Directed Assembly of Homopentameric Cholera Toxin B-Subunit Proteins into Higher-Order Structures Using Coiled-Coil Appendages

James F. Ross,[†,‡,§,#] Gemma C. Wildsmith,[†,‡] Michael Johnson,[†,‡] Daniel L. Hurdiss,[†,§] Kristian Hollingsworth,[†,‡] Rebecca F. Thompson,[†,§] Majid Mosayebi,[∥,¶] Chi H. Trinh,[†,§] Emanuele Paci,[†,§] Arwen R. Pearson,*,[⊥] Michael E. Webb,*,[†,‡] and W. Bruce Turnbull*,[†,‡]

†Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, United Kingdom
‡School of Chemistry, University of Leeds, Leeds LS2 9JT, United Kingdom
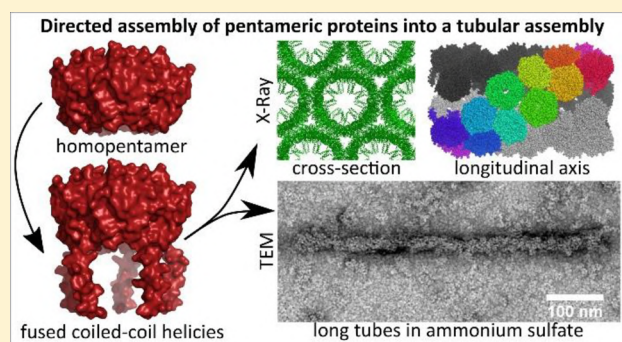§School of Molecular and Cellular Biology, University of Leeds, Leeds LS2 9JT, United Kingdom
∥School of Mathematics, University of Bristol, Bristol BS8 1TW, United Kingdom
¶BrisSynBio, Life Sciences Building, University of Bristol, Bristol BS8 1TQ, United Kingdom
⊥Institute for Nanostructure and Solid State Physics, Universität Hamburg, Hamburg D-22761, Germany

S Supporting Information

ABSTRACT: The self-assembly of proteins into higher order structures is ubiquitous in living systems. It is also an essential process for the bottom-up creation of novel molecular architectures and devices for synthetic biology. However, the complexity of protein−protein interaction surfaces makes it challenging to mimic natural assembly processes in artificial systems. Indeed, many successful computationally designed protein assemblies are prescreened for "designability", limiting the choice of components. Here, we report a simple and pragmatic strategy to assemble chosen multisubunit proteins into more complex structures. A coiled-coil domain appended to one face of the pentameric cholera toxin B-subunit (CTB) enabled the ordered assembly of tubular supra-molecular complexes. Analysis of a tubular structure determined by X-ray crystallography has revealed a hierarchical assembly process that displays features reminiscent of the polymorphic assembly of polyomavirus proteins. The approach provides a simple and straightforward method to direct the assembly of protein building blocks which present either termini on a single face of an oligomer. This scaffolding approach can be used to generate bespoke supramolecular assemblies of functional proteins. Additionally, structural resolution of the scaffolded assemblies highlight "native-state" forced protein−protein interfaces, which may prove useful as starting conformations for future computational design.

Directed assembly of pentameric proteins into a tubular assembly

homopentamer
fused coiled-coil helicies
X-Ray cross-section
longitudinal axis
TEM
long tubes in ammonium sulfate
100 nm

INTRODUCTION

Macromolecular self-assembly is essential and ubiquitous in the phenomena of biology;[1] for example, the homomeric assembly of microtubules to provide tracks along which vesicles are transported via the transient heteromeric association of myosins;[2] the assembly of capsids;[3] and protein nanocompartment assembly, encapsulating specific enzymes to create bacterial organelles.[4,5] Harnessing protein self-assembly to prepare functional, artificial, supramolecular structures has become an ambition for the bionanoscience and synthetic biology research communities.[6−13]

The programmed self-assembly of oligonucleotides[14,15] and peptides[16−22] to form discrete nanostructures has been the subject of considerable effort. Attention has also been focused on the directed assembly of globular proteins to form novel capsid and tube-like supramolecular structures.[23−29] Most

strategies to achieve assembly of globular proteins have typically employed combinations of dimeric and trimeric building blocks to form the edges and vertices of the target polyhedra.[30] Early pioneering work in this field was conducted in the Yeates group, where a trimeric bromoperoxidase was fused, via a rigid α-helix, to the dimeric M1 matrix protein of influenza.[31] These proteins were identified via a computational screen of the PDB for proteins possessing an appropriate geometry of *N*- or *C*-terminal α-helices. Once assembled, 15 nm particles were observed under electron microscopy. Later, a similar approach was used to fuse further dimers and trimers to assemble porous cubic cages.[32] Alternatively, high-throughput docking algorithms have been employed to identify dimeric,

trimeric, and pentameric proteins that can be appropriately redesigned to assemble into defined homo and heterotypic polyhedral structures.[25,28,29] Directed evolution has also been used to generate and identify mutant proteins that can assemble into novel three-dimensional architectures.[33,34] Naturally occurring capsids, such as Lumazine Synthase (LS) from *Aquifex aeolicus*, can also be redesigned to assemble into expanded capsids displaying tetrahedral and supra-icosahedral symmetry by altering the charge of the particles internal surface.[23] Interestingly, for these expanded LS capsids, each of the structures possess the same C_2 interaction between pentamers which mediates the native interaction, however the planar angle between pentamers diverges, as the internal charge alters. Each of the methods highlighted thus far relies heavily on computational design of a protein, extensive experimental screening or the pre-existence of a capsid particle to achieve the higher-order assembly.

However, a simple and generic method to assemble proteins can be envisioned when some flexibility between the polyhedral vertices is allowed, eliminating the need for precise interface design. This approach has been exemplified by the Marsh group. Initial studies employing trimeric proteins fused to either half of a heterodimeric antiparallel coiled-coil, led to polydisperse mixtures of heterodimeric, tetrahedral, and octahedral assemblies;[35,36] however, assembly of a trimeric protein using parallel coiled-coils can lead to octahedra[37] or tetrahedra,[38] depending on whether homotetrameric or homotrimeric coiled-coils are employed. However, when fused to larger protein domains the oligomeric state adopted by a coiled-coil may differ from that expected.[39]

Herein we report a biologically inspired strategy for the self-assembly of functional pentameric proteins, which intentionally exploits coiled-coils that display plasticity in their oligomerization state. Our strategy is illustrated using the cholera toxin B-subunit (CTB), which is the nontoxic cell-surface binding portion of the cholera toxin, which has found widespread interest for biotechnological applications in targeted drug delivery or vaccine production.[40–42] Appended C-terminal parallel coiled-coils are used to actively encourage interactions between the CTB pentamers. Upon clustering the globular subunits via the coiled-coil domains, weaker secondary interactions are identified between the clusters, highlighting important residues and geometries for hierarchical self-assembly into protein capsules and nanotubes.

CTB is readily overexpressed as a monodisperse homo-pentameric ring of ∼58 kDa, in which its glycolipid-binding sites are arranged on the widest face of a pentagonal frustum (Figure 1a,b).[43] We postulated that it might be feasible to assemble these pentagonal units into a variety of supramolecular structures in a manner analogous to the assembly of polyomavirus capsid protein VP1, which can be directed to form icosahedral, octahedral or tubular assemblies.[44–46] The remarkable VP1 capsid protein makes use of C-terminal peptide extensions to organize its pentameric building blocks into both pentagonal and hexagonal arrangements by mediating several distinct bilateral and trilateral interactions between adjacent pentamers. We reasoned that appropriate peptide extensions on the C-termini of the CTB-pentamer (i.e., distal to the glycolipid binding face) could mediate self-assembly in a conceptually analogous manner (Figure 1b). We therefore sought a protein motif that would have the potential to form either dimeric (Figure 1c) or trimeric (Figure 1d) interactions between CTB pentamers.
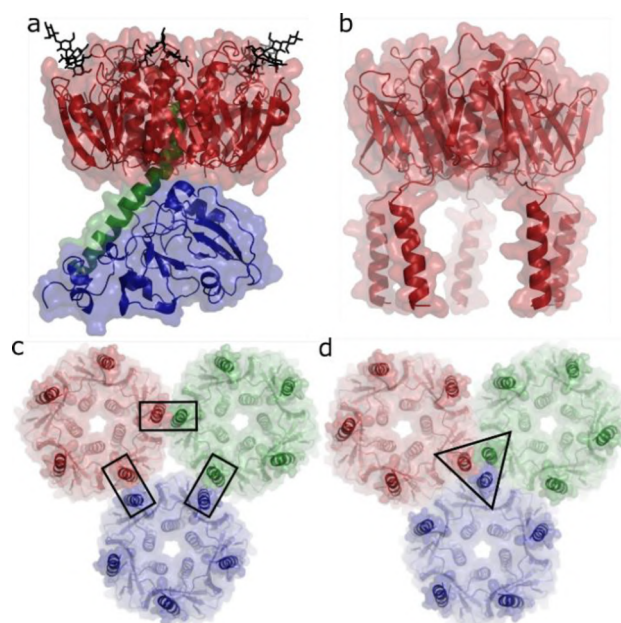

Figure 1. Design of Cholera Toxin B-subunit (CTB) fusions to allow assembly into higher-order structures. (a) CTB, in red, with the A-subunit (CTA) in blue, these domains are associated through the CTA2 α helix, in green. The oligosaccharide of the GM1 ganglioside is presented in black. (b) Five coiled-coil domains appended to the C-terminus of CTB, distal to the GM1 ganglioside binding face. (c and d) Potential binding conformations, with either dimeric or trimeric coiled-coils, respectively, of the CTB molecules allowing higher-order assemblies.

As discussed, coiled-coil peptides present attractive and tractable motifs for protein oligomerization.[47,48] These amphipathic α helices share an hpphppp heptad amino-acid sequence repeat denoted *abcdefg*, wherein h and p are hydrophobic and polar amino acids, respectively.[49] The hydrophobic amino acids, *a* and *d*, come together to form the core of the assembly which is further reinforced through salt bridges between adjacent helices at the *e* and *g* positions, which can be used to direct the specific assembly of heteromeric or homomeric coiled-coils.[50] Further stabilization of the coiled-coils can be achieved by altering the residues in positions *b*, *c* and *f*.[51] Whereas the C-terminal peptide extensions on polyomavirus VP1 interact with the globular domains of adjacent VP1 pentamers, coiled-coil peptides interact specifically with each other, and can thus provide an oligomerization domain that could be transplanted to any protein of interest. In this work we transplant the oligomerizing potential of coiled-coils onto the pentameric protein CTB, thereby transferring potential 2-fold and/or 3-fold symmetry axes to the current 5-fold symmetric CTB pentamers. This combination of symmetries allows the formation of curvature and thus the assembly of tubular and potentially spherical nanostructures.

■ RESULTS

A homomeric, three-heptad coiled-coil sequence was selected to scaffold the CTB interaction. This homomeric coiled-coil (CC2, Table 1, and Supporting Information, SI, Figure S1) was loosely based on a known synthetic heteromeric coiled-coil,[52] however, by altering the pattern of charged residues we aimed to promote homomeric assembly as a dimer (CC2). The

Table 1. Coiled-Coil Sequence: The Amino Acid Sequences of CC2 Aligned to the *abcdefg* Coiled-Coil Notation

notation	g	abcdefg	abcdefg	abcdefg	abcdef
CC2	G	VSALEKE	VSALKEK	VSALEF	
CC2a	G	VSALEKE	VSALKEK	VSALEKE	VSALKF

coiled-coil was fused to maltose-binding protein to allow preliminary studies of its propensity to self-assemble (SI Figure S2). While no oligomerization was observed for the three-heptad MBP-CC fusion, a four-heptad version of this sequence allowed the formation of dimers and trimers in a concentration-dependent manner, however the oligomers observed did not follow the stoichiometry predicted by the sequence knob into hole interactions. We anticipated that the observed promiscuity might allow an assembling system to choose between either dimeric or trimeric interactions in analogy to the assembly of VP1 proteins. Also in analogy with assembly of natural viruses,[3] we anticipated that the pentameric structure of CTB would give rise to multivalent interactions, thus enhancing the stability of the coiled-coils appended to the pentamers. Therefore, we anticipated that introduction of the lower affinity three-heptad version of the coiled-coil sequence at the C-terminus of CTB might allow more efficient reversibility of the interactions and thus avoid kinetic trapping of partially assembled species.

Both the CC2 and CC2a sequences were introduced at the C-terminus of the CTB expression construct. Overexpression and purification of the CTB-CC2 fusion proteins generated monodisperse pentameric proteins, however, the CTB-CC2a fusion showed evidence of dimers of pentamers when assessed by SEC (SI Figure S3). As we considered the easy formation of lower-order assemblies to be undesirable, we chose to continue with only the lower affinity three-heptad version of the coiled-coil. Extended incubation of CTB-CC2 in PBS at room temperature led to the development of a high weight peak at the void volume (Figure 2a). This peak was analyzed by DLS and found to have a hydrodynamic diameter of ~12−50 nm (Figure 2b). In contrast, wild-type CTB without coiled-coil extensions did not give rise to any high weight peaks during SEC even after extended periods of incubation (Figure 2a), and DLS returned a hydrodynamic diameter of 5.6 nm as expected for the CTB pentamer (Figure 2b). However, TEM failed to reveal any convincing ordered higher structures for CTB-CC2 incubated in PBS, only small amorphous aggregates were seen. Therefore, we investigated the assembly of CTB-CC2 in alternative buffers. Polyomavirus VP1 capsid proteins are known to assemble efficiently into icosahedral virus-like particles and other architectures upon incubation in 0.5−1 M ammonium sulfate solution.[45] When CTB-CC2 was exposed to these conditions, extended rod-like species with a diameter of 24 ± 4 nm were observed by TEM (Figure 2c,d and SI Figure S4) and tubular structures were observed in vitreous ice by cryo-EM (SI Figure S5). The assembled species were not sufficiently well ordered to permit TEM averaging techniques for structure determination studies. However, we anticipated that crystallization of CTB-CC2 might lead to the formation of similar extended assemblies.

Crystallization trials conducted under a broad range of conditions gave rise to various crystalline forms of the CTB-CC2 protein, including hexagonal plates that diffracted to a resolution of 2.45 Å in space group $P6_3$. The CTB proteins were found to assemble into nanotubes with an outer diameter
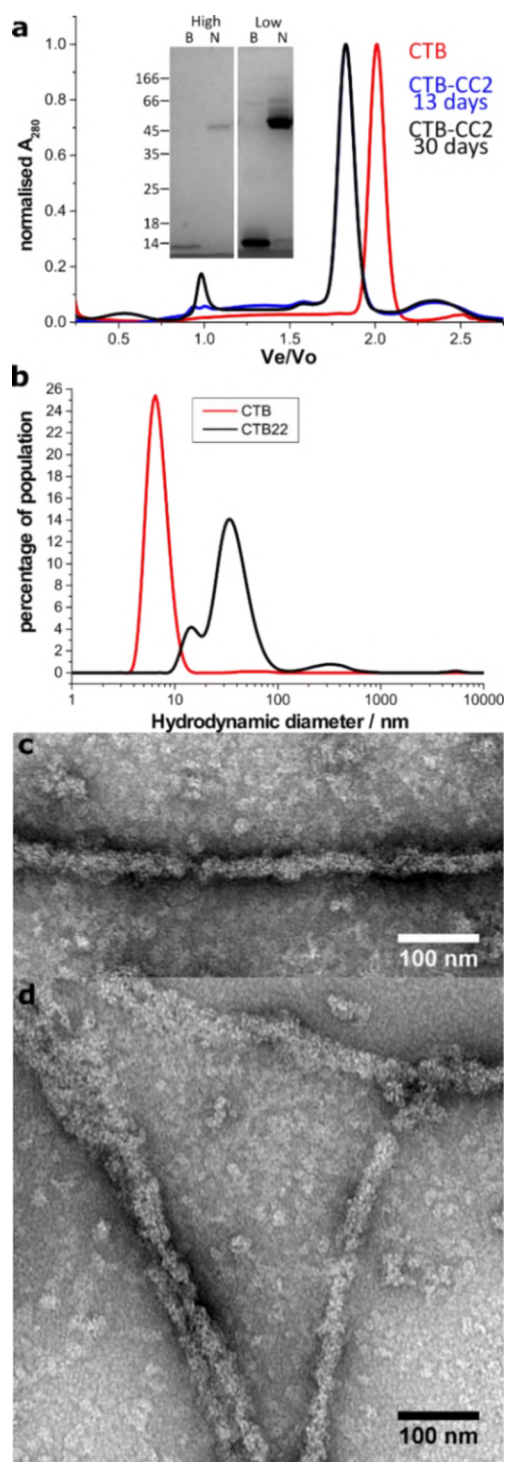
Figure 2. Characterization of CTB-CC2 by size exclusion chromatography (SEC), dynamic light scattering (DLS), and transmission electron microscopy (TEM). (a) SEC of CTB-CC2 revealed the development of a high weight species after prolonged incubation (blue and black), whereas CTB without the coiled-coil showed no increase in retention volume, inset is a SDS PAGE sample from the high weight peak compared to the main CTB peak which display the same profile (B = boiled, N = not boiled). (b) DLS of the high weight species revealed particles of approximately 13 and 32 nm in diameter, whereas native CTB pentamers gave a diameter of 5 nm. (c and d) Incubation of 58 μM CTB-CC2 in 0.5 M ammonium sulfate gave rise to tubular structures under TEM with a diameter of approximately 24 nm and various lengths up to 600 nm.

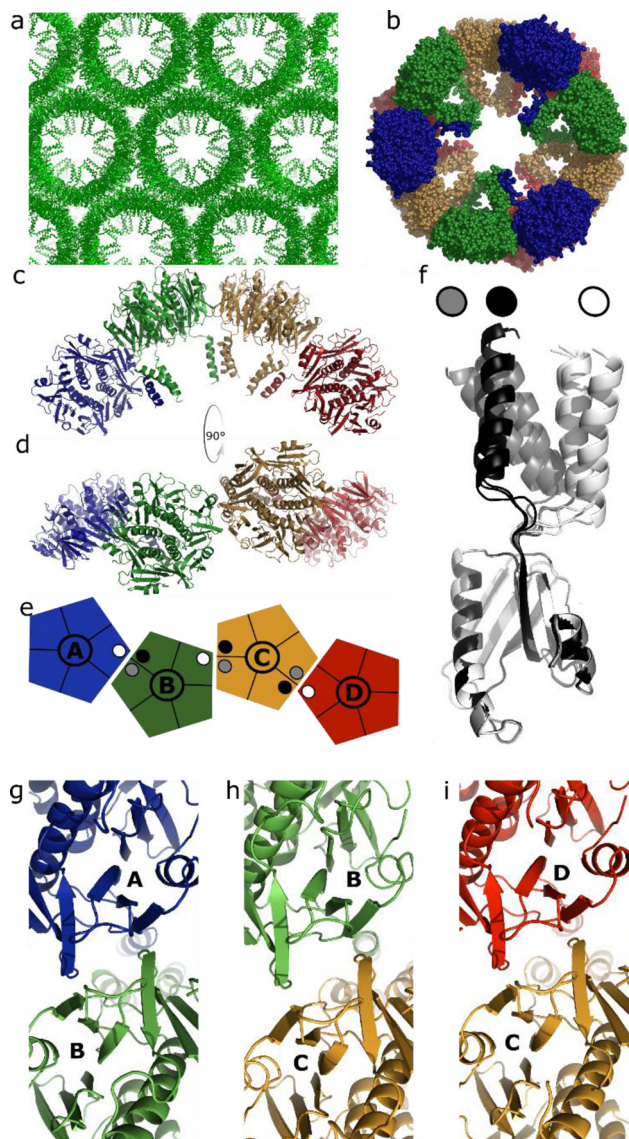of ~19.5 nm and inner diameter of ~6 nm (Figure 3a,b). The asymmetric unit comprises four pentamers of CTB (colored

Figure 3. Crystal structure of CTB-CC2 tubes; the asymmetric unit. (a) A section of the crystal structure looking down the *z*-axis, showing packing of the tubes and the inner lining of the coiled-coils. (b) Three asymmetric units, each consisting of four pentamers, form a ring structure. (c–e) There are four pentamers per asymmetric unit (blue, green, beige, and red), between each associated pentamer there is a trimeric coiled-coil with two coils donated from one pentamer, binding with a third coiled-coil from the neighboring pentamer. (e,f) These coiled-coils adopt specific structures relative to each other and its parent CTB monomer, colored white, gray, and black. (g–i) The interface, mediated by the coiled-coils, between each of the CTB pentamers are very similar, hinting at a pseudo 2-fold symmetry axis.

blue (A), green (B), beige (C), and red (D), Figure 3b−e) connected by three trimeric coiled-coils to form a short, twisted ribbon. The coiled-coils are aligned on the inside of the tube with the CTB GM1 ligand-binding sites on the outer face of the tube. The helical peptides in the trimeric coiled-coils can be classified into three different conformations (colored black, white, and gray in Figure 3f), and each coiled-coil has one of each conformation. The resolution of the structure around the

coiled-coils is lower than the resolution for the globular part of the pentamer, but there is convincing electron density for each of the three coiled-coils that connect the pentamers in the unit cell (SI Figure S6). It is possible that there may be additional coiled-coils present in the structure but the electron density for those was not of sufficient quality to model (SI Figure S7). The coiled-coils bring the pentamers together to create new protein−protein interaction interfaces (Figure 3g−i) which have local C2 symmetry that is broken only by the presence of the coiled-coil. While the A−B and C−D interfaces (Figure 3g,i) are essentially identical to one another, they differ slightly from the B−C interface (Figure 3h) which has a slightly greater twist. In the absence of the coiled-coil structure, the B−C interface would be an additional crystallographic axis of symmetry. For these nine modeled α-helices, the side chain positions were modeled using Rosetta[53,54] to enable their contribution to the interaction energy to be estimated. With coiled-coils present (and side chains modeled) the AB, BC, and CD interfaces were each estimated to have an interaction energy of ~−14 kcal/mol with PISA,[55] however, when the coils were removed, the interaction energy dropped to ~−0.4 kcal/mol, which suggests that the coiled-coils are the driving factor for the formation of the asymmetric unit (full PISA analysis is available in SI Figure S8).

The asymmetric units then assemble in a hierarchical fashion to produce the tubes. Each asymmetric unit is a short, twisted ribbon, three of which assemble in a staggered, right-handed helix to form a ring comprising 12 pentamers (Figures 3b and 4a), these rings then stack together to create the tube (Figure 4b). In bringing together the short, twisted ribbons, two new protein−protein interaction interfaces are created between the A−B pentamers of one asymmetric unit with the C−D pentamers of two adjacent asymmetric units, as highlighted in Figure 4c. The green and cyan interface (Figure 4d) forms a relatively tight interaction between the pentamers, whereas there is a looser interaction between the green and yellow tetramers (Figure 4e). The diamond-shaped tetramers ($A^0B^0-C^{+1}D^{+1}$ and $C^{-1}D^{-1}-A^0B^0$) formed by these new protein−protein interaction interfaces present two alternative depictions of the unit cell which differ in their local topology. Looking at the structure across $C^{+1}-A^0B^0$ (Figure 4f), one sees that the surface of the tetramer follows the overall curvature of the tube. However, looking across the $D^{+1}-A^0$ interface, the tetramer curves in the opposite direction, thereby forming a saddle-shaped surface. The surface of the yellow and green tetramer also follows the curvature of the tube across the $A^0-C^{-1}D^{-1}$ direction (Figure 4g), but the B^0-C^{-1} interface is essentially flat, giving a local summit-like topology. These alternative diamond-shaped depictions of the unit cell give rise to a series of overlapping saddles and summits which form long helical ribbons (Figure 4c, dotted lines). Three of these long helical ribbons, connected by the B−C coiled-coils, form the tube. The list of the normal vectors, planar angle of intersection and distance between subunit centers of mass can be found in SI Figure S9

Computational analysis of the interaction energies within these alternative asymmetric units allows us to estimate the contribution of individual interactions at these interfaces. For the saddle-shaped junction between $A^0B^0-C^{+1}D^{+1}$, the interaction ΔG was −15.5 kcal/mol), with much of the interaction energy contributed by a phosphate flanked by two molecules of the crystallization liqueur 2-methyl-2,4-pentane-diol (MPD), seen in the center of Figure 4d (described in
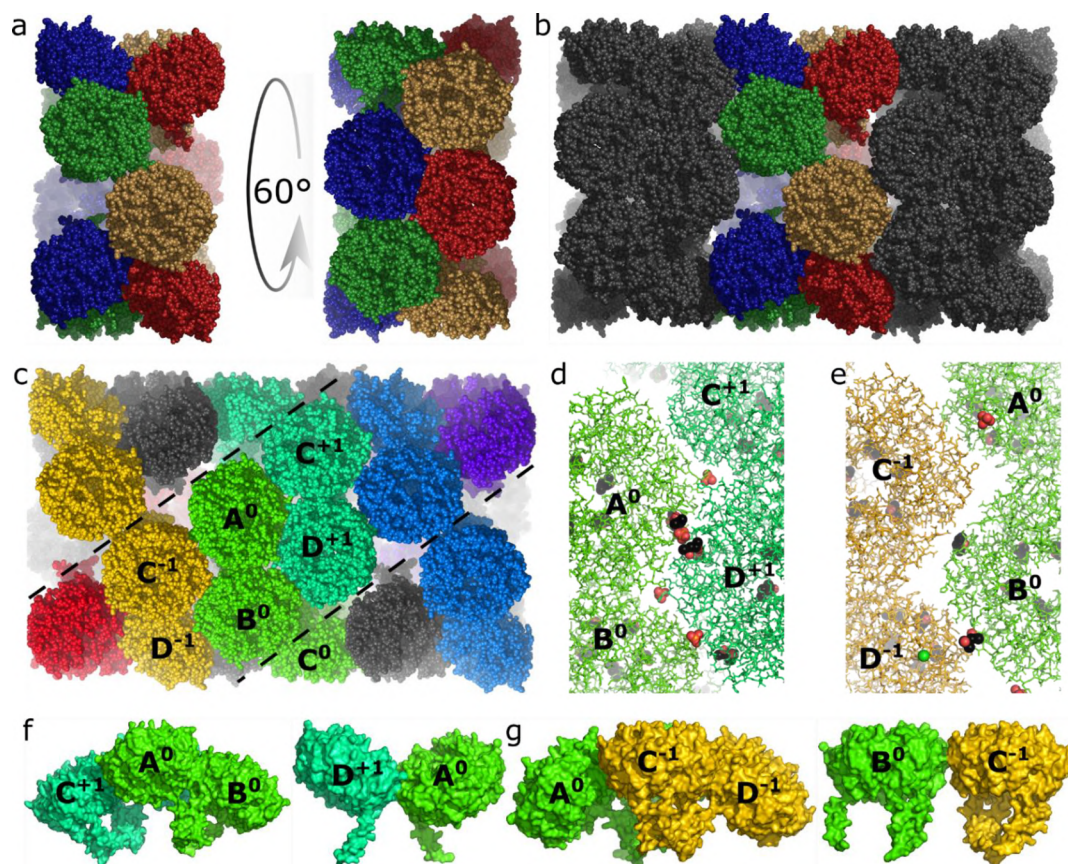
Figure 4. Crystal structure of CTB-CC2 tubes; assembly of the tube. (a) Two views of the ring structure made from three asymmetric units, the first showing the B−C coiled-coil connection in the center and the next presenting an alternate depiction of the asymmetric unit. (b) The ring structures stack together to form a tube. (c) Alternatively, the tube can be thought of three "ribbons" as shown between the dashed lines. (d and f) The interaction between the A^0B^0−$C^{+1}D^{+1}$ interface (with 2-methyl-2,4-pentanediol (MPD) in black spheres and phosphates in spheres), this interface forms the ring structure and is mediated by strong interactions from the crystallization buffer, it forms a saddle shape. (e and g) The interaction between the A^0B^0−$C^{-1}D^{-1}$ interface (with MPD in black spheres and phosphates in spheres), this is the interface between the rings which propagates the tube, and it forms a summit shape.

more detail in SI Figure S10). When solute molecules are excluded from the interaction, the contribution from the protein−protein interaction across this interface sums to −2.1 kcal/mol. For the summit-junction (between A^0B^0−$C^{-1}D^{-1}$) the subunits are less tightly packed and fewer solute molecules take part in the interaction, giving ΔG for the interaction of −2.6 kcal/mol, in the absence of solute molecules the interaction energy is −1.2 kcal./mol. Most importantly, in the absence of both coiled-coils and solutes of the crystallization liquor, the interaction energy between pentamers is essentially zero (−0.5 kcal/mol), suggesting the driving forces for the formation of the tube arise from the coiled-coils and the solute. Using this analysis, we hypothesize that the tubes assemble in a hierarchical fashion. Four pentamers might initially assemble into the asymmetric unit mediated by three $\sim$−14 kcal/mol interactions. This could be followed by formation of the saddle junction, mediated by the solute molecules which, with the asymmetric unit, form rings. The combined A^0B^0−$C^{+1}D^{+1}$ interface totals −15.5 kcal/mol, which is located in three positions around the ring—these rings would then nucleate tube assembly either by individual recruitment of additional tetramers of pentamers or by stacking of the rings.

DISCUSSION

In this work we have shown that appending a given oligomeric protein with homomeric coiled-coil motifs allows the assembly of higher order structures, with CTB-CC2 forming tubes under crystallization conditions and in the presence of an ammonium sulfate assembly buffer. As the coiled-coil motifs were appended to a single face of the protein, the coiled-coils effectively clustered this surface, presenting the opposite surface, which contains the functional GM1 ganglioside binding site, to the bulk solvent. Although all the CTB residues that participate in the protein−protein interaction interfaces are present in the wild-type protein, it does not naturally self-assemble. It is the presence of the coiled-coils that drives assembly of the pentamers. However, it is the oligomeric state, shape of the protein, and the necessity to find the lowest energy arrangement of the forced protein−protein interaction that dictates the geometry of the assembly. The opportunity to make the A^0B^0−$C^{+1}D^{+1}$ interface leads to the saddle structure and probably also drives the small distortion of the B−C interface compared to the geometry that is adopted by the A−B and C−D coiled-coil interfaces (SI Figure S9). The saddles have a higher interaction energy than the summits, based on the modeled coiled-coils, which may be one reason why the tubes appear to form in preference to spherical particles.

It is important to note that only nine of 20 possible coiled-coil helices could be modeled well using the observed electron density. Due to the geometry of the coiled-coils in the asymmetric unit, it is not possible to form a coiled-coil interaction across the saddle interface, the helices proximal to this interface are occupied by the coiled-coils of the asymmetric unit (SI Figure S7a). However, some of the unmodelled helices are sufficiently close to one another to potentially form two further coiled-coil interactions across the summit interface (SI Figure S7) and electron density for coiled-coils at these positions is apparent, but not within the confidence limits for modeling. It is therefore possible that a wide range of structurally inhomogeneous interactions (i.e., multiple conformations) between the unmodelled helices may also contribute to the stability of the observed structures. If these interactions do occur (they might not be observed due to being in multiple conformations), then they would make the summit conformation more energetically stable than the saddle conformation. While this need not necessarily change our hypothesis for the order of assembly steps, it would suggest a potential pathway to assembly of spherical particles.

The polyomavirus VP1 proteins, which are also pentameric, can also form either capsids or tubes depending on the assembly conditions. It is interesting to note that in the "ribbon" architecture of the assembly (Figure 4c), the A^0 pentamer is surrounded by four additional pentamers, C^{-1}, B^0, C^{+1}, and D^{+1}, however there is room for a further two pentamers to interact with A^0. Indeed, by sliding an adjacent ribbon along the dashed lines, two further pentamers can indeed be brought into proximity. This new arrangement would give a central pentamer surrounded by six additional pentamers which is similar to the arrangement necessary to facilitate the assembly of $T = 7d$ particles such as the VP1 capsid. Modeling the CTB pentamers in a $T = 7d$ conformation gives a theoretical capsid diameter of ∼32 nm which is within the bounds of the particle sizes detected by DLS in the high-weight peak from SEC (Figure 2a,b).

■ CONCLUSIONS

Appending coiled-coils to a chosen functional protein provides a pragmatic approach for the creation of novel supramolecular protein assemblies. With careful consideration of the appendage site for the coiled-coils, it is possible to control the directed assembly of the proteins so that their functional sites are presented in a desired orientation which would allow their continued use when assembled. Crystallographic characterization of the tubular assembly has revealed low-energy "native interactions" that offer a starting point to allow future rational design of protein–protein interaction interfaces with enhanced stability. We suggest that this strategy therefore represents a valuable addition to the currently available approaches for engineering of protein assemblies.

■ METHODS

Molecular Biology and Protein Purification. CTB and CTB-CC2 genes were constructed by assembly PCR from single stranded oligonucleotides (Integrated DNA Technologies), see SI section S12 for sequences. Assembly PCR was performed with a final concentration of 1× Pfu Buffer, 0.6 U Pwo Polymerase, 0.8 mM dNTP mix, 1 μM terminal priming oligonucleotides, 10 nM oligonucleotide parts (one each of parts p1–p8), and made to a final volume of 50 μL with water. Amplified genes were spliced into the pSAB2.2 plasmid (SI section S12) via restriction cloning with SbfI

and PstI (NEB). The correct gene sequence of plasmid extracts (Qiagen Miniprep Kit) from transformations (Agilent XL10-Gold) of the ligation product were confirmed via DNA sequencing (GATC-Biotech). Plasmids were transformed into C41-DE3 cells (Lucigen) and 5 mL overnight starter cultures were grown. Three mL of starter culture was used to inoculate 1 Lt of LB media, which was grown until an OD_{600} of 0.6−0.8 was achieved. Protein expression was induced with 1 mL of 500 mM IPTG overnight at 30 °C, shaking at 220 rpm. Cells were pelleted at 10 000g for 10 min, 570 g of ammonium sulfate was slowly added with stirring to 1 Lt of the cleared expression media, which was then incubated at room temperature for 2 h with stirring. The precipitation media was then centrifuged at 17 800g for 20 min, and the pellet was resuspended in Nickel Wash Buffer (20 mM imidazole in PBS), centrifuged at 10 000g for 10 min and then passed through a vacuum filter (0.45 μm pore size). Nickel affinity purification was performed with a step elution on gravity drip columns containing Ni-NTA agarose (Qiagen) using Nickel Elution Buffer (300 mM imidazole in PBS). Pure samples were pooled and concentrated with Amicon centrifugal filters (Merck-Millipore).

For the MBP-CC genes, CC2/2a, and Tri were assembled in the same assembly PCR manner as the CTB-CC genes (one each of parts p1−p3), see SI section S13 for sequences. These were, however, cloned into the pMal-c5x plasmid (NEB) via BamHI and PstI (NEB) restriction cloning. The correct gene sequence of plasmid extracts (Qiagen Miniprep Kit) from transformations (E. coli XL10-Gold) of the ligation product were confirmed via DNA sequencing (GATC-Biotech). Plasmids were transformed into E. coli C41-DE3 cells (Lucigen) and 5 mL overnight starter cultures were grown. Three mL of starter culture was used to inoculate 1 L of LB media, which was grown until an OD_{600} of 0.6−0.8 was achieved. Protein expression was then induced with 1 mL of 500 mM IPTG overnight at 30 °C, shaking at 220 rpm. Cells were pelleted at 10 000g for 10 min, the pellet was resuspended in PBS. The cells were then disrupted with two passes through a Constant Systems cell disrupter. The clarified lysate was washed through an amylose affinity column (GE Healthcare) via an AKTA Explorer and eluted under 10 mM maltose in PBS. Pure samples were pooled and concentrated with Amicon centrifugal filters (Merck-Millipore).

Protein Characterization. Size exclusion chromatography (SEC) was conducted in order to both detect and isolate higher oligomeric species of CTB, both with and without coiled-coil extensions using either a Superdex 200 10/300 or a Superose 6 10/300 on an AKTA purifier (GE Healthcare) equilibrated with PBS, pH 7.4. 100 μL of concentrated protein (typically 5−10 mg/mL) was injected.

Dynamic light scattering (DLS) was performed to determine the hydrodynamic radii of CTB pentamers and CTB-coiled-coil-mediated oligomeric assemblies. 100 μL of 1 mg/mL protein samples in PBS, pH 7.4 were analyzed with a Zetasizer Nano ZS (Malvern) via 3 replicates of 3 measurements for 10 s each, the data were processed via the "protein analysis" function of the Malvern software by non-negative least-squares analysis followed by L-curve.

Transmission electron microscopy (TEM) was used to gain insight into the finer structure of the assemblies observed by DLS. Carbon coated TEM grids were prepared by charging the grids under UV light for 30 min, grids were then allowed 5 min to rest and then 5 μL of protein sample were applied and allowed 30 s to bind before being wicked off, 5 μL of 1% uranyl acetate was then added and immediately wicked, and prepared grids were allowed to dry and then observed with either a FEI Tecnai T12 G2 Spirit TWIN/BioTWIN or a FEI Tecnai F20 G2-Spirit electron microscope.

For cryoEM sample preparation, Quantifoil R1.2/1.3 or Lacy carbon grids with 2 nm continuous support were subjected to glow discharged (60 s, 20 mA). Three μL of sample was applied to the grid, which was vitrified using a Vitrobot Mk IV. CryoEM imaging was carried out on a Titan Krios microscope with Falcon-3 direct electron detector.

Protein crystals were grown by manually prepared sitting drop vapor diffusion with 1 μL of 10 mg/mL protein and 1 μL of mother liquor (0.2 M ammonium phosphate monobasic, 0.1 M tris (pH 8.5), 50% v/v (±)-2-methyl-2,4-pentanediol). Crystals were flash cooled in

liquid nitrogen with no additional cryoprotectant added. Diffraction data were collected at 100 K on the I24 beamline at Diamond Light Source, integrated in the $P6_3$ space group using xia2 (ccp4, distl, labelit, pointless, xds),[56−60] and then scaled and merged using Aimless.[61] Molecular replacement, using PDB 3CHB was carried out using Molrep.[62] Four pentamers of cholera toxin were found per asymmetric unit that assembled via symmetry to give a tubular structure. The asymmetric unit was refined using iterative cycles of reciprocal space refinement (TLS and restrained refinement) with REFMAC5[63] following the generation of secondary structure restraints using PROSMART, and manual model rebuilding using COOT.[64] Each CTB pentamer (residues 1−103, chains ABCDE, FGHIJ, KLMNO, PQRST) and each of the three triple helices (residues 104−124, chains EFG, JLM, NOP) was described with a single TLS group. Restrained refinement used REFMAC's automatically generated local NCS restraints. Several small molecule ligands derived from the crystallization mother liquor were found by examination of Fo−Fc electron density maps and modeled accordingly. The triple helices (residues 104−124 were modeled as ideal helices using a poly ala model. While Fo−Fc electron density did show evidence for side chain positions in the triple helices, modeling of the side chains resulted in the working R and R_{free} diverging, suggesting overfitting. A table of the crystallographic statistics can be found in SI section S11, PDB accession code 6HSV.

Assessment of CTB interaction energy: Two models were assessed to estimate the binding energies between the CTB pentamers in the crystal structure. In the first model, the coiled−coil regions were removed leaving only residues 1−103 for each of the monomers in the asymmetric unit. In the second model, to assess a more realistic interaction energies between the CTB subunits, side chains of the modeled coiled−coil sections were added using the fixbb routine in Rosetta.[53] This allowed the addition of the missing side chains while maintaining rigid atom positions for the rest of the structure. For each model, the nomenclature of the chain IDs was changed in PyMOL so that the CTB pentamer chains ABCDE were relabeled as chain A, chains FGHIJ were relabeled as chain B, chains KLMNO were relabeled as chain C, and chains PQRST were relabeled as chain D. Residue numbers for each new chain were incremented to remove ambiguously labeled residues. The two structures were submitted to analysis by PISA.[55] These data are presented in SI section S8.

Obtaining the Angle between Two Pentamers. We define a normal vector, $\vec{n}$, for each pentamer subunit as a unit vector in the direction of the vector that connects the center of mass of five T78 residue at the bottom to the center of mass of five K63 residues at the top of the pentamer. We estimate the angle between two subunit planes with normal vectors $\vec{n_1}$ and $\vec{n_2}$, as $\theta = \cos^{-1}(\vec{n_1} \cdot \vec{n_2})$ that gives an angle in the interval $[0, \pi]$. These data are presented in SI section S9.

ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.8b11480.

Coiled-coil sequences and oligomer characterization; X-ray crystallography electron density maps showing coiled-coils; PISA subunit interaction analysis; geometrical description of the tube assembly; crystallographic statistics; and nucleotide sequences for coiled-coils (PDF)

AUTHOR INFORMATION

Corresponding Authors

*arwen.pearson@cfel.de
*m.e.webb@leeds.ac.uk
*w.b.turnbull@leeds.ac.uk

ORCID ⓘ
James F. Ross: 0000-0002-4784-103X
Emanuele Paci: 0000-0002-4891-2768
Michael E. Webb: 0000-0003-3574-4686
W. Bruce Turnbull: 0000-0002-7352-0360

Present Address
#Present address of J.F.R.: Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

Author Contributions
All authors have given approval to the final version of the manuscript

Notes
The authors declare no competing financial interest.
The data associated with this paper are openly available from the University of Leeds Data Repository at DOI: 10.5518/575.

REFERENCES

(1) Marsh, J. A.; Teichmann, S. A. Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annu. Rev. Biochem.* **2015**, *84* (1), 551−575.

(2) Subramanian, R.; Kapoor, T. M. Building Complexity: Insights into Self-Organized Assembly of Microtubule-Based Architectures. *Dev. Cell* **2012**, *23* (5), 874−885.

(3) Mateu, M. G. Assembly, Stability and Dynamics of Virus Capsids. *Arch. Biochem. Biophys.* **2013**, *531* (1−2), 65−79.

(4) Bobik, T. A.; Lehman, B. P.; Yeates, T. O. Bacterial Microcompartments: Widespread Prokaryotic Organelles for Isolation and Optimization of Metabolic Pathways. *Mol. Microbiol.* **2015**, *98* (2), 193−207.

(5) Kerfeld, C. A.; Aussignargues, C.; Zarzycki, J.; Cai, F.; Sutter, M. Bacterial Microcompartments. *Nat. Rev. Microbiol.* **2018**, *16* (5), 277−290.

(6) Bromley, E. H. C.; Channon, K.; Moutevelis, E.; Woolfson, D. N. Peptide and Protein Building Blocks for Synthetic Biology: From Programming Biomolecules to Self-Organized Biomolecular Systems. *ACS Chem. Biol.* **2008**, *3* (1), 38−50.

(7) Frank, S.; Lawrence, A. D.; Prentice, M. B.; Warren, M. J. Bacterial Microcompartments Moving into a Synthetic Biological World. *J. Biotechnol.* **2013**, *163*, 273−279.

(8) Jutz, G.; Van Rijn, P.; Santos Miranda, B.; Böker, A. Ferritin: A Versatile Building Block for Bionanotechnology. *Chem. Rev.* **2015**, *115* (4), 1653−1701.

(9) Yeates, T. O.; Liu, Y.; Laniado, J. The Design of Symmetric Protein Nanomaterials Comes of Age in Theory and Practice. *Curr. Opin. Struct. Biol.* **2016**, *39*, 134−143.

(10) Ljubetič, A.; Gradišar, H.; Jerala, R. Advances in Design of Protein Folds and Assemblies. *Curr. Opin. Chem. Biol.* **2017**, *40*, 65−71.

(11) Yeates, T. O. Geometric Principles for Designing Highly Symmetric Self-Assembling Protein Nanomaterials. *Annu. Rev. Biophys.* **2017**, *46* (1), 23−42.

(12) Kobayashi, N.; Arai, R. Design and Construction of Self-Assembling Supramolecular Protein Complexes Using Artificial and

Fusion Proteins as Nanoscale Building Blocks. *Curr. Opin. Biotechnol.* 2017, 46, 57−65.

(13) Arai, R. Hierarchical Design of Artificial Proteins and Complexes toward Synthetic Structural Biology. *Biophys. Rev.* 2018, 10 (2), 391−410.

(14) Chen, J.; Seeman, N. C. Synthesis from DNA of a Molecule with the Connectivity of a Cube. *Nature* 1991, 350, 631−633.

(15) Yin; Ke, Y.; Ong, L. L.; Shih, W. M. Three-Dimensional Structures Self-Assembled from DNA Bricks. *Science* 2012, 338 (6111), 1177−1184.

(16) Fletcher, J. M.; Boyle, A. L.; Bruning, M.; Bartlett, G. J.; Vincent, T. L.; Zaccai, N. R.; Armstrong, C. T.; Bromley, E. H. C. C.; Booth, P. J.; Brady, R. L.; et al. A Basis Set of de Novo Coiled-Coil Peptide Oligomers for Rational Protein Design and Synthetic Biology. *ACS Synth. Biol.* 2012, 1 (6), 240−250.

(17) Fletcher, J. M.; Harniman, R. L.; Barnes, F. R. H.; Boyle, A. L.; Collins, A.; Mantell, J.; Sharp, T. H.; Antognozzi, M.; Booth, P. J.; Linden, N.; et al. Self-Assembling Cages from Coiled-Coil Peptide Modules. *Science* 2013, 340 (6132), 595−599.

(18) Gradišar, H.; Božič, S.; Doles, T.; Vengust, D.; Hafner-Bratkovič, I.; Mertelj, A.; Webb, B.; Šali, A.; Klavžar, S.; Jerala, R. Design of a Single-Chain Polypeptide Tetrahedron Assembled from Coiled-Coil Segments. *Nat. Chem. Biol.* 2013, 9 (6), 362−366.

(19) Egelman, E. H.; Xu, C.; Dimaio, F.; Magnotti, E.; Modlin, C.; Yu, X.; Wright, E.; Baker, D.; Conticello, V. P. Structural Plasticity of Helical Nanotubes Based on Coiled-Coil Assemblies. *Structure* 2015, 23 (2), 280−289.

(20) Chen, K. H.; Corro, K. A.; Le, S. P.; Nowick, J. S. X-Ray Crystallographic Structure of a Giant Double-Walled Peptide Nanotube Formed by a Macrocyclic β-Sheet Containing Aβ16−22. *J. Am. Chem. Soc.* 2017, 139 (24), 8102−8105.

(21) Thomson, A. R.; Wood, C. W.; Burton, A. J.; Bartlett, G. J.; Sessions, R. B.; Brady, R. L.; Woolfson, D. N. Computational Design of Water-Soluble α-Helical Barrels. *Science* 2014, 346 (6208), 485−488.

(22) Ljubetič, A.; Lapenta, F.; Gradišar, H.; Drobnak, I.; Aupič, J.; Strmšek, Ž.; Lainšček, D.; Hafner-Bratkovič, I.; Majerle, A.; Krivec, N.; et al. Design of Coiled-Coil Protein-Origami Cages That Self-Assemble in Vitro and in Vivo. *Nat. Biotechnol.* 2017, 35 (11), 1094−1101.

(23) Sasaki, E.; Böhringer, D.; Van De Waterbeemd, M.; Leibundgut, M.; Zschoche, R.; Heck, A. J. R.; Ban, N.; Hilvert, D. Structure and Assembly of Scalable Porous Protein Cages. *Nat. Commun.* 2017, 8, 14663.

(24) Azuma, Y.; Herger, M.; Hilvert, D. Diversification of Protein Cage Structure Using Circularly Permuted Subunits. *J. Am. Chem. Soc.* 2018, 140 (2), 558−561.

(25) King, N. P.; Bale, J. B.; Sheffler, W.; McNamara, D. E.; Gonen, S.; Gonen, T.; Yeates, T. O.; Baker, D. Accurate Design of Co-Assembling Multi-Component Protein Nanomaterials. *Nature* 2014, 510 (7503), 103−108.

(26) Miranda, F. F.; Iwasaki, K.; Akashi, S.; Sumitomo, K.; Kobayashi, M.; Yamashita, I.; Tame, J. R. H.; Heddle, J. G. A Self-Assembled Protein Nanotube with High Aspect Ratio. *Small* 2009, 5 (18), 2077−2084.

(27) Nguyen, T. K.; Negishi, H.; Abe, S.; Ueno, T. Construction of Supramolecular Nanotubes from Protein Crystals. *Chem. Sci.* 2019, 10, 1046−1051.

(28) Hsia, Y.; Bale, J. B.; Gonen, S.; Shi, D.; Sheffler, W.; Fong, K. K.; Nattermann, U.; Xu, C.; Huang, P. S.; Ravichandran, R.; et al. Design of a Hyperstable 60-Subunit Protein Icosahedron. *Nature* 2016, 535 (7610), 136−139.

(29) Bale, J. B.; Gonen, S.; Liu, Y.; Sheffler, W.; Ellis, D.; Thomas, C.; Cascio, D.; Yeates, T. O.; Gonen, T.; King, N. P.; Baker, D.; et al. Accurate Design of Megadalton-Scale Two-Component Icosahedral Protein Complexes. *Science* 2016, 353 (6297), 389−394.

(30) Aumiller, W. M.; Uchida, M.; Douglas, T. Protein Cage Assembly across Multiple Length Scales. *Chem. Soc. Rev.* 2018, 47 (10), 3433−3469.

(31) Padilla, J. E.; Colovos, C.; Yeates, T. O. Nanohedra: Using Symmetry to Design Self Assembling Protein Cages, Layers, Crystals, and Filaments. *Proc. Natl. Acad. Sci. U. S. A.* 2001, 98 (5), 2217−2221.

(32) Lai, Y. T.; Reading, E.; Hura, G. L.; Tsai, K. L.; Laganowsky, A.; Asturias, F. J.; Tainer, J. A.; Robinson, C. V.; Yeates, T. O. Structure of a Designed Protein Cage That Self-Assembles into a Highly Porous Cube. *Nat. Chem.* 2014, 6 (12), 1065−1071.

(33) Wörsdörfer, B.; Woycechowsky, K. J.; Hilvert, D. Directed Evolution of a Protein Container. *Science* 2011, 331 (6017), 589−592.

(34) Cornell, T. A.; Ardejani, M. S.; Fu, J.; Newland, S. H.; Zhang, Y.; Orner, B. P. A Structure-Based Assembly Screen of Protein Cage Libraries in Living Cells: Experimentally Repacking a Protein-Protein Interface to Recover Cage Formation in an Assembly-Frustrated Mutant. *Biochemistry* 2018, 57 (5), 604−613.

(35) Patterson, D. P.; Desai, A. M.; Holl, M. M. B.; Marsh, E. N. G. Evaluation of a Symmetry-Based Strategy for Assembling Protein Complexes. *RSC Adv.* 2011, 1 (6), 1004−1012.

(36) Patterson, D. P.; Su, M.; Franzmann, T. M.; Sciore, A.; Skiniotis, G.; Marsh, E. N. G. Characterization of a Highly Flexible Self-Assembling Protein System Designed to Form Nanocages. *Protein Sci.* 2014, 23 (2), 190−199.

(37) Sciore, A.; Su, M.; Koldewey, P.; Eschweiler, J. D.; Diffley, K. A.; Linhares, B. M.; Ruotolo, B. T.; Bardwell, J. C. A.; Skiniotis, G.; Marsh, E. N. G. Flexible, Symmetry-Directed Approach to Assembling Protein Cages. *Proc. Natl. Acad. Sci. U. S. A.* 2016, 113 (31), 8681−8686.

(38) Badieyan, S.; Sciore, A.; Eschweiler, J. D.; Koldewey, P.; Cristie-David, A. S.; Ruotolo, B. T.; Bardwell, J. C. A.; Su, M.; Marsh, E. N. G. Symmetry-Directed Self-Assembly of a Tetrahedral Protein Cage Mediated by de Novo-Designed Coiled Coils. *ChemBioChem* 2017, 18 (19), 1888−1892.

(39) Cristie-David, A. S.; Sciore, A.; Badieyan, S.; Escheweiler, J. D.; Koldewey, P.; Bardwell, J. C. A.; Ruotolo, B. T.; Marsh, E. N. G. Evaluation of: De Novo -Designed Coiled Coils as off-the-Shelf Components for Protein Assembly. *Mol. Syst. Des. Eng.* 2017, 2 (2), 140−148.

(40) De Haan, L.; Hirst, T. R. Cholera Toxin: A Paradigm for Multi-Functional Engagement of Cellular Mechanisms (Review). *Mol. Membr. Biol.* 2004, 21 (2), 77−92.

(41) Tinker, J. K.; Erbe, J. L.; Holmes, R. K. Characterization of Fluorescent Chimeras of Cholera Toxin and Escherichia Coli Heat-Labile Enterotoxins Produced by Use of the Twin Arginine Translocation System. *Infect. Immun.* 2005, 73 (6), 3627−3635.

(42) Barrett, L. B.; Berry, M.; Ying, W. B.; Hodgkin, M. N.; Seymour, L. W.; Gonzalez, A. M.; Read, M. L.; Baird, A.; Logan, A. Ctb Targeted Non-Viral CDNA Delivery Enhances Transgene Expression in Neurons. *J. Gene Med.* 2004, 6 (4), 429−438.

(43) Merritt, E. A.; Kuhn, P.; Sarfaty, S.; Erbe, J. L.; Holmes, R. K.; Hol, W. G. J. The 1.25 Å Resolution Refinement of the Cholera Toxin B-Pentamer: Evidence of Peptide Backbone Strain at the Receptor-Binding Site. *J. Mol. Biol.* 1998, 282 (5), 1043−1059.

(44) Salunke, D. M.; Caspar, D. L.; Garcea, R. L. Polymorphism in the Assembly of Polyomavirus Capsid Protein VP1. *Biophys. J.* 1989, 56 (5), 887−900.

(45) Kanesashi, S. N.; Ishizu, K. I.; Kawano, M. A.; Han, S. I.; Tomita, S.; Watanabe, H.; Kataoka, K.; Handa, H. Simian Virus 40 VP1 Capsid Protein Forms Polymorphic Assemblies in Vitro. *J. Gen. Virol.* 2003, 84 (7), 1899−1905.

(46) Kiselev, N. A.; Klug, A. The Structure of Viruses of the Papilloma-Polyoma Type V. Tubular Variants Built of Pentamers. *J. Mol. Biol.* 1969, 40, 155−171.

(47) Woolfson, D. N.; Bartlett, G. J.; Bruning, M.; Thomson, A. R. New Currency for Old Rope: From Coiled-Coil Assemblies to α-Helical Barrels. *Curr. Opin. Struct. Biol.* 2012, 22 (4), 432−441.

(48) Truebestein, L.; Leonard, T. A. Coiled-Coils: The Long and Short of It. *BioEssays* 2016, 38 (9), 903−916.

(49) Walshaw, J.; Woolfson, D. N. SOCKET: A Program for Identifying and Analysing Coiled-Coil Motifs within Protein Structures. *J. Mol. Biol.* **2001**, *307* (5), 1427−1450.

(50) Crick, F. H. C. The Packing of α-Helices: Simple Coiled-Coils. *Acta Crystallogr.* **1953**, *6* (8), 689−697.

(51) Drobnak, I.; Gradišar, H.; Ljubetič, A.; Merljak, E.; Jerala, R. Modulation of Coiled-Coil Dimer Stability through Surface Residues While Preserving Pairing Specificity. *J. Am. Chem. Soc.* **2017**, *139* (24), 8229−8236.

(52) De Crescenzo, G.; Litowski, J. R.; Hodges, R. S.; O'Connor-McCourt, M. D. Real-Time Monitoring of the Interactions of Two-Stranded de Novo Designed Coiled-Coils: Effect of Chain Length on the Kinetic and Thermodynamic Constants of Binding. *Biochemistry* **2003**, *42* (6), 1754−1763.

(53) Alford, R. F.; Leaver-Fay, A.; Jeliazkov, J. R.; O'Meara, M. J.; DiMaio, F. P.; Park, H.; Shapovalov, M. V.; Renfrew, P. D.; Mulligan, V. K.; Kappel, K.; et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **2017**, *13* (6), 3031−3048.

(54) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science* **2003**, *302* (5649), 1364−1368.

(55) Krissinel, E.; Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **2007**, *372* (3), 774−797.

(56) The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1994**, *50* (5), 760−763.

(57) Kabsch, W. XDS. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 125−132.

(58) Sauter, N. K.; Grosse-Kunstleve, R. W.; Adams, P. D. Robust Indexing for Automatic Data Collection. *J. Appl. Crystallogr.* **2004**, *37*, 399−409.

(59) Winter, G. Xia2 : An Expert System for Macromolecular Crystallography Data Reduction. *J. Appl. Crystallogr.* **2010**, *43*, 186−190.

(60) Zhang, Z.; Sauter, N. K.; Van Den Bedem, H.; Snell, G.; Deacon, A. M. Automated Diffraction Image Analysis and Spot Searching for High-Throughput Crystal Screening. *J. Appl. Crystallogr.* **2006**, *39* (1), 112−119.

(61) Evans, P. R.; Murshudov, G. N. How Good Are My Data and What Is the Resolution? *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2013**, *69* (7), 1204−1214.

(62) Vagin, A.; Teplyakov, A. MOLREP: An Automated Program for Molecular Replacement. *J. Appl. Crystallogr.* **1997**, *30* (6), 1022−1025.

(63) Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the Refinement of Macromolecular Crystal Structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67* (4), 355−367.

(64) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66* (4), 486−501.