from identical interacting subunits<sup>1,14–16</sup>, the design of assemblies with multiple identical or nearly identical chains in non-symmetry equivalent positions is an outstanding challenge.

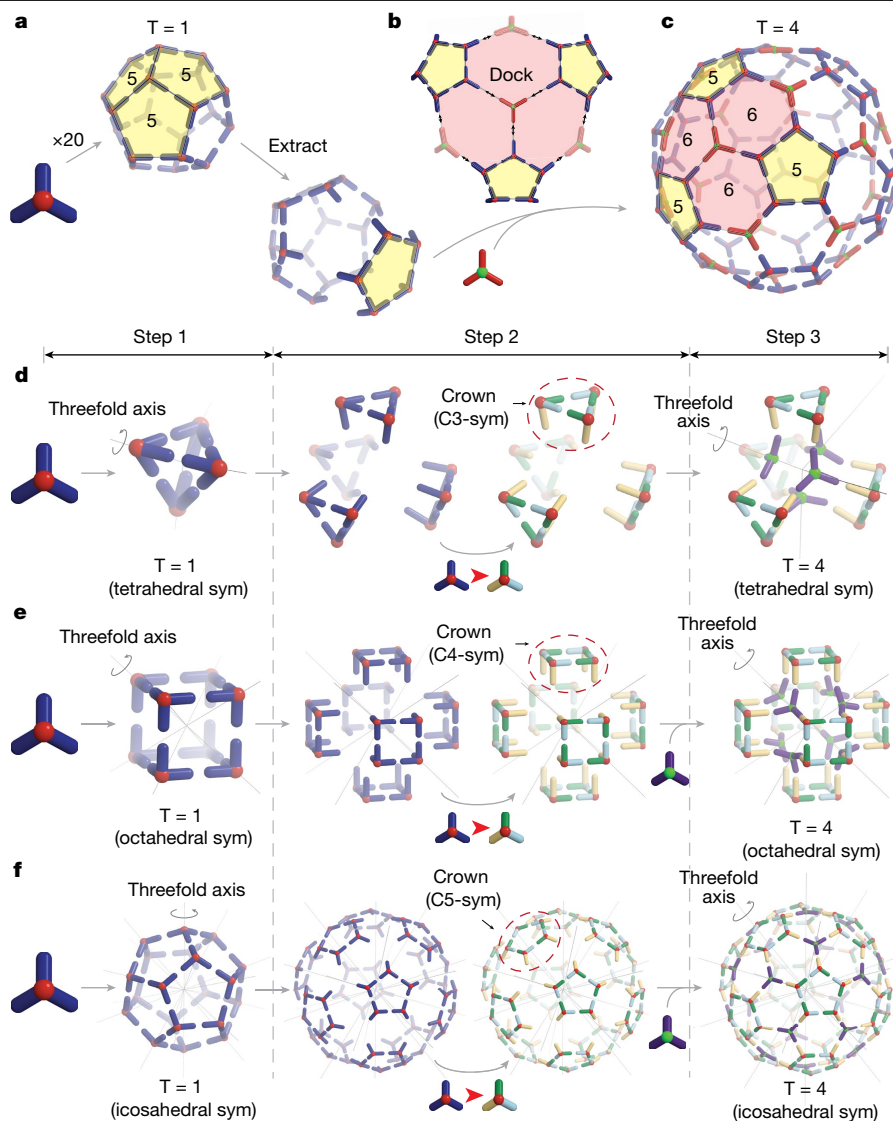
We set out to develop a systematic approach to design higher T-number protein assemblies that could generate not only icosahedral (I-sym) architectures but also tetrahedral (T-sym) and octahedral (O-sym) higher T-number nanostructures, which, to our knowledge, have not been found in nature. Of the two routes to breaking symmetry described above, we reasoned that using closely related but distinct subunits (pseudosymmetry) would have the advantage over

quasisymmetry as it avoids the complexity of designing single subunits with multiple distinct states and sets of interactions. We reasoned further that a design strategy that started with creating T = 1 architectures from homotrimers, and then extended these to T = 4 by inserting pseudosymmetric heterotrimers, would be more robust than directly designing for higher T-number cages, as subsets of the cage forming interfaces could be experimentally validated independently.

Our design strategy is illustrated in Fig. 1a for the icosahedral case. A T = 1 cage with icosahedral point group symmetry can be built from 20 homotrimers aligned along the icosahedral threefold symmetry axes (Fig. 1a). The centres of the homotrimers can be visualized as the vertices of 12 pentagonal faces that form the sides of a dodecahedron (Fig. 1a). Higher T-number cages conserve the number of pentagons (pentons), which are always aligned along the 12 icosahedral fivefold symmetry axes. T = 4 cages with hexagons bridging the 12 pentons could be generated if it were possible to extract pentons from the T = 1 cages and insert homotrimers along the icosahedral threefold symmetry axes connecting three pentons (Fig. 1b,c). Extracting pentons from T = 1 cages requires that one of the three interfaces joining each trimer with its nearest neighbours be eliminated. This can be accomplished by substituting the homotrimers with geometrically identical (that is, pseudosymmetric) heterotrimers composed of three distinct chains with different amino acid sequences. This symmetry breaking allows the interface facing outwards from the penton to be specifically eliminated, isolating the penton as a free-standing structure that can be used to build larger structures.

We reasoned that this approach could be used to design T = 4 tetrahedral, octahedral and icosahedral nanostructures using pseudosymmetric heterotrimers to precisely programme the six distinct interfaces

<sup>1</sup>Department of Biochemistry, University of Washington, Seattle, WA, USA. <sup>2</sup>Institute for Protein Design, University of Washington, Seattle, WA, USA. <sup>3</sup>Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA. <sup>4</sup>Department of Chemical Engineering, Pohang University of Science and Technology (POSTECH), Pohang, Republic of Korea. <sup>5</sup>Division of Basic Sciences, Fred Hutchinson Cancer Center, Seattle, WA, USA. <sup>6</sup>These authors contributed equally: Sangmin Lee, Ryan D. Kibler. ✉e-mail: [dabaker@uw.edu](mailto:dabaker@uw.edu)



**Fig. 1 | Overview of the design strategy.** **a–c**, General design route to  $T = 4$  icosahedral cages using substructures extracted from a  $T = 1$  cage. Twelve pentagonal substructures (pentons) from a  $T = 1$  cage (**a**) are docked with 20 homotrimers to form a closed cage structure (**b**), which creates hexagonal local structures (shaded red) placed between pentons (shaded yellow; **c**). **d–f**, Schematic of the route to  $T = 4$  tetrahedral (**d**), octahedral (**e**) and icosahedral (**f**) cages. Step 1:  $T = 1$  cages are designed starting from C3

symmetric trimeric building blocks. Step 2: the trimers constituting each face of the  $T = 1$  cages are displaced away from the origin along the symmetry axis orthogonal to the face (left) and replaced by ABC-type pseudosymmetric heterotrimers to produce crowns in which one of the three components (yellow) is free to be designed to dock to other building blocks (right). Step 3: the crowns are docked with a new set of homotrimers aligned along the threefold symmetry axis of each architecture, which produces  $T = 4$  cages.

in such structures. We adopted the three-step hierarchical approach outlined in Fig. 1d–f. We first designed  $T = 1$  cages by arranging C3 symmetric homotrimers along the threefold symmetry axis of each architecture<sup>1,2</sup> (Fig. 1d–f, step 1). Following experimental validation, we next extracted cyclic ‘crown’-like substructures from these cages (pentons in the icosahedral case) by replacing the symmetric homotrimers with ABC-type heterotrimers that lack the cage-forming trimer–trimer interface on one of the subunits, and redesigning the crown-forming trimer–trimer interface to be heterodimeric (Fig. 1d–f, step 2). Following experimental validation, these crown-like structures are arranged along one of the threefold symmetry axes of T-sym, the fourfold symmetry axis of O-sym or the fivefold symmetry axis of I-sym architectures, and docked and designed to interact with an additional symmetric homotrimer placed on the remaining threefold axis (Fig. 1d–f, step 3). This generates four component  $T = 4$  cages for each cage symmetry, with a hexameric motif between edges of the crowns ( $h = 2$  and  $k = 0$  in the Caspar and Klug nomenclature<sup>8</sup>).

## Design of pseudosymmetric heterotrimers

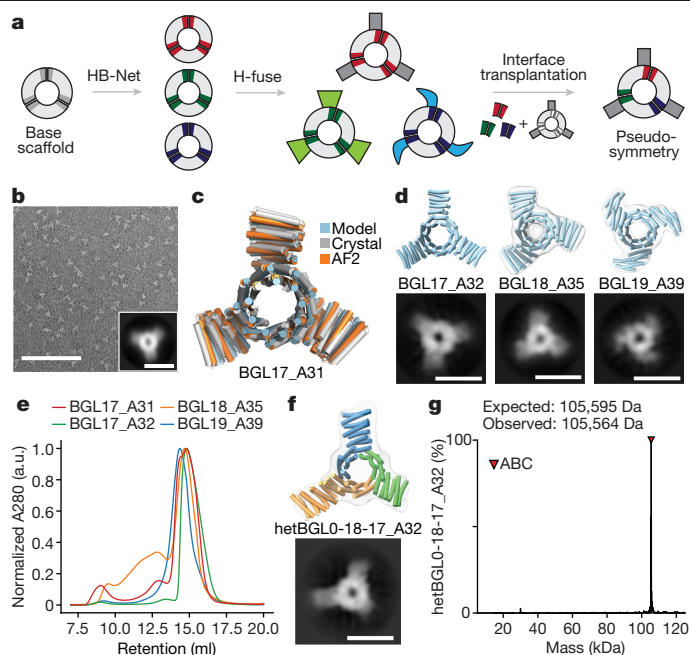
Our design strategy requires homotrimeric building blocks that can be replaced by pseudosymmetric heterotrimeric blocks with distinct intra-trimer interfaces but identical outwards-facing interfaces that mediate nanocage assembly. We chose to utilize a previously designed family of ring-shaped homotrimers and heterotrimers<sup>17</sup>, which have the same structure but different amino acid sequences at the protomer–protomer interfaces. The heterotrimer designs contain all three chains in equal stoichiometries, and both homotrimer and heterotrimer designs assemble into symmetric cyclic rings very close to the computational design models<sup>17</sup>. To generate a wider diversity of geometries to enable docking of the rings into closed architectures through helix–helix interfaces (Fig. 2a), we built on previous work<sup>17–19</sup> and rigidly fused various helical repeat protein extensions onto the termini<sup>19</sup> (referred to as ‘arms’ below), yielding homotrimers of approximately 10 nm in diameter (Fig. 2b–e, Supplementary Table 2, Supplementary Fig. 1

and Supplementary information 1). Of the designs, 37 of 39 were soluble following expression in *Escherichia coli* and purification from immobilized metal affinity chromatography, 33 of 39 had size-exclusion chromatography (SEC) retention volumes consistent with the design models and the SEC profiles were monodisperse for 22 of the designs (Supplementary Figs. 2 and 3). The homotrimeric state was confirmed by native mass spectrometry (nMS) for 27 of 39 designs (Supplementary Figs. 4–6), and small-angle X-ray scattering profiles were closely consistent with profiles computed from the design models for 19 (Supplementary Figs. 7 and 8). Negative-stain electron microscopy (nsEM) characterization of the overall shapes of the designs was again consistent with the design models, with curved, straight, wide and narrow arm arrangements evident in 2D class averages and 3D reconstructions (Fig. 2b,d and Supplementary Fig. 9). We obtained a crystal structure of one of the designs, BGL17\_A31 (see Supplementary information 1.3 for protein naming), which was very close to the design model and AlphaFold2 (AF2) prediction (Fig. 2c and Extended Data Table 2).

To design pseudosymmetric heterotrimers, we used an interface transplantation approach (Supplementary Figs. 10 and 11), preserving the overall structural C3 symmetry (Supplementary information 2). A homotrimer was selected as the host scaffold and three different protomer–protomer interfaces from different homo-oligomers (guests) were transplanted onto the subunit–subunit interfaces (Fig. 2a and Supplementary Table 1), conserving the residues at the arm junction. We first checked the compatibility between a host and a guest in the homo-oligomer context by symmetrically transplanting the guest interface into the host. We selected highly expressed and homogenous homotrimers for hybridization: seven with arms (BGLO\_A10, BGLO\_A11, BGL17\_A31, BGL17\_A32, BGL18\_A35, BGL19\_A38 and BGL19\_A39) as host scaffolds and four without arms (BGLO, BGL17, BGL18 and BGL19)<sup>17</sup> as guest interfaces. We experimentally characterized 18 of the 28 combinations (Supplementary Table 3); 14 of 18 had SEC peaks at the correct oligomeric size (Supplementary Fig. 12) and 11 of 18 had strong nMS signals at the correct masses (Supplementary Fig. 13). The BGL17\_A32 host backbone was found to be compatible with multiple guest interfaces; we constructed heterotrimers by splicing interfaces from different guests together in different combinations<sup>17</sup> (Supplementary Table 4). To enable assignment of chain type by electron microscopy, we varied the number of repeat units on the arms protruding from each chain (–1 repeat for the A component and +1 to the C component). Three of five heterotrimers (hetBGL0-17-19\_A32, hetBGL0-18-17\_A32 and hetBGL0-19-17\_A32; see Supplementary information 2 for protein naming) formed ABC-type heterotrimers as shown by SEC, SDS–PAGE and nMS (Fig. 2f and Supplementary Figs. 14–17). For two of these (hetBGL0-18-17\_A32 and hetBGL0-19-17\_A32), the expected differences in arm lengths were clearly evident in 2D nsEM averages and 3D reconstructions (Fig. 2f and Supplementary Fig. 18). This approach to creating pseudosymmetric hetero-oligomers should be readily applicable to symmetric protein oligomers quite generally, if the amino acid sequence of the protomer–protomer interface can be diversified.

## Cage design

We generated base T = 1 tetrahedral, octahedral and icosahedral cages from the BGL17\_A32 homotrimer using RPXdock<sup>20,21</sup> to sample rotational and translational displacements of the trimer C3 axis along the threefold cage axes (Supplementary information 3). We found that for the different symmetries, different numbers of repeat units of arms (1.5, 2.5 and 3.5 repeat units for tetrahedral, octahedral and icosahedral cages, respectively) were optimal for docking of the trimers with good shape complementarity (Fig. 3a,b,e,f,i,j and see Supplementary information 3.1). The newly generated cage interfaces were designed using ProteinMPNN<sup>22</sup>, and designs for which the AF2 prediction of the arm–arm interface was less than 2.0 Å root-mean-square deviation

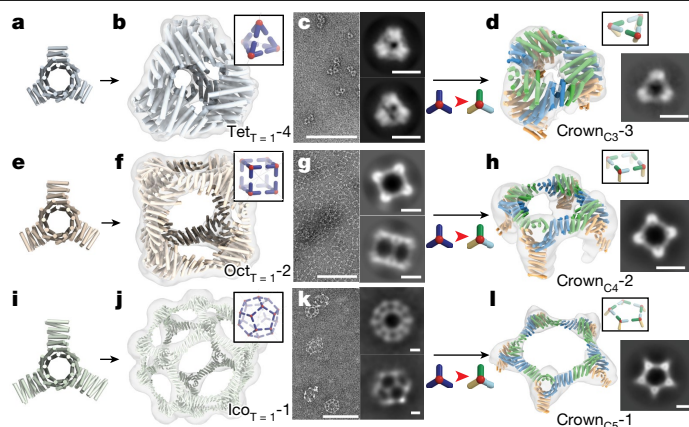


**Fig. 2 | Design of pseudosymmetric ABC heterotrimers with extensible arms.** **a**, Schematic of the design strategy for homotrimers with arms and pseudosymmetric heterotrimers. **b**, nsEM micrograph of the BGL17\_A31 homotrimer and a 2D class average (inset) along a threefold symmetry axis. **c**, Structure of BGL17\_A31 solved by X-ray crystallography (white) compared with its design model (light blue) and AF2 prediction (orange). **d**, Superpositions of the 3D-reconstructed nsEM map (transparent cloud) to the structure model (blue) of BGL17\_A32 (no map; top left), BGL18\_A35 (top middle) and BGL19\_A39 (top right). 2D class averages along threefold symmetry axis are also shown (bottom). **e**, SEC results of homotrimers. **f**, ABC-type pseudosymmetric heterotrimer (hetBGL0-18-17\_A32) with arms of different lengths extending from each component. Top, the 3D-reconstructed nsEM map (transparent cloud) superimposed on the design model (colours indicate the three different chains). Bottom, a 2D class average along the pseudo-threefold symmetry axis. **g**, nMS analysis of hetBGL0-18-17\_A32; only the heterotrimeric ABC species is observed. Scale bars, 100 nm (**b**) and 10 nm (**b** (inset), **d**, **f**). See Extended Data Table 1 for the amino acid sequences of the oligomers.

(RMSD) from the design model were selected for experimental characterization; these comprise seven tetrahedral cages (Tet<sub>T=1</sub>-1 to Tet<sub>T=1</sub>-7; Supplementary Table 5) with 12 subunits and diameters of approximately 13 nm, 8 octahedral cages (Oct<sub>T=1</sub>-1 to Oct<sub>T=1</sub>-8; Supplementary Table 6) with 24 subunits and diameters of approximately 20 nm, and 4 icosahedral cages (Ico<sub>T=1</sub>-1 to Ico<sub>T=1</sub>-4; Supplementary Table 7) with 60 subunits and diameters of approximately 40 nm. All of the tetrahedral designs had peaks at the expected retention volume (Supplementary Fig. 20) on SEC, and four (Tet<sub>T=1</sub>-1, Tet<sub>T=1</sub>-2, Tet<sub>T=1</sub>-4 and Tet<sub>T=1</sub>-6) were structurally homogeneous by nsEM (Extended Data Fig. 1 and Supplementary Fig. 21), with 2D class averages and 3D-reconstructed nsEM maps (Fig. 3c and Supplementary Fig. 21) matching the design models (Fig. 3b,c). Seven of the octahedral cages had single peaks on SEC, and two (Oct<sub>T=1</sub>-2 and Oct<sub>T=1</sub>-4) showed homogenous structures matching the design models by nsEM (Fig. 3f,g and Supplementary Figs. 22 and 23). One of the icosahedral cages, Ico<sub>T=1</sub>-1, had a single peak on SEC and was close to the design model by nsEM (Fig. 3j,k), although imperfectly formed cages were also observed (Supplementary Fig. 24).

We next extracted C3, C4 and C5 symmetric cyclic oligomers (which we refer to as crowns because of their shape) from the structurally confirmed T = 1 cages by substituting in the structurally identical pseudosymmetric hetBGL0-18-17\_A32 heterotrimer in place of the BGL17\_A32 homotrimer (Fig. 1, third and fourth columns). The chain A (ch\_A) and ch\_B interfaces were redesigned using ProteinMPNN to





**Fig. 3 | Extraction of homotrimer cycles (crowns) from  $T = 1$  cages by pseudosymmetrization.** **a–c, e–g, i–k**, BGL17\_A32 with 1.5 (**a**), 2.5 (**e**) and 3.5 (**i**) repeat unit arms docked into tetrahedral ( $Tet_{T=1-4}$ ; **b,c**), octahedral ( $Oct_{T=1-2}$ ; **f,g**) and icosahedral ( $Ico_{T=1-1}$ ; **j,k**)  $T = 1$  cages. Superpositions of the 3D-reconstructed nsEM map (transparent cloud) on the cage design model (colours) are shown (**b,f,j**). nsEM micrographs (left) and characteristic 2D class averages (right) of the cages are also displayed (**c,g,k**). **d,h,i**, C3 (crown<sub>C3-3</sub>), C4 (crown<sub>C4-2</sub>) and C5 (crown<sub>C5-1</sub>) crowns made from pseudosymmetric heterotrimers. Left, superpositions of the 3D-reconstructed nsEM map on the crown design model (ch\_A (green), ch\_B (blue) and ch\_C (orange)). Right, 2D class averages along threefold, fourfold and fivefold symmetry axes. The diameters of the crowns are 11 nm (C3), 20 nm (C4) and 35 nm (C5). Scale bars, 100 nm (**c,g,k** (left)) and 10 nm (**c,g,k** (right), **d,h,i**). See Extended Data Tables 3 and 4 for the amino acid sequences of the  $T = 1$  cages and crowns.

interact at the crown trimer–trimer interface; this was necessary to avoid a potential off-target structure possible with the original C2 interface (Supplementary Fig. 25). The surface of the arm of ch\_C, which points outwards from the crown, was redesigned to be entirely polar. This isolates the crowns as free-standing building blocks that can be used to build  $T = 4$  cages in the next step. We selected designs for which AF2 predicted the ch\_A–ch\_B interface with RMSD  $< 2$  Å and did not predict the ch\_A–ch\_A or ch\_B–ch\_B homodimer interfaces to form (Supplementary Fig. 19). We obtained genes encoding 19 sets of crowns that passed these filters (crown<sub>C3-1</sub> to crown<sub>C3-5</sub> for C3 crowns, crown<sub>C4-1</sub> to crown<sub>C4-7</sub> for C4 crowns and crown<sub>C5-1</sub> to crown<sub>C5-7</sub> for C5 crowns; Supplementary Tables 8–10), and the three chains for each crown were expressed separately in independent *E. coli* cultures. Following expression, the amount of each protein was estimated by SDS–PAGE gel densitometry, and appropriate amounts of culture were combined to achieve mixtures with stoichiometric amounts of the three chains, which were co-lysed and co-purified. By SEC, four of five C3, five of seven C4 and two of seven C5 crowns had peaks at the expected elution volumes (Supplementary Figs. 26–31) containing three distinct bands by SDS–PAGE (Supplementary Fig. 32), indicating that the complexes were heterotrimers. nsEM 2D class averages and 3D-reconstructed nsEM maps matched well with the crown design models (Fig. 3d,h,i and Extended Data Fig. 1). Thus, symmetric substructures can be extracted from larger symmetric assemblies by substituting homotrimers with pseudosymmetric heterotrimers.

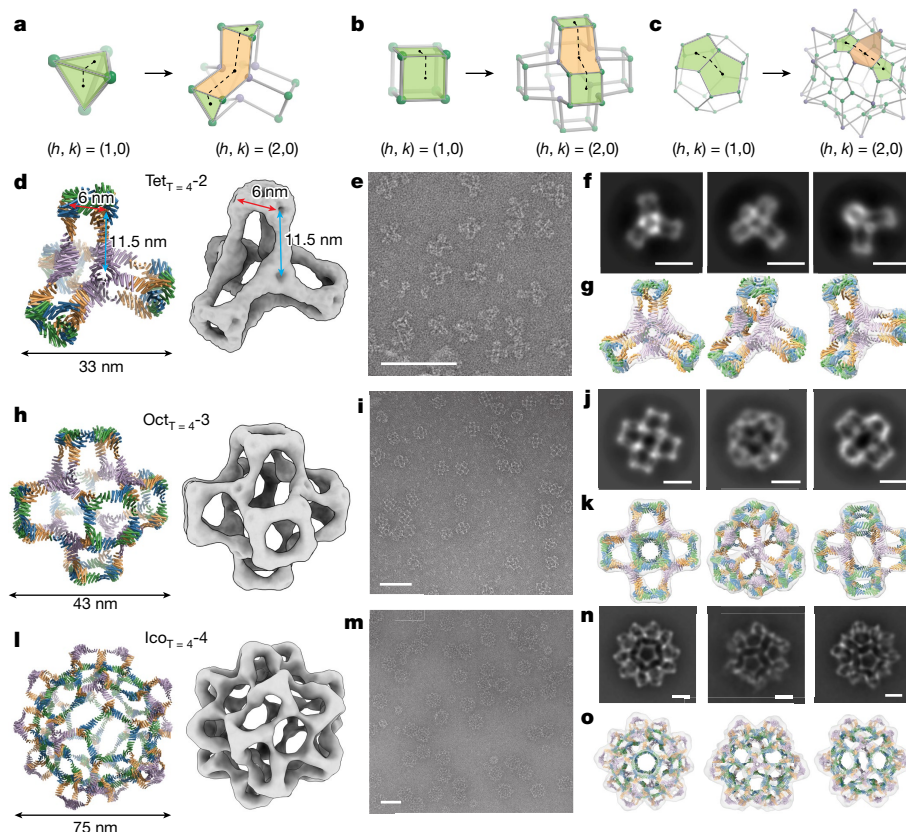
In the final step of our hierarchical design approach, we designed  $T = 4$  cages by combining the experimentally confirmed crowns with the BGL17\_A32 homotrimer (Fig. 1, last column, and Supplementary information 3). The C3, C4 and C5 crowns were aligned with the threefold, fourfold and fivefold axes of tetrahedral, octahedral and icosahedral architectures, and BGL17\_A32 was aligned to the remaining threefold axis. This generated assemblies in which the heterotrimer arms pointing outwards from the crowns (ch\_C of heterotrimer) interact with the arms of the homotrimer (ch\_ho). To find optimal docking interfaces, we used RPXdock, sampling the lengths of the interacting arms and

the rotations and translations along the common axis, and designed sequences using proteinMPNN for the highest RPX scoring models for each symmetry. Designs were filtered based on the formation of the designed interface, and lack of formation of the self interfaces, in AF2 predictions (Supplementary Fig. 19). Very few initial designs for the octahedral architecture did not form predicted self interfaces; thus, to decrease the probability of self-interaction, we performed explicit negative design using proteinMPNN<sup>22</sup> against the predicted self interfaces. We experimentally tested 14 sets of  $T = 4$  cage designs ( $Tet_{T=4-1}$  to  $Tet_{T=4-5}$ ,  $Oct_{T=4-1}$  to  $Oct_{T=4-4}$  and  $Ico_{T=4-1}$  to  $Ico_{T=4-5}$ ; Supplementary Tables 11–13) that passed the AF2 filters. The four components were expressed independently in different *E. coli* cultures, mixed with 1:1:1:1 stoichiometry and co-lysed. The lysed samples were purified using immobilized metal affinity chromatography and SEC, and the cage structures were characterized by nsEM (Extended Data Fig. 1 and Supplementary Figs. 35–47). As described in the following paragraphs, the major species in each case was the designed  $T = 4$  structure; we also observed minor species of smaller off-target  $T = 1$ -like cages (Supplementary Figs. 37, 42 and 45).

The  $T = 4$  tetrahedral cage ( $Tet_{T=4-2}$ ) has a tetrapod shape (diameter of 33 nm) with the four C3 crowns pointing outwards, and the homotrimers bridging the crowns closer to the centre of the cage and facing inwards (Fig. 4a,d,g and Supplementary Fig. 36). The homotrimer–heterotrimer distance (11.5 nm) is almost twice the heterotrimer–heterotrimer distance (6.0 nm) due to the arm length difference between components (1.5 repeat units for ch\_A and ch\_B, and 3 repeat units for ch\_C and ch\_ho), and the interior volume is a tetrahedral channel 6.0 nm in width (Fig. 4d). Overall, the structure maps to a  $T = 4$  Goldberg polyhedra with tetrahedral symmetry<sup>23</sup>, in which the hexagonal motifs between triangles are highly elongated (Fig. 4a). These structural features are evident in the nsEM map (Fig. 4d), micrographs (Fig. 4e) and 2D average classes (Fig. 4f), and the design model is closely consistent with the reconstructed nsEM map (Fig. 4g). The design model could be readily relaxed to fit the nsEM 3D map with all four components clearly within density (Supplementary Fig. 36); overall, the relaxed model matches well with the design model, with the exception of a slight twist of the overall structure resulting from curvature in the arm near the homotrimer–heterotrimer interface (Fig. 4d and Supplementary Fig. 36).

The  $T = 4$  octahedral cage ( $Oct_{T=4-3}$ ) has a 3D cross shape structure (diameter of 43 nm) with the original cubic shape of the  $T = 1$  structure repeated six times and shifted away from the origin to positive and negative values of  $x$ ,  $y$  and  $z$  along the three fourfold symmetry axes (Fig. 4b,h). Six C4 crowns form the outward faces of the structure along the  $x$ ,  $y$  and  $z$  axes, which are connected by eight homotrimers placed in a cubic arrangement closer to the centre of the cage; as for the tetrahedral cage, the homotrimers and heterotrimers face in opposite directions (Supplementary Fig. 40). The overall architecture is that of a  $T = 4$  Goldberg polyhedra with octahedral symmetry<sup>23</sup>, with an elongated hexagon bridging the square faces (Fig. 4b) and a 10-nm cavity in the centre. Homogeneous populations of cages are observed in nsEM micrographs (Fig. 4i), and 2D class averages along the twofold, threefold and fourfold symmetry axes are closely consistent with the design model (Fig. 4j). The nsEM 3D reconstruction is very close to the design model (Fig. 4k), but a slightly curved connection between crowns and homotrimers leading to a slight twist of the overall structure (Supplementary Fig. 40). To characterize the structure of  $Oct_{T=4-3}$  at higher resolution, we collected cryo-electron microscopy (cryo-EM) data and generated a 3D reconstruction, which, following refinement, resulted in a 3D cryo-EM map with 6.87 Å resolution with clear secondary structure features (Fig. 5a–d and Supplementary Fig. 51). Following relaxation via molecular dynamics<sup>24</sup>, the design model fits well into the cryo-EM map (Fig. 5e–g), with the crown and homotrimer substructures and the individual chains clearly defined (Supplementary Fig. 41). Around the fourfold symmetry axes, ch\_A and ch\_B of the hetBGL0-18-17\_A32





**Fig. 4 | nsEM characterization of designed T = 4 tetrahedral, octahedral and icosahedral protein cages.** **a–c**, Ball-and-stick models of T = 1 (left) and T = 4 (right) cages for each symmetry, defined by the Caspar and Klug nomenclature<sup>8</sup>. **d–o**, Four component T = 4 tetrahedral cage with 48 subunits (**d–g**), octahedral cage with 96 subunits (**h–k**) and icosahedral cage with 240 subunits (**l–o**). Design models of the T = 4 cages (left) and a 3D-reconstructed nsEM map (right) are shown (**d,h,l**). In the design models, each chain is shown in a distinct

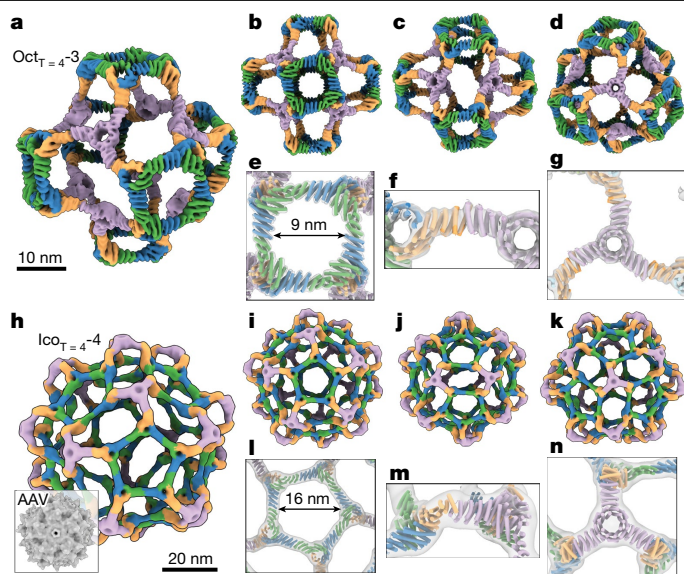
colour (ch\_A (green), ch\_B (blue), ch\_C (orange) and ch\_ho (purple)). nsEM micrographs (**e,i,m**), characteristic 2D class averages of nsEM (**f,j,n**) and superpositions of the 3D-reconstructed nsEM map (grey density) on the cage design model (colours; **g,k,o**) are also displayed. Scale bars, 100 nm (**e,i,m**) and 20 nm (**f,j,n**). See Extended Data Table 5 for the amino acid sequences of the T = 4 cages.

heterotrimers (Fig. 5b,e, green and blue) form square motifs; the five helices in each arm in both chains are clearly evident in the cryo-EM map (Fig. 5e). BGL17\_A32 homotrimers are placed along the threefold symmetry axes (Fig. 5d); the arm of each subunit has six helices and forms an interface with ch\_C of the heterotrimer at the end (Fig. 5f,g, purple and orange). The slight twist of the C4 crown substructures relative to the design model arises from shifts at the homotrimer–heterotrimer interface (Supplementary Fig. 41).

The T = 4 icosahedral cage ( $Ico_{T=4-4}$ ) consists of 12 C5 crowns (pentons) connected by 20 outwards-facing homotrimers (Fig. 4c,l). Largely homogeneous 75-nm-sized cages were identified by nsEM and dynamic light scattering (DLS; Fig. 4m and Supplementary Fig. 44). SDS–PAGE showed clear bands corresponding to each component, suggesting that all four chains are present (Supplementary Fig. 43). The 2D class averages (Fig. 4n) and 3D nsEM reconstructions (Fig. 4l,o) have the overall designed shape but the orientations of the C5 crowns, and homotrimers appeared inverted from the design model (Supplementary Figs. 34 and 47). We collected cryo-EM images, and 3D reconstruction and refinement of the cryo-EM data yielded a 3D cryo-EM map with 13.15 Å resolution (Fig. 5h–k and Supplementary Fig. 51), in which the holes at the centre of trimers and the orientations of trimers are clearly identified. The design model with inverted components fits well into the cryo-EM density following relaxation (Fig. 5l–n and Supplementary Fig. 48). The overall structure has the architecture of a T = 4 Goldberg polyhedron with icosahedral symmetry<sup>23</sup>, with boat-type hexagonal motifs placed between pentagons (Fig. 4c). Surrounding the fivefold symmetry axes are pentons formed from hetBGL0-18-17\_A32 heterotrimers

(Fig. 5h,i, green, blue and orange). The pentons are bridged by BGL17\_A32 homotrimers, which form tripod-like protrusions on the threefold symmetry axis (Fig. 5h,k, purple). On the twofold axes are boat-type distorted hexagons with two homotrimers and four heterotrimers on the vertices; two of the edges are formed by interacting heterotrimer subunits, and four edges are formed by interacting homotrimer and heterotrimer subunits (Fig. 5j). The outer diameter of the  $Ico_{T=4-4}$  cage is about three times larger than that of the adeno-associated virus capsid (Fig. 5h), and the inner diameter of the empty pore at the centre of the  $Ico_{T=4-4}$  cage is approximately 50 nm (volume of approximately  $6.55 \times 10^4 \text{ nm}^3$ ), which can be used to package diverse cargos such as nucleic acids and enzymes.

We tested the thermal stability and pH tolerance of the  $Oct_{T=4-3}$  and  $Ico_{T=4-4}$  cages using DLS (Supplementary information 4.5). The Z-average diameter ( $d_z$ ) of the two cages was consistent with the expected diameters of the intact cages (approximately 80 nm for  $Ico_{T=4-4}$  and approximately 50 nm for  $Oct_{T=4-3}$ ) at 25 °C (the smaller T = 1 contaminants observed by electron microscopy were not resolved by DLS probably due to the screening effect of the larger of two similarly sized particles<sup>25</sup>), was unchanged from 25–70 °C and then decreased above 75 °C (Fig. 6a,b). To test pH tolerance of the cages, we measured  $d_z$  while decreasing the pH of the buffer. The  $Ico_{T=4-4}$   $d_z$  of around 80 nm was unchanged from pH 8.0 to pH 5.3, but increased at pH 4.7, probably due to aggregation (Fig. 6c), whereas the  $Oct_{T=4-3}$  of approximately 50 nm was unchanged between pH 8.0 and pH 6.4, but increased at pH 5.9 or lower (Fig. 6d). Thus, the cages are quite thermostable (melting temperatures higher than 70 °C), which should facilitate



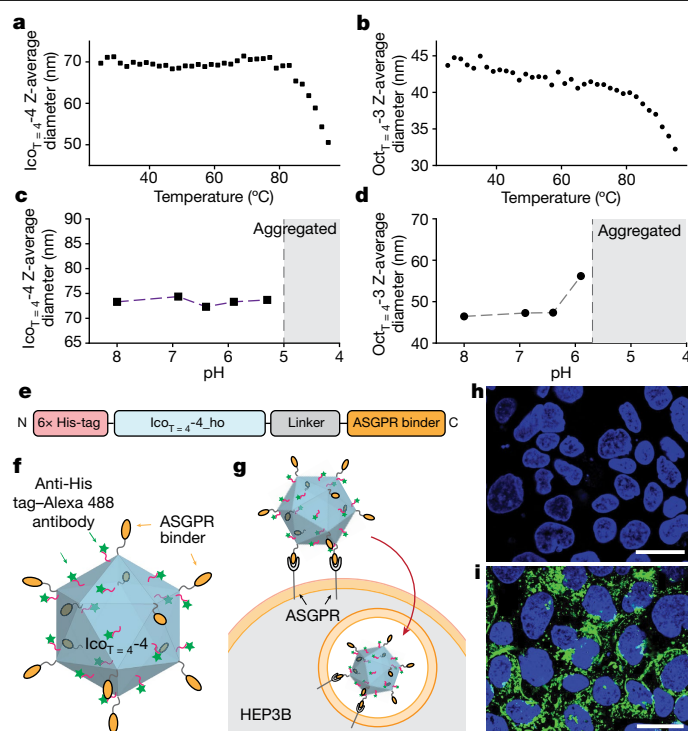
**Fig. 5 | Cryo-EM characterization of  $T = 4$  octahedral and icosahedral protein cages.** **a–d**, 3D cryo-EM map of  $\text{Oct}_{T=4-3}$  from different views. **e–g**, Overlay between the cryo-EM map (grey transparent) and the design model relaxed into the map (colours) for each substructure: C4 crown (**e**), the homotrimer–heterotrimer interface (**f**) and the homotrimer (**g**). **h–k**, 3D cryo-EM map of  $\text{Ico}_{T=4-4}$  from different views. The inset (**h**) is an adeno-associated virus (AAV) capsid shown at the same scale for size comparison. **l–n**, Overlay between the cryo-EM map (continuous density) and the relaxed design model (colours) for each substructure: C5 crown (**l**), the homotrimer–heterotrimer interface (**m**) and the homotrimer (**n**). Both the cryo-EM map and the protein model are coloured by chain (ch\_A (green), ch\_B (blue), ch\_C (orange) and ch\_ho (purple)). See Extended Data Table 7 for cryo-EM data collection information.

their development as delivery vehicles, and undergo pH-dependent transitions at the pH sampled in the endocytic pathway, which may be exploitable for endosomal escape.

As a first step towards exploring the use of the cages for delivery applications, we evaluated internalization by fusing one of the four components of  $\text{Ico}_{T=4-4}$  (ch\_ho) to a designed binder<sup>26</sup> of asialoglycoprotein receptor (ASGPR) (Fig. 6e and Extended Data Table 6). ASGPR is an endocytic internalizing receptor that is exclusively expressed on the surface of liver cells. This receptor has been utilized for delivery of therapeutic oligonucleotides such as small interfering RNA<sup>27</sup> and liver-specific LYTAC-mediated protein degradation<sup>26,28</sup>. Sixty ASGPR mini-binders were displayed on the outer surface of the  $\text{Ico}_{T=4-4}$  cage through genetic fusion (Fig. 6f). We labelled the ASGPR cages with anti-His tag–Alexa-488 antibody (the subunits each have His tags) and incubated them with the hepatocellular carcinoma cell line HEP3B. By confocal imaging, we observed a substantial intracellular accumulation of the cage in a punctate localization pattern, probably indicating vesicle localization, whereas treatment with only anti-His–Alexa-488 antibody without the cage resulted in no intracellular signal (Fig. 6h,i). This result suggests that these large designed  $T = 4$  cages can be internalized by cells for delivery or LYTAC applications.

## Conclusion

Our  $T = 4$  assemblies are a considerable advance in compositional and structural complexity over previous nanocage designs, which have one or two distinct components and up to three distinct interfaces; the designs described here have four distinct core structural components and six distinct interfaces (Supplementary Fig. 33). Our design approach is not limited to  $T = 4$  cages but can be extended to higher



**Fig. 6 | Thermal stability and pH tolerance of cages and internalization of the  $T = 4$  icosahedral cage.** **a–d**, Z-average diameter of  $\text{Ico}_{T=4-4}$  and  $\text{Oct}_{T=4-3}$  at varying temperature (**a**, **b**) and pH (**c**, **d**), measured by DLS. **e**, An ASGPR binding cage was generated by fusing the ch\_ho chain of  $\text{Ico}_{T=4-4}$  to an ASGPR-binding protein. **f**, **g**, Schematic of the ASGPR cage (**f**) and internalization of the cage in HEP3B cells (**g**). **h**, **i**, Confocal microscope images of HEP3B cells treated without (**h**) and with (**i**) the cages. Nuclei were stained with DAPI (blue), and nanocages with anti-His–Alexa-488 are shown in green. Scale bars, 30  $\mu\text{m}$  (**h**, **i**).

$T$ ; for example,  $T = 7$  cages with 420 subunits can be constructed by incorporating a second heterotrimer, based on either another BGL homodimer or using a different set of designed heterotrimers<sup>29</sup> (see Supplementary Figs. 49 and 50 for routes to  $T = 7$  and  $T = 9$ ). Our design approach can be applied to any set of protein oligomers that can be pseudosymmetrized and have dockable interfaces for the target point group symmetry. To our knowledge, pseudosymmetric tetrahedral and octahedral assemblies have not been observed in nature; with four distinct chains and eight termini available for functionalization, these provide starting points for sophisticated multicomponent materials. With a 75-nm diameter, our  $T = 4$  icosahedral design has considerably more interior volume available for packaging nucleic acid and other cargoes than previously designed cages; the accompanying paper by Dowling et al.<sup>30</sup> (see ref. 31 for a high-level comparison of approaches) describes a route to even larger, albeit less homogeneous, structures using two-component heterotrimers. For vaccine and delivery applications, the four distinct structural components of our  $T = 4$  assemblies provide many more opportunities for further functional elaboration than the one-component and two-component nanoparticles designed to date; for vaccines, presentation of multiple antigens and immunomodulators; and for targeted delivery, incorporation of distinct designed modules for targeting and internalization.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-024-07814-1>.

1. King, N. P. et al. Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **336**, 1171–1174 (2012).
2. Hsia, Y. et al. Design of a hyperstable 60-subunit protein icosahedron. *Nature* **535**, 136–139 (2016).
3. Rayment, I., Baker, T. S., Caspar, D. L. D. & Murakami, W. T. Polyoma virus capsid structure at 22.5 Å resolution. *Nature* **295**, 110–115 (1982).
4. Liddington, R. C. et al. Structure of simian virus 40 at 3.8-Å resolution. *Nature* **354**, 278–284 (1991).
5. Johnson, J. E. Functional implications of protein–protein interactions in icosahedral viruses. *Proc. Natl Acad. Sci. USA* **93**, 27–33 (1996).
6. Prasad, B. V. V. et al. X-ray crystallographic structure of the Norwalk virus capsid. *Science* **286**, 287–290 (1999).
7. Baker, T. S., Olson, N. H. & Fuller, S. D. Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. *Microbiol. Mol. Biol. Rev.* **63**, 862–922 (1999).
8. Caspar, D. L. D. & Klug, A. Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* **27**, 1–24 (1962).
9. De Colibus, L. et al. Assembly of complex viruses exemplified by a halophilic euryarchaeal virus. *Nat. Commun.* **10**, 1456 (2019).
10. Marcandalli, J. et al. Induction of potent neutralizing antibody responses by a designed protein nanoparticle vaccine for respiratory syncytial virus. *Cell* **176**, 1420–1431.e17 (2019).
11. Boyoglu-Barnum, S. et al. Quadrivalent influenza nanoparticle vaccines induce broad protection. *Nature* **592**, 623–628 (2021).
12. Tong, G. J., Hsiao, S. C., Carrico, Z. M. & Francis, M. B. Viral capsid DNA aptamer conjugates as multivalent cell-targeting vehicles. *J. Am. Chem. Soc.* **131**, 11174–11178 (2009).
13. Banskota, S. et al. Engineered virus-like particles for efficient in vivo delivery of therapeutic proteins. *Cell* **185**, 250–265.e16 (2022).
14. Lai, Y.-T., Cascio, D. & Yeates, T. O. Structure of a 16-nm cage designed by using protein oligomers. *Science* **336**, 1129 (2012).
15. King, N. P. et al. Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510**, 103–108 (2014).
16. Lai, Y.-T. et al. Structure of a designed protein cage that self-assembles into a highly porous cube. *Nat. Chem.* **6**, 1065–1071 (2014).
17. Kibler, R. D. et al. Stepwise design of pseudosymmetric protein hetero-oligomers. Preprint at *bioRxiv* <https://doi.org/10.1101/2023.04.07.535760> (2023).
18. Brunette, T. et al. Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580–584 (2015).
19. Hsia, Y. et al. Design of multi-scale protein complexes by hierarchical building block fusion. *Nat. Commun.* **12**, 2294 (2021).
20. Fallas, J. A. et al. Computational design of self-assembling cyclic protein homo-oligomers. *Nat. Chem.* **9**, 353–360 (2017).
21. Sheffler, W. et al. Fast and versatile sequence-independent protein docking for nanomaterials design using RPxDock. *PLoS Comput. Biol.* **19**, e1010680 (2023).
22. Dauparas, J. et al. Robust deep learning–based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022).
23. Schein, S. & Gayed, J. M. Fourth class of convex equilateral polyhedron with polyhedral symmetry related to fullerenes and viruses. *Proc. Natl Acad. Sci. USA* **111**, 2920–2925 (2014).
24. Kidmose, R. T. et al. Namdinator — automatic molecular dynamics flexible fitting of structural models into cryo-EM and crystallography experimental maps. *IUCrJ* **6**, 526–531 (2019).
25. Tomaszewska, E. et al. Detection limits of DLS and UV-Vis spectroscopy in characterization of polydisperse nanoparticles colloids. *J. Nanomater.* **2013**, 313081 (2013).
26. Huang, B. et al. Designed endocytosis-inducing proteins degrade targets and amplify signals. *Nature* <https://doi.org/10.1038/s41586-024-07948-2> (2024).
27. Springer, A. D. & Dowdy, S. F. GalNAc–siRNA conjugates: leading the way for delivery of RNAi therapeutics. *Nucleic Acid Ther.* **28**, 109–118 (2018).
28. Ahn, G. et al. LYACs that engage the asialoglycoprotein receptor for targeted protein degradation. *Nat. Chem. Biol.* **17**, 937–946 (2021).
29. Bermeo, S. et al. De novo design of obligate ABC-type heterotrimeric proteins. *Nat. Struct. Mol. Biol.* **29**, 1266–1276 (2022).
30. Dowling, Q. et al. Hierarchical design of pseudosymmetric protein nanocages. *Nature* <https://doi.org/10.1038/s41586-024-08360-6> (2024).
31. Lee, S. et al. Expanding protein nanocages through designed symmetry-breaking. *Baker Lab* <https://www.bakerlab.org/wp-content/uploads/2024/04/Expanding-protein-nanocages-through-designed-symmetry-breaking.pdf> (2024).

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



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

© The Author(s) 2024



## Methods

### Computational design

All computational design methods are described in Supplementary information 1–3.

### Protein expression

Plasmids (100 ng) were transformed into a chemically competent *E. coli* expression strain, BL21(DE3) or BL21(DE3)Star, for protein expression following the manufacturer's protocol, with the exception of using 10  $\mu$ l competent cells per reaction. Following transformation and recovery, the entire transformation products were used to inoculate 1 ml LB medium containing 100  $\mu$ g ml<sup>-1</sup> kanamycin and grown at 37 °C with shaking at 225 rpm overnight. Of overnight cultures, 500  $\mu$ l was diluted into 50 ml TBM-5052 supplemented with 100  $\mu$ g ml<sup>-1</sup> kanamycin in 250 ml baffled flasks and incubated at 37 °C with shaking at 225 rpm for 18–24 h.

For systems using hetero-oligomers (pseudosymmetric hetero-oligomers, crowns and T = 4 cages), we used two different protocols: protocol 1 and protocol 2.

**Protocol 1.** Each component was transformed and incubated independently. After the incubation, we first checked relative expression level between components using SDS–PAGE (Supplementary information 4.5) based on the intensity of bands. Then, the incubated *E. coli* cultures of each component were mixed with a correct stoichiometry for target assemblies (1:1:1 for heterotrimers and crowns; 1:1:1:1 for T = 4 cages). After mixing the *E. coli* cultures, we harvested proteins following Supplementary information 4.4.

**Protocol 2.** Plasmids of all components were transformed into *E. coli* together. For example, in heterotrimer expression (three components: A, B and C), 300 ng of plasmids (100 ng of A + 100 ng of B + 100 ng of C) were transformed into 10  $\mu$ l *E. coli* cells and incubated as described in this section. Then, proteins were harvested following Supplementary information 4.4.

### Protein purification

**Immobilized metal affinity chromatography.** Cultures were harvested by centrifugation at 4,000g for 10 min, culture supernatant was decanted, and pellets were resuspended to 30 ml in lysis buffer. Of PMSF, 300  $\mu$ l (100 mM in 100% ethanol) was added immediately before sonication at 70% power for 5 min. 'Lysate' fractions were saved, and then lysates were clarified by ultracentrifugation at 18,000g at 12 °C for at least 30 min and applied to 1.5-ml Ni-NTA resin (Qiagen) pre-equilibrated with lysis buffer and packed into Econo-Pac columns (Bio-Rad) for gravity chromatography. The columns were washed twice with 15 ml wash buffer and eluted with 10 ml elution buffer. Hetero-oligomers were purified according to a similar procedure, except they used the 'Het' variants of the lysis, wash and elution buffers, as this was found to improve the yield of single-His-tagged complexes over nonspecific multiple-His-tagged complexes, which can arise at high concentrations and probably dominate binding to the Ni-NTA resin. Samples prepared for crystallization were treated similarly, except 500-ml cultures were used and lysate was divided among six gravity columns.

**SEC.** Immobilized metal affinity chromatography (IMAC)-eluted samples were concentrated using 10k MWCO spin concentrators and were purified using a Superdex 200 10/300 increase (for oligomers) or Superdex 6 10/300 increase (for crowns and cages) columns (Cytiva) in SEC buffer using an ÄKTA pure system (Cytiva). SEC traces were also used to qualitatively determine homogeneity and quantitatively measure total yield by A280 absorbance integrated over the collected fractions using Unicorn (Cytiva).

Note that, for T = 4 icosahedral cages (Ico<sub>T=4</sub>-4), we kept the IMAC-eluted sample at room temperature for 24 h before concentrating and running SEC. The 24-h equilibration was important to generate homogeneous samples, and we think that an even longer equilibration time could be helpful to further increase the level of homogeneity. Given the complexity of this sample (4 components, 6 different interfaces and 240 subunits), a relatively slow assembly kinetics seems reasonable; thus, giving an enough equilibration time (more than 24 h) is critical to get homogeneous samples.

For all other systems, the equilibration time is not a critical issue, so we concentrated the elution sample right after IMAC and ran SEC within a few hours. However, giving enough equilibration time (more than 24 h) may not harm the sample quality, but rather possibly increase the homogeneity.

**TEVp cleavage.** Purification and mass tags were buffer exchanged into TEV buffer and cleaved with TEVp at a ratio of 1 mg TEV per 100 mg substrate for 24–72 h at room temperature. After TEV cleavage, samples were exchanged into lysis buffer and passed over a Ni-NTA gravity column and washed with 10 ml lysis buffer. Flowthrough was collected, concentrated using 10k MWCO spin concentrators and purified once again by SEC.

### Sample analysis

**SDS–PAGE.** Samples were diluted 1:1 with 2× Laemmli sample buffer (Bio-Rad) without  $\beta$ -mercaptoethanol, and 15  $\mu$ l was loaded onto AnykD Criterion TGX Precast Midi Protein Gels (Bio-Rad). The ladder was 10  $\mu$ l of Precision Plus Protein Kaleidoscope Prestained Protein Standards (Bio-Rad). Gels were run at 300 V for 18 min, then stained using the eStain L1 Protein Staining System (Genscript). Stained gels were imaged using Chemidoc XRS+ (Bio-Rad).

**Liquid chromatography–mass spectrometry.** To identify the molecular mass of each protein and thus verify sample identity and integrity, intact mass spectra were obtained via reverse-phase light chromatography–mass spectrometry on an Agilent G6230B TOF on an AdvanceBio RP-Desalting column, and subsequently deconvoluted by way of Bioconfirm using a total entropy algorithm.

**nMS.** The oligomeric state of SEC-purified and light chromatography–mass spectrometry-verified samples was analysed by on-line buffer exchange mass spectrometry in 200 mM ammonium acetate using a Vanquish ultra-high performance light chromatography 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; Bio-Rad). The recorded mass spectra were deconvoluted with UniDec version 4.2+.

**nsEM.** SEC-purified samples were diluted (0.005 mg ml<sup>-1</sup> for oligomers and 0.05–0.1 mg ml<sup>-1</sup> for crowns and cages) using SEC buffer immediately before application for 45 s to glow-discharged thick carbon film-coated 400 mesh copper grids (CF400-CU TH, Electron Microscopy Sciences). Grids were then stained and dried immediately twice using 2% uranyl formate. Dried grids were screened on a 120 kV Talos L120C transmission electron microscope. The E. Pluribus Unum (EPU; FEI Thermo Scientific) software was used for automated data collection. Data processing was carried out in CryoSPARC (Structura Biotechnology Inc).

**Cryo-EM sample preparation.** To prepare cryo-EM sample grids, 3  $\mu$ l of 0.5–1.0 mg ml<sup>-1</sup> of protein in 25 mM Tris and 300 mM NaCl at pH 8.0 was applied to glow-discharged Quantifoil R 2/2 300 mesh copper grids overlaid with a thin layer of carbon. Vitrification was performed on a Mark IV Vitrobot with a wait time of either 5 s or 7.5 s, a blot time of 0.5 s and a blot force of either 0 or –1 before being immediately

# Article

plunged frozen into liquid ethane. The sample grids were clipped following standard protocols before being loaded into the microscope for imaging.

Note that we were not able to conduct cryo-EM for  $T = 4$  tetrahedral cages, because the cages were aggregated when they are concentrated higher than  $0.3 \text{ mg ml}^{-1}$ . They were stable below  $0.15 \text{ mg ml}^{-1}$ , which was too dilute for conducting cryo-EM.

**Cryo-EM data collection.** Data collection was performed automatically using either Leginon<sup>32</sup> or SerialEM to control either a Thermo Fisher Titan Krios 300 kV TEM equipped with a standalone K3 Summit direct electron detector with an energy filter or a Thermo Fisher Glacios 200 kV equipped with a standalone K3 Summit direct electron detector<sup>33</sup>. Both samples were collected using counting mode, with random defocus ranges spanning between  $-0.7 \mu\text{m}$  and  $-1.8 \mu\text{m}$  using image shift, with one shot per hole on the Glacios for  $\text{Oct}_{T=4}-3$  or multiple shots per hole on the Titan Krios for  $\text{Ico}_{T=4}-4$ . Movies ( $n = 1,320$  and  $1,048$ ) were collected with a pixel size of  $0.44 \text{ \AA}$  for  $\text{Oct}_{T=4}-3$  with a total dose of approximately  $50 \text{ e}^- \text{ \AA}^{-2}$ , and  $6,678$  movies were collected with a pixel size of  $0.84 \text{ \AA}$  for  $\text{Ico}_{T=4}-4$  with a total dose of approximately  $61 \text{ e}^- \text{ \AA}^{-2}$ .

## Reporting summary

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

## Data availability

Crystallographic data of BGL17\_A31 have been deposited in the Protein Data Bank (ID 8FLX). Electron microscopy maps have been deposited in the Electron Microscopy Data Bank (accession codes EMD-40260, EMD-40267, EMD-40268 and EMD-40269).

## Code availability

The source code from RXPdock for the protein docking is freely available at <https://github.com/willsheffler/rpxdock>. The source code from ProteinMPNN for the sequence design is freely available at <https://github.com/dauparas/ProteinMPNN>.

32. Suloway, C. et al. Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).
33. Sun, M. et al. Practical considerations for using K3 cameras in CDS mode for high-resolution and high-throughput single particle cryo-EM. *J. Struct. Biol.* **213**, 107745 (2021).

**Acknowledgements** We thank E. C. Yang, Q. M. Dowling and N. P. King for helpful discussions; H. Eisenach for advice on explicit negative design; X. Li for help with mass spectrometry analysis of proteins; and C. Liu for advice on cryo-EM data processing. nMS measurements were provided by F. Busch, A. Norris, N. Horvath and S. Cleary of the NIH-funded Resource for Native Mass Spectrometry Guided Structural Biology at The Ohio State University (NIH P41 GM128577 awarded to V. Wysocki). Small-angle X-ray scattering was conducted at the Advanced Light Source, a national user facility operated by Lawrence Berkeley National Laboratory on behalf of the US Department of Energy (DOE), Office of Basic Energy Sciences, through the Integrated Diffraction Analysis Technologies program, supported by the DOE Office of Biological and Environmental Research. Additional support comes from the NIH project ALS-ENABLE (P30 GM124169) and the High-End Instrumentation grant S10OD018483. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a US DOE Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under contract no. DE-AC02-05CH11231 using the NERSC award BER-ERCA0022018. We acknowledge funding from the Howard Hughes Medical Institute (S.L. and D.B.); the Audacious Project at the Institute for Protein Design (S.L., R.D.K., A.J.B., A.P. and D.B.); the Open Philanthropy Project Improving Protein Design Fund (Y.H. and D.B.); the Open Philanthropy Project Universal Flu Vaccine Fund (Y.H. and D.B.); the Bill and Melinda Gates Foundation (OPP1156262 to A.J.B.); the Defense Advanced Research Projects Agency (DARPA) Biostasis (Y.H.); and the NIH's National Institute on Aging (R01AG063845 to A.J.B.). This work was also supported by Outstanding Young Scientist grants (RS-2024-00345974) and the Nano & Material Technology Development Program (RS-2024-00409405) through the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT)(no. RS-2024-00411809). This work was supported by the Institute for Protein Design Breakthrough Fund, for “De novo design of 100nm scale protein assemblies” (S.L., R.D.K., Y.H. and D.B.).

**Author contributions** S.L., R.D.K., Y.H. and D.B. conceptualized the study and designed the protocol. R.D.K., S.L. and Y.H. curated the computational design of oligomers. S.L. and R.D.K. came up with the computational design of cages. S.L. and R.D.K. performed protein synthesis and characterizations. G.A., S.L., R.D.K. and B.H. performed the functionalization of protein cages and cell experiments. A.J.B., A.P., R.D.K. and S.L. undertook cryo-EM. M.A.K. and B.S. performed X-ray crystallography. D.B. supervised the study. S.L., R.D.K. and D.B. wrote the original draft of the manuscript. All authors read and contributed to the manuscript.

**Competing interests** D.B., S.L., R.D.K., Y.H. and G.A. are inventors on a provisional patent application (#49728.01US1) submitted by the University of Washington for the design, composition and function of the proteins created in this study.

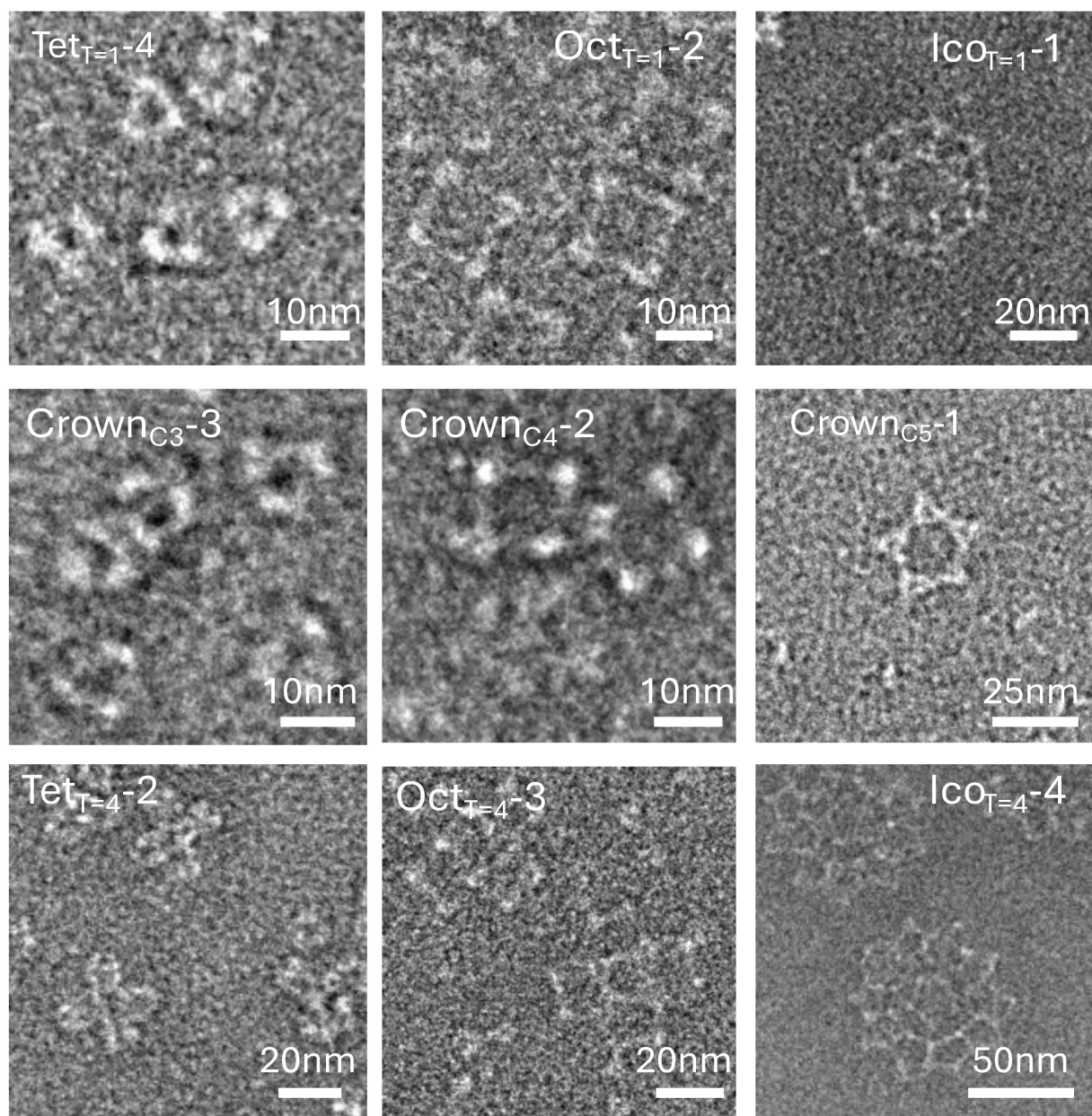
## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41586-024-07814-1>.

**Correspondence and requests for materials** should be addressed to David Baker.

**Peer review information** Nature thanks the anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



**Extended Data Fig. 1 | Zoomed-in view of nsEM images.** T = 1 cages (top), crowns (middle), T = 4 cages (bottom).



Extended Data Table 1 | Designed protein sequence

ID	Structure Types	Main Figure	Protomer sequence
BGL17_A31	Homo-trimer	Fig. 2b	GSPELFLQDLRSLVEAARILARLARQRGDEHALERAARWAEQAARQAERLARQA RKEGNLELALKALQILVNAAYVLAEIARDRGNEELLEYYAARLAEEAARQAIEIWAQ AARRGNQQLRTKAAHILLRAAEVLLEIARDRGNQELLEKAQRIVEAVAAAQQVAAL ALRLAEELDSEEAKKAVRAIAEAAAAALLAALQKGDEVAKLALKVLKEAIELAKENR SEEALKVVLEIARAAAAAARAAEEGKTEVAKLALKVLEEAIELAKENRSEEALKVVL EIAARAALAAQAAEEGKSDEARDALRRLEEAIIEAKENRSKESLEKVREEAKEAE QQAEDAREGKGWSENLYFQGS GSHHHHHH
BGL17_A32	Homo-trimer	Fig. 2d	GSPELFLQDLRSLVEAARILARLARQRGDEHALERAARWAEQAARQAERLARQA RKEGNLELALKALQILVNAAYVLAEIARDRGNEELLEYYAARLAEEAARQAIEIWAQ MEEGNQQLRTKAAHILLRAAEVLLEIARDRGNQELLEKAASLVDAAALQAAAAAIL EGDVEKAVRAAQEAVKAAKEAGDNDMLRAVAIAALRIAKEAEKQGNVEVAVKAA RVAVEAAKQAGDNDVLRKVAEQALRIAKEAEKQGNVEVAVKAA RVAVEAAKQAG DQDVLRKVSEQAERISKEAKKQGNSEVSEEARKVADEAKKQTGDG WSENLYFQ GS GSHHHHHH
BGL18_A35	Homo-trimer	Fig. 2d	GSPELVLKALENMVRAAHTLAEIARDNGNEEWLEAAARLAEVAKAAEELAREAR KEGNLELALKALQILVNAAYVLAEIARDRGNEELLEYYAARLAEEAARQAIEIASKAA KKGNFELALEALRILNEAARVLARIAHHRGNQELLKKAWLLTAMSAAMSALIALAA KSTDEEVRKIAARVAEELAQLAEALRSGSEECIRLAEVVIKAVLLLAALAWKATDE EVIRVALEVARELIRLAQEACRSNDDECLRLASEVVKAVQELVKLAEQATDEEVIR VALEVARELIRLAQEACRSNDDECLREASEVVKVQELVKEAEKSTDEEEIRELLQ RAEERIREAQERCREGDG WSENLYFQGS GSHHHHHH
BGL19_A39	Homo-trimer	Fig. 2d	GSPRWYLQHLETLVRAAEILAEARQRGDEEAEAAARMAEHAARAAEKLAREA RKEGNLELALKALQILVNAAYVLAEIARDRGNEELLEYYAARLAEEAARQAIEIAAQA LEEGDMQLAIKALRILLEAAHVLMQIARDRGNEELLEKAKRLVRLAARISEMVMEL QRAKEEARRKKDSEEAEAYWAQKAIAAALAALAAQALEEGDEDARRVAEELLRQ AEEAARKKNSEAEAVYWAAEAVLAALAELEQAKREGDEDARRVAEELLRQAE AARKKNPEEARAVYEAARDVLEALQRLEEAKRRGDEEERREAEERLRQAEERAR KKN G WSENLYFQGS GSHHHHHH
hetBGL0-18-17_A32	Hetero-trimer (Chain A)	Fig. 2f	MGLELALKALQILVNAAYVLAEIARDRGNEELLEKAARLAEEAARQAERLARQARK EGNLELALKALQILVNAAYVLAEIARDRGNEELLEYYAARLAEEAARQAIEIWAQAM EEGNQQLRTKAAHILLRAAEVLLEIARDRGNQELLEKAASLVDAAALQAAAAIIE GDVEKAVRAAQEAVKAAKEAGDNDMLRAVAIAALRIAKEAEKQGNVEVAVKAA VAVEAAKQAGDQDVLRKVSEQAERISKEAKKQGNSEVSEEARKVADEAKKQTGD G
	Hetero-trimer (Chain B)		MGPRLVLRALENMVRAAHTLAEIARDNGNEEWLEAARLAEEVARRAERLAREA RKEGNLELALKALQILVNAAYVLAEIARDRGNEELLEYYAARLAEEAARQAIEIAAQA MEEGNLELALKALQIIVNAAYVLAEIARDRGNEELLEKAASLAEEAALAEIAAILE GDVEKAVRAAQEAVKAAKEAGDNDMLRAVAIAALRIAKEAEKQGNVEVAVKAA VAVEAAKQAGDNDVLRKVAEQALRIAKEAEKQGNVEVAVKAA RVAVEAAKQAGD QDVLRKVSEQAERISKEAKKQGNSEVSEEARKVADEAKKQTGDG
	Hetero-trimer (Chain C)		MGPELFLQDLRSLVEAARILARLARQRGDEHALERAARWAEQAARQAERLARQA RKEGNLELALKALQILVNAAYVLAEIARDRGNEELLEYYAARLAEEAARQAIEIAAQA MEEGNFELALEALEIINEAARVLARIAHHRGNQELLEKAASLTHASAALSRAIAAILE GDVEKAVRAAQEAVKAAKEAGDNDMLRAVAIAALRIAKEAEKQGNVEVAVKAA VAVEAAKQAGDNDVLRKVAEQALRIAKEAEKQGNVEVAVKAA RVAVEAAKQAGD QDVLRKVAEQALRIAKEAEKQGNVEVAVEAARVAVEAAKQAGDQDVLRKVSEQA ERISKEAKKQGNSEVSEEARKVADEAKKQTGDG SENLYFQGS SHHHHHHG

Designed sequences of C3 homotrimers and pseudo-symmetric heterotrimer shown in the main figures.

Extended Data Table 2 | Crystallographic data

BGL17_A31	
Data collection	
Space group	I23
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	153.47, 153.47, 153.47
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 90,
Resolution (Å)	4.5 (4.5-4.66)
<i>R</i> <sub>sym</sub> or <i>R</i> <sub>merge</sub>	0.087 (0.821)
<i>I</i> / $\sigma$ <i>I</i>	30.25 (2)
Completeness (%)	97.4 % (94.5 %)
Redundancy	18.7 (13.4)
Refinement	
Resolution (Å)	4.5
No. reflections	3609
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.301 / 0.344
No. atoms	
Protein	342
Ligand/ion	0
Water	0
<i>B</i> -factors	
Protein	170.9
Ligand/ion	-
Water	-
R.m.s. deviations	
Bond lengths (Å)	0.002
Bond angles (°)	0.50

BGL17\_A31 (PDB ID: 8FLX).

Extended Data Table 3 | Designed protein sequence

ID	Structure Types	Main Figure	Protomer sequence
Tet <sub>T=1</sub> -4	T=1 Cage	Fig. 3b	MGSPFLQLDLRSLVEAARILARLARQRGDEHALERAARWAEQAARQAERLARQ ARKEGNLELALKALQILVNAAYVLAIEIARDRGNEELLEYYAARLAEEAARQAIEIWAN AMIEGNQQLRTKAAHIILRAAEVLLEIARDRGNQELLEKAASLVDAVAALQQAAAAI LEGDVEKAVRAAQEAVKAAKEAGDNDMLNAVASAAARIALEAMQQGNLEVALKA LRVAEEALKQAGGSGGSHHHHHH
Oct <sub>T=1</sub> -2	T=1 Cage	Fig. 3f	MGSPFLQLDLRSLVEAARILARLARQRGDEHALERAARWAEQAARQAERLARQ ARKEGNLELALKALQILVNAAYVLAIEIARDRGNEELLEYYAARLAEEAARQAIEIWAQ AMEEGNQQLRTKAAHIILRAAEVLLEIARDRGNQELLEKAASLVDAVAALQQAAAA ILEGDVEKAVRAAQEAVKAAKEAGDNDMLRAVAIAALRIAKEAEKQGNVEVAVKA ARVAVEAAKQAGDQDVLAKVADQAYRIAHEALKQGNLEVALEASKVAMEALKQT GGSGGSHHHHHH
Ico <sub>T=1</sub> -1	T=1 Cage	Fig. 3j	GSPFLQLDLRSLVEAARILARLARQRGDEHALERAARWAEQAARQAERLARQA RKEGNLELALKALQILVNAAYVLAIEIARDRGNEELLEYYAARLAEEAARQAIEIWAQA MEEGNQQLRTKAAHIILRAAEVLLEIARDRGNQELLEKAASLVDAVAALQQAAAAIL EGDVEKAVRAAQEAVKAAKEAGDNDMLRAVAIAALRIAKEAEKQGNVEVAVKAA RVAVEAAKQAGDNDVLRKVAEQALRIAKEAEKQGNVEVAVKAARVAVEAAKQAG DNDVLRKVARQAFQIAGKALEQGDIGVAKKAFDVAVEAASQGGGSGGSHHHHH H

T=1 cages shown in the main figures.



**Extended Data Table 4 | Designed protein sequence**

ID	Structure Types	Main Figure	Protomer sequence
Crown <sub>C3</sub> -3	C3 Crown (Chain A)	Fig. 3d	MGSLELALKALQILVNAAYVLA EIARDRGNEELLEKAARLAEEAARQAERIARQAR KEGNLELALKALQILVNAAYVLA EIARDRGNEELLEYAARLAEEAARQAIEIWAQA MVEGNQQLRTKAAHIILRAAEVLLEIARDRGNQELLEKAASLVDVAALQAAAAAIL EGDVEKAVRAAQEAVKAAKEAGDNDMLIAVAAAAARIAVEAMRQGNVEVALKAL RVASEALKQAGGSGGSHHHHHH
	C3 Crown (Chain B)		MGSPRLVLRALENMVRAAHTLAEIARDNGNEEWLERAARLAEEVARRAERLARE ARKEGNLELALKALQILVNAAYVLA EIARDRGNEEELEYAARLAEEAARQAIEIAAT AMVEGNLELALKALQIIVNAAYVLA EIARDRGNEELLEKAASLAEEAAAALAEIAAIL EGDVEKAVRAAQEAVKAAKEAGDNDMLTAVAAAAARIAVEALLQGNTEVALKALR VASEALKQAGGSGGSHHHHHH
	C3 Crown (Chain C)		Same with the Chain C of hetBGL0-18-17_A32
Crown <sub>C4</sub> -2	C4 Crown (Chain A)	Fig. 3h	MGSLELALKALQILVNAAYVLA EIARDRGNEELLEKAARLAEEAARQAERIARQAR KEGNLELALKALQILVNAAYVLA EIARDRGNEELLEYAARLAEEAARQAIEIWAQA MEEGNQQLRTKAAHIILRAAEVLLEIARDRGNQELLEKAASLVDVAALQAAAAAIL EGDVEKAVRAAQEAVKAAKEAGDNDMLRAVAIAALRIAKEAEKQGNVEVAVKAA RVAVEAAKQAGDQDVLKQVATQAARIALEAIKQGNLEVKLEALKVAQEALKQTGG SGGSHHHHHH
	C4 Crown (Chain B)		MGSPRLVLRALENMVRAAHTLAEIARDNGNEEWLERAARLAEEVARRAERLARE ARKEGNLELALKALQILVNAAYVLA EIARDRGNEEELEYAARLAEEAARQAIEIAAQ AMEEGNLELALKALQIIVNAAYVLA EIARDRGNEELLEKAASLAEEAAAALAEIAAIL EGDVEKAVRAAQEAVKAAKEAGDNDMLRAVAIAALRIAKEAEKQGNVEVAVKAA RVAVEAAKQAGDQDVLKQVAAQAARIMLEAIKQGNTEVALEALKVAQEALKQTGG SGGSHHHHHH
	C4 Crown (Chain C)		Same with the Chain C of hetBGL0-18-17_A32
Crown <sub>C5</sub> -1	C5 Crown (Chain A)	Fig. 3i	MGSLELALKALQILVNAAYVLA EIARDRGNEELLEKAARLAEEAARQAERIARQAR KEGNLELALKALQILVNAAYVLA EIARDRGNEELLEYAARLAEEAARQAIEIWAQA MEEGNQQLRTKAAHIILRAAEVLLEIARDRGNQELLEKAASLVDVAALQAAAAAIL EGDVEKAVRAAQEAVKAAKEAGDNDMLRAVAIAALRIAKEAEKQGNVEVAVKAA RVAVEAAKQAGDNDVLRKVAEQALRIAKEAEKQGNVEVAVKAARVAVEAAKQAG DNDVLRKVADQALEIAKAALEQGDIDVAQKAMDVAVEALTQAGGSGGSHHHHHH
	C5 Crown (Chain B)		MGSPRLVLRALENMVRAAHTLAEIARDNGNEEWLERAARLAEEVARRAERLARE ARKEGNLELALKALQILVNAAYVLA EIARDRGNEEELEYAARLAEEAARQAIEIAAQ AMEEGNLELALKALQIIVNAAYVLA EIARDRGNEELLEKAASLAEEAAAALAEIAAIL EGDVEKAVRAAQEAVKAAKEAGDNDMLRAVAIAALRIAKEAEKQGNVEVAVKAA RVAVEAAKQAGDNDVLRKVAEQALRIAKEAEKQGNVEVAVKAARVAVEAAKQAG DNDVLRKVAEQALEIAKAAEQGDVGVMQKAMDVALRAAGQAGGSGGSHHHHHH H
	C5 Crown (Chain C)		Same with hetBGL0-18-17_A32_chC.

Crowns shown in the main figures.

Extended Data Table 5 | Designed protein sequence

ID	Structure Types	Main Figure	Protomer sequence
Tet <sub>T=4-2</sub>	T=4 Cage (Chain A)	Fig. 4d	Same with Crown <sub>C3-3</sub> _chA
	T=4 Cage (Chain B)		Same with Crown <sub>C3-3</sub> _chB
	T=4 Cage (Chain C)		MGSPFLQDLRSLVEAARILARLARQRGDEHALERAARWAEQAARQAERLARQ ARKEGNLELALKALQILVNAAYVLAIEIARDRGNEELLEYYAARLAEAAARQAIEIAAQ AMEEGNFELALEALEIINEAARVLARIAHHRGNQELLEKAASLTHASAALSRAIAAIL EGDVEKAVRAAQEAVKAAKEAGDNDMLRAVAIAALRIAKEAEKQGNVEVAVKAA RVAVEAAKQAGDNDVLRKVAEQALRIAKEAEKQGNVHVALLQALAVAGEAASAAG DTDVLRKVLEQFRRIVKELEKQG
	T=4 Cage (Chain ho)		MGSPFLQDLRSLVEAARILARLARQRGDEHALERAARWAEQAARQAERLARQ ARKEGNLELALKALQILVNAAYVLAIEIARDRGNEELLEYYAARLAEAAARQAIEIWAQ AMEEGNQQLRTKAAHIILRAAEVLLEIARDRGNEELLEKAASLVDAAALQAAAAA ILEGDVEKAVRAAQEAVKAAKEAGDNDMLRAVAIAALRIAKEAEKQGNVEVAVKA ARVAVEAAKQAGDNDVLRKVAEQALRIAKEAEKQGNVHVALLALDVAATAAAVAG DTDVLRKVLEQFQRIVKELEKQGGSGGSHHHHHH
Oct <sub>T=4-3</sub>	T=4 Cage (Chain A)	Fig. 4h	Same with Crown <sub>C4-2</sub> _chA
	T=4 Cage (Chain B)		Same with Crown <sub>C4-2</sub> _chB
	T=4 Cage (Chain C)		MGSPFLQDLRSLVEAARILARLARQRGDEHALERAARWAEQAARQAERLARQ ARKEGNLELALKALQILVNAAYVLAIEIARDRGNEELLEYYAARLAEAAARQAIEIAAQ AMEEGNFELALEALEIINEAARVLARIAHHRGNQELLEKAASLTHASAALSRAIAAIL EGDVEKAVRAAQEAVKAAKEAGDNDMLRAVAIAALRIAKEAEKQGNVEVAVKAA RVAVEAAKQAGDNDVLRKVAEQALRIAKEAEKQGNVIVAMKAIDVAVEAAGQAGD IDVIKKVADQTRIVDEWLKQG
	T=4 Cage (Chain ho)		MGSPFLQDLRSLVEAARILARLARQRGDEHALERAARWAEQAARQAERLARQ ARKEGNLELALKALQILVNAAYVLAIEIARDRGNEELLEYYAARLAEAAARQAIEIWAQ AMEEGNQQLRTKAAHIILRAAEVLLEIARDRGNEELLEKAASLVDAAALQAAAAA ILEGDVEKAVRAAQEAVKAAKEAGDNDMLRAVAIAALRIAKEAEKQGNVEVAVKA ARVAVEAAKQAGDNDVLRKVAEQALRIAKEAEKQGNVYVAAKAVQVAAEAAKQA GDIDVLKKVIGQTQIRIDEWVKQGGSGGSHHHHHH
Ico <sub>T=4-4</sub>	T=4 Cage (Chain A)	Fig. 4i	Same with Crown <sub>C5-1</sub> _chA
	T=4 Cage (Chain B)		Same with Crown <sub>C5-1</sub> _chB
	T=4 Cage (Chain C)		MGSPFLQDLRSLVEAARILARLARQRGDEHALERAARWAEQAARQAERLARQ ARKEGNLELALKALQILVNAAYVLAIEIARDRGNEELLEYYAARLAEAAARQAIEIAAQ AMEEGNFELALEALEIINEAARVLARIAHHRGNQELLEKAASLTHASAALSRAIAAIL EGDVEKAVRAAQEAVKAAKEAGDNDMLRAVAIAALRIAKEAEKQGNVEVAVKAA RVAVEAAKQAGDNDVLRKVAEQALRIAKEAEKQGNVEVAVKAARVAVEAAKQAG DNDVLRKVAEQALRIAKEAEKQGNVIVAVRAALVAAQAATEAGDQDVLKRVDEQL ERILKEARKQGGSGGSHHHHHH
	T=4 Cage (Chain ho)		MGSPFLQDLRSLVEAARILARLARQRGDEHALERAARWAEQAARQAERLARQ ARKEGNLELALKALQILVNAAYVLAIEIARDRGNEELLEYYAARLAEAAARQAIEIWAQ AMEEGNQQLRTKAAHIILRAAEVLLEIARDRGNEELLEKAASLVDAAALQAAAAA ILEGDVEKAVRAAQEAVKAAKEAGDNDMLRAVAIAALRIAKEAEKQGNVEVAVKA ARVAVEAAKQAGDNDVLRKVAEQALRIAKEAEKQGNVEVAVKAARVAVEAAKQA GDNDVLRKVAEQALRIAKEAEKQGNVVVATQAAIVAIQAARLAGDQDVGRKVIEQ EERIAKEARKQGGSGGSHHHHHH

T=4 cages shown in the main figures.

Extended Data Table 6 | Designed protein sequence

ID	Structure Types	Main Figure	Protomer sequence
(ASGPR-minibinder-fused) Ico <sub>T=4</sub> -4	T=4 Cage (Chain A)	Fig. 6e	Same with Ico <sub>T=4</sub> -4_chA
	T=4 Cage (Chain B)		Same with Ico <sub>T=4</sub> -4_chB
	T=4 Cage (Chain C)		Same with Ico <sub>T=4</sub> -4_chC
	T=4 Cage (Chain ho)		MGSHHHHHHGGSGGSPFLQDLRSLVEAARILARLARQRGDEHALE RAARWAEQAARQAERLARQARKEGNLELALKALQILVNAAYVLAEIAR DRGNEELLEYAARLAEEAARQAIEIWAQAMEEGNQQLRTKAAHIILRAA EVLLEIARDRGNQELLEKAASLVDAVAALQQAAAAILEGDVEKAVRAA QEAVKAAKEAGDNDMLRAVAIAALRIAKEAEKQGNVEVAVKAARVAVE AAKQAGDNDVLRKVAEQALRIAKEAEKQGNV VVATQAAIVAIQAARLAGDQD AGDNDVLRKVAEQALRIAKEAEKQGNV VVATQAAIVAIQAARLAGDQD VGRKVIEQEERIAKEARKQGGSGGSGGSGGSAAEEAIRRLKELNTPEA EDLLDFIETLLWGRPEEEALRKAAALARRWGEPEARVIEEALKSLS

The ASGPR-cage in which one of the four components (chain\_ho) of Ico<sub>T=4</sub>-4 is fused to an ASGPR minibinder.



Extended Data Table 7 | CryoEM data collection information

	Oct <sub>T=4</sub> -3 (T=4) (EMDB-40260)	Ico <sub>T=4</sub> -4 (T=4) (EMDB-40268)
Data collection and processing		
Magnification	45,000	210,000
Voltage (kV)	200	300
Electron exposure (e-/Å <sup>2</sup> )	50	61
Defocus range (µm)	Between -0.8 and -1.8	Between -0.8 and -1.8
Pixel size (Å)	0.44	0.42
Symmetry imposed	Octahedral symmetry	Icosahedral symmetry
Initial particle images (no.)	35,828 + 21,472	12,780
Final particle images (no.)	29,069	6,896
Map resolution (Å)	6.87	12.78

	Oct <sub>T=4</sub> -3 (T=1) (EMDB-40269)	Ico <sub>T=4</sub> -4 (T=1) (EMDB-40267)
Data collection and processing		
Magnification	45,000	210,000
Voltage (kV)	200	300
Electron exposure (e-/Å <sup>2</sup> )	50	61
Defocus range (µm)	Between -0.8 and -1.8	Between -0.8 and -1.8
Pixel size (Å)	0.44	0.42
Symmetry imposed	Octahedral symmetry	Icosahedral symmetry
Initial particle images (no.)	35,828 + 21,472	12,780
Final particle images (no.)	15,852	1,172
Map resolution (Å)	6.34	6.54

T=4 cages (Oct<sub>T=4</sub>-3 and Ico<sub>T=4</sub>-4) and off target small cages (T=1 cages) found in Oct<sub>T=4</sub>-3 and Ico<sub>T=4</sub>-4 systems.

## Reporting Summary

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

### Statistics

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

n/a Confirmed

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

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

### Software and code

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

- |                 |   |
|-----------------|---|
| Data collection | Provide a description of all commercial, open source and custom code used to collect the data in this study, specifying the version used OR state that no software was used.  |
| Data analysis   | The source code of RPXdock for the protein docking is freely available at <a href="https://github.com/willsheffler/rpxdock">https://github.com/willsheffler/rpxdock</a> . The source code of ProteinMPNN for the sequence design is freely available at <a href="https://github.com/dauparas/ProteinMPNN">https://github.com/dauparas/ProteinMPNN</a> . |

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

### Data

Policy information about [availability of data](#)

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

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

Crystallographic data of BGL17\_A31 have been deposited in the PDB (ID: 8FLX). EM maps have been deposited in the EMDB (accession codes: EMD-40260, EMD-40267, EMD-40268, and EMD-40269).

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

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

## Field-specific reporting

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

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

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

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

Sample size	N/A
Data exclusions	N/A
Replication	N/A
Randomization	N/A
Blinding	N/A

## Reporting for specific materials, systems and methods

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

### Materials & experimental systems

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

### Methods

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

## Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	HEP3B cells (ATCC)
Authentication	Authenticated by ATCC
Mycoplasma contamination	Tested by ATCC
Commonly misidentified lines (See <a href="#">ICLAC</a> register)	<i>Name any commonly misidentified cell lines used in the study and provide a rationale for their use.</i>

## Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A