LETTER

Design of a hyperstable 60–subunit protein icosahedron

Yang Hsia[1,2,3]*, Jacob B. Bale[1,2,4]*, Shane Gonen[1,2,3,5], Dan Shi[5], William Sheffler[1,2], Kimberly K. Fong[1], Una Nattermann[1,2,3], Chunfu Xu[1,2], Po-Ssu Huang[1,2], Rashmi Ravichandran[1,2], Sue Yi[1,2], Trisha N. Davis[1], Tamir Gonen[5], Neil P. King[1,2] & David Baker[1,2,6]

The icosahedron is the largest of the Platonic solids, and icosahedral protein structures are widely used in biological systems for packaging and transport[1,2]. There has been considerable interest in repurposing such structures[3–5] for applications ranging from targeted delivery to multivalent immunogen presentation. The ability to design proteins that self-assemble into precisely specified, highly ordered icosahedral structures would open the door to a new generation of protein containers with properties custom-tailored to specific applications. Here we describe the computational design of a 25-nanometre icosahedral nanocage that self-assembles from trimeric protein building blocks. The designed protein was produced in *Escherichia coli*, and found by electron microscopy to assemble into a homogenous population of icosahedral particles nearly identical to the design model. The particles are stable in 6.7 molar guanidine hydrochloride at up to 80 degrees Celsius, and undergo extremely abrupt, but reversible, disassembly between 2 molar and 2.25 molar guanidinium thiocyanate. The icosahedron is robust to genetic fusions: one or two copies of green fluorescent protein (GFP) can be fused to each of the 60 subunits to create highly fluorescent 'standard candles' for use in light microscopy, and a designed protein pentamer can be placed in the centre of each of the 20 pentameric faces to modulate the size of the entrance/ exit channels of the cage. Such robust and customizable nanocages should have considerable utility in targeted drug delivery[6], vaccine design[7] and synthetic biology[8].

Programming protein subunits to self-assemble into well defined complexes is a promising route to the custom design of macromolecular machines. Protein assemblies have been engineered using metals[9,10], disulfide bonds[11–14], genetic fusions[12,15–17], and ideal helix–helix interactions[11,13,16], but these approaches have generally yielded polydisperse or unanticipated products. Recently, symmetric modelling coupled with computational protein–protein interface design has accurately generated protein assemblies with tetrahedral and octahedral symmetry[18,19], but these relatively small (<16 nm diameter) nanocages have limited use for packaging or delivery applications because they have little internal volume.

Icosahedral point group symmetry contains two-, three-, and five-fold axes of rotation (Fig. 1a). To generate novel icosahedral protein assemblies, trimeric protein scaffolds of known structure were arranged with icosahedral symmetry (the three-fold axes of the trimers aligned with the three-fold axes of an icosahedron) and the two remaining degrees of freedom—the distance r from the icosahedron centre to the centre of mass of each trimer, and the angle ω of rotation of each trimer about its axis—were optimized for close packing without steric clashes (Fig. 1b, c). The amino acid sequences at the newly formed interfaces between the trimer building blocks

were then optimized using RosettaDesign[20,21], and 17 designs were selected for experimental characterization on the basis of properties of the designed interface, including shape complementarity[22], predicted binding energy, and the number of buried unsatisfied hydrogen-bond donors and acceptors (see Methods).

Genes encoding the designs were assembled from oligonucleotides and cloned into the pET29b+ vector for expression in *E. coli*. Most of the designs were found in the insoluble fraction upon cell lysis; of the three soluble designs, two (both based on a KDPG aldolase[23,24]) showed substantial shifts in migration relative to the wild-type scaffold when analysed by native (non-denaturing) polyacrylamide gel electrophoresis (PAGE), suggesting higher-order assembly. We selected the one with fewer mutations, I3-01, for further analysis. Five substitutions (E26K, E33L, K61M, D187V and R190A) were made to generate the designed interface between trimers (Fig. 1d; the amino acid sequences are provided in the Supplementary Information).

I3-01 was purified using immobilized metal affinity and size exclusion chromatography (SEC), yielding a single peak with an apparent molecular weight much larger than that of the wild-type trimeric protein and consistent with the expected elution volume for the 60-subunit assembly (Fig. 1e). A mutant bearing a leucine-to-arginine substitution predicted to disrupt the designed interface eliminated the high-molecular-weight species and returned the elution volume to that of the wild-type scaffold (Fig. 1e). Dynamic light scattering (DLS) measurements of I3-01 showed a monodisperse population of particles with a hydrodynamic radius of 14 nm, consistent with the design model (Fig. 1f). No disassembly to the trimeric building block was observed at 80 °C or, remarkably, in 6.7 M guanidine hydrochloride (GuHCl) (Extended Data Fig. 1). This hyperstability is a property of both the trimeric scaffold from which I3-01 was derived and of the designed interface: both are completely resistant to GuHCl denaturation. An exceptionally sharp disassociation into the constituent trimers was observed between 2 M and 2.25 M guanidinium thiocyanate (GITC): at 2 M the dominant species is the icosahedron, while at 2.25 M only the trimeric building block is observed (Fig. 1g, Extended Data Fig. 2). Importantly for cargo packaging applications, the disassociation is fully reversible: the hydrodynamic radius of particles formed by diluting disassembled protein in 3 M GITC down to 1 M GITC is identical to those originally produced in *E. coli* (Fig. 1h).

We investigated the structure of I3-01 using cryo-electron microscopy (cryo-EM). The individual particles in large fields of view were homogenous in size and shape (Fig. 2a), and in class averages from 6,461 particles, the three projections along the symmetry axes and the overall icosahedral architecture are clearly discernible (Fig. 2b, c). A three-dimensional model calculated from the cryo-EM data matches

[1]Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA. [2]Institute for Protein Design, University of Washington, Seattle, Washington 98195, USA. [3]Graduate Program in Biological Physics, Structure and Design, University of Washington, Seattle, Washington 98195, USA. [4]Graduate Program in Molecular and Cellular Biology, University of Washington, Seattle, Washington 98195, USA. [5]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia 20147, USA. [6]Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA.
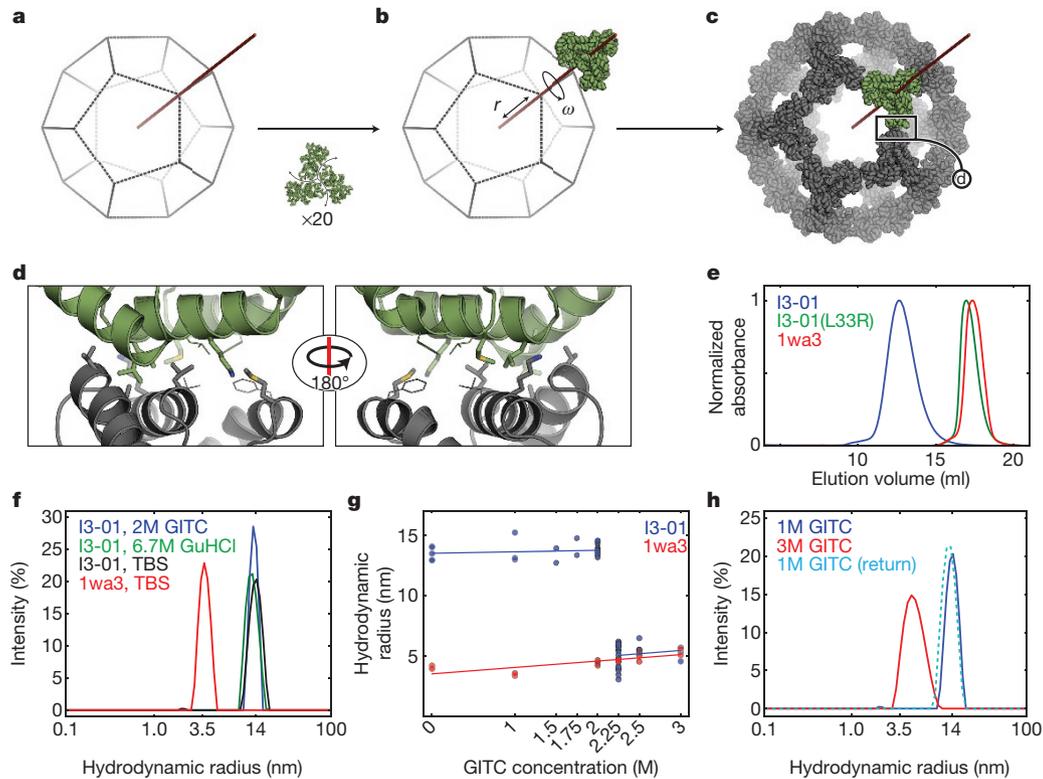*These authors contributed equally to this work.

Figure 1 | Design methodology and biochemical characterization.
 a, **b**, Icosahedral three-fold axis in red and aligned trimeric building block in green. **c**, Optimization of r and ω yields closely opposed interfaces between subunits. **d**, Sequence design yields low-energy interfaces; in the I3-01 case, composed of five designed residues (thick representations) and two native residues (thin representations). **e**, I3-01 appears larger by SEC than the similarly sized I3-01(L33R) and wild-type trimer (1wa3). **f**, DLS measurement of hydrodynamic radius (note logarithmic scale in **f** and **h**) of 1wa3 (3.5 nm) and I3-01 (14 nm). I3-01 remains assembled in 6.7 M GuHCl and in 2 M GITC. **g**, Extremely sharp disassociation to trimeric building blocks at 2.25 M GITC. Data points represent independent measurements. **h**, I3-01 icosahedron disassembles into the trimeric building blocks at 3 M GITC, and reassembles following dilution to 1 M.

the I3-01 design model very well with a correlation coefficient of 0.92 at 20 Å and 1.5σ (Fig. 2d, e), clearly indicating that I3-01 forms the designed structure: an icosahedron with a diameter of 25 nm and an interior volume of approximately 3,000 nm^3, values that are within the range of those observed in small viral capsids[25].

To probe the robustness of I3-01 to genetic fusion, we fused superfolder GFP (sfGFP)[26] to one or both termini of the monomeric subunit and produced the resulting proteins in *E. coli*. SEC analysis showed that the fusion proteins had hydrodynamic radii consistent with cage formation (Extended Data Fig. 3). Analysis of I3-01 with a carboxy
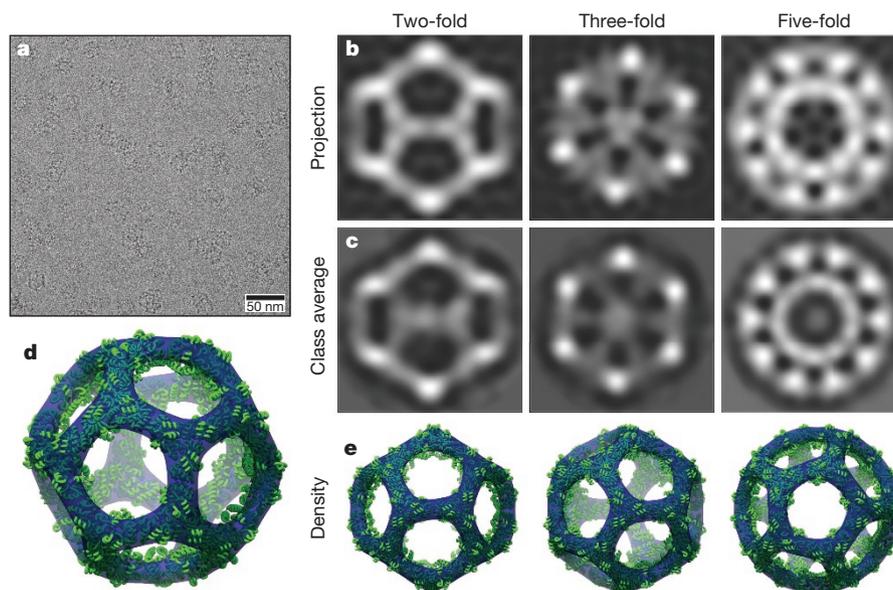

Figure 2 | Cryo-EM. a, Field-of-view cryo-EM micrograph showing homogeneous icosahedral particles in various orientations. **b**, Back-projections of I3-01 from the design model. **c**, Cryo-EM class averages closely match the design projections along all three symmetry axes.
d, **e**, The calculated initial, unrefined density (blue, 3.22σ) closely matches the design model (green).
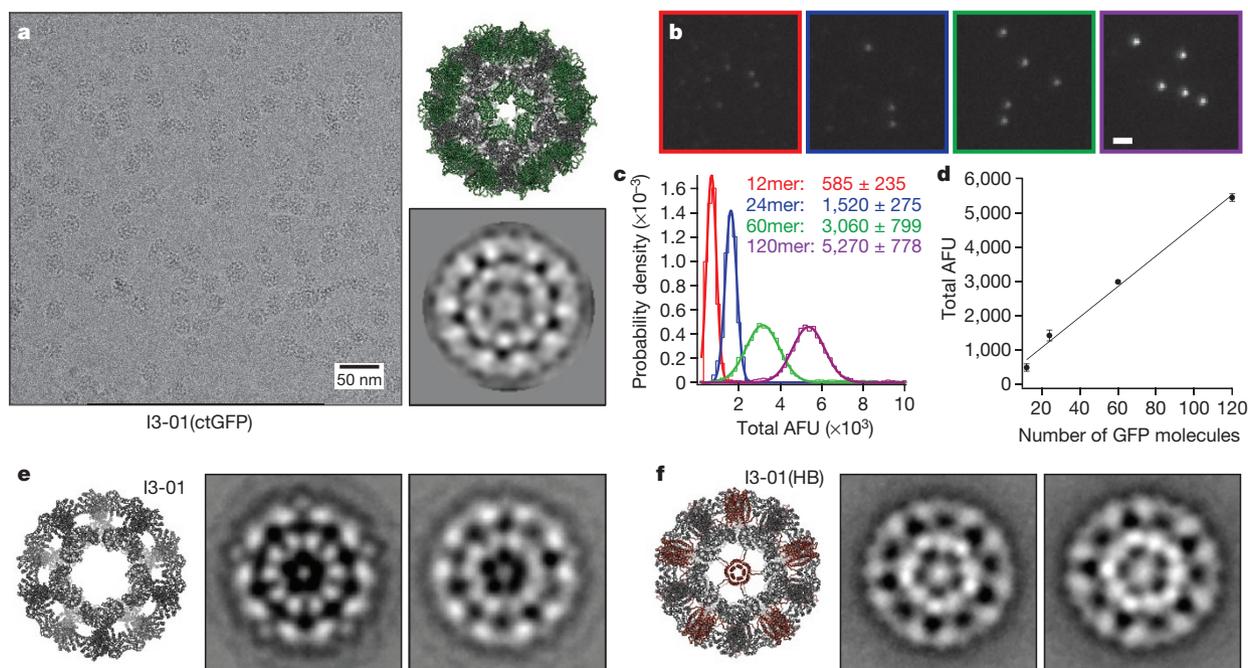
Figure 3 | Tuning nanocage structure and function with genetic fusions.
a, The left panel shows a cryo-EM micrograph of I3-01(ctGFP);
the top right panel shows a computational model with sfGFP in green;
the bottom right panel shows the class average along the five-fold axis.
b, Fluorescence microscopy fields of view. **c**, Fluorescence intensity
histograms. AFU, arbitrary fluorescence units; ± standard deviation.
d, Correlation between the mean fluorescence intensity and sfGFP

copy number for nanoparticles with different numbers of fused sfGFP
molecules. Error bars are s.e.m. ($n = 3$). **e, f**, Computational model and
class averages along the five-fold axis of negatively stained I3-01 (**e**) and
I3-01(HB) (**f**); the helical bundle is shown in red. Weak density in the
centre of the pentameric faces in I3-01 may reflect randomly packaged
material. There is clear density in the centre of the pentameric faces in the
I3-01(HB) class averages consistent with the model.

(C)-terminal sfGFP fusion—called I3-01(ctGFP)—by cryo-EM
revealed icosahedral particles with overall shapes very similar to those
of the original design. Class averages of 13,297 particles revealed con-
siderable internal density compared to the original I3-01 averages,
consistent with computational models of the fusion complex (Fig. 3a).
The I3-01 sfGFP fusions are robust to denaturation of the amino (N)-
or C-terminal fused sfGFP in GuHCl; the particles remain assembled
as GFP signal is lost[27] (Extended Data Fig. 4).

It is at present challenging to infer subunit copy number in GFP-
tagged assemblies from their fluorescence intensity. What is needed
are 'standard candles' with known fluorescent protein copy numbers
that can be used to correlate fluorescence intensity to copy number. To
complement the icosahedra with 60 and 120 copies of sfGFP described
above, we fused sfGFP to one or both components of a previously
described two-component tetrahedron (T33-21; ref. 19) to generate
assemblies with 12 or 24 copies of sfGFP (Extended Data Fig. 3).
Intensity histograms obtained for each of the sfGFP-nanocage
constructs using widefield fluorescence microscopy were well fitted
with Gaussians (Fig. 3b, c), and the mean fluorescence intensity for
each cage was found to be linearly proportional ($r^2 = 0.9925$) to sfGFP
copy number (Fig. 3d). The fluorescent properties of the particles
were readily manipulated by substituting sfGFP with mTurqoise2 and
sYFP2 (Extended Data Fig. 5). In addition to serving as genetically
encoded, water-soluble fluorescent standard candles, the fluorescent
protein cage fusions could be useful for correlative light and electron
microscopy[28] since the icosahedral shape is quite distinctive.

In I3-01, the trimeric building blocks are aligned with the three-fold
axes while the designed interface is along the icosahedral two-folds.
To explore the possibility of symmetry-matched fusions to designed
nanocages, we modelled a designed pentameric helical bundle[29] into
the centre of the large 9-nm pore at the five-fold axis with a C-terminal
linker; this fusion was named I3-01(HB). Negative-stain electron
microscopy showed monodisperse particles of the expected size and
symmetry; the incorporation of the pentamer does not interfere with
icosahedron assembly. Particle averages showed a structure similar to

that of the original icosahedron, with additional density at the centre of
each five-fold axis, consistent with computational models of the fusion
protein (Fig. 3e, f). The capability of incorporating symmetry-matched
substructures into designed nanocages offers considerable flexibil-
ity and modularity; for example, pentamers filling otherwise open
pentameric faces could control the release of cargo contained within
the nanocage.

The designed I3-01 icosahedron is exceptionally stable, robust to
genetic fusion, and has a considerably larger internal volume than
previously designed nanocages with well defined and prespecified
structures[14,17,19]. Enzymatic activity is retained in the assembled ico-
sahedron (Extended Data Fig. 6), suggesting a route to custom nano-
reactors. The ability to accurately design icosahedral protein structures
opens the door to new approaches to vaccine generation and targeted
drug delivery.

Online Content Methods, along with any additional Extended Data display items and
Source Data, are available in the online version of the paper; references unique to
these sections appear only in the online paper.

1. Zandi, R., Reguera, D., Bruinsma, R. F., Gelbart, W. M. & Rudnick, J. Origin of
   icosahedral symmetry in viruses. *Proc. Natl Acad. Sci. USA* **101**, 15556–15560
   (2004).
2. Ritsert, K. *et al.* Studies on the lumazine synthase/riboflavin synthase complex
   of *Bacillus subtilis*: crystal structure analysis of reconstituted, icosahedral
   beta-subunit capsids with bound substrate analogue inhibitor at 2.4 Å
   resolution. *J. Mol. Biol.* **253**, 151–167 (1995).
3. Howorka, S. Rationally engineering natural protein assemblies in
   nanobiotechnology. *Curr. Opin. Biotechnol.* **22**, 485–491 (2011).
4. Roldão, A., Mellado, M. C. M., Castilho, L. R., Carrondo, M. J. T. & Alves, P. M.
   Virus-like particles in vaccine development. *Expert Rev. Vaccines* **9**, 1149–1176
   (2010).
5. Effio, C. L. & Hubbuch, J. Next generation vaccines and vectors: designing
   downstream processes for recombinant protein-based virus-like particles.
   Biotechnol. J. **10**, 715–727 (2015).

6. Ma, Y., Nolte, R. J. M. & Cornelissen, J. J. L. M. Virus-based nanocarriers for drug delivery. *Adv. Drug Deliv. Rev.* **64,** 811–825 (2012).
7. Smith, M. L. *et al.* Modified tobacco mosaic virus particles as scaffolds for display of protein antigens for vaccine applications. *Virology* **348,** 475–488 (2006).
8. Bauler, P, Huber, G., Leyh, T. & McCammon, J. A. Channeling by proximity: the catalytic advantages of active site colocalization using Brownian dynamics. *J. Phys. Chem. Lett.* **1,** 1332–1335 (2010).
9. Brodin, J. D. *et al.* Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. *Nat. Chem.* **4,** 375–382 (2012).
10. Der, B. S. *et al.* Metal-mediated affinity and orientation specificity in a computationally designed protein homodimer. *J. Am. Chem. Soc.* **134,** 375–385 (2012).
11. Fletcher, J. M. *et al.* Self-assembling cages from coiled-coil peptide modules. *Science* **340,** 595–599 (2013).
12. Usui, K. *et al.* Nanoscale elongating control of the self-assembled protein filament with the cysteine-introduced building blocks. *Protein Sci.* **18,** 960–969 (2009).
13. Raman, S., Machaidze, G., Lustig, A., Aebi, U. & Burkhard, P. Structure-based design of peptides that self-assemble into regular polyhedral nanoparticles. *Nanomedicine* **2,** 95–102 (2006).
14. Raman, S. *et al.* Design of peptide nanoparticles using simple protein oligomerization domains. *Open Nanomed. J.* **2,** 15–26 (2009).
15. Sinclair, J. C., Davies, K. M., Vénien-Bryan, C. & Noble, M. E. M. Generation of protein lattices by fusing proteins with matching rotational symmetry. *Nat. Nanotechnol.* **6,** 558–562 (2011).
16. Boyle, A. L. *et al.* Squaring the circle in peptide assembly: from fibers to discrete nanostructures by de novo design. *J. Am. Chem. Soc.* **134,** 15457–15467 (2012).
17. 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).
18. King, N. P. *et al.* Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **336,** 1171–1174 (2012).
19. King, N. P. *et al.* Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510,** 103–108 (2014).
20. Leaver-Fay, A. *et al.* ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487,** 545–574 (2011).
21. DiMaio, F., Leaver-Fay, A., Bradley, P., Baker, D. & André, I. Modeling symmetric macromolecular structures in Rosetta3. *PLoS One* **6,** e20450 (2011).
22. Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234,** 946–950 (1993).
23. Griffiths, J. S. *et al.* Cloning, isolation and characterization of the *Thermotoga maritima* KDPG aldolase. *Bioorg. Med. Chem.* **10,** 545–550 (2002).
24. Fullerton, S. W. B. *et al.* Mechanism of the class I KDPG aldolase. *Bioorg. Med. Chem.* **14,** 3002–3010 (2006).
25. Perlmutter, J. D. & Hagan, M. F. Mechanisms of virus assembly. *Annu. Rev. Phys. Chem.* **66,** 217–239 (2015).
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. Andrews, B. T., Schoenfish, A. R., Roy, M., Waldo, G. & Jennings, P. A. The rough energy landscape of superfolder GFP is linked to the chromophore. *J. Mol. Biol.* **373,** 476–490 (2007).
28. Cortese, K., Diaspro, A. & Tacchetti, C. Advanced correlative light/electron microscopy: current methods and new developments using Tokuyasu cryosections. *J. Histochem. Cytochem.* **57,** 1103–1112 (2009).
29. Huang, P.-S. *et al.* High thermodynamic stability of parametrically designed helical bundles. *Science* **346,** 481–485 (2014).

METHODS

Computational design. Crystal structures of 300 trimers with resolution better than 2.5 Å and lacking long loops were selected from the Protein Data Bank (PDB) to use as building blocks. For each scaffold, 20 trimeric building blocks were arranged in icosahedral symmetry by aligning the three-fold rotational axis of each trimer with one of the three-fold icosahedral symmetry axes. While preserving symmetry, the building blocks were then docked together by enumeratively sampling their rotations (ω) about the three-fold symmetry axes and translating (r) them into contact along the aligned axes. Configurations in which backbone atoms from different building blocks were less than 3.5 Å apart were discarded. Non-clashing design models were ranked based on the number of pairs of β-carbons in adjacent subunits within 12 Å and further sampling was carried out around the top 208 docked configurations on a 0.5 Å by 0.2 Å grid. Symmetric RosettaDesign[20,21] calculations were then used to generate low-energy, symmetric hydrophobic interfaces, and the resulting designs were filtered based on shape complementarity[22] (sc), interface surface area (sasa), buried unsatisfied hydrogen bonds (uhb), and binding energy (ddg). Designed substitutions that did not substantially contribute to the interface were reverted to their original identities. All Rosetta scripts used are available upon request; the full 60-subunit design model of I3-01 is provided in the Supplementary Information.

Cloning, screening, and protein purification. Codon-optimized genes encoding the wild-type and the designed molecules were generated by recursive polymerase chain reaction (PCR) from sets of synthetic oligonucleotides (Integrated DNA Technologies). Five mutations were incorporated into I3-01: E26K, E33L, K61M, D187V and R190A. All genes were cloned into the pET29b+ plasmid with kanamycin resistance and expressed in BL21 Star (DE3) *E. coli* cells (Invitrogen) induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) for 4 h at 37 °C. Cell lysis was accomplished in Tris-buffered saline (TBS; 50 mM Tris, 500 mM NaCl) with lysozyme (0.25 mg ml^{-1}) and sonication (Fisher Scientific) at 20 W for 5 min total 'on' time, using cycles of 10 s on, 10 s off.

For initial screening, all constructs were labelled with the CoA-488 fluorophore (NEB) by the addition of AcpS[30] (NEB) using an A1 peptide tag, allowing the solubility and assembly state of each design to be analysed using SDS–PAGE and native-PAGE (Bio-Rad), following procedures previously described[18]. All subsequent experiments were performed on either (His)$_6$-tagged protein or remained untagged.

After lysis and centrifugation at 20,000g for 30 min, the soluble fraction of (His)$_6$-tagged proteins were passed through 2 ml of nickel nitrilotriacetic acid agarose (Ni-NTA) (Qiagen), washed with 30 mM imidazole, and eluted with 500 mM imidazole. Pure proteins were collected after elution from a Superose 6 10/300 GL SEC column (GE Healthcare) at 9–11 ml, depending on the fusion variant.

For non-(His)$_6$-tagged proteins, cells were lysed as above, and the cleared lysates were treated with serial ammonium sulfate precipitation treatments (20%, 60% w/v). During each step, solid ammonium sulfate was added to the lysate to the desired percentage, and equilibrated at room temperature for 1 h. Ammonium sulfate precipitated protein was then collected by centrifugation at 20,000g for 30 min at 25 °C. After treatment at 60%, the pellet was solubilized in TBS and heated at 80 °C for 10 min. The soluble fraction was then collected and further purified through SEC as described.

KDPG enzyme assay. The reaction was carried out in 25 mM HEPES, 20 mM NaCl buffer at pH 7.0 with the presence of NADH (0.1 mM), L-lactate dehydrogenase (LDH, 0.11 U μl^{-1}), and 2-keto-3-deoxy-6-phosphogluconate (KDPG, 1 mM) at 25 °C, based on previously described methods[23]. Native 1wa3, I3-01, or I3-01(K129A) was added at 0.02 μM final concentration to each well and immediately monitored for 339 nm ultraviolet absorbance over time.

Dynamic light scattering. Purified protein was measured using a DynaPro NanoStar (Wyatt) DLS setup. 0.5 mg ml^{-1} of I3-01 and 1wa3 were measured at 25 °C, then the temperature was ramped up to 90 °C, then ramped back down to 25 °C for temperature scans at 2 °C min^{-1}. Measurements were taken in the presence of TBS: 25 mM Tris, 500 mM NaCl; buffered GuHCl: 25 mM Tris, 500 mM NaCl, 1–6.7 M GuHCl; or buffered GITC: 25 mM Tris, 500 mM NaCl, 1–4 M GITC. Different concentrations of GITC equilibrated samples were achieved by combining stocks of 0 M and 4 M equilibrated solutions in different ratios while GuHCl equilibrated samples were equilibrated individually. Each sample was allowed to equilibrate in their respective buffer for at least 24 h before measurement. Re-annealing experiments were performed by diluting I3-01 equilibrated in 3 M GITC down to 1 M GITC final concentration (0.166 mg ml^{-1} protein). Data analysis was performed using DYNAMICS v7 (Wyatt), reporting regularization fits (with D10/D50/D90) except for temperature ramp experiments, where cumulant fits were used. The ~1 nm radius particle consistent with GITC buffer alone was disregarded for analysis, and monodispersity was assumed when peak polydispersity was below 15% (refs 31 and 32).

Negative-stain electron microscopy. 3 μl of purified I3-01 and I3-01(ctGFP) at 0.1 mg ml^{-1} were applied to glow discharged, carbon-coated 200-mesh copper grids (Ted Pella), washed with Milli-Q water and stained with 0.75% uranyl formate as described previously[33]. Grids were visualized for assembly validation and stability and subsequently optimized for cryo-EM data collection. Screening was performed on a 120 kV Tecnai Spirit T12 transmission electron microscope (FEI) with a bottom-mount TVIPS F416 CMOS 4k camera.

6 μl of purified I3-01 and I3-01(HB) at 0.05–0.1 mg ml^{-1} were applied to glow discharged, carbon-coated 400-mesh copper grids (Ted Pella), washed with Milli-Q water and stained with 0.75% uranyl formate. Grids were visualized for assembly validation and optimized for data collection. Screening and sample optimization was performed on a 100 kV Morgagni M268 transmission electron microscope (FEI) equipped with an Orius charge-coupled device (CCD) camera (Gatan). Data collection was performed on a 120 kV Tecnai G2 Spirit transmission electron microscope (FEI). All final images were recorded using an Ultrascan 4000 4 k × 4 k CCD camera (Gatan) at 52,000× magnification at the specimen level. Coordinates for 6,576 I3-01 and 4,131 I3-01(HB) unique particles were obtained for averaging using EMAN2[34]. Boxed particles were used to obtain two-dimensional class averages by refinement in EMAN2. Additional image analysis was performed using ImageJ[35].

Cryo-EM. 5 μl of purified untagged I3-01 and I3-01(ctGFP), diluted to ~0.1 mg ml^{-1} using TBS buffer (25 mM Tris pH 8.0, 150 mM NaCl) with an additional 2 mM dithiothreitol were applied to glow discharged 1.2/1.3 Quantifoil grids, blotted and plunged into liquid ethane using a Vitrobot (FEI). Screening and grid optimization was performed on a 200 kV TF20 transmission electron microscope (FEI) with a bottom-mount TVIPS F416 CMOS 4k camera. 4–6 s movies were recorded on a 300 kV Titan Krios (FEI) using a Gatan K2 direct detector at either 29,000× or 37,000× magnification at the specimen level at ~10 electrons per pixel per second.

Movies were motion-corrected using previously described methods[36]. Coordinates for 6,461 (I3-01) and 13,297 (I3-01(ctGFP)) unique particles were obtained for averaging using EMAN2[34]. Extracted frames of these particles were used to calculate class averages by refinement in IMAGIC[37] using multiple rounds of multivariate statistical analysis and multi-reference alignment. An initial density model was calculated based on the calculated averages using EMAN2[34] and the fitting of the model and correlation were calculated using UCSF Chimera[38]. Low-resolution (17–30 Å) volumes from the I3-01 design model were calculated using SPIDER[39] and inspected in UCSF Chimera[38]. Back-projection images were computed in SPIDER[39] on the low-resolution volumes and visualized using WEB[39]. The contrast of all micrographs was enhanced in Fiji[40].

Symmetrical linker modelling. RosettaRemodel[41] was used to model I3-01(ctGFP) and to generate linkers for I3-01(HB). For I3-01(ctGFP), I3-01 was held static while the linker was sampled via fragment insertion, placing the sfGFP molecules at the end of the linker. The overall model was sampled symmetrically with icosahedral symmetry.

For I3-01(HB), I3-01 was held static while linkers of different lengths (7–12 residues) were sampled via fragment insertion. The resulting placement of the helical bundle at the end of the linker was filtered with pentameric assembly constraints to determine linker lengths that could satisfy formation of the pentameric helical bundle. The shorter linkers that allowed unstrained helical assembly were selected for experimental testing. Example scripts are supplied in the Supplementary Information.

Fluorescence microscopy. Different constructs used for fluorescence microscopy were generated by genetically fusing sfGFP to the termini of nanocages. For T33-21, the sfGFP was fused to either the C terminus of the first component (12 sfGFP molecules), or the C terminus of both components (24 sfGFP molecules). For I3-01, the sfGFP was fused to either terminus of I3-01 (60 sfGFP molecules), or both termini of I3-01 (120 sfGFP molecules). For mTurquoise2 and SYFP2 versions, sfGFP was replaced with the sequence of the respective fluorescent protein bearing additional surface mutations identical to sfGFP[26].

GFP nanocages were mounted on agarose pads for microscopy as previously described[42]. Images of the GFP nanocages were obtained using a DeltaVision system (Applied Precision) with an IX70 inverted microscope (Olympus), a U Plan Apo 100× objective (1.35 NA) and a CoolSnap HQ digital camera (Photometrics). GFP images were taken with a 0.4 s exposure, in a single focal plane, and binned 1 × 1.

The fluorescence intensities of GFP puncta were identified and quantified using custom Matlab programs as previously described[43]; programs are available upon request. Fluorescent intensity histograms of individual sfGFP-fused cages were fitted with Gaussian distributions, shown with mean total arbitrary fluorescence unit (AFU) intensity ± one standard deviation.

30. Zhou, Z. *et al.* Genetically encoded short peptide tags for orthogonal protein labeling by Sfp and AcpS phosphopantetheinyl transferases. *ACS Chem. Biol.* **2,** 337–346 (2007).
31. Baalousha, M. & Lead, J. R. Nanoparticle dispersity in toxicology. *Nat. Nanotechnol.* **8,** 308–309 (2013).
32. Zulauf, M. & D'Arcy, A. Light scattering of proteins as a criterion for crystallization. *J. Cryst. Growth* **122,** 102–106 (1992).
33. Nannenga, B. L., Iadanza, M. G., Vollmar, B. S. & Gonen, T. in *Current Protocols in Protein Science* (eds Coligan, J. E., Dunn, B. M., Speicher, D. W. & Wingfield, P. T.) Ch. 17.15 (John Wiley & Sons, 2013).
34. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157,** 38–46 (2007).
35. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9,** 671–675 (2012).
36. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10,** 584–590 (2013).
37. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116,** 17–24 (1996).
38. Pettersen, E. F. *et al.* UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
39. Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116,** 190–199 (1996).
40. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9,** 676–682 (2012).
41. Huang, P.-S. *et al.* RosettaRemodel: a generalized framework for flexible backbone protein design. *PLoS One* **6,** e24109 (2011).
42. Muller, E. G. D. *et al.* The organization of the core proteins of the yeast spindle pole body. *Mol. Biol. Cell* **16,** 3341–3352 (2005).
43. Shimogawa, M. M., Wargacki, M. M., Muller, E. G. & Davis, T. N. Laterally attached kinetochores recruit the checkpoint protein Bub1, but satisfy the spindle checkpoint. *Cell Cycle* **9,** 3619–3628 (2010).

a) TBS
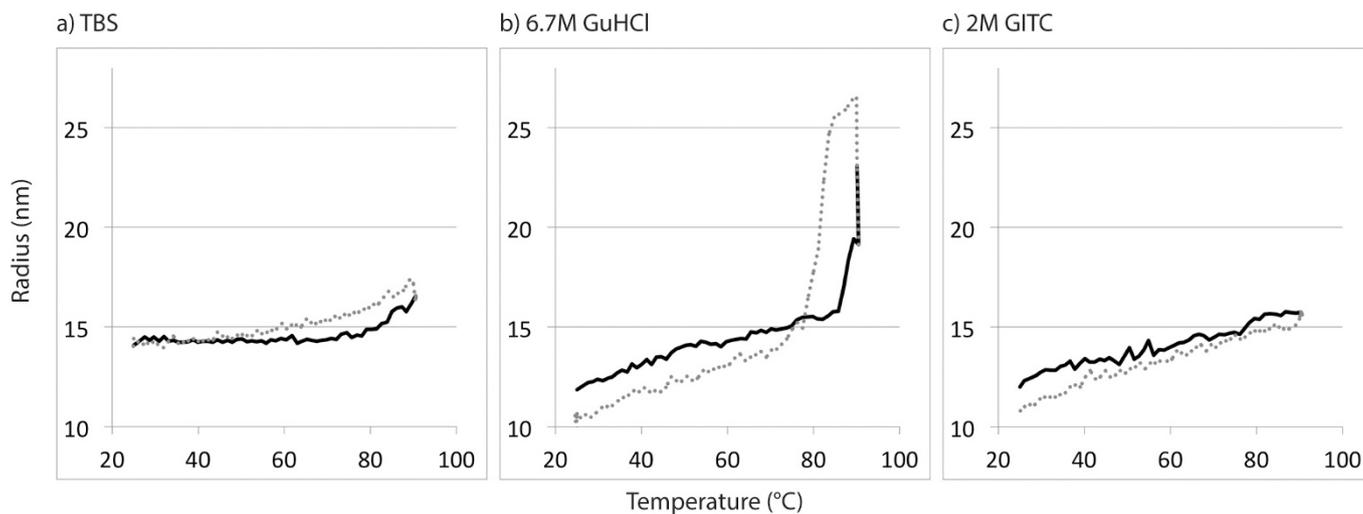
b) 6.7M GuHCl

c) 2M GITC

Extended Data Figure 1 | I3-01 tolerance to temperature. DLS measurements as I3-01 is subjected to heating to 90 °C (solid line), then cooling to 25 °C (dotted line) in TBS (**a**), 6.7 M GuHCl (**b**) and 2 M GITC (**c**). Under all three conditions, any indications of aggregation or increase in size due to temperature appear to be completely reversible.
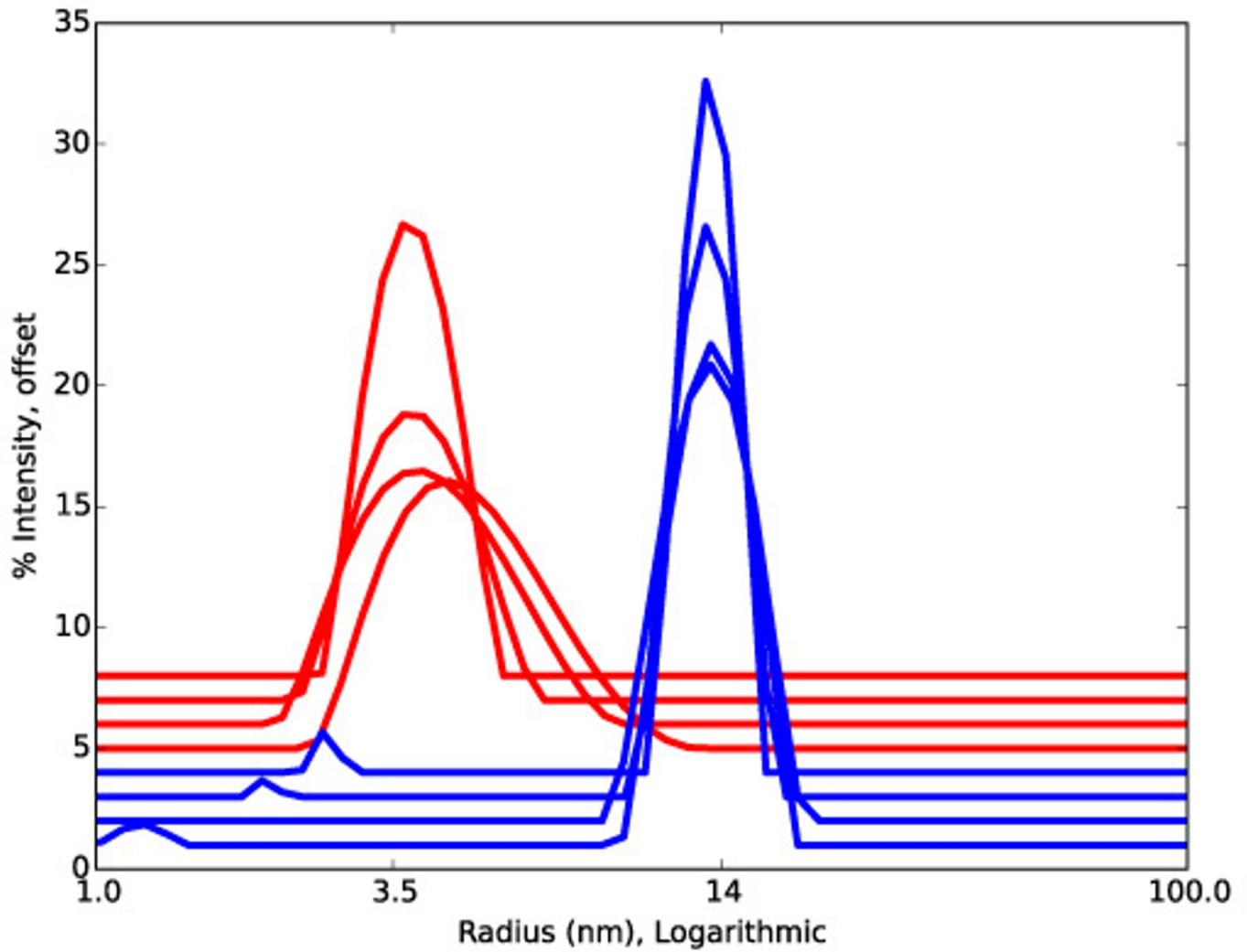
Extended Data Figure 2 | Reproducibility of I3-01 transition in 2 M to 2.25 M GITC. Four examples each of independent measurements at 2 M (blue) and 2.25 M (red) GITC using DLS show the reproducibility of the cage disassociation. Histograms are plotted offset by 1% intensity from each other for clarity.
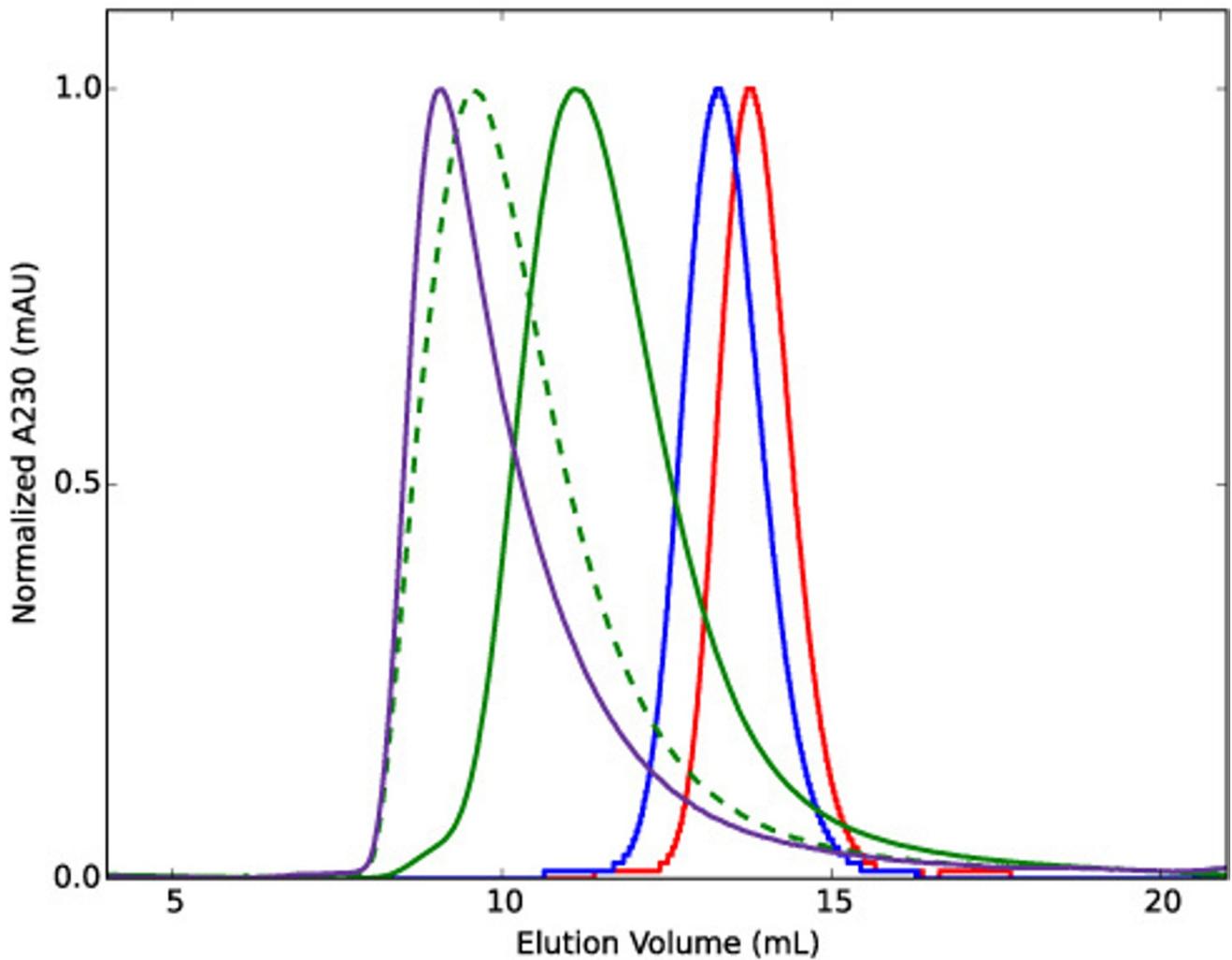
Extended Data Figure 3 | SEC of T33-21 and I3-01 fused with sfGFP. Size exclusion chromatography traces for T33-21 (12mer in red and 24mer in blue) and I3-01 (60mer in green and 120mer in purple) sfGFP fusions, display increased particle sizes with increasing copies of GFP, but retain monodispersed populations. The N-terminal fusion of sfGFP (dashed line) is expected to extend mostly outward from the icosahedron, thus greatly increasing the hydrodynamic radius while the C-terminal fusion is predicted to occupy the internal void space. A230, ultraviolet absorbance at 230 nm; mAU, milli-absorbance units.
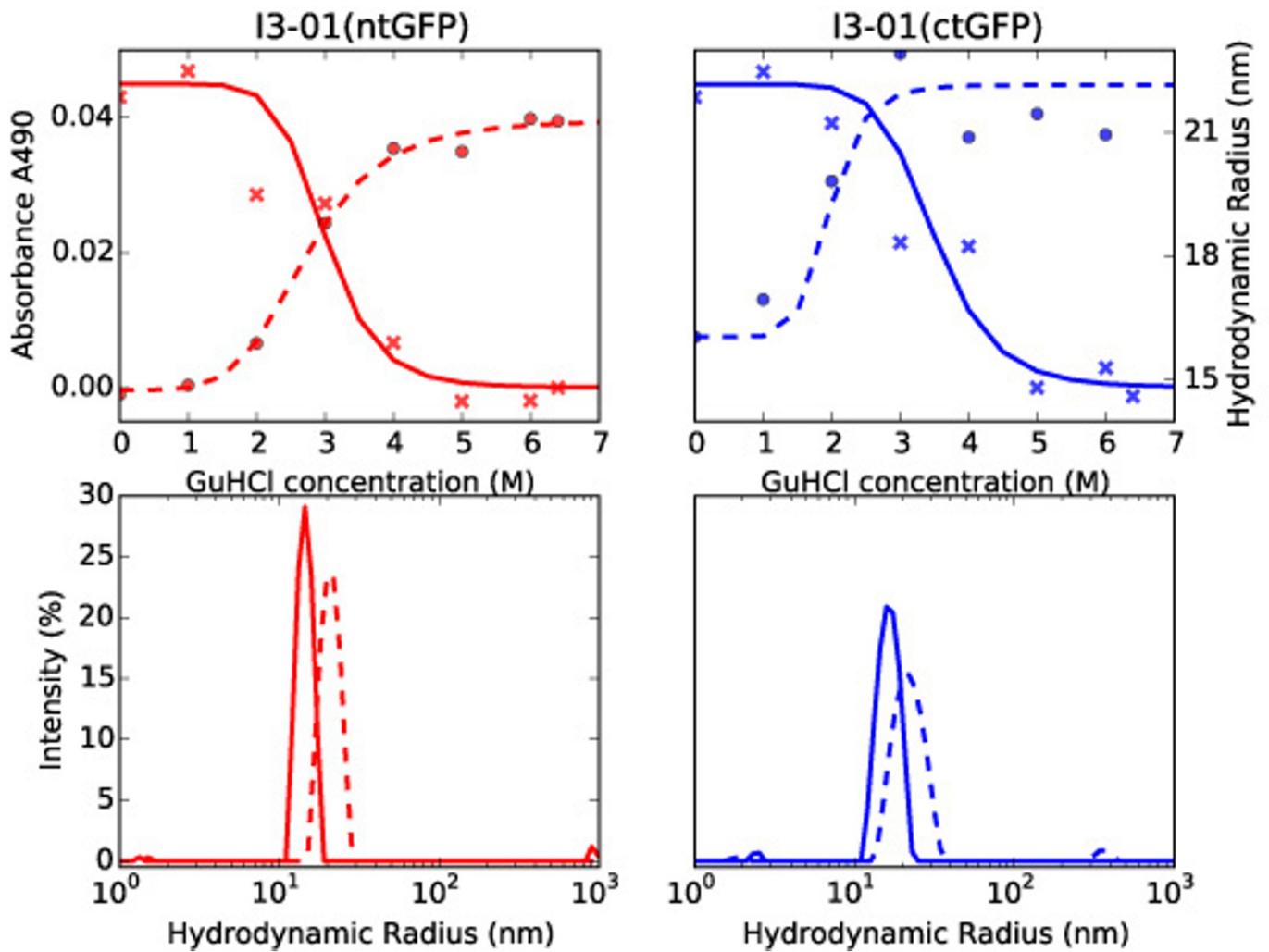
Extended Data Figure 4 | Tolerance of I3-01–sfGFP fusions to GuHCl.
N-terminal (red) and C-terminal (blue) sfGFP fusions were equilibrated
to 0–6.4 M GuHCl. Ultraviolet absorbance at 490 nm (A490) monitors the
unfolding of sfGFP (top, solid line and crosses). DLS experiments (top,
dotted line and dots) reveal as sfGFP unfolds, the hydrodynamic radius
increases slightly, and then stabilizes. The bottom panels show that in
1 M GuHCl (solid line) and in 6 M GuHCl (dotted line), the icosahedral
assemblies remain relatively monodisperse.
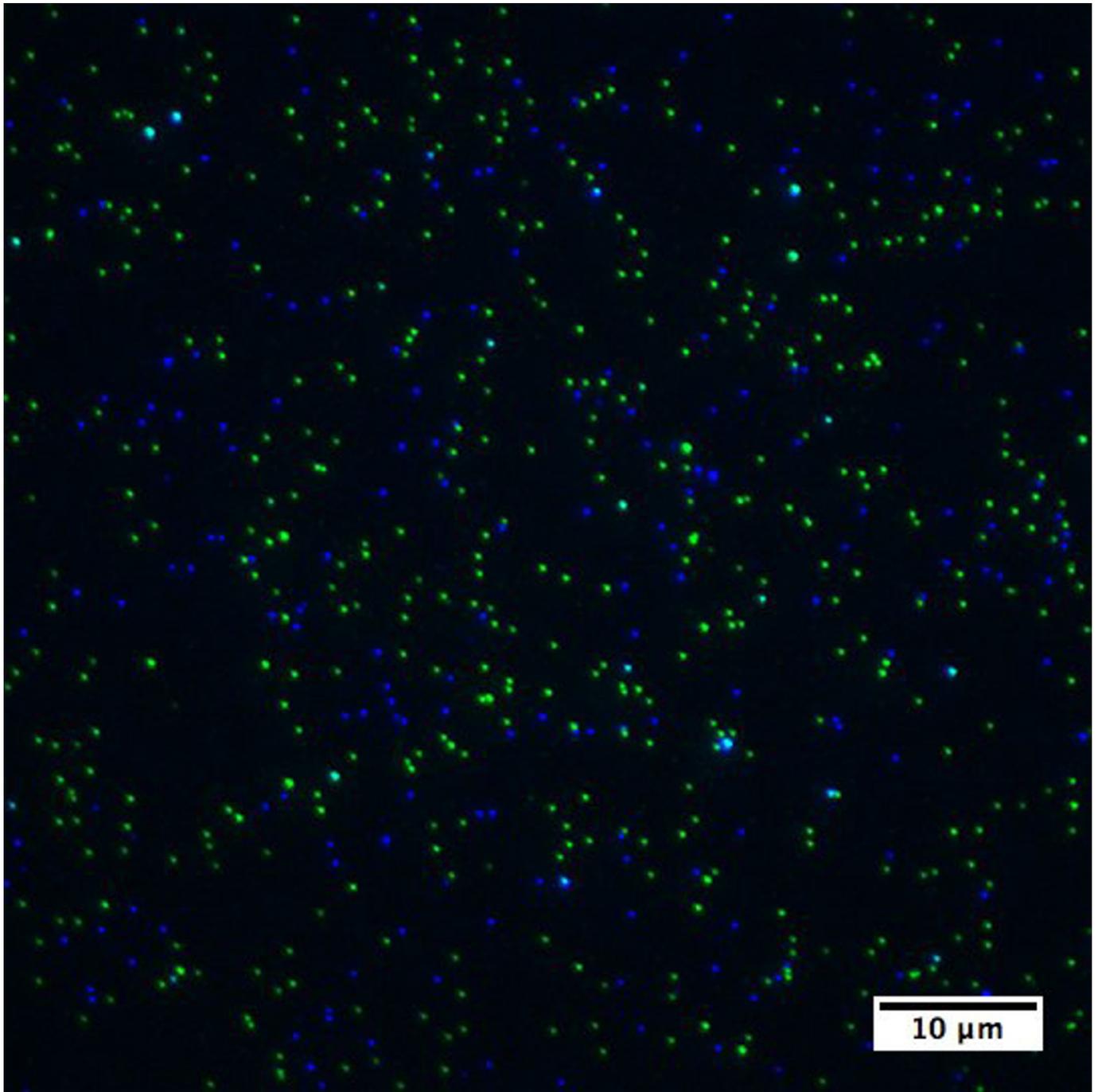
Extended Data Figure 5 | I3-01 C-terminal fusions with other fluorescent proteins. Fluorescent proteins mTurquoise2 (in blue) or sYFP2 (in green) were fused to the C terminus of I3-01. The field of view using widefield fluorescence microscopy shows distinct signals of each type when the two types are mixed together.
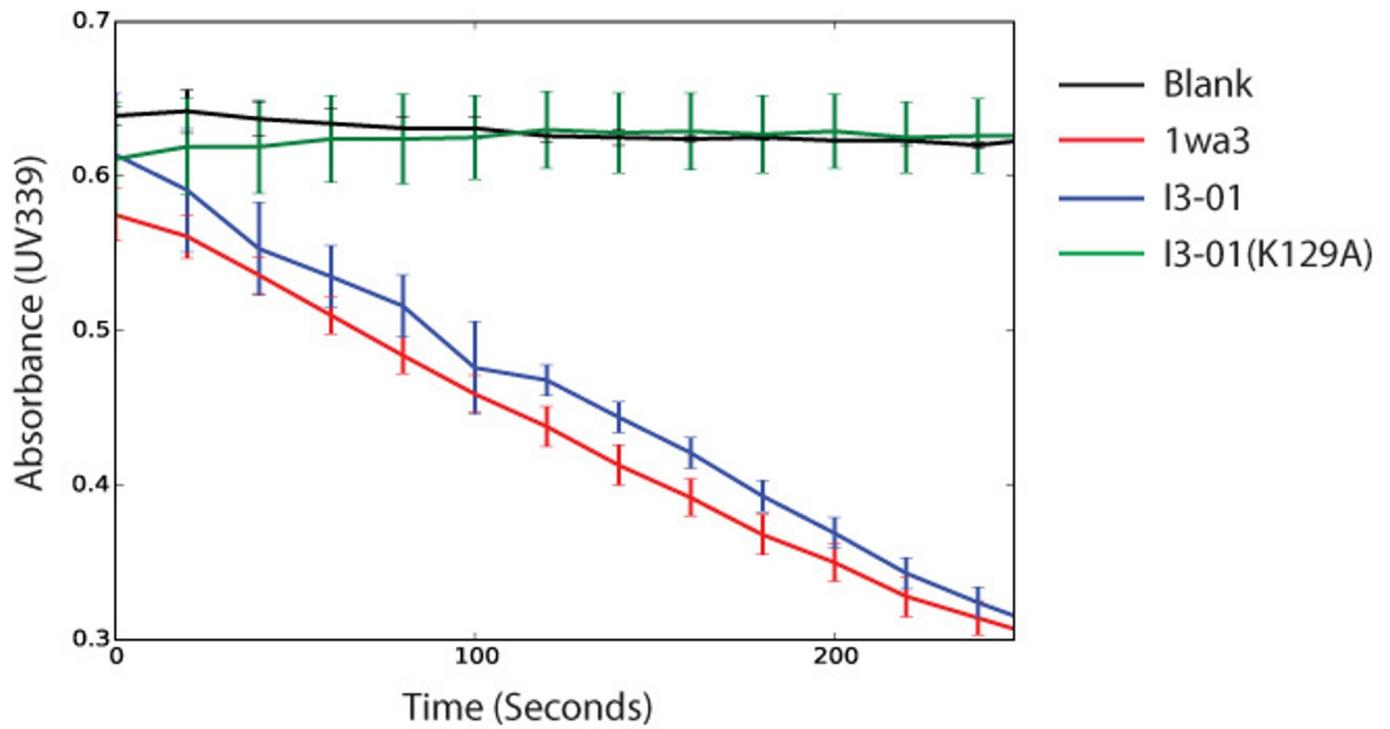
Extended Data Figure 6 | I3-01 retains native enzyme activity. Coupled KDPG aldolase assay showing native-like enzymatic activity in I3-01. The K129A knockout shows no enzyme activity, similar to buffer alone. UV339, absorbance at 339 nm; error bars are standard deviation.

CORRECTIONS & AMENDMENTS

Corrigendum: Design of a hyperstable 60-subunit protein icosahedron

Yang Hsia, Jacob B. Bale, Shane Gonen, Dan Shi, William Sheffler, Kimberly K. Fong, Una Nattermann, Chunfu Xu, Po-Ssu Huang, Rashmi Ravichandran, Sue Yi, Trisha N. Davis, Tamir Gonen, Neil P. King & David Baker

In this Letter, the 60-subunit protein assembly presented more strongly resembles a wireframe dodecahedron than an icosahedron. We thank the reader who drew this to our attention and we agree that all instances of the word "icosahedron" should be changed to "dodecahedron". In addition, "20 pentameric faces" in the boldface first paragraph should have read "12 pentagonal faces", and throughout the text "pentameric faces" should have read "pentagonal faces". The original Letter has not been corrected.