Title: Symmetry-directed Self-assembly of a Tetrahedral Protein Cage Mediated by De Novo-designed Coiled Coils

Authors: Somayesadat Badieyan; Aaron Sciore; Joseph Eschweiler; Philipp Koldewey; Ajitha S Cristie-David; Brandon T Ruotolo; James C.A. Bardwell; Min Su; Neil Marsh

To be cited as: ChemBioChem 10.1002/cbic.201700406

Link to VoR: https://doi.org/10.1002/cbic.201700406

Symmetry-directed Self-assembly of a Tetrahedral Protein Cage Mediated by *De Novo*-designed Coiled Coils

Somayesadat Badieyan,[a,‡] Aaron Sciore,[a,‡] Joseph D. Eschweiler,[a] Philipp Koldewey,[b] Ajitha S. Cristie-David,[a] Brandon T. Ruotolo,[a] James C. A. Bardwell,[b,c,d] Min Su,[e] and E. Neil G. Marsh[a,c*]

Abstract: The organization of proteins into new hierarchical forms is an important challenge in synthetic biology. However, engineering new interactions between protein subunits is technically challenging and typically requires extensive redesign of protein-protein interfaces. We have developed a conceptually simple approach, based on symmetry principles, that uses short coiled-coil domains to assemble proteins into higher-order structures. Here we demonstrate the assembly of a trimeric enzyme into a well-defined tetrahedral cage. This is achieved by genetically fusing a trimeric coiled-coil domain to its C-terminus through a flexible poly-glycine linker sequence. The linker length and coiled-coil strength were the only parameters that needed to be optimized to obtain a high yield of correctly assembled protein cages.

The assembly of protein subunits into higher-order structures is ubiquitous in Nature and often appears to be guided by the principles of symmetry.[1] These natural protein assemblies have inspired the fields of synthetic biology and nanotechnology to apply symmetry principles to the design of new self-assembling protein systems. Among the architectures that have been achieved are: 1-D protein fibers and rings; 2-D and 3-D protein lattices; and tetrahedral, octahedral and icosahedral protein cages.[2]

The principle challenge to designing protein cages and lattices with well-defined architectures lies in precisely aligning the two rotational symmetry axes at the correct angle necessary to specify the intended geometry.[3] One approach involves engineering new C_2-symmetric interfaces into oligomeric proteins, generally C_3-symmetric trimers, to provide the second symmetry axis needed to mediate assembly. The rigid interface ensures that the C_2 and C_3 symmetry axes are aligned with the precision necessary to specify the desired geometry. However, this approach requires extensive computational resources to evaluate candidate designs *in silico*, and screening of many trial designs to identify successfully assemblies. The extensive mutations introduced to create the new protein-protein interface may also compromise the protein's biological activity.

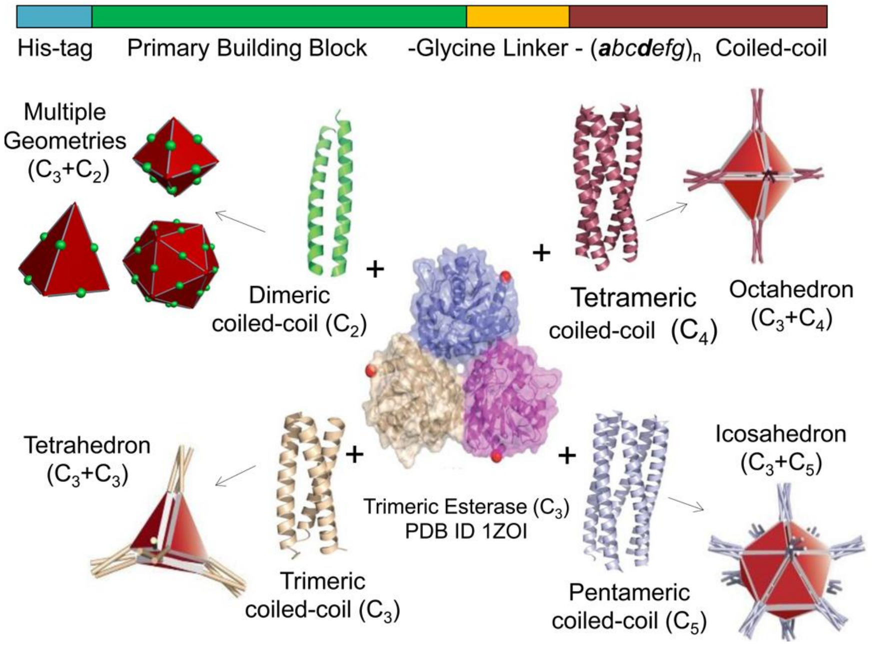


Figure 1. Design scheme: By fusing the C-terminus (red spheres) of a trimeric protein through a flexible linker to coiled-coils with different oligomerization states, protein cages of different geometries can be assembled. The geometry is specified by the combination of symmetry axes.

We have focused on an alternative approach to protein assembly that uses a flexible linker sequence to connect the two symmetry elements, e.g. by fusing a C_2 dimeric protein to a C_3 trimeric protein.[4] Normally, this results in poly-disperse assemblies due to the limited control over the orientation of the two domains. However, we recognized that using protein elements with rotational symmetries of C_3 or higher significantly restricts the number of compatible geometries so that unique structures can, in principle, form. Relaxing the requirement that the rotational symmetry axes need to be rigidly aligned greatly simplifies the design process.

Coiled-coil proteins are attractive components for this 'flexible' approach as they are among the simplest and best-understood protein-protein interactions. There are a large

---

[a] Dr. S Badieyan, Dr. A Sciore, Dr. JD Eschweiler, Ms. AS Cristie-David, Prof. BT Ruotolo, Prof. ENG Marsh*
Department of Chemistry
University of Michigan
Ann Arbor, MI 48109
Email: nmarsh@umich.edu
[b] Dr. P Koldewey, Prof. JCA Bardwell
Department of Molecular, Cellular and Developmental Biology
University of Michigan
Ann Arbor, MI 48109
[c] Prof. JCA Bardwell, Prof. ENG Marsh
Department of Biological Chemistry
University of Michigan
Ann Arbor, MI 48109
[d] Prof. JCA Bardwell
Howard Hughes Medical Institute
[e] Dr. M Su,
Life Sciences Institute
University of Michigan
Ann Arbor, MI 48109
[‡] These authors contributed equally.

Supporting information for this article is given via a link at the end of the document.

number of well-characterized, *de novo*-designed coiled-coil available as “off-the-shelf” components that can be fused to larger, natural protein subunits to mediate assembly.[5] This design strategy naturally lends itself to a modular approach for constructing protein cages: by mixing and matching protein building blocks and coiled-coil domains with different rotational symmetries it should be possible to assemble protein cages with different geometries (Figure 1).

Previously,[4a] we selected a trimeric esterase (PDB ID: 1ZOI),[6] as a C_3-symmetric protein building block for the assembly of an octahedral cage assembled with a tetrameric coiled-coil. This protein was chosen because each subunit comprises a single domain and the C-terminus is located near the apex of the triangle formed by the three subunits - this should minimize possible steric interference associated with assembling the trimeric subunits around the coiled-coil domain. Other than these, no other structural criteria were applied in the selection of the protein.

In this study, we aimed to investigate whether the octahedral cage could be converted to a tetrahedral cage by replacing the C_4 symmetric 4-stranded coiled-coil domain with a C_3 symmetric 3-stranded coiled-coil. We selected a crystallographically characterized *de novo*-designed trimeric, parallel coiled-coil (PDB ID: 4DZL)[5] for the purpose. This coiled-coil comprises 4 repeating heptads of sequence IAAIKQE, with the trimer structure being principally maintained by Ile residues at the hydrophobic ‘a’ and ‘d’ positions.[5]

To estimate the minimal length of the linker needed to connect the C-terminus of the esterase to the N-terminus of the coiled-coil, the C_3 axes of the esterase and coiled-coil were computationally aligned along the C_3 axes defining the tetrahedral point group. Using an algorithm implemented in Rosetta,[7] the distance from the origin of each protein and its angle of rotation about its symmetry axis were allowed to vary whilst maintaining tetrahedral symmetry (Figure S1).[4a] Configurations with inter-subunit backbone atom distances shorter than 4 Å, representing steric clashes, were discarded. In this manner, the minimum distance between the N- and C-termini was estimated as ~ 12 Å, a distance that could be spanned by a 4-residue Gly linker.

Based on our model, a gene encoding a fusion protein in which the C-terminus of the esterase was linked to the N-terminus of the trimeric coiled-coil domain through a 4-Gly linker sequence was assembled by standard methods (Figure S2). This protein, designated Tet4, was overexpressed in *E. coli* BL21 and purified to homogeneity using an N-terminal His-tag. Tet4 expressed as soluble protein, however initial analysis by native PAGE and size exclusion chromatography (SEC) indicated that it formed an ensemble of oligomeric species with M_r > 600 kDa (Figure 2), significantly larger than the M_r of ~ 430 kDa expected for a tetrahedral cage.

As it seemed that a 4-Gly linker might be too short to allow the correct protein-protein interactions to form, we increased the length of the linker to 6 and 8 Gly residues to give constructs Tet6 and Tet8. Analysis by SEC and native PAGE indicated that a complex of the correct molecular weight was formed by Tet8. SEC-purified Tet8 was further analyzed by negative stain electron microscopy (EM) and native mass spectrometry (native-MS) which indicated that complexes of the expected morphology and molecular weight were being formed (Figure S3). However, the yield of correctly assembled protein cages was only ~ 25 %, as judged by native PAGE, and further variation of the linker length provided no additional selectivity for tetrahedral complexes.

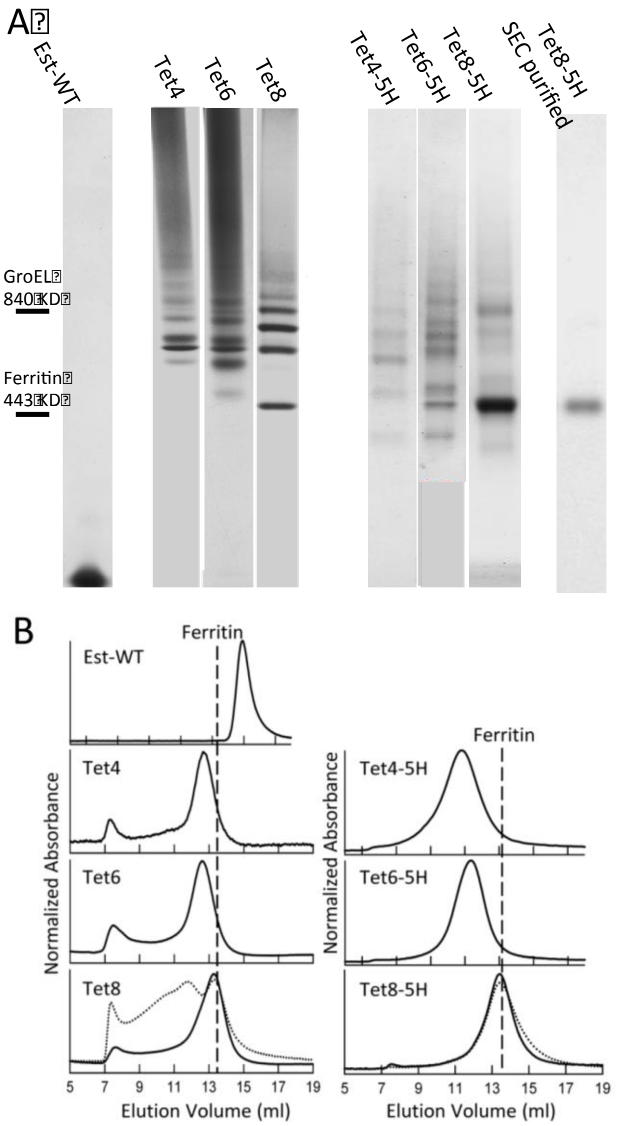


Figure 2. Screening Tet constructs with different linker lengths and coiled-coil strengths for their ability to assemble into cages. A) Characterization by native-PAGE (migration of standard proteins indicated on left; Est-WT: parent esterase) B) SEC of Tet constructs. The dotted traces represent the elution profiles of Tet8 and Tet8-5H after 4 months storage at 4°C.

To optimize the design, we investigated the effect of strengthening the coiled-coil interaction by increasing the number of IAAIKQE heptad repeats from 4 to 5. We synthesized constructs in which the trimeric esterase was fused to the 5-heptad coiled-coil through a 4-, 6-, 8- or 10-Gly residue linker, which are referred to as Tet4-5H, Tet6-5H, Tet8-5H, and Tet10-5H, respectively. Analysis of the Tet4-5H and Tet6-5H constructs by native PAGE indicated that they formed a mixture of oligomerization states with a similar size distribution to Tet4 and Tet6. However, Tet8-5H was far more homogenous (Figure 2). It comprised mainly one oligomerization state after nickel-affinity chromatography and could be obtained in high yields of 10-20 mg purified protein per liter of culture. Further purification by SEC yielded material with the approximate M_r expected for a tetrahedron that was nearly homogenous as judged by native PAGE (Figure 2). Surprisingly, whereas Tet4-5H, Tet6-5H and Tet8-5H were stable for months at 4°C, Tet10-5H spontaneously formed insoluble aggregates in a matter of days.

Based on the promising characteristics of Tet8-5H, its structure was further analyzed by analytical ultracentrifugation (AUC), native MS, negative stain EM and cryo-EM.

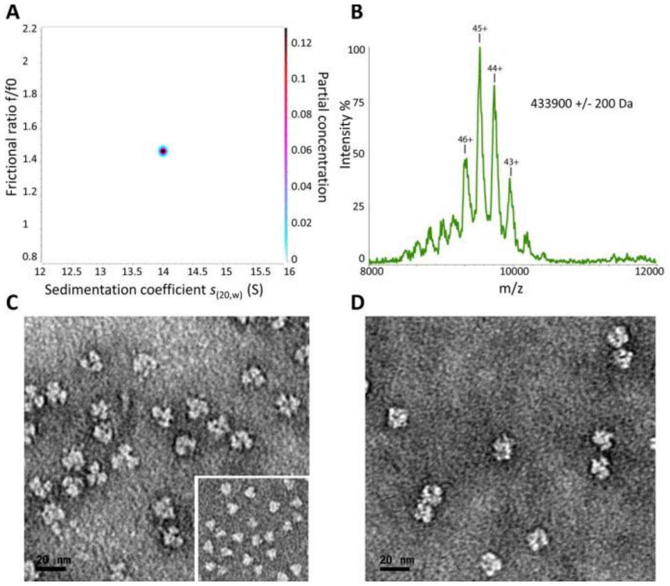


Figure 3. Characterization of Tet8-5H. (A) Analytical ultracentrifugation, (B) Native MS. (C) Negative stain EM of Tet8-5H before and (D) after crosslinking. Inset: TEM of wild-type trimeric esterase.

Sedimentation velocity AUC of Tet8-5H was performed at 39,000 rpm and the sedimentation traces subjected to 2D sedimentation spectrum analysis[8] (2DSA) which fits the velocity traces to the Lamm equation. This technique determines sedimentation coefficient, s, and frictional ratio f/f_0 of sedimenting species. The 2DSA plot of Tet8-5H (Figure 3A) shows a single hydrodynamic species with sedimentation coefficient, $s_{20,w}$ = 14 S and frictional ratio f/f_0 = 1.5 (Table S1). The frictional ratio is within the range measured for proteins such as ferritin, f/f_0 = 1.3,[9] and the E2 complex of pyruvate dehydrogenase, f/f_0 = 2.5 which also adopt porous cage-like structures.[10] The M_r of Tet8-5H calculated from these data was 430 kDa, closely matching the expected molecular weight of 439 kDa for a tetrahedral assembly comprising 12 protein subunits.

As a complementary technique to determine the oligomerization state of Tet8-5H we examined its structure by native ESI MS.[11] This technique uses very mild ionization and desolvation conditions that preserve non-covalent protein-protein interaction in the gas phase, thereby allowing subunit composition to be determined. The native MS spectrum of Tet8-5H (Figure 3B) exhibited an envelope of well-resolved peaks with m/z ranging from 8,000 to 10,000. Deconvolution of this spectrum gave a mass of 434 ± 0.2 kDa for Tet8-5H, in good agreement the 12-subunit assembly expected for a tetrahedron.

Negative stain EM of Tet8-5H clearly showed particles with the 3-fold symmetry characteristic of a tetrahedral structure (Figure 3C), although many of the particles had collapsed and splayed out in the carbon support during sample preparation. Initial cryo-EM images indicated that freezing disrupted the particles, a common problem with cryo-EM sample preparation.[12] Thus, samples of Tet8-5H were first reacted with a lysine-specific cross-linking agent to rigidify the particles (Figures 3D and S4). The cross-linked Tet8-5H particles were flash-frozen and visualized by cryo-EM. Particles were selected automatically and subjected to reference-free alignment and classification.[13] The 2-D class averages revealed a variety of distinct views with well-resolved structural features (Figure S5). Many class averages showed thin projections emanating from the tetrahedral core that could be ascribed to the coiled-coil domains.

35,217 particles, extracted from fully assembled and well-defined class average images were used to reconstruct a 3D model of Tet8-5H (See SI for details). The final model, constrained by tetrahedral symmetry, had an indicated resolution of ~ 13 Å (Figure S6). The map clearly reveals the arrangement of trimeric esterase domains and peripheral electron density attributable to the coiled-coil domains. The crystal structures of the trimeric esterase and the *de novo* designed trimeric coiled-coil could be modeled into the electron density map as shown in Figure 4. The resolution of the map did not allow us to unambiguously determine which face of the trimer faces outwards. The correlation coefficients between simulated map and EM density map, determined using the Chimera “fit in map” tool, yield essentially the same value (0.6355 for the current docking, and 0.6346 for the alternative face docking). However, in one orientation the 6xHis tag used to purify the protein is sequestered within the cage, whereas in the other it points outward. Given that the protein bound tightly to the Ni-NTA column used to purify it, we consider the latter orientation to be more likely and this determined our choice of model.

With the structures modeled into the electron density, the distance between the C-terminus of the esterase and N-terminus of the coiled-coil is ~25 Å (This distance was similar with the trimer docked in either orientation). This distance is slightly longer than could be spanned by the 8-Gly linker sequence, but given the resolution of the model and possible structural reorganization of the terminal residue, the modeled geometry appears reasonable.

The esterase activity of purified Tet8-5H, with p-nitrophenol acetate as substrate, was found to be very similar to the parent esterase enzyme, and both the parent esterase and Tet8-5H showed similar thermal stability profiles, with T_m ~78 °C (Figures S7 and S8). These observations indicate that the structure of the trimeric esterase is not significantly perturbed by the addition of the coiled-coil domain and subsequent assembly of the protein into a cage.

Our results demonstrate that it is possibly to apply the principles of symmetry to direct the assembly of a generic trimeric protein into large-scale geometrical structures *without explicitly specifying the orientation between symmetry axes*.

Thus, the same protein may be assembled into either octahedral or tetrahedral cages, dependent simply on the choice of the coiled-coil domain. These results provide a striking demonstration of the utility of coiled-coils as modular components for protein assembly.

To achieve the desired tetrahedral assembly, it was necessary to optimize both the Gly-linker length and the coiled-coil strength. In particular, whereas a 4-heptad coiled-coil was satisfactory for assembling an octahedral cage,[4a] a 5-heptad coiled-coil was needed to achieve good yields of the tetrahedral cage. This may reflect the smaller hydrophobic core associated with the trimeric as opposed to the tetrameric coiled-coil design.

Also noteworthy is the sensitivity of the assembly process to the Gly linker length. In this case an 8-Gly linker gave a much more homogenous preparation than a 6-Gly or 4-Gly linker, whereas a 10-Gly linker resulted in an assembly that was unstable and prone to aggregation. The factors that determine the optimal linker length remain poorly understood. Presumably, a linker that is too short results in unfavorable steric interactions between the large esterase domains preventing the protein assembling as intended, whereas an overly long linker may allow coiled-coils attached to esterase subunits in the same trimer to self-associate.

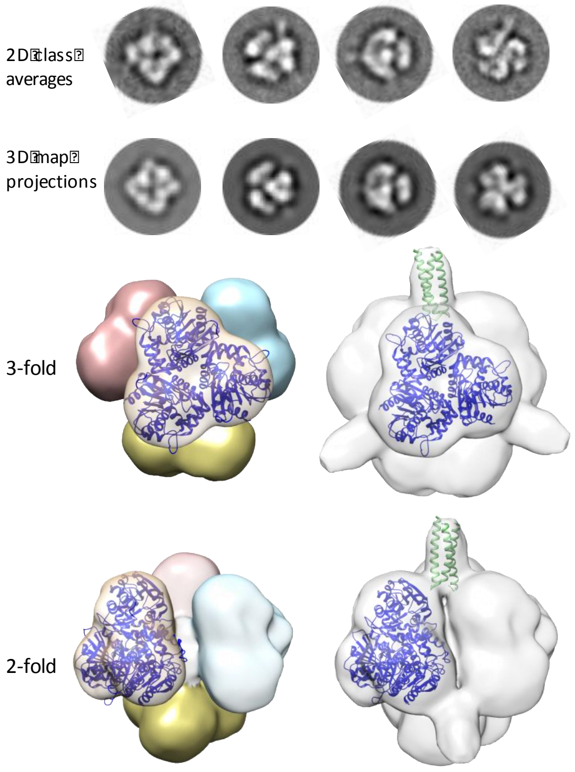


Figure 4. *Top*: Representative 2D class-averaged images of Tet8-5H and re-projections generated from the 3D electron density map. *Bottom:* Reconstructed electron density for Tet8-5H viewed along the threefold and twofold axes with one esterase trimer modeled into the electron density., At a lower threshold (grey) the 3D density map shows density for the coiled-coils allowing them to be docked into the map.

These results suggest that other trimeric proteins could be assembled into octahedral and tetrahedral cages by selecting the appropriate coiled-coil and sampling a fairly sparse matrix of coiled-coil and linker lengths to optimize the assembly process. Furthermore, this approach could be extended to assemble proteins into a wider range of geometrical structures by selecting protein building blocks and coiled-coils with different combinations of rotational symmetries. For example, by combining of C_4 and C_5 systemic proteins with trimeric coiled-coils, cubic and dodecahedral constructs could in principle be formed.

Acknowledgements

This work was supported by grants from the Army Research Office W911NF-11-1-0251 and W911NF-16-1-0147 to E.N.G.M. J.C.A.B acknowledges support from the Howard Hughes Foundation. Supercomputer resources at the Texas Advanced Computing Center, and the San Diego Supercomputer Center were used to analyze AUC data.

Keywords: protein design • self-assembly • Coiled-coil • symmetry • protein cages

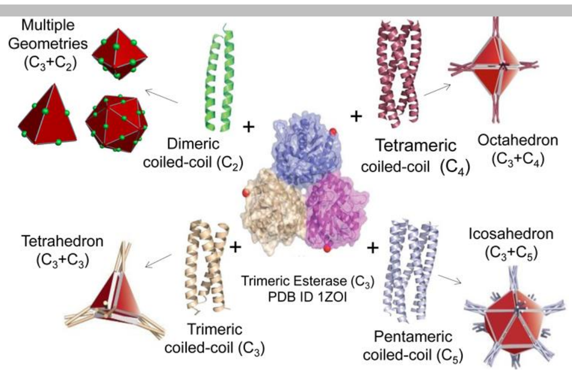
[1] a T. O. Yeates, M. C. Thompson, T. A. Bobik, *Curr. Opin. Struct. Biol.* **2011**, *21*, 223-231; b J. A. Marsh, S. A. Teichmann, *Annu. Rev. Biochem.* **2015**, *84*, 551-575.
[2] a J. B. Bale, S. Gonen, Y. Liu, W. Sheffler, D. Ellis, C. Thomas, D. Cascio, T. O. Yeates, T. Gonen, N. P. King, D. Baker, *Science* **2016**, *353*, 389-394; b Y. Hsia, J. B. Bale, S. Gonen, D. Shi, W. Sheffler, K. K. Fong, U. Nattermann, C. Xu, P. S. Huang, R. Ravichandran, S. Yi, T. N. Davis, T. Gonen, N. P. King, D. Baker, *Nature* **2016**, *535*, 136-139; c Y. T. Lai, G. L. Hura, K. N. Dyer, H. Y. Tang, J. A. Tainer, T. O. Yeates, *Science Adv.* **2016**, *2*, e1501855; d T. O. Yeates, Y. Liu, J. Laniado, *Curr. Opin. Struct. Biol.* **2016**, *39*, 134-143; e J. D. Brodin, X. I. Ambroggio, C. Tang, K. N. Parent, T. S. Baker, F. A. Tezcan, *Nat. Chem.* **2012**, *4*, 375-382; f J. D. Brodin, S. J. Smith, J. R. Carr, F. A. Tezcan, *J. Am. Chem. Soc.* **2015**, *137*, 10468-10471.
[3] a P. S. Huang, S. E. Boyken, D. Baker, *Nature* **2016**, *537*, 320-327; b C. H. Norn, I. Andre, *Curr. Opin. Struct. Biol.* **2016**, *39*, 39-45; c Y. T. Lai, E. Reading, G. L. Hura, K. L. Tsai, A. Laganowsky, F. J. Asturias, J. A. Tainer, C. V. Robinson, T. O. Yeates, *Nat. Chem.* **2014**, *6*, 1065-1071.
[4] a A. Sciore, M. Su, P. Koldewey, J. D. Eschweiler, K. A. Diffley, B. M. Linhares, B. T. Ruotolo, J. C. Bardwell, G. Skiniotis, E. N. G. Marsh, *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 8681-8686; b D. P. Patterson, M. Su, T. M. Franzmann, A. Sciore, G. Skiniotis, E. N. G. Marsh, *Protein Sci.* **2014**, *23*, 190-199; c D. P. Patterson, A. M. Desai, M. M. B. Holl, E. N. G. Marsh, *RSC Advances* **2011**, *1*, 1004-1012; d A.S. Cristie-David, A. Sciore, S. Badieyan, J.D. Eschweiler, P. Koldewey, B.T. Ruotolo, J.C.A. Bardwell, and E.N.G. Marsh (2017) *Molec. Syst. Design Engin.* **2,** 140 – 148; e F. B. L. Cougnon, *ChemBioChem* **2016**, *17*, 2296-2298.
[5] J. M. Fletcher, A. L. Boyle, M. Bruning, G. J. Bartlett, T. L. Vincent, N. R. Zaccai, C. T. Armstrong, E. H. C. Bromley, P. J. Booth, R. L. Brady, A. R. Thomson, D. N. Woolfson, *ACS Synth. Biol.* **2012**, *1*, 240-250.
[6] F. Elmi, H. T. Lee, J. Y. Huang, Y. C. Hsieh, Y. L. Wang, Y. J. Chen, S. Y. Shaw, C. J. Chen, *J. Bacteriol.* **2005**, *187*, 8470-8476.
[7] R. Das, D. Baker, *Ann. Rev. Biochem., Vol. 77*, **2008**, pp. 363-382.
[8] a B. Demeler, H. Saber, J. C. Hansen, *Biophys. J.* **1997**, *72*, 397-407; b B. Demeler, K. E. van Holde, *Anal.l Biochem.* **2004**, *335*, 279-288.
[9] G. Jutz, P. van Rijn, B. S. Miranda, A. Boeker, *Chem. Rev.* **2015**, *115*, 1653-1701.
[10] H. J. Bosma, A. Dekok, B. W. Vanmarkwijk, C. Veeger, *Eur. J. Biochem.* **1984**, *140*, 273-280.
[11] B. T. Ruotolo, J. L. Benesch, A. M. Sandercock, S. J. Hyung, C. V. Robinson, *Nat. Protocols* **2008**, *3*, 1139-1152.
[12] R. F. Thompson, M. Walker, C. A. Siebert, S. P. Muench, N. A. Ranson, *Methods* **2016**, *100*, 3-15.

[13] S. H. Scheres, *J. Struct. Biol.* **2012**, *180*, 519-530.

Entry for the Table of Contents

COMMUNICATION

A modular approach for assembling proteins into large-scale geometric structures was developed in which coiled-coil domains act as “twist ties” facilitate assembly. The geometry of the cage is specified primarily by the rotational symmetries of the coiled-coil and building-block protein and is largely independent of protein structural details.



*Somayesadat Badieyan, Aaron Sciore, Joseph D. Eschweiler, Philipp Koldewey, Ajitha S. Cristie-David, Brandon T. Ruotolo, James C. A. Bardwell, Min Su and E. Neil G. Marsh**

Page No. – Page No.

Symmetry-directed Self-assembly of a Tetrahedral Protein Cage Mediated by *De Novo*-designed Coiled-Coils