Article

De novo design of quasisymmetric two-component protein cages

https://doi.org/10.1038/s41586-026-10464-0

Received: 1 June 2025

Accepted: 26 March 2026

Published online: 20 May 2026

Check for updates

Shunzhi Wang[1,2,3]✉, Ying Xie[3], David Chemielewski[1,2], Connor Weidle[1,2], Tong Shu[3], Green Ahn[1,2], Ryan D. Kibler[1,2], Cindy Hernandez[3], Wei Chen[1,2], David Camilo Duran[3], Ann Carr[1,2], Asim K. Bera[1,2], Sangmin Lee[1,2,4], Justin Decarreau[1,2], Alex Kang[1,2], Evans Brackenbrough[1,2], Emily Joyce[1,2], Kejia Wu[1,2], Andrew J. Borst[1,2], Andrew Favor[2,5], Buwei Huang[2,5], Frank DiMaio[1,2], Liam J. Holt[3] & David Baker[1,2,5,6]✉

Quasisymmetric icosahedral viral capsids achieve larger sizes than possible with strictly symmetric icosahedra by tessellating pentagons and hexagons using a single subunit that adopts different conformations in symmetrically non-equivalent locations[1,2]. Recapitulating such quasisymmetric architectures through computational design is a considerable challenge in nanomaterials engineering. Here we introduce a computational design strategy based on geometric frustration to generate two-component, quasisymmetric protein cages with customizable properties. We designed complementary trimeric and dimeric protein components that co-assemble into positively curved local hexagonal assemblies. Hexagonal lattices cannot tile spherical surfaces; instead, the components form closed sphere-like cage assemblies through incorporation of curvature-inducing pentagonal defects, as evidenced by electron microscopy. By designing dimers that encode different local curvatures, we programmed cage dimensions ranging from 40 to over 200 nm in diameter and with molecular weights from 2 MDa to over 50 MDa, comparable with natural virus capsids. We further functionalized these large cages with additional protein domains to enable ribonucleoprotein cargo loading and cellular uptake. Fluorescently labelled cage assemblies expressed in mammalian cells function as rheological probes and cargo recruiters, enabling a systematic study of size-dependent cytoplasmic diffusion and protein localization. Thus, the quasi-symmetry that has long fascinated structural biologists can now be achieved by computational protein design, with immediate applications to biologics delivery and molecular cell biology.

Although symmetry is widespread in nature, it is not always perfectly respected, as illustrated by fullerenes[3], quasicrystals[4] and clathrin cages[5]. In the 1960s, Caspar and Klug showed that viruses generate larger capsids by relaxing perfect icosahedral symmetry among subunits with subtle deviations[1]. Such structural plasticity offers evolutionary advantages, enabling viruses to expand their capsid shells and accommodate larger genome cargos without a substantial increase in genetic encoding[2]. De novo protein design has generated well-defined self-assembling icosahedral structures[6,7], which show considerable promise for vaccines[8,9] and other applications[10,11]. However, the structures realized so far are largely limited to regular point-group symmetries with discrete copy numbers, with the exception of pseudosymmetric structures[12–14], in which multiple chemically distinct yet structurally similar chains occupy unique environments. Genetically encodable quasisymmetric structures—in which a single sequence encodes multiple local conformations—remain a formidable design challenge, as multiple conformations must be accessible to a single sequence and the correct ones must be adopted during assembly to avoid large-scale aggregation. Developing a strategy that intentionally breaks perfect symmetry in a programmable manner could unlock a vast design space for quasisymmetric structures, leading to structural and materials properties at organelle and subcellular scales (10–1,000 nm), which are currently out of reach by current approaches.

We sought to develop a general strategy based on geometric frustration[15] for designing self-limiting quasisymmetric particles. Instead of targeting protein monomers that inherently fold into multiple states, we focused on designing proteins capable of undergoing local conformational changes to accommodate a desired globally low-energy architecture[16]. To drive spherical particle closure, we imposed positive curvature on hexagonal building blocks. As hexagons alone cannot tile a spherical surface, pentagonal defects (+60° wedge disclinations) can emerge to relieve accumulated strain as these components co-assemble

[1]Department of Biochemistry, University of Washington, Seattle, WA, USA. [2]Institute for Protein Design, University of Washington, Seattle, WA, USA. [3]Institute for Systems Genetics, New York University Langone Medical Center, New York, NY, USA. [4]Department of Chemical Engineering, Pohang University of Science and Technology (POSTECH), Pohang, Republic of Korea. [5]Molecular Engineering and Sciences Institute, University of Washington, Seattle, WA, USA. [6]Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA. ✉e-mail: shunzhi.wang@nyulangone.org; dabaker@uw.edu

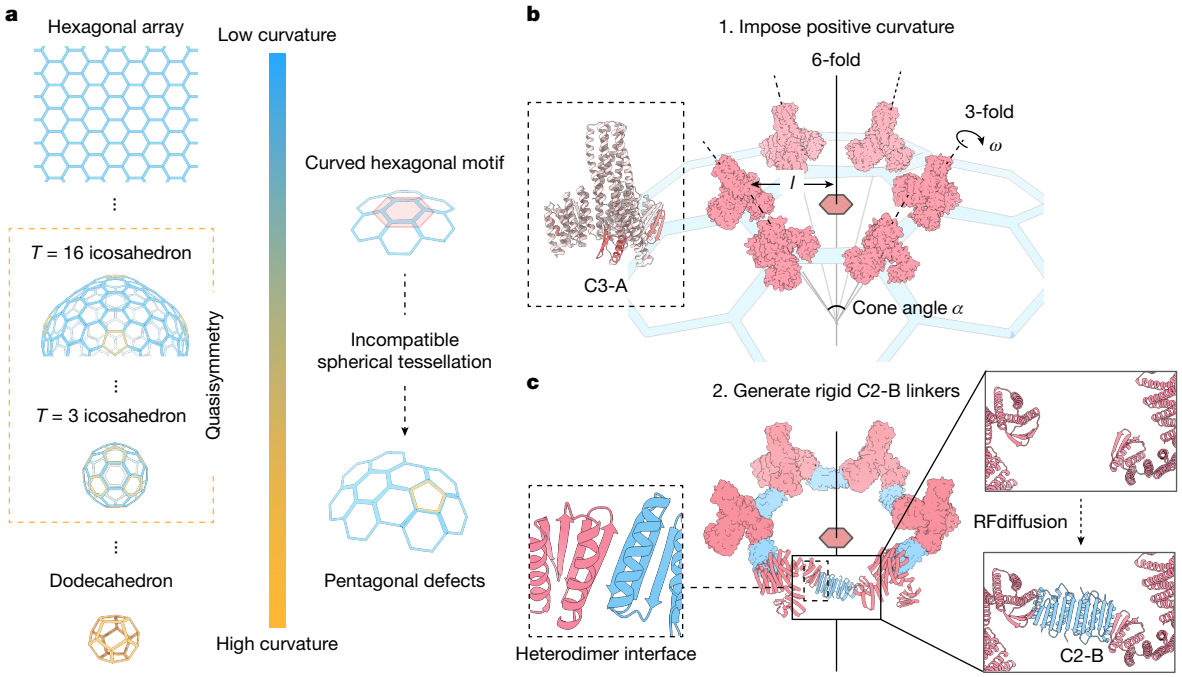


Fig. 1 | Computational design strategy for generating quasisymmetric particles. a, Regular hexagons tessellate flat two-dimensional lattices with zero curvature. By contrast, twelve regular pentagons form a dodecahedron with positive curvature, creating an angular curvature of 36° at each vertex. Quasisymmetric icosahedral cages do not maintain constant curvature; instead 12 pentagons are combined with varying numbers of hexagons to form closed particles with different triangulation numbers. The ratio of pentagons to hexagons determines the overall curvature. **b**, We approximated quasisymmetric particles using positively curved, symmetric hexagonal building motifs. The inset shows six copies of the homotrimer symmetrically arranged into a hexagonal pattern and bent inwards to impose positive curvature. **c**, Next, we used RFdiffusion—a generative protein design method—to create dimeric linkers that rigidify the designed curvature. The inset shows a previously validated heterodimeric interface driving binding between the trimer and the designed dimeric linkers, ensuring proper particle self-assembly.

into closed spherical particles (a total wedge disclination of 4π, which corresponds with 12 pentagons; Fig. 1a). We reasoned that symmetry breaking could arise in such a system if the energetic cost of monomer deformation—which determines the energy difference between pentameric and hexameric states—could be overcome by the gain in overall contact energy following full particle closure (Extended Data Fig. 1). Such symmetry breaking to generate both pentagonal and hexagonal faces, driven by the formation of energetically favourable interactions upon assembly closure, also occurs in carbon fullerenes; however, in fullerenes the cost of the required C–C bond angle deformations is fully determined by carbon hybridization chemistry, whereas in designed protein assemblies the interfacial bending angles can be continuously tuned to minimize local strain and stabilize the desired target structure.

We aimed to design two-component protein assemblies that approximate fullerene topology starting from a spherical hexagonal tessellation. In our design, each vertex is occupied by a previously designed protein homotrimer[17] called C3-A, and each edge by a homodimer designed with RFdiffusion[18] called the C2-B linker, which has a pre-specified curvature (Methods and Extended Data Fig. 2). We used a previously validated heterodimer[19] as the molecular interface between C3-A and C2-B to ensure precise assembly into the specified geometry. As shown in Fig. 1b, we placed the threefold axes of each C3-A at the vertices of an ideal hexagon of edge length l. We then tilted each axis towards the global sixfold axis so that they lie on a cone of angle α; l and α together set the target sphere size. Finally, we rotated each C3-A by an angle ω about its own threefold axis to control the relative orientation of the C2-B binding domains. We chose to design C2-B linkers with α–β folds for improved structural rigidity compared with all α-helical folds[20], to achieve the small deviations in cone angles required for different T number cages (Fig. 1c). The strong inter-subunit binding affinity ($K_d \approx 10$ nM; ref. 19) enforces interface specificity and minimizes unintended off-target interactions, driving complete assembly closure to maximize binding interface pairing. This two-component system allows for straightforward separate purification of building blocks and subsequent in vitro experimental validation under various assembling conditions.

Encoding cage size with hexagon curvature

We began by seeking to design a wide range of cage sizes by varying the cone angles. To target Goldberg icosahedra[21] ranging from triangulation numbers $T = 1$ to $T = 25$, we designed a set of C2-B linkers with cone angles α ranging from 40° to 4° and a fixed edge length of $l = 125$ Å (Fig. 2a–c and Supplementary Table 1). We used RFdiffusion[18] to generate C2-symmetric homodimers that rigidly scaffold the heterodimer interface motif at a pre-specified geometry (Methods), and selected 36 for experimental testing. These 36 were expressed in *Escherichia coli*; 32 could be purified via immobilized metal-affinity chromatography, and 26 eluted as a single peak at the expected elution volume on size-exclusion chromatography (SEC; the protein yields are shown in Supplementary Table 2). After mixing with C3-A at an equal molar ratio (50 μM), we observed the formation of larger assemblies as the major product fraction for 18 designs (Supplementary Fig. 2 and Supplementary Table 3). We then used negative-stain electron microscopy (nsEM) to screen for particle formation, and observed eight designs that show particle formation with low polydispersity (the corresponding SEC profiles and dynamic light scattering data are shown in Supplementary Fig. 1). The remaining ten designs formed particles with large size distributions, which we attributed to linker geometries unfavourable for particle closure. Analysis of particle diameters derived from nsEM measurement showed a clear negative correlation between designed cone angle and particle size distributions (Fig. 2d). The smaller cages

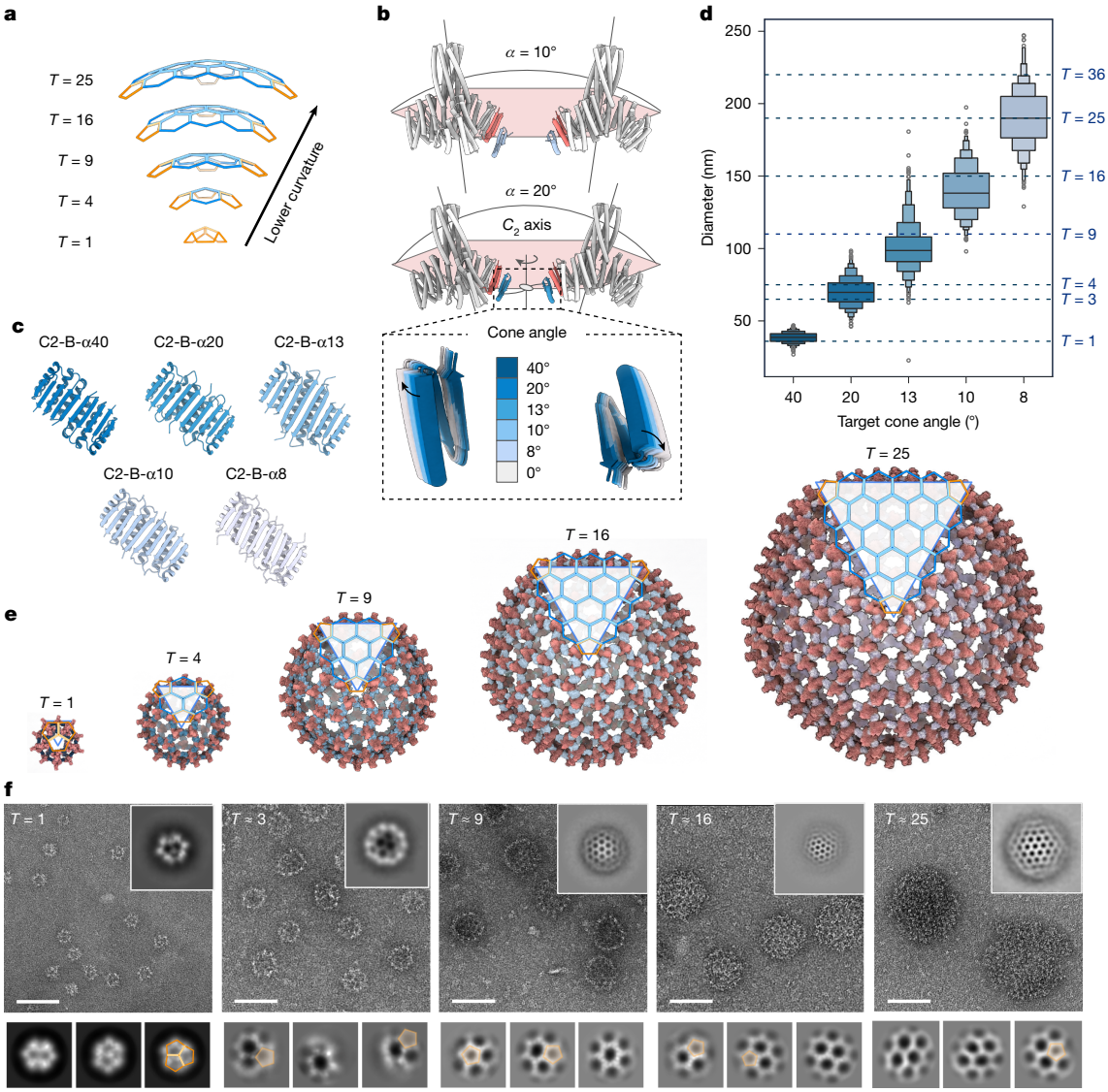


Fig. 2 | Cage size can be programmed through the curvature of designed dimeric linkers. a, Cartoon depiction of high-*T* cages, illustrating reduced local curvature at individual hexagonal facets. **b**, Local curvature of the hexagonal building motif is controlled by a pre-specified cone angle, defined as the intersecting angle between two adjacent trimers linked by a single C2-B linker. A magnified view highlights the close spatial proximity of C2-B interface placements targeting different cone angles (inset). **c**, Computational models of five designed dimeric linkers spanning a range of target cone angles. **d**, Particle size distributions were measured by nsEM and plotted as a function of cone angles (8° to 40°), shown as letter-value box plots. The centre lines indicate the median; from here, successive boxes outwards represent the middle 50%, 75%, 87.5% and 93.75% of the data, respectively (n = 254, 272, 476, 317 and 270 particles for 40, 20, 13, 10 and 8° target cone angles, respectively). Individual outliers are indicated by dots. **e**, Cartoon depiction of ideal two-component icosahedral cages with *T* values ranging from 1 to 25. **f**, Representative nsEM micrographs of the five designed dimeric linkers assembled with the shared trimeric component to form quasisymmetric cages (N = 3 independent experiments), with insets displaying selected two-dimensional class averages. The bottom row shows nsEM two-dimensional class averages, where pentagonal structures (highlighted in orange) are clearly visible across all designs. Scale bars, 100 nm.

tended to be more monodisperse than the larger cages. By contrast to the weak subunit interactions that drive assembly of quasisymmetric clathrin cages[22], the strong heterodimeric interface between C3-A and C2-B drives assembly in our designed cages, which could lead to greater heterogeneity due to kinetic trapping.

We set out to structurally characterize the range of particles formed by designs with different cone angles using nsEM (Fig. 2d–f and Supplementary Figs. 3 and 4). We first examined a design with a cone angle of $\alpha = 40°$, which, as intended, generated a $T = 1$ regular dodecahedron (~40 nm), as evidenced by two-dimensional nsEM classes, which predominantly revealed pentagons. Reducing the cone angle to $\alpha = 30°$ yielded a broader distribution of polyhedral assemblies, comprising a mixture of $T = 1$ and larger $T = 1$–3 cages, consistent with a partial relaxation of curvature or closure constraints (Extended Data Fig. 3). Fitting the design model into the experimental three-dimensional maps indicated that particle assembly based on heterodimer interactions was sufficient to bend the C3-A and C2-B linkers from the original hexagonal geometry to achieve strained five-membered ring closure. C2-B-α20 linkers, with a cone angle $\alpha = 20°$, were found to generate spherical particles approximately 50–70 nm in diameter (expected *T* number ≈ 3–4; Extended Data Fig. 4). C2-B-α13 linkers with $\alpha = 13°$ formed spherical particles about 100 nm in diameter. Although we could not precisely determine the exact *T* number, the observed size range suggests a triangulation number of $T \approx 9$. A C2-B-α10 linker formed particles about 130 nm in diameter ($T \approx 16$), and three designed $\alpha = 8°$ linkers yielded particles approaching 200 nm in diameter, corresponding to $T \approx 25$

(Supplementary Fig. 5). The small angular deviation in the designed C2-B linkers (C2-B-α40 versus C2-B-α10) results in a significant change in cage triangulation number ($T \approx 1$ versus $T \approx 16$; Supplementary Fig. 6). Owing to the heterogeneous particle assembly, we were not able to obtain well-defined two-dimensional class averages for full particles with high T number (for example, $T > 9$). The morphology of our assembled particles appeared rounded rather than faceted, consistent with the expected spherical shape of Goldberg polyhedra. Smaller cone angles generated roughly flat hexagonal lattices, and we observed irreversible aggregation upon mixing the two components when targeted cone angles were smaller than 6° (Supplementary Fig. 7). Local structural analyses based on nsEM two-dimensional class averages revealed the presence of pentagonal defects in all samples (Fig. 2f, bottom row and inset), confirming our hypothesis that cooperative deformation resulting in pentagonal defects drives particle closure.

Cryo-EM characterization of $T = 3$ and high-T cages

We sought to characterize in vitro-assembled cages with cryo-EM to understand the structural basis of quasisymmetric particle closure. We chose to first analyse the cage design targeting $T = 3$—the simplest Goldberg icosahedron—which is structurally equivalent to a C60 fullerene and has the general shape of a soccer ball with 12 pentagonal and 20 hexagonal facets. There are a total of 60 vertices (C3-A) and 90 edges (C2-B). We obtained a crystal structure at 2.1 Å of the designed C2-B-α20 linker, which closely agrees with the design model (root mean square deviation, r.m.s.d. = 0.7 Å; Fig. 3a). For the complete cage, we obtained both single-particle (Supplementary Fig. 8) and cryo-electron tomography (cryoET; Fig. 3b) data. Subtomogram three-dimensional classification of all particles into three classes without symmetry averaging revealed half-complete cages with arrangements of pentagon and hexagon assemblies matching the expected $T = 3$ architecture (Fig. 3b, top). Further refinement of the largest class with icosahedral symmetry applied resulted in an approximately 30 Å map with $T = 3$ icosahedral structure (Fig. 3c, left). Starting from the designed hexamer model, we used Rosetta with a make_symmdef_file (ref. 23) to build a C5 complex preserving the subunit-subunit interface in the design. These rings were docked into the density using Chimera, the asymmetric unit (a full C3-A and three adjacent C2-B-α20 linkers) was extracted, and Rosetta relax with density constraints[24] was used to refine the full assembly into the cryo-EM density (Protein Data Bank (PDB) ID: 9OP9). This resulted in a full assembly model with a truncated icosahedron shape (diameter = 61 nm; Methods and Fig. 3c, right), where the outward-pointing C3-A helical bundles are bridged by C2-B-α20 dimers. The extracted experimental hexagonal and pentagonal structural motifs are both slightly smaller than the designed hexagonal and computationally derived pentagonal models, respectively (Fig. 3d). In the particle, the homotrimeric protein C3-A and the associated C2-B-α20 partner assume two distinct local environments: two adjacent hexagons (denoted 6–6) or a hexagon adjacent to a pentagon (denoted 6–5). This splitting causes the homotrimeric protein C3-A to break its original C3 symmetry and become an asymmetric trimer (Fig. 3e and Supplementary Fig. 9). The termini of the two different conformations (6–6 to 6–5) differ in position by 20 Å when the components are aligned on the trimeric core (Fig. 3f). The designed helical repeat region of the trimer (residues 90–234) undergoes the most significant conformational shift, contributing to symmetry breaking. We identified two conformational states for the relaxed C2-B-α20 linkers with different bending angles and broken C2 symmetry, corresponding to the hexagonal 6–6 and pentagonal 6–5 ring configurations, respectively. The 6–5 conformation is closer to the crystal structure of the linker alone (r.m.s.d. = 1 Å) than the 6–6 conformation (r.m.s.d. = 2 Å) (Fig. 3g).

To investigate larger cages with higher triangulation numbers, a single-particle cryo-EM dataset was collected for the in vitro-assembled $T \approx 16$ cages (C2-B-α10). Most particles were observed to be spherical and closed, with no apparent gaps (Fig. 3h). The variations in particle sizes and shapes suggest cage assemblies do not strictly follow Goldberg icosahedral architecture. The observed structural heterogeneity prevents averaging full particles; instead we combine two-dimensional class averages and local region reconstructions to characterize the curved lattices and the spatial distribution of pentagonal assemblies. In two-dimensional class averages, we observed hexagonal array patterns and confirmed the presence of pentagonal assemblies (Fig. 3i). We did not observe two pentagons within the same two-dimensional class regardless of particle box size, and hence the spatial distribution of pentagons is most likely to be random and not evenly distributed (Supplementary Fig. 10). We also identified heptagonal defects adjacent to pentagonal rings and distorted hexagons, both of which have been observed as topological defects in large fullerenes[25], highlighting the flexibility of the lattice to accommodate both positive and negative curvatures (Extended Data Fig. 5 and Supplementary discussion). The reconstructed three-dimensional map of a curved hexagon and its nearest neighbours show good agreement with the corresponding region of an idealized $T = 16$ icosahedron (Fig. 3j). Electron cryo-tomography of the in vitro assembled $T \approx 25$ cage reveals a near-spherical morphology with a hollow interior (Supplementary Fig. 11).

Modular cargo loading and cellular uptake

Modular functionalization of de novo cages could provide control over their interactions with biological systems and enable downstream biomedical applications. Both C3-A and C2-B of our designed cages have exposed termini that support straightforward genetic fusions with natural and de novo proteins; for example, fluorescent proteins[26], de novo peptide binders[27], nanobodies[28], HaloTag[29], SpyCatcher/SpyTag[30] and ApoE3 (ref. 31). C3-A-nSpyCatcher and C2-B-α20-nMCherry fusion proteins were expressed in *E. coli* and purified by SEC as soluble oligomers with the expected molecular weights (Fig. 4a and Supplementary Fig. 12). Robust in vitro assembly of functionalized C3-A and C2-B resulted in monodisperse cages at high yield as indicated by SEC elution profiles and nsEM (Supplementary Fig. 13 and Extended Data Fig. 6). Owing to their internal cavities, these designed cages have the potential to encapsulate large cargoes such as ribonucleoprotein gene editors, which remains challenging for both natural viral particles and previous designs. As a step in this direction, we prepared $T = 3$ cages with C3-A and C2-B linkers fused to an anti-green-fluorescent-protein (anti-GFP) nanobody[27] (C2-B-α20_nGFPnb). Upon mixing with GFP-labelled Cas9, we observed efficient cargo encapsulation, as evidenced by co-elution of the GFP signal with the nanobody-functionalized cages using SEC (Supplementary Fig. 14). NsEM analysis revealed no appreciable morphological changes following encapsulation (Fig. 4b).

Next, we sought to functionalize cages with designed binding proteins[31] that target different endocytic cell surface receptors to facilitate uptake. We genetically fused the N terminus of different C2-B-α20 linkers with GFP and three de-novo-designed proteins that promote endocytosis (EndoTags) through the asialoglycoprotein receptor (ASGPR), the insulin-like growth factor 2 receptor (IGF2R) and sortillin[32] (ASGPR_EndoTag, IGF_EndoTag and Sort_EndoTag). In-vitro-assembled cages were purified by SEC as a single peak with no appreciable change in morphology according to nsEM. Using confocal microscopy, we observed intracellular puncta for all three types of EndoTag cage conjugate following incubation with HEP3B cells for 24 h (Fig. 4c). Quantitative uptake was assessed by flow cytometry (Supplementary Fig. 15); $T = 3$ cages (diameter ≈ 60 nm) were functionalized with ASGPR, and sortilin binding proteins showed cellular uptake, as evidenced by the partial co-localization of GFP signals with the lysosomal tracker LAMP1 signal. GFP-labelled (no endotags) cages, by contrast, showed minimal internalization.

We also evaluated cellular uptake of ASGPR_EndoTag and Sort_EndoTag functionalized cages of three different sizes (60 nm, C2-B-α20; 80 nm,

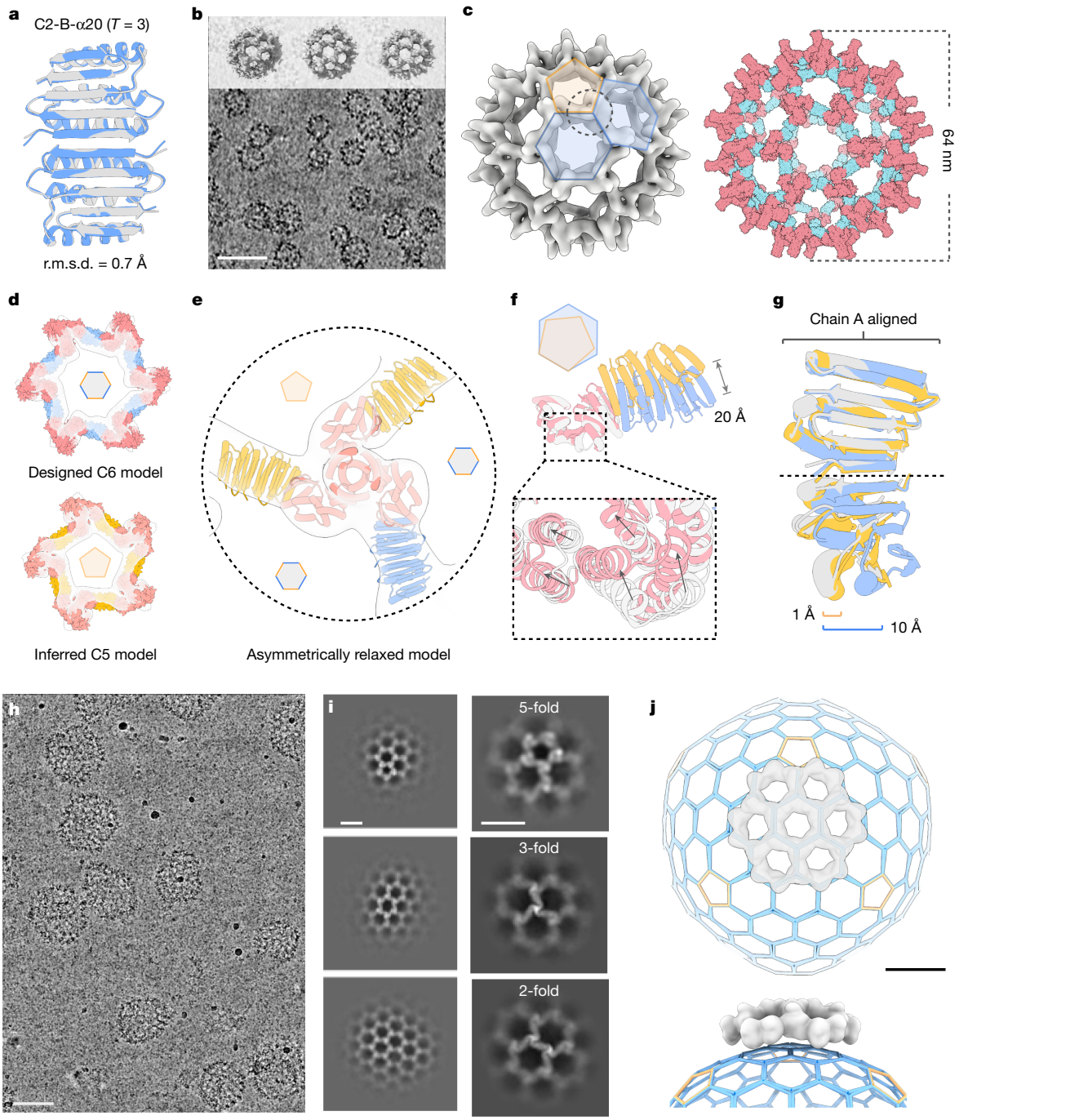


Fig. 3 | Cryo-EM characterization of in vitro assembled cages. a, The crystal structure of the designed C2-B-α20 linker for the $T = 3$ cage (blue) aligned to the design model (grey). **b**, Cryo-ET tomogram Z-slice and three representative ab initio model reconstructions based on subtomogram-averaged density without applied symmetry (inset). Scale bar, 100 nm. **c**, A cryo-EM three-dimensional density map at 31.6 Å resolution refined with applied icosahedral symmetry (left) and the corresponding relaxed full particle model (right). The dihedral angle between the two hexagonal facets highlighted blue (6–6) is 138°, and 142° between a hexagonal and pentagonal facet (6–5). **d**, An overlay of the experimental map (grey) with hexagonal ring design model (top) and calculated pentameric ring model (bottom). **e**, A magnified view of one C3-A homotrimer and three adjacent C2-B-α20 linkers relaxed into experimental density. C3 symmetry is broken to form asymmetric local structures as a result of particle assembly. **f**, Alignment of 6–6 (grey and blue) to 6–5 (pink and orange) conformations showing a gap distance of 20 Å between the termini of two dimers. The inset shows a magnified view of the designed helical repeat region. **g**, Comparison of 6–6 (blue) to 6–5 (orange) conformations of the C2-B-α20 linker aligned to one end of its crystal structure (grey). **h**,**i**, Cryo-EM micrograph (successfully repeated for $N = 3$ experiments) (**h**) and local two-dimensional class averages of a $T \approx 16$ cage (C2-B-α10) (**i**). Scale bar, 100 nm (**h**); 20 nm (**i**). **j**, Top and side view of three-dimensional reconstructed maps corresponding to curved local hexagons. Scale bar, 20 nm.

C2-B-α10; and 100 nm, C2-B-α8). Compared with unfunctionalized C2-B linkers, GFP-fused and EndoTag-fused constructs consistently formed smaller assemblies under identical monomer and salt concentrations, which we attribute to charge repulsion from the highly anionic GFP domains. Flow-cytometry and time-course measurements (Supplementary Fig. 15) revealed that uptake increased over 24 h for all constructs, but the influence of cage diameter was observed to be receptor-dependent: sortilin-targeted cages favoured larger diameters under the tested

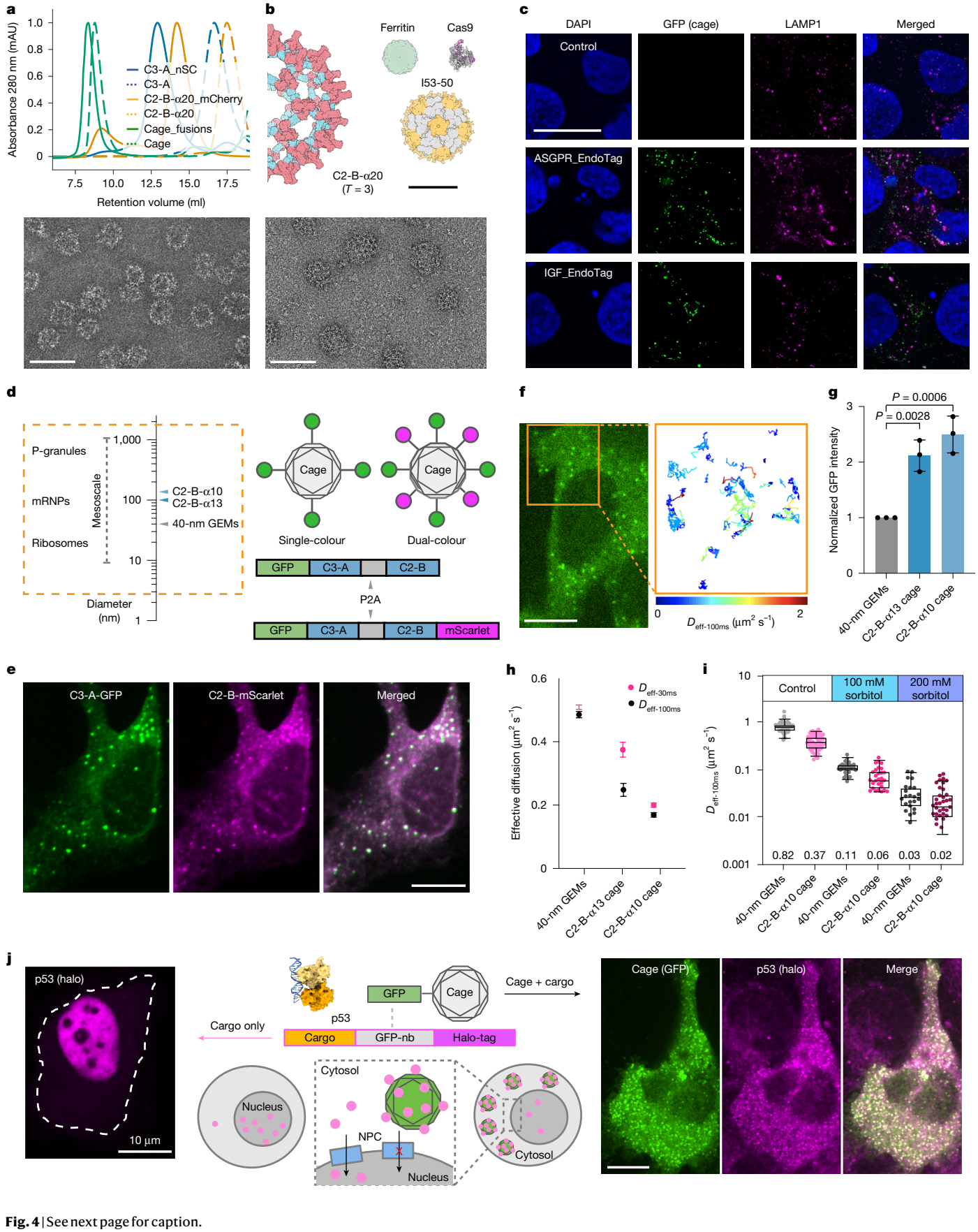


Fig. 4 | See next page for caption.

Fig. 4 | Cage assembly in vitro and in living cells for cargo packaging and rheological probes. a, Size-exclusion chromatography elution profiles (top) and nsEM micrograph (bottom) comparing cages assembled in vitro from the original building blocks (dashed lines) and from fusion-protein building blocks (solid lines). Scale bar, 100 nm. **b**, Bulky RNP cargo (Cas9) does not fit in most natural (ferritin) and de novo cages (I53-50); nsEM shows GFP-nanobody-functionalized $T = 3$ cages encapsulating GFP-Cas9 (bottom, the reported morphology was observed across multiple fields of view from the same preparation). Scale bars, 20 nm (top), 100 nm (bottom). **c**, Confocal fluorescence images of fixed HEP3B cells incubated with ASGPR_EndoTag-functionalized and IGF_EndoTag-functionalized GFP-cages. Scale bar, 10 µm. **d**, Probe sizes relative to cellular length scales (left) and schematic of single-colour and dual-colour cages (right). mRNP, messenger ribonucleoprotein. **e**, Dual-colour cages (GFP/mScarlet) in U2OS cells. Scale bar, 10 µm. **f**, Single-particle tracking of single-colour C2-B-α10 cages (GFP) in U2OS cells. Individual particle trajectories (magnified view) were colour coded on the basis of their effective diffusivity ($D_{\text{eff-100ms}}$). Scale bar, 10 µm. **g**, Normalized probe intensity in fixed U2OS cells (data are presented as mean ± s.d. from $N = 3$ independent experiments; 40-nm GEMs, $n = 329$; C2-B-α13 cages, $n = 358$; C2-B-α10 cages, $n = 235$; a one-way ANOVA with Dunnett's multiple comparisons test was used). **h**, The effective diffusivity at 30 ms ($D_{\text{eff-30ms}}$) and 100 ms ($D_{\text{eff-100ms}}$) decreases with probe size of the particles (40-nm GEMs, $n = 3{,}358$ particles; C2-B-α13 cages, $n = 769$ particles; C2-B-α10 cages, $n = 3{,}157$ particles, median ± 95% CI). **i**, $D_{\text{eff-100ms}}$ for 40-nm GEMs and C2-B-α10 cages following 10 min sorbitol treatment (each data point represents one cell, $N = 3$ independent experiments, 40-nm GEMs, control $n = 55$, 100 mM sorbitol $n = 31$, 200 mM sorbitol $n = 23$; C2-B-α10 cage, control $n = 56$, 100 mM sorbitol $n = 28$, 200 mM sorbitol $n = 34$). Box bounds represent the 25th or 75th percentiles; whiskers represent the minimum or maximum; centre lines are the median. **j**, Cargo-loading schematic and nuclear import test based on confocal imaging: p53 localizes to the nucleus alone (left) but is retained in the cytosol when bound to cages (right). Scale bars, 10 µm. NsEM micrographs and confocal images are representative; key observations were reproduced in at least three independent experiments (**a**,**c**,**e**,**f**,**j**). All *P*-values were determined from Dunn's multiple comparison test (**g**).

conditions, whereas ASGPR-targeted cages showed higher uptake for smaller cages. Receptor engagement thus impacts cellular internalization, with size effects context-dependent rather than universal.

Intracellular cage assembly as rheological probes

Single-particle tracking nanorheology is a powerful strategy to probe the complex mesoscale physical properties of intracellular environments. These properties are dynamic and vary in a non-linear fashion across length- and timescales[33]. Passive rheological probes have been developed based on genetically encoded multimeric nanoparticles (GEMs) from natural scaffolds, including *Pyrococcus furiosus* (40-nm GEMs; ref. 34) and *Quasibacillus thermotolerans* (50-nm GEMs[35]); however, it has been difficult to make larger nanoparticles. Condensates such as µNS particles (50–150 nm) have been used to investigate longer length scales[36], on which distinct cellular biophysical properties emerge[37], but these condensates do not have a defined size. Thus, there is a need for well-defined rheological probes at the approximately 100 nm length scale relevant to cell biological processes such as vesicle transport. Given their tunable size and robust assembly, we hypothesized that our quasisymmetric protein cages could provide an orthogonal and genetically encoded approach for expanding the nanorheology toolbox to study physical properties in mammalian cells.

To study cytoplasmic physical properties at the mesoscale[34], we prepared two types of fluorescently labelled cage (Fig. 4d) and expressed them in human bone osteosarcoma epithelial cells (U2OS cells). To confirm that the observed fluorescence signals originated from particles formed with the designed two components instead of random aggregates, we co-expressed the C3-A protein fused with GFP and the C2-B-α10 linker protein fused to the mScarlet fluorescence protein[38] (Fig. 4e, Supplementary Fig. 16 and Supplementary Table 4). Using identical TrackMate analysis across samples within each condition, our C3-A-nGFP_C2-B-α10 cages displayed narrow Gaussian particle-intensity distributions—slightly wider than that of 40-nm GEMs—in contrast to the broad, heterogeneous distributions characteristic of synDrops (Supplementary Fig. 17). When assembled in test tubes, the C2-B-α10 construct forms 140-nm cages ($T \approx 16$) with a high particle assembly yield; 60% of the GFP-labelled particles were co-localized with mScarlet-labelled particles (the particle positions in the two channels were within 180 nm), whereas almost no GFP puncta were in close proximity with randomized spot positions, which is consistent with the cages being assembled from the two designed components.

We performed single-particle tracking analyses on C2-B-α13 and C2-B-α10 single-colour cages generated by fusing GFP to the N terminus of the scaffold C3-A protein and co-expressing them with two different C2-B linkers (Fig. 4f). Compared with previously developed 40-nm GEMs[33], an increase in GFP mean intensity was observed for both cages, consistent with their larger number of subunits (Fig. 4g and Supplementary Video 1). We analysed particle trajectories that were continuously monitored for more than ten frames and deduced the effective diffusivity at 30 ms ($D_{\text{eff-30ms}}$) and 100 ms ($D_{\text{eff-100ms}}$) from mean square displacements (MSDs). We found that the effective diffusivity of different particles negatively correlates with their in silico-designed size, making them useful rheological tools to infer physical parameters in mammalian cells. All three particle types had higher effective diffusivity at 30 ms than at 100 ms, indicating timescale-dependent physical constraints (Fig. 4h). This effect was more pronounced for cages designed in this study, suggesting greater intracellular constraints at larger length scales. To explore whether these probes can effectively sense cellular environmental changes, we used sorbitol treatment to acutely reduce intracellular water content, decrease cell volume and thereby increase macromolecular crowding. The diffusivity $D_{\text{eff-100ms}}$ of both 40-nm GEMs and C2-B-α10 cages decreased with increasing sorbitol concentrations, indicating that their diffusion is sensitive to physical property perturbations (Fig. 4i), as would be expected for a rheological tracer. We found that the particles can also form in *Saccharomyces cerevisiae* (yeast; Supplementary Fig. 18), indicating that they may be useful in diverse organisms.

We reasoned that tagging intracellular cargo molecules of interest could enable their recruitment into our quasisymmetric cages expressed and assembled in cells. To explore this, we created a stable cell line that expresses the C2-B-α10 cages fused to GFP. As a model system, we genetically fused the C terminus of the p53 protein with a anti-GFP nanobody and a HaloTag[29], and transiently transfected U2OS cells with the plasmid (Fig. 4j). The assembled C2-B-α10 cages effectively recruited the p53 protein through GFP binding to the GFP nanobody in U2OS cells, as evidenced by pronounced co-localization of the GFP-labelled cage and the halo-tagged p53 following addition of the Halo ligand. p53 is normally localized in the nucleus (Fig. 4j, left), but upon cage binding, the constructs become too large to enter the nucleus or invade mitotic chromatin and are thus restricted to the cytosol[39] (Fig. 4j, right). Introducing interface-disrupting C3-A mutations (L267K/R) impaired cage assembly and reduced p53 translocation, confirming that cargo redistribution requires intact cage formation (Supplementary Fig. 19). Thus, our designed multivalent protein cages can be loaded with cargo in cells, which could be useful for delivery applications, and for controlling subcellular protein localization.

Discussion

We demonstrate the computational design of two-component quasisymmetric protein assemblies. A key conceptual advance is the

departure from traditional explicit modelling strategies, which require multiple explicitly defined conformational states. Instead, our implicit multistate design approach uses locally flexible structures as inputs that approximate conformational ensembles, leading to particle closure through spontaneous symmetry breaking (Extended Data Fig. 1 and Supplementary discussion). By encoding the curvature of the overall assembly in the local geometry of the subunits, we are able to specify the overall size distribution. In the accompanying manuscript, Lee et al.[40] show that the same principle can be used to generate one-component quasisymmetric cages, demonstrating the generality of the approach. The tunable self-assembly of our quasisymmetric protein materials contrasts with the uncontrolled high-energy carbon vaporization approaches used in the commercial synthesis of fullerenes which make larger fullerenes difficult to access[41]. A fundamental limitation in carbon-based cages is that the C–C bond angle is fixed by atomic hybridization and cannot be reprogrammed; in our protein assemblies, by contrast, the effective inter-subunit bond angle is an explicit design parameter, enabling predictable and continuous tuning of particle size. The two component quasisymmetric cages designed in this work assemble robustly and reproducibly, driven by strong binding affinities between the heterodimeric interfaces grafted on the building blocks. Rapid nucleation and growth contribute to size heterogeneity, which can be qualitatively mitigated by modulating assembly conditions, such as ionic strength, temperature and crowding agents (Supplementary Figs. 20–22), which probably improve particle uniformity by slowing down nucleation and growth.

Our ability to predictably control particle size, from 40 to over 200 nm, and to functionalize components through genetic fusion, sets these de novo cages apart from naturally occurring viral capsids and other engineered protein particles. The modularity of our assembly approach enables precise control over individual components and straightforward cargo loading for biological applications. Our assemblies provide tools for passively probing the cytosolic biophysical environment, actively recruiting both intracellular and extracellular receptors, and as protein vehicles for cell-based delivery. The tagging of intracellular proteins with these particles could open new approaches in spatial proteomics[42] and CryoET-based structure determination[43]. More generally, these genetically encodable cages, along with methodologies for creating new ones, open up new approaches in synthetic biology and metabolic engineering[44]; for example engineered cells could use our assemblies as scaffolds to host customizable enzymatic cascades, analogous to bacterial microcompartments[45]. Our methodology, alongside rapid advancements in the de novo design of protein building blocks, lays the groundwork for a next generation of programmable biomaterials tailored to intracellular delivery and synthetic biology applications.

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/s41586-026-10464-0.

1. Caspar, D. L. D. & Klug, A. Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* **27**, 1–24 (1962).
2. Johnson, J. E. & Speir, J. A. Quasi-equivalent viruses: a paradigm for protein assemblies. *J. Mol. Biol.* **269**, 665–675 (1997).
3. Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. C60: buckminsterfullerene. *Nature* **318**, 162–163 (1985).
4. Shechtman, D., Blech, I., Gratias, D. & Cahn, J. W. Metallic phase with long-range orientational order and no translational symmetry. *Phys. Rev. Lett.* **53**, 1951–1953 (1984).
5. Kirchhausen, T., Owen, D. & Harrison, S. C. Molecular structure, function, and dynamics of clathrin-mediated membrane traffic. *Cold Spring Harb. Perspect. Biol.* **6**, a016725 (2014).
6. Hsia, Y. et al. Design of multi-scale protein complexes by hierarchical building block fusion. *Nat. Commun.* **12**, 2294 (2021).
7. Bale, J. B. et al. Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **353**, 389–394 (2016).
8. Walls, A. C. et al. Elicitation of potent neutralizing antibody responses by designed protein nanoparticle vaccines for SARS-CoV-2. *Cell* **183**, 1367–1382 (2020).
9. Boyoglu-Barnum, S. et al. Quadrivalent influenza nanoparticle vaccines induce broad protection. *Nature* **592**, 623–628 (2021).
10. Divine, R. et al. Designed proteins assemble antibodies into modular nanocages. *Science* **372**, eabd9994 (2021).
11. Yang, E. C. et al. Computational design of non-porous pH-responsive antibody nanoparticles. *Nat. Struct. Mol. Biol.* **31**, 1404–1412 (2024).
12. Sigl, C. et al. Programmable icosahedral shell system for virus trapping. *Nat. Mater.* **20**, 1281–1289 (2021).
13. Lee, S. et al. Four-component protein nanocages designed by programmed symmetry breaking. *Nature* **638**, 546–552 (2025).
14. Dowling, Q. M. et al. Hierarchical design of pseudosymmetric protein nanocages. *Nature* **638**, 553–561 (2025).
15. Sadoc, J.-F. & Mosseri, R. *Geometrical Frustration* (Cambridge Univ. Press, 2006).
16. Perlmutter, J. D. & Hagan, M. F. Mechanisms of virus assembly. *Annu. Rev. Phys. Chem.* **66**, 217–239 (2015).
17. Wang, S. et al. Bond-centric modular design of protein assemblies. *Nat. Mater.* **24**, 1644–1652 (2025).
18. Watson, J. L. et al. De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).
19. Sahtoe, D. D. et al. Reconfigurable asymmetric protein assemblies through implicit negative design. *Science* **375**, eabj7662 (2022).
20. Khmelinskaia, A. et al. Local structural flexibility drives oligomorphism in computationally designed protein assemblies. *Nat. Struct. Mol. Biol.* **32**, 1050–1060 (2025).
21. 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).
22. Morris, K. L. et al. Cryo-EM of multiple cage architectures reveals a universal mode of clathrin self-assembly. *Nat. Struct. Mol. Biol.* **26**, 890–898 (2019).
23. DiMaio, F., Leaver-Fay, A., Bradley, P., Baker, D. & André, I. Modeling symmetric macromolecular structures in Rosetta3. *PLoS ONE* **6**, e20450 (2011).
24. Wang, R. Y.-R. et al. Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. *eLife* **5**, e17219 (2016).
25. Saito, R., Dresselhaus, G. & Dresselhaus, M. S. Topological defects in large fullerenes. *Chem. Phys. Lett.* **195**, 537–542 (1992).
26. Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**, 79–88 (2006).
27. Wu, K. et al. Design of intrinsically disordered region binding proteins. *Science* **389**, eadr8063 (2025).
28. Kubala, M. H., Kovtun, O., Alexandrov, K. & Collins, B. M. Structural and thermodynamic analysis of the GFP:GFP-nanobody complex. *Protein Sci.* **19**, 2389–2401 (2010).
29. Los, G. V. et al. HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol.* **3**, 373–382 (2008).
30. Zakeri, B. et al. Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl Acad. Sci. USA* **109**, E690–E697 (2012).
31. Chen, J., Li, Q. & Wang, J. Topology of human apolipoprotein E3 uniquely regulates its diverse biological functions. *Proc. Natl Acad. Sci. USA* **108**, 14813–14818 (2011).
32. Huang, B. et al. Designed endocytosis-inducing proteins degrade targets and amplify signals. *Nature* **638**, 796–804 (2025).
33. Luby-Phelps, K., Taylor, D. L. & Lanni, F. Probing the structure of cytoplasm. *J. Cell Biol.* **102**, 2015–2022 (1986).
34. Delarue, M. et al. mTORC1 controls phase separation and the biophysical properties of the cytoplasm by tuning crowding. *Cell* **174**, 338–349 (2018).
35. Hernandez, C. M., Duran-Chaparro, D. C., Van Eeuwen, T., Rout, M. P. & Holt, L. J. Development and Characterization of 50 nanometer diameter genetically encoded multimeric nanoparticles. Preprint at *bioRxiv* https://doi.org/10.1101/2024.07.05.602291 (2024).
36. Parry, B. R. et al. The bacterial cytoplasm has glass-like properties and is fluidized by metabolic activity. *Cell* **156**, 183–194 (2014).
37. Xie, Y. et al. Polysome collapse and RNA condensation fluidize the cytoplasm. *Mol. Cell* **84**, 2698–2716 (2024).
38. Bindels, D. S. et al. mScarlet: a bright monomeric red fluorescent protein for cellular imaging. *Nat. Methods* **14**, 53–56 (2017).
39. Cuylen-Haering, S. et al. Chromosome clustering by Ki-67 excludes cytoplasm during nuclear assembly. *Nature* **587**, 285–290 (2020).
40. Lee, S. et al. Design of one-component quasisymmetric protein nanocages. *Nature* https://www.doi.org/10.1038/s41586-026-10554-z (2026).
41. Mojica, M., Alonso, J. A. & Méndez, F. Synthesis of fullerenes. *J. Phys. Org. Chem.* **26**, 526–539 (2013).
42. Lundberg, E. & Borner, G. H. H. Spatial proteomics: a powerful discovery tool for cell biology. *Nat. Rev. Mol. Cell Biol.* **20**, 285–302 (2019).
43. Fung, H. K. H. et al. Genetically encoded multimeric tags for subcellular protein localization in cryo-EM. *Nat. Methods* **20**, 1900–1908 (2023).
44. Volk, M. J. et al. Metabolic engineering: methodologies and applications. *Chem. Rev.* **123**, 5521–5570 (2023).
45. Sutter, M., Greber, B., Aussignargues, C. & Kerfeld, C. A. Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell. *Science* **356**, 1293–1297 (2017).

Methods

Computational design of C2-B linkers

The C3-A protein used in this study is a designed homotrimeric protein[17] with C-terminal rigid fusions of split heterodimer LHD101 A chain[19] (residues 38–75, PDB ID: 7MWR), which forms a specific non-covalent interface with its binding partner, the LHD101 B chain. Given a target input parameter set (arc length l and cone angle α), the design calculation of the C2-B linker was initialized by symmetrically arranging six copies of C3-A proteins along the z axis in a cyclic C6 symmetry. We then placed LHD101 B chains adjacent to C3-A proteins as scaffolding motifs. The rotational degree of freedom ω of C3-A along its threefold axis was sampled to control the relative position of C2 binding domain B, for which we targeted a small gap distance to avoid excessive flexibility. A sample Python script was used for model construction and a test input protein model is available via Zenodo[46]. RFdiffusion generates C2-symmetric homodimers that maintain the pre-specified heterodimer binding sites (Extended Data Fig. 2). Following sequence design with ProteinMPNN[47] (the '--tied_positions' option was selected to ensure that two chains are identical), the designs were confidently predicted using AlphaFold2 (ref. 48) to assemble as C2-symmetric oligomers (pLDDT > 92, r.m.s.d. < 1.2 Å), and to scaffold the heterodimer interface with high accuracy. Finally, to facilitate alternative pentameric ring closure, we docked AlphaFold2-predicted structures of C2-B in both six-membered and five-membered ring configurations in silico, and subselected for designs that require a terminal bending of below 4 Å to close the five-membered ring.

Protein expression and purification from *E. coli*

The full list of protein constructs designed and tested in this study is provided in Supplementary Table 5 and the Supplementary DNA Sequences. Synthetic genes for computationally filtered designs were purchased from Integrated DNA Technologies and cloned into the pET29b+ vector using NdeI and XhoI restriction sites. Constructs were expressed in BL21* (DE3) *E. coli*-competent cells with a C-terminal polyhistidine tag. For small-scale expression, transformed cells were cultured in 50 ml of terrific broth supplemented with 200 mg l^{-1} kanamycin and induced for 24 h at 37 °C under the T7 promoter. Cells were harvested by centrifugation, resuspended in Tris-buffered saline, and lysed with five minutes of sonication. The lysates were then subjected to nickel-affinity chromatography, washed with ten column volumes of 40 mM imidazole and 500 mM NaCl, and eluted with 500 mM imidazole and 300 mM NaCl. Successful complex formation was confirmed by the presence of both oligomers on sodium dodecyl sulfate–polyacrylamide gel electrophoresis following Ni-NTA pulldown. Proteins of the correct molecular weights were analysed by electron microscopy. Selected designs were scaled up to 0.5 l for expression and purification under the same conditions.

In vitro cage assembly

For standard in vitro assembly, purified cage components were mixed at equimolar ratios to 40 μM per subunit in SEC elution buffer (25 mM Tris-HCl, pH 8.0; 300 mM NaCl) at room temperature. Assemblies also formed across a broad range of conditions: component concentrations of 5–200 μM and a NaCl concentration of 150 mM to 1.5 M, which modulated particle size and structural uniformity in a construct-dependent manner. For many assemblies, cage uniformity improved when NaCl was increased to 0.6–1.0 M and components were pre-incubated at 50 °C followed by rapid mixing. We also observed enhanced uniformity when components were directly mixed in the immobilized metal affinity chromatography elution buffer (25 mM Tris-HCl, pH 8.0; 500 mM imidazole; 300 mM NaCl), which probably facilitates assembly by stabilizing the subunits and suppressing premature aggregation.

Negative stain electron microscopy

Size-exclusion chromatography purified sample fractions obtained from in vitro assembly were diluted to a final concentration of about 0.5 μM (monomer equivalent) for nsEM analysis. For each fraction, a 6 μl sample was applied to glow-discharged 400-mesh copper grids coated with formvar and carbon (Ted Pella) and left to adsorb for over 2 min. Excess sample was blotted, and the grid was stained with 6 μl of 2% uranyl formate. After blotting, an additional 6 μl of uranyl formate was applied for 20 s before a final blotting step. Imaging was performed using a Talos L120C transmission electron microscope operating at 120 kV. Automated nsEM datasets were collected using the E. Pluribus Unum software (FEI Thermo Fisher Scientific), and further processed using CryoSPARC software[49]. Micrographs were uploaded to the CryoSPARC platform, through which contrast transfer function corrections were performed. Two-dimensional class averages and three-dimensional maps were generated using CryoSPARC. Particles from selected classes were used to generate an ab initio model, which was subsequently refined using C1 symmetry and further adjusted for I symmetry as needed.

Identification of off-targets assemblies from nsEM two-dimensional class averages

In addition to the major assembly product of $T = 3$ cage (structural analogue of C60), we also observed a distribution of minor species under nsEM micrographs two-dimensional class averages (Extended Data Fig. 4a), including particles of different sizes and anisotropic shapes. Based on careful examination of two-dimensional class averages, we isolated a subset of particles that correspond to a $T = 4$ icosahedron (C80) through a three-dimensional nsEM map; however, we could not reconstruct and identify other particles due to polydispersity, and we speculate that these assemblies correspond to a variety of fullerene cage analogues, including C70 (Extended Data Fig. 4b,c). The formation of these non-uniform off-target assemblies could be kinetically controlled, which we aim to optimize in future studies. For both $T = 3$ and $T = 4$ cages, all pentagonal rings are located at the vertices of an icosahedron, however they no longer occupy vertices when lower symmetry particles are formed.

X-ray crystallography

C2-B-α20 linker crystals were grown using protein purification as described above. A concentrated sample was used for crystallization. The crystallization screening was performed using a Mosquito LCP by STP Labtech. Crystals grew in 2.0 M ammonium sulfate and 0.1 M sodium acetate (pH 4.6). Crystals were harvested directly from a screening tray, and flash cooled in liquid nitrogen. Glycerol (20%) was used as a cryoprotectant. X-ray diffraction was performed at NSLS2 beamline 17-ID-2; the crystal diffracted to 2 Å. Data were processed using the X-ray detector software[50], and merged or scaled using the pointless or aimless functions in the CCP4 program suite[51]. The structure was phased by molecular replacement using the designed structure as the search model in Phaser[52], and then refined with Phenix[53]. Following molecular replacement, the models were improved, and efforts were made to reduce model bias. The final structure was refined in Phenix. Model building was performed using COOT[54]. The final model was evaluated using MolProbity[55]. Data collection and refinement statistics are recorded in Extended Data Table 1. Data deposition, atomic coordinates and structure factors reported in this paper have been deposited in the PDB under accession code 9NDL.

Cryo-EM and cryo-ET data collection and processing

Three microlitres of purified cages (~40 μM) in 25 mM Tris-HCl and 300 mM NaCl (pH 8) were applied to glow-discharged 300 mesh R2/2 holey carbon C-flat grids. The grids were blotted using a Vitrobot Mark IV (Thermo Fisher Scientific) for 4 s at room temperature and about

100% chamber humidity. For data collection, the grids were imaged on a Glacios cryo-electron microscope (Thermo Fisher Scientific) operated at 200 kV using SerialEM data collection software[56] and images were recorded on a K3 Summit electron detector at a magnification of 22,000×, at 2× binning corresponding to a calibrated sampling of 1.82 Å per pixel. Tilt series were acquired using a dose-symmetric collection scheme[57] at 3° increments with low dose settings and a defocus range of −2 to −5 µm. Each tilt series was composed of 41 individual tilt images, with a total accumulated dose of 110 e^- Å^{-2}. Tomogram analysis steps were performed using the EMAN2 (v.2.3) Tomo pipeline. Tilt image movie stacks were motion-corrected using MotionCor2[58], aligned and reconstructed into three-dimensional tomogram volumes using e2_tomogram.py. Contrast transfer function estimation for each tilt image was performed using the EMAN program e2spt_tomoctf.py. Data deposition, atomic coordinates and structure factors reported in this paper have been deposited in the PDB under accession code 9OP9 (Extended Data Table 2)

For single particle cryo-EM analysis, the C2-B-α20 sample dataset ($T = 3$ cage) was collected automatically using SerialEM[56] and used to control a ThermoFisher Titan Krios 300 kV. The microscopes were equipped with a K3 Summit direct electron detector[59] and BioQuantum Gif energy filter and operated in counting mode. Random defocus ranges spanned between −0.8 and −1.8 µm using image shift. Then, 7,920 movies with a pixel size of 0.843 and total dose of 48.3 e^- Å^{-2} were recorded with a total exposure of 5 s over 99 frames. All data were processed in CryoSPARC[49]. Default parameters were used for refinement unless otherwise noted. The video frames were aligned using Patch Motion, and defocus and astigmatism values were estimated using the patch contrast transfer function. The 'curate exposures' function was used to remove 1,295 micrographs, leaving 6,625 good-quality micrographs (Supplementary Fig. 8a). Micrographs were then denoised in using the 'micrograph denoiser' function with the pretrained model. Then, 210 particles were picked on 21 micrographs to generate templates. Using these templates, the 'template picker' function was used with particle diameter set to 500 Å; the maximum number of local maxima to consider was set to 100; 'pick on denoised micrographs' was set to true; 278,552 particles were picked, and after an 'inspect picks' job, 258,422 particles were remaining. In total, 142,674 particles were extracted with a box size of 1,400 pixels, Fourier cropped to 350 pixels. Two-dimensional classification was run with 300 classes, with the number of online-electron-microscopy iterations set to 40, and then number of final full iterations set to 10. Four classes were selected with 49,467 particles and another round of two-dimensional classification was run with: 50 classes; the number of online-electron-microscopy iterations set to 40; and the number of final full iterations set to 10. Eight classes with 49,224 particles were selected (Supplementary Fig. 8b). An ab initio was run with 5 volumes using Icosahedral symmetry (Supplementary Fig. 8c). 22,924 particles generated an ab initio volume that closely matched the expected design structure and tomography reconstruction. This volume and particles were used for a non-uniform refinement (Supplementary Fig. 8d–f). All volumes were shown using ChimeraX[60]. Data deposition, atomic coordinates, and structure factors reported in this paper have been deposited in the PDB under accession code 9OM3.

Construction of cage models based on an electron microscope density map

Initial hexamer and pentamer models were created by taking the designed model as the hexamer model, and creating a C5 pentamer model using Rosetta's make_symmdef_file (ref. 22) to build a C5 complex that preserved as much as possible of the inter-subunit geometry from the original hexameric design. These C5 and C6 rings were then docked into the cryo-EM density using Chimera. When C5 and C6 subunits overlapped, the C6 subunits were used as a reference. This gave us a rough initial model of the cage.

Then the full cage structure was refined using Rosetta's cryo-EM refinement protocol[23]. One asymmetric unit was extracted from the initial rough model. We then used Rosetta's relax with icosahedral symmetry restraints to refine the entire complex. Furthermore, during this refinement, weak atom-pair constraints were used for all subunit interfaces in the asymmetric unit to prevent the various interfaces from diverging too much. All interface alpha carbons within 9 Å were constrained to maintain their distances from the original hexamer design. A single repeat of relax yielded the final model shown in Fig. 3c.

Cage cellular uptake experiments

For confocal imaging: HEP3B cells (ATCC HB-8064, authenticated and tested for mycoplasma contamination by provider) were plated at 35,000 cells per well in an eight-well Labtek plate a day before treatment. GFP-labelled IGF2R, ASGPR and a sortilin cage (250 nM each) were added to HEP3B cells for a 24 h treatment. Cells were then fixed with 4% paraformaldehyde and stained with LAMP1 for lysosomal staining and DAPI for nuclear staining. Cells were washed with Dulbecco's phosphate-buffered saline (DPBS) and imaged with a Nikon A1R confocal microscope using a Plan Fluor ×60, 1.30-NA oil objective.

For flow cytometry: HEP3B cells were plated at 7,500 cells per well in a 96-well plate a day before the treatment. GFP-labelled cages (250 nM) were added for a 3 h, 6 h and 24 h treatment. Cells were then trypsinized, quenched and transferred to a 96-well-U-bottomed plate. Cells were washed three times with DPBS and were analysed on an Attune Flow Cytometer (Thermo Fisher Scientific).

In cellulo expression of cages

Mammalian cell lines: U2OS cells were cultured in Dulbecco's modification of Eagle's medium (DMEM; Gibco) containing high glucose and sodium pyruvate supplemented with 10% foetal bovine serum (Gemini bio-products), 50 U ml^{-1} penicillin and 50 µg ml^{-1} streptomycin (Gibco), and maintained at 37 °C in a humidified incubator with 5% CO_2. Cells were regularly split in fresh medium upon reaching 80–90% confluency. All cells were routinely tested for mycoplasma by PCR screening of the conditioned medium.

Transient transfection of mammalian cell lines: to express particles in mammalian cells, we introduced plasmids encoding the particle-forming two-component monomer proteins into cells. Gene fragments were synthesized by GenScript and assembled into plasmid backbones via Gibson assembly. All constructs were sequence-verified by nanopore sequencing services (Plasmidsaurus). Plasmids were then delivered by either transient transfection or lentiviral transduction, enabling intracellular protein expression and self-assembly into particles inside cells.

For transfection, U2OS cells were seeded as 60–70% confluency in a six-well glass-bottomed plate (Cellvis) on the day before transfection and were transfected with 1 µg of plasmid DNA per well using FuGENE HD (Promega) reagent per manufacturer guidelines. Fresh DMEM was replaced 24 h after transfection. Imaging experiments were usually performed between 48 to 72 h post-transfection.

Lentivirus production and cell transduction for stable cell lines

Human embryonic kidney cells (HEK293T cells, 6×10^6 per 10 cm dish) were plated in antibiotic-free DMEM supplemented with 10% FaBS. The next day, cells were transfected with transgene plasmid (pLH2601 or pLH2704) together with lentivirus packaging plasmids pMD2.G (a gift from D. Trono; Addgene plasmid no. 12259) and psPAX2 (a gift from D. Trono; Addgene plasmid no. 12260), using FuGENE HD transfection reagent following manufacturer's protocol. After 24 h, 6 ml antibiotic-free DMEM was replaced and the supernatants were collected at both 48 and 72 h post-transfection, and stored at 4 °C. Fresh virus titres (1.5 ml) were added to a 35-mm dish culture of U2OS cells.

The rest of the virus titres were concentrated by centrifugation at 4,000 rcf for 30 min in an Amicon Ultra-15 30-KDa centrifugal filter (MilliporeSigma). Concentrated viral suspensions were aliquoted and stored at −80 °C until later use. After cell lines stabilized, they were expanded in culture to 10 cm dishes, and enriched for high GFP expression cell populations by fluorescence-activated cell sorting. The cell lines were frozen in 10% DMSO (Sigma-Aldrich) in DMEM, and thawed for use in experiments whenever needed.

Passive rheological probes imaging and single-particle tracking

To image 40-nm GEMs and de novo cages in U2OS cells, Andor Yokogawa CSU-X confocal spinning disc on a Nikon Ti2 X1 microscope was used at 488 nm excitation with 100% laser power. Fluorescence was recorded with a sCMOS Prime 95B camera (Photometrics) with a 60× objective (pixel size = 0.18 μm), at a 100 fps image capture rate for a total of 2 s. The tracking of particles was performed with the Mosaic suite of FIJI, using the following parameters: radius = 3, cutoff = 0, per/abs = 0.5, a link range of 1, and a maximum displacement of 5 px, assuming Brownian dynamics. All trajectories were then analysed using the GEM single particle analysis software package being developed in house[51]. Mean-square displacement (MSD) was calculated for every two-dimensional trajectory, and trajectories continuously followed for more than ten time points were used to fit with linear time dependence based on the first ten time intervals to quantify time-averaged MSD: $\mathrm{MSD}(T) = 4D_{\mathrm{eff}}T$, where T is the imaging time interval and D_{eff} is the effective diffusivity (units are $\mu m^2\ s^{-1}$). To determine the ensemble-time-averaged MSD, all trajectories were fitted with $\mathrm{MSD}(\tau)_{\mathrm{T-ens}} = 4D\tau^{\alpha}$, where α is the anomalous exponent, with $\alpha = 1$ being Brownian motion, $\alpha < 1$ suggests sub-diffusive motion and $\alpha > 1$ as super-diffusive motion. All analysed trajectories within one cell were averaged to present single cell D_{eff} for each condition (Supplementary Fig. 16).

Measurement of particle intensity and co-localization

To fix U2OS cells, 4% paraformaldehyde in PBS was added to each of the imaging wells for 10 min at room temperature. The cells were then washed three times in 1× PBS before imaging. Andor Yokogawa CSU-X confocal spinning disc on a Nikon Ti2 X1 microscope was used to capture particles in the cell at one focal plane. The particles were identified using the LoG detector algorithm with manual thresholding and quality examination performed using the trackmate plugin, Fiji[61]; the mean intensity and the x,y coordinates of each particle were measured. The mean intensity of the cytosolic regions without particles was also measured across cells to determine the background intensity. Thereafter, the background-subtracted GFP mean intensity was calculated for each particle. The median value of all the quantified particles was denoted for each type of particle in individual experiment, and the fold change of particle intensity was calculated by normalizing the median value of de novo cages to 40-nm GEMs.

A custom Python script was developed to quantify the co-localization between GFP and mScarlet particles. Owing to potentially immobile bright particles being detected in the mScarlet channel but not in the GFP channel, we set up criteria to pair exclusive green and red particles before measuring the distance between them. Initially, all green particles were assigned to the nearest-neighbour red particles; if more than one green particle were paired with the same red particle, they were not considered an exclusive pairing and thereby removed from further analysis. The identified pairs of green and red particles were considered for co-localization only if their locations were within 180 nm (1 pixel). As a control analysis, randomized green particle locations were generated and assigned within the cytosolic region of the cell, and the same function was used to identify pairings of green and red particles, and then measure the distance between them. We reported the percentage of co-localized green particle numbers over the total exclusive pairings of green–red particles in individual cells.

Reporting summary

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

Data availability

The principal data supporting the findings of this work are available in the figures and the Supplementary Information. Design models for the cage building blocks are available via Zenodo at https://doi.org/10.5281/zenodo.18892841 (ref. 46). The heterodimer protein pair LHD101 was derived from ref. 19 (PDB ID 7MWR). Coordinates and structure factors are available in the PDB with the following accession codes: 9NDL (C2-B-α20), 9OM3 ($T = 3$ cage, cryo-EM structure) and 9OP9 ($T = 3$ cage, cryo-ET structure). Source data are provided with this paper.

Code availability

RFDiffusion code can be downloaded from https://github.com/RosettaCommons/RFdiffusion. ProteinMPNN is available from https://github.com/sokrypton/ColabDesign/. AlphaFold2 is available from https://github.com/google-deepmind/alphafold. Scripts for generating curved hexagon are available via Zenodo at https://doi.org/10.5281/zenodo.18892841 (ref. 46)

46. Shunzhi, W. QS-2compCage-BBs_Hexagon-Generator. *Zenodo* https://doi.org/10.5281/zenodo.18892840 (2026).
47. Dauparas, J. et al. Robust deep learning–based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022).
48. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
49. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
50. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
51. Winn, M. D. et al. Overview of the *CCP* 4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
52. McCoy, A. J. et al. *Phaser* crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
53. Adams, P. D. et al. *PHENIX*: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
54. Emsley, P. & Cowtan, K. *Coot*: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
55. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
56. Mastronarde, D. N. SerialEM: a program for automated tilt series acquisition on tecnai microscopes using prediction of specimen position. *Microsc. Microanal.* **9**, 1182–1183 (2003).
57. Hagen, W. J. H., Wan, W. & Briggs, J. A. G. Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *J. Struct. Biol.* **197**, 191–198 (2017).
58. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
59. Keegan, S., Fenyö, D. & Holt, L. J. GEMspa: a Napari plugin for analysis of single particle tracking data. Preprint at *bioRxiv* https://doi.org/10.1101/2023.06.26.546612 (2023).
60. Meng, E. C. et al. UCSF ChimeraX: tools for structure building and analysis. *Protein Sci.* **32**, e4792 (2023).
61. Ershov, D. et al. TrackMate 7: integrating state-of-the-art segmentation algorithms into tracking pipelines. *Nat. Methods* **19**, 829–832 (2022).

Acknowledgements We thank D. Juergens for their help with computational design and scripts; F. Praetorius and D. Sahtoe for contributing de-novo-designed LHD heterodimer building blocks; Q. Dowling and W. Sheffler for helpful discussions; S. Gerben for their assistance with protein purification; N. Bethel for their assistance with molecular dynamics simulations; Y. Hsia for their assistance with field flow fractionation; and S. Dickinson and J. Quipse for their help with maintaining and operating the electron microscopes. Molecular graphics and cage models were generated with UCSF ChimeraX. This work was supported by the Defense Threat Reduction Agency (grant no. HDTRA1-19-1-0003 to S.W.) and the Burroughs Wellcome Fund (S.W.); The Audacious Project at the Institute for Protein Design (S.W., A.F., A.J.B., A.B., J.D., A.K., D.B.); a gift from Microsoft (D.C.); the National Science Foundation (grant no. CHE-2226466 to F.D.); the Human Frontiers Science Program (grant no. RGP0061/2019 to F.D.); the Bill and Melinda Gates Foundation (grant no. INV-043758 to C.W., J.D., A.K., A.B. and no. OPP1156262 to A.J.B.); the Howard Hughes Medical Institute (C.W., A.B., A.J.B., D.B.); the National Institutes of Health's National Institute on Aging (grant no. R01AG063845 to A.B., A.F. A.J.B. and B.H.); Spark Therapeutics/Computational Design of a Half Size Functional (grant no. ABCA4 to K.W.); and The Nordstrom Barrier Institute for Protein Design Directors Fund (B.H.).

Crystallographic work was done at the National Synchrotron Light Source II on beamline FMX (17-ID-2). The Center for Bio-Molecular Structure (CBMS) is primarily supported by the NIH-NIGMS through a Center Core P30 Grant (grant no. P30GM133893), and by the DOE Office of Biological and Environmental Research (grant no. KP1607011). NSLS-II is a US DOE Office of Science User Facility (operated under contract no. DE-SC0012704). This publication resulted from data collected using the beamtime obtained through NECAT BAG proposal no. 311950. L.J.H. was funded by NIH (grant nos. R01 GM132447 and R37 CA240765) the NIH Director's Transformative Research Award (no. TR01 NS127186), a Hypothesis Fund Award and a Research Grant from HFSP (grant no. RGP0016/2022).

Author contributions S.W. and D.B. conceptualized the study. S.W. developed the computational pipeline, performed the design calculations and experimentally characterized designs. D.C., C.W., S.W. and A.J.B. performed cryo-EM characterization and analyses. Y.X., T.S., C.H., D.C.D, J.D. and L.J.H. designed and developed experimental approaches for studying cage assemblies in live cells and performed imaging analyses. F.D. constructed the cage model from electron microscope density. W.C. designed the cargo-loading experiments. G.A. and B.H. designed and performed cellular uptake experiments. A.B., A.K., E.B. and E.J. determined the crystal structures. S.L. performed and analysed the molecular dynamics simulations. R.D.K., A.C., K.W. and A.F. performed protein purification experiments. L.J.H. and D.B. supervised research. S.W. and D.B. wrote the original manuscript with input from Y.X., L.J.H. and D.C. All authors read and edited the manuscript.

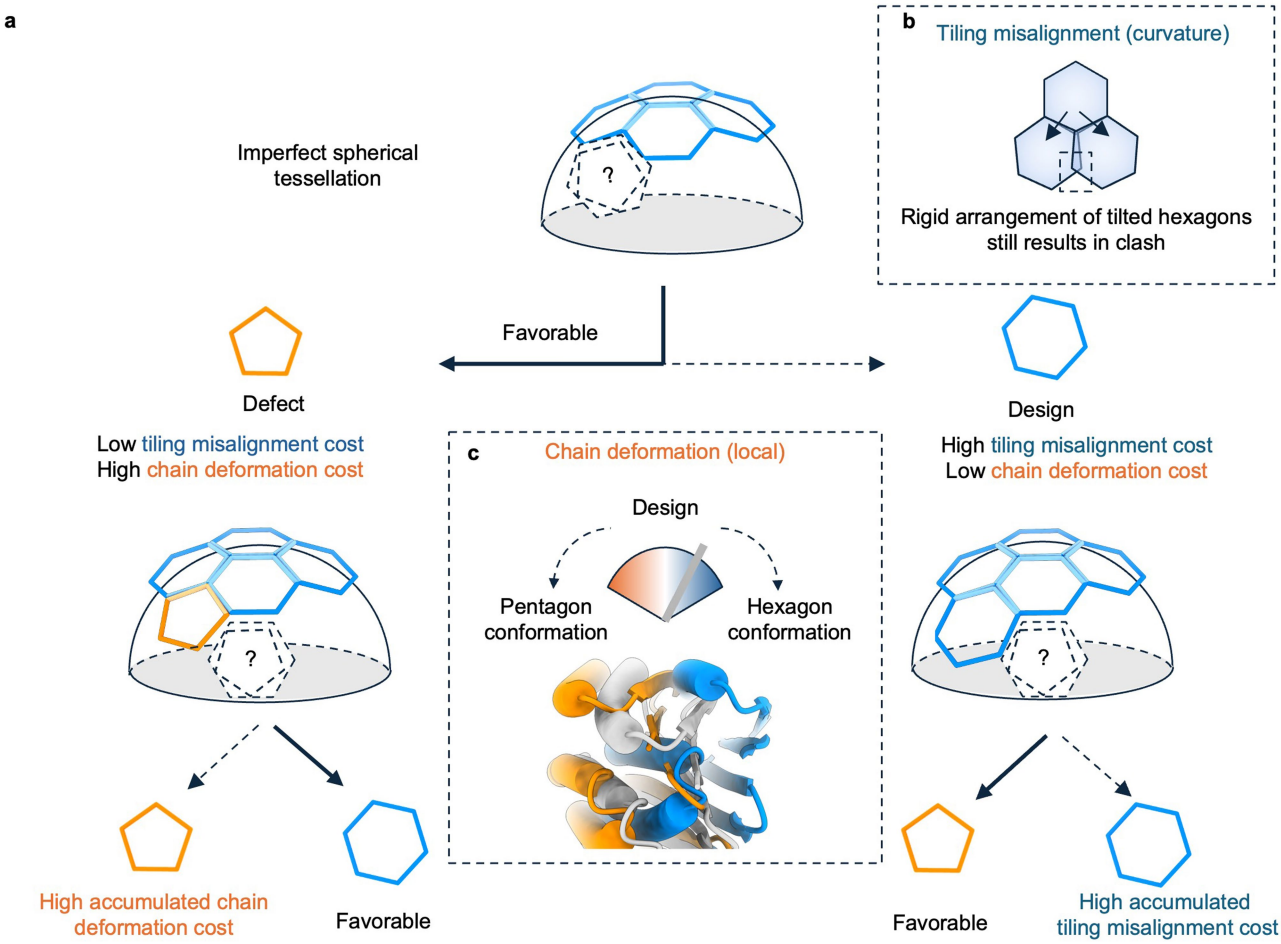
Competing interests D.B., S.W., S.L., D.C. and R.D.K. are co-inventors on a provisional patent application that incorporates discoveries described in this manuscript. The remaining authors declare no competing interests.

Additional information

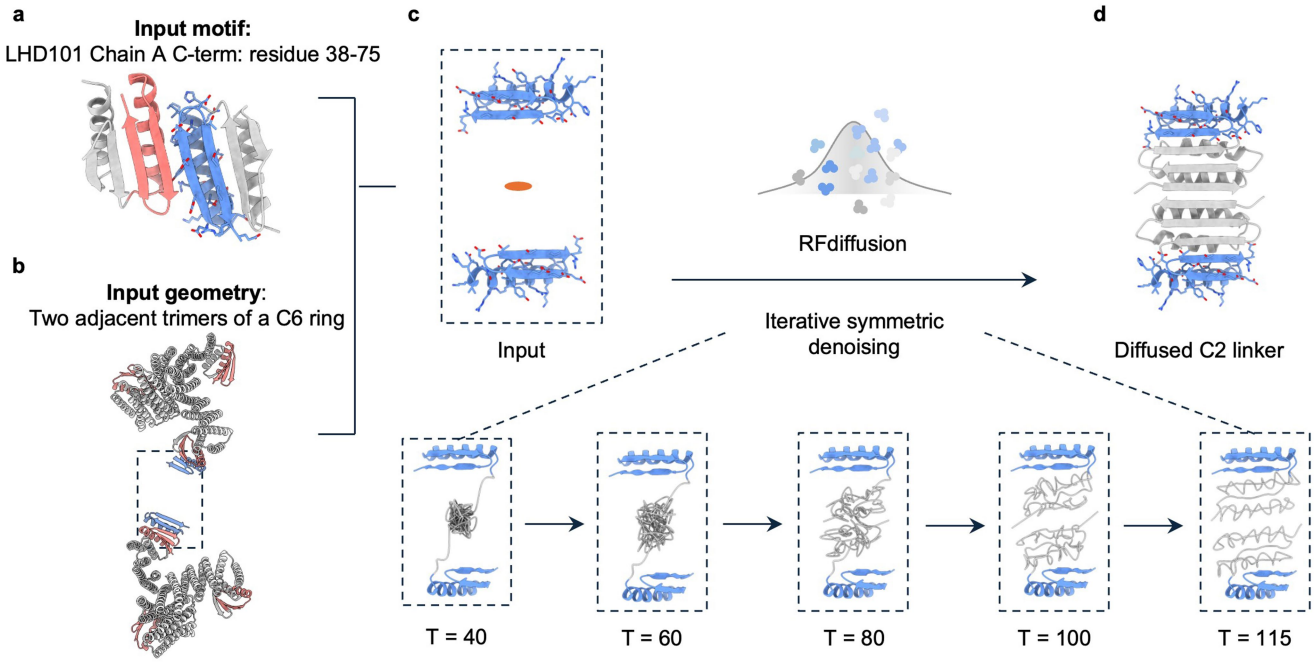
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-026-10464-0.

Correspondence and requests for materials should be addressed to Shunzhi Wang or David Baker.

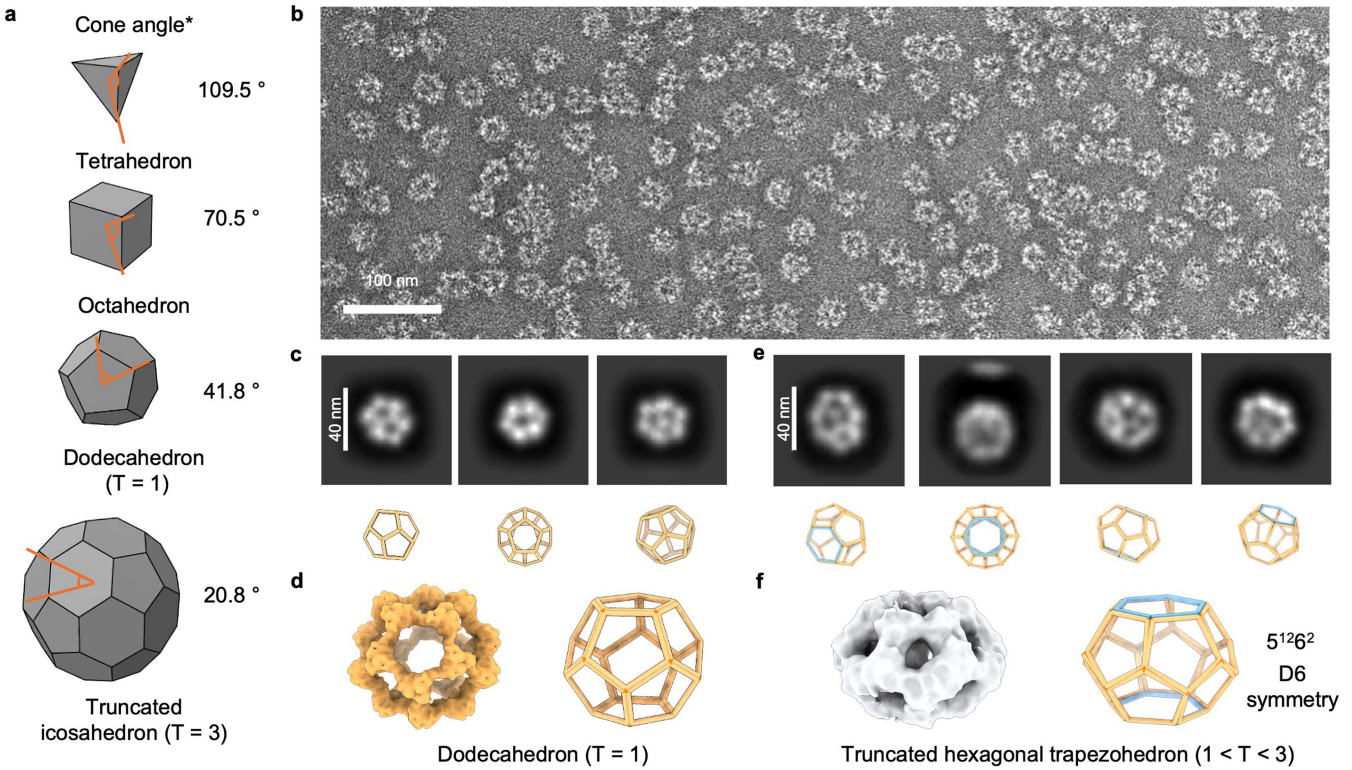
Peer review information *Nature* thanks Guang Hu and the other, anonymous, reviewers for their contribution to the peer review of this work. Peer reviewer reports are available



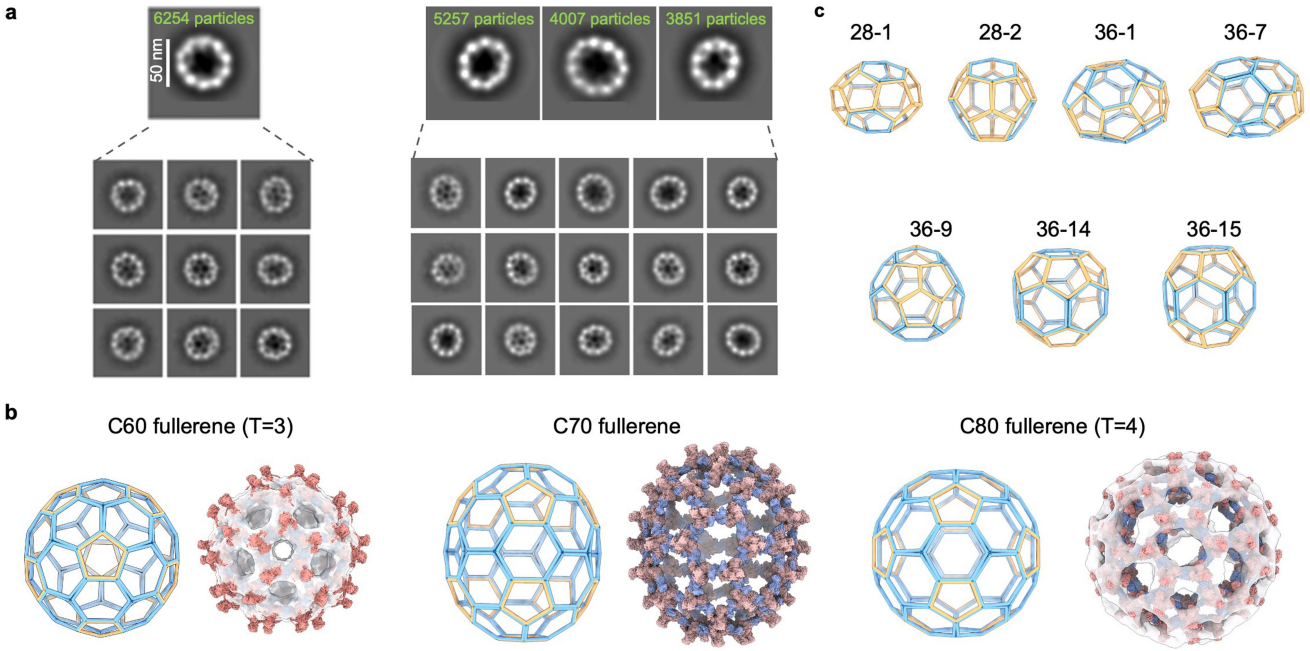
Extended Data Fig. 1 | Geometric frustration encountered during cage closure. a) Symmetry breaking as a result of two contrasting additions to a self-limiting growing assembly: left, formation of pentagonal defect that accounts for the sphere's positive curvature, right, formation of designed hexagonal assembly which minimizes local chain deformation. The different types of energy costs are noted below each case: **b**) Tiling misalignment cost is a result of rigid arrangement of tilted hexagons which cannot perfectly tile spherical surfaces, and therefore need to bend to avoid clashes. **c**) Chain deformation cost is the energy required for the designed building block to bend into desired local pentagonal or hexagonal conformation.



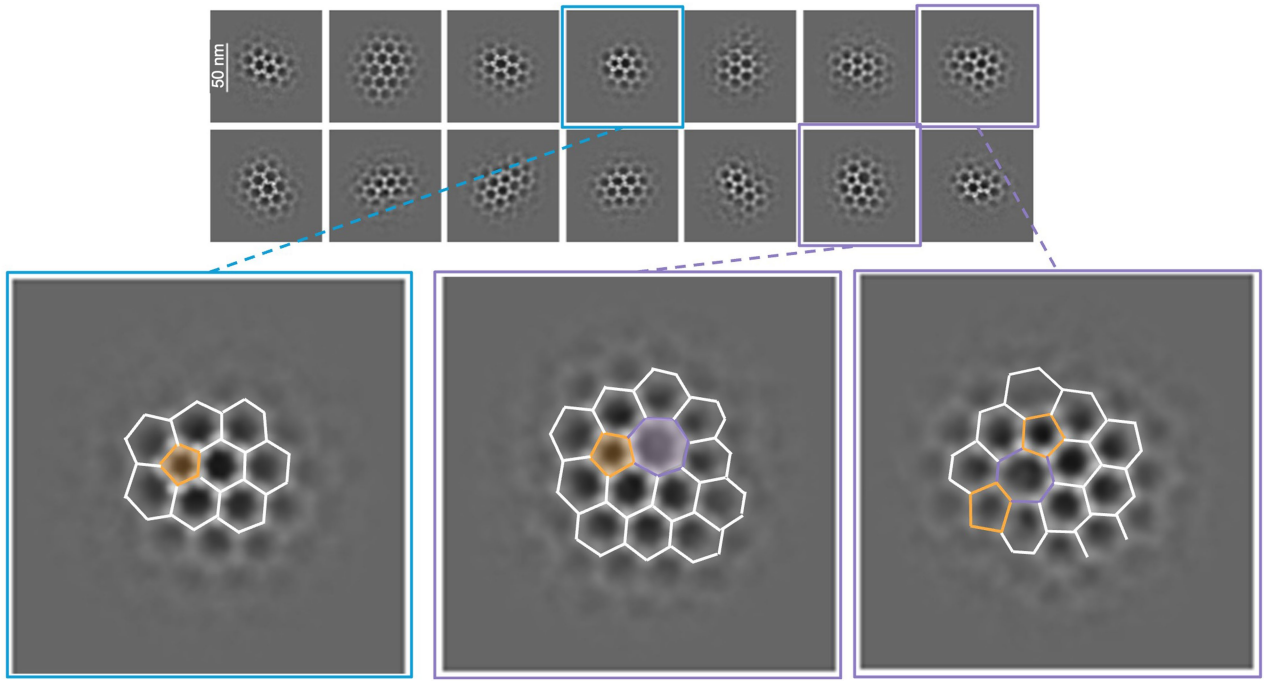
Extended Data Fig. 2 | Computational generation of C2 linker Backbones with RFdiffusion. a) Cartoon representation of input motif (blue, LHD101 chain A, residue 38–75) extracted from a previously designed heterodimer LHD101. The trimeric building block C3-A is rigidly fused with the complementary binding partner LHD101B (red). **b**) As C3-A trimers are symmetrically arranged in a tilted hexagon at a defined cone angle, two adjacent input motifs (blue) are extracted to create a C2 dimer with missing core. **c**) Symmetric RFdiffusion generatively designs a C2 backbone scaffold to stabilize input motifs in the fixed pre-specified geometry. **d**) Generated C2-B linkers were further sequence designed with ProteinMPNN and filtered by AlphaFold2.



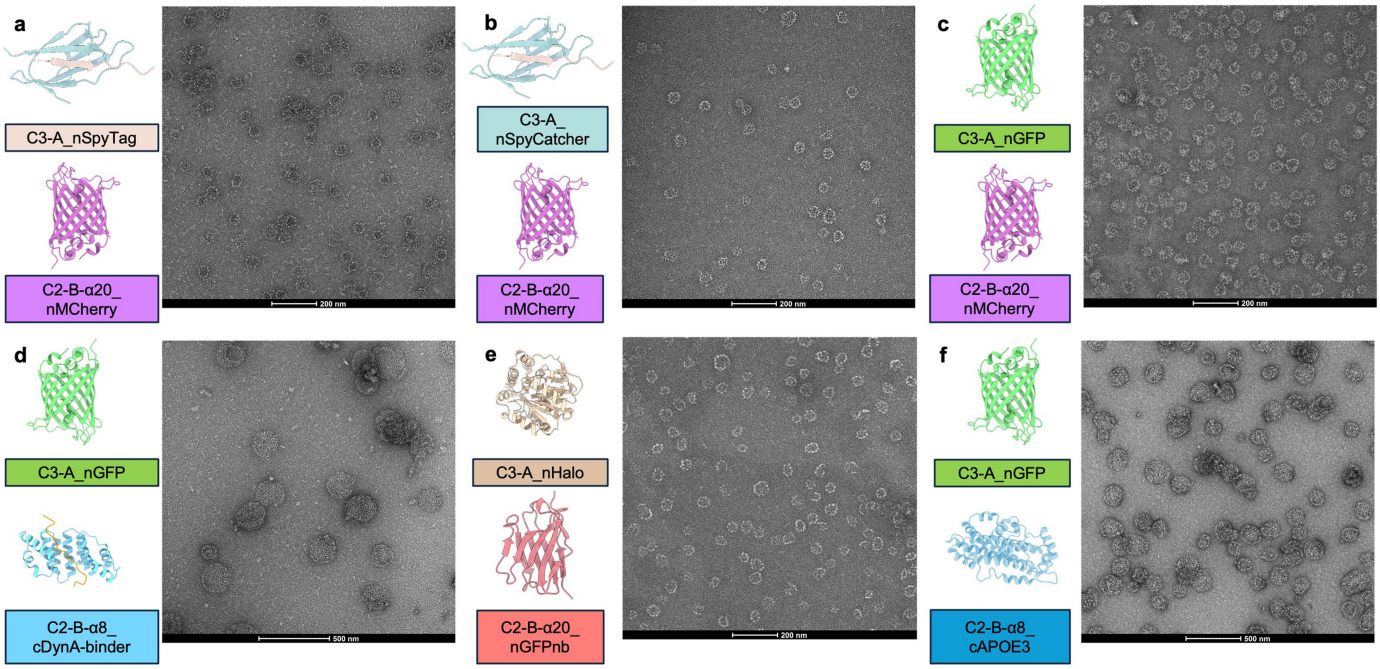
Extended Data Fig. 3 | Vertex angle–guided polyhedral cage formation. a) Schematic of different polyhedral cages and their corresponding intersecting angles between adjacent vertices. In vitro mixing of C3-A with the C2-B-α30 linker yields assemblies comprising a mixture of T = 1 and T = 1–3 cages. **b**) Representative nsEM micrograph of the assembled cages, the reported morphology was observed across three independent experiments. **c**) nsEM 2D class averages revealing assemblies consistent with dodecahedral geometry. **d**) Comparison between the experimentally reconstructed nsEM 3D density map (left, with imposed Icosahedral symmetry) and a regular dodecahedron (right). **e**) Selected 2D class averages showing additional cage populations, potentially corresponding to a larger cage containing 12 pentagonal faces and 2 hexagonal faces. **f**) nsEM 3D reconstruction map without imposed symmetry (left) compared with a cage model with D6 symmetry (right).



Extended Data Fig. 4 | nsEM characterization of the T ~ 3 cage. a) Selected subsets of nsEM 2D class averages showing distinct cage architectures observed from for the T ~ 3 cage sample. Scale bar = 50 nm. **b**) A model of C60 fullerene cage (T = 3) and an experimental nsEM 3D reconstruction map (with applied Icosahedral symmetry, left). An elongated cage with high aspect ratio could potentially be C70 fullerenes (middle). C80 fullerene topology (T = 4) and experimental nsEM 3D reconstruction (with icosahedral symmetry applied, right) agrees with the design model. **c**) Other possible cage topologies from hexagon-pentagon mixed tiling that have been reported.



Extended Data Fig. 5 | Local defects analyses based on cryo-EM 2D class averages. For cages with T < 9, only pentagonal defects were observed. For high T cages, both pentagonal (orange) and heptagonal defects (purple) were observed. Although heptagonal defects generate negative Gaussian curvatures that are energetically unfavourable, they could be formed during cage closure under kinetic control.



Extended Data Fig. 6 | NsEM characterization of functionalized protein cage assemblies. Representative nsEM images show the formation of cage structures utilizing various building block combinations terminally fused to diverse proteins of interest. Similar particles were observed across at least three independent experiments.

Extended Data Table 1 | X-ray diffraction data collection and refinement statistics for C2-B-α20 linker

	C2-B-α20 (N9DL)
Resolution range	28.99 - 2.15 (2.21 - 2.15)
Space group	*C 2*
Unit cell	61.19, 93.30, 60.69; 90, 119.94, 90
Unique reflections	15664 (1104)
Multiplicity	4.3 (4.3)
Completeness (%)	98.00 (94.5)
Mean I/sigma (I)	8.70 (2.70)
Wilson B-factor	36
R-merge	0.094 (0.524)
$CC_{1/2}$	0.993 (0.863)
Reflections used in refinement	15518 (1388)
R-work	0.2029 (0.2583)
R-free	0.2523 (0.3319)
Number of non-hydrogen atoms	2463
macromolecules	2427
solvent	36
Protein residues	318
RMS (bonds)	0.003
RMS (angles)	0.530
Average B-factor	
macromolecules	55
solvent	52

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

Extended Data Table 2 | Cryo-EM data collection, refinement and validation statistics

	#1 Two-component protein cage (T=3) (EMDB-70605) (PDB 9OM3)	#2 Two-component protein cage (T=3) (EMDB-70685) (PDB 9OP9)
Data collection and processing		
Magnification	105,000 ×	22,000 ×
Voltage (kV)	300	200
Electron exposure (e–/Å^2)	9.66	120
Defocus range (μm)	0.8-1.8	2–4
Pixel size (Å)	0.843	0.885
Symmetry imposed	Icosahedral	Icosahedral
Initial particle images (no.)	107,197	599
Final particle images (no.)	11,155	462
Map resolution (Å)	17.45	31.6
FSC threshold	0.143	0.143
Map resolution range (Å)	20.02 – 15.91	N/A
Refinement		
Initial model used (PDB code)	Design model	Design model
Model resolution (Å)	17.45	31.6
FSC threshold	0.143	0.143
Model resolution range (Å)	20.02 – 15.91	N/A
Map sharpening *B* factor (Å^2)	0	N/A
Model composition		
Non-hydrogen atoms	553,560	553,560
Protein residues	71,640	71,640
Ligands	NA	NA
B factors (Å^2)		
Protein	98.11	98.11
Ligand	NA	NA
R.m.s. deviations		
Bond lengths (Å)	0.017	0.019
Bond angles (°)	1.435	1.565
Validation		
MolProbity score	0.86	0.93
Clashscore	1.35	1.72
Poor rotamers (%)	0	0
Ramachandran plot		
Favored (%)	100	100
Allowed (%)	0	0
Disallowed (%)	0	0

nature portfolio

Corresponding author(s): Shunzhi Wang, David Baker

Last updated by author(s): Mar 6, 2026

Reporting Summary

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

Statistics

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

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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Open source softwares were used to generate design models: RFdiffusion (https://github.com/RosettaCommons/RFdiffusion), ProteinMPNN (https://github.com/dauparas/ProteinMPNN), and RPXdock (https://github.com/willsheffler/rpxdock, version 2). Rosetta was used for the construction of cage models based on EM density map (https://rosettacommons.org/software/).
Data analysis	Data was analyzed and plotted using python3.12.3 and GraphPad Prism9. ImageJ1.5 was used to measure particle size. CryoSPARC v3. UCSF ChimeraX 1.6. Computational filtering of designed sequences was achieved with AlphaFold 2 (https://github.com/google-deepmind/alphafold). Scripts for generating curved hexagons are available via Zenodo at: https://doi.org/10.5281/zenodo.18892841. GEM single particle analysis was performed using the GEMspa software package we developed: https://github.com/liamholtlab/GEMspa/releases/tag/v0.11-beta.

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

Coordinates and structure factors are available in the Protein Data Bank with PDB ID 9NDL (C2-B-α20). All data are available in the main text or the supplementary materials. Design models are available at: https://doi.org/10.5281/zenodo.18892841. Example commands for symmetric RFdiffusion motif scaffolding and automated sequence design pipeline of assemblies are available on request.

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

Life sciences study design

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

Sample size	No statistical method was used to determine the sample size. Sample sizes were chosen based on previous experience or similar studies. (Cell: PMID: 29937223; Molecular Cell: PMID: 39059370)
Data exclusions	No data were excluded for data analysis.
Replication	At least three replicates were performed or included for each experiment, The reproducibility of all experiments are ensured.
Randomization	For cell-based experiments, culture dishes from the same preparation batch were randomly allocated to conditions. To control for potential covariates, all plates were maintained under identical environmental conditions.
Blinding	Experiment design and sample collection were performed by two different investigators. Data analysis was performed by reported Python package automatically, with limited human evaluation.

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
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	Cell lines: U-2 OS (ATCC HTB-96), Hep 3B (ATCC HB-8064)
Authentication	Cell line authentication was not performed. All cell lines were expanded from the original vials and used for experiments within twenty passages.
Mycoplasma contamination	All cell lines were regularly tested negative for Mycoplasma contamination by PCR.
Commonly misidentified lines (See ICLAC register)	We did not use commonly misidentified lines listed by ICLIAC.

Plants

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

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	HEP3B cells were seeded at 7,500 cells per well in a 96-well plate one day prior to treatment. GFP-labeled protein cages were added at a final concentration of 250 nM and incubated for 3, 6, or 24 hours. Following incubation, cells were trypsinized, neutralized, and transferred to a 96-well U-bottom plate. Cells were washed three times with DPBS prior to flow cytometry analysis.
Instrument	Sony SH800 / Thermo Attune NxT
Software	FlowJo v9
Cell population abundance	At least 5,000 cells were collected and resuspended in flow focusing buffer per sample.

Gating strategy | The cells were gated based on the major population based on SSC and FSC.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.