

LJUBETIČ, A., LAPENTA, F., GRADIŠAR, H., DROBNAK, I., AUPIČ, J.,
STRMŠEK, Ž., LAINŠČEK, D., HAFNER BRATKOVIČ, I., MAJERLE, A.,
KRIVEC, N., BENČINA, M., PISANSKI, T., ČIRKOVIĆ-VELIČKOVIĆ, T.,
ROUND, A., CARAZO, J. M., MELERO, R., JERALA, R.

Design of coiled-coil protein-origami cages that self-assemble *in vitro* and *in vivo*

Nature biotechnology, ISSN 1087-0156, 2017, 35, str. 1094-1101

Author's Accepted Manuscript

<http://www.nature.com/articles/nbt.3994>

<https://doi.org/10.1038/nbt.3994>

Published (AOP): 16th October 2017

Design of coiled-coil protein origami cages that self-assemble *in vitro* and *in vivo*

Ajasja Ljubetič^{*1}, Fabio Lapenta^{*1,2}, Helena Gradišar^{1,3}, Igor Drobnak¹, Jana Aupič^{1,4}, Žiga Strmšek¹, Duško Lainšček¹, Iva Hafner-Bratkovič^{1,3}, Andreja Majerle¹, Nuša Krivec¹, Mojca Benčina¹, Tomaž Pisanski⁵, Tanja Čirković Veličković⁶, Adam Round⁷, José María Carazo⁸, Roberto Melero⁸ and Roman Jerala^{1,3†}

¹ Department of Synthetic Biology and Immunology, National Institute of Chemistry, Ljubljana, Slovenia.

² Graduate School of Biomedicine, University of Ljubljana, Ljubljana, Slovenia.

³ EN-FIST Centre of Excellence, Ljubljana, Slovenia.

⁴ Doctoral Study Programme in Chemical Sciences, University of Ljubljana, Ljubljana, Slovenia.

⁵ University of Primorska, Koper, Slovenia and IMF, Ljubljana, Slovenia.

⁶ Center of Excellence for Molecular Food Sciences, Faculty of Chemistry, University of Belgrade, Serbia.

⁷ The European Synchrotron Radiation Facility (ESRF), Grenoble, France.

⁸ Centro Nacional de Biotecnología (CNB-CSIC), Madrid, Spain.

† Correspondence to: roman.jerala@ki.si; tel. +38614760335; fax +38614760300;

National Institute of Chemistry, Hajdrihova 19, SI-1000 Ljubljana, Slovenia.

* Shared equal contribution.

Polypeptides and polynucleotides are natural programmable biopolymers that can self-assemble into complex tertiary structures. We describe a system analogous to designed DNA nanostructures in which protein coiled-coil (CC) dimers serve as building blocks for modular *de novo* design of polyhedral protein cages that efficiently self-assemble *in vitro* and *in vivo*. We produced and characterized more than 20 single-chain protein cages in three shapes—tetrahedron, four-sided pyramid, and triangular prism—with the largest containing >700 amino-acid residues with a hydrodynamic size of 11 nm. Their stability and folding kinetics were similar to those of natural proteins. Solution small-angle X-ray scattering (SAXS), electron microscopy (EM), and biophysical analysis confirmed agreement of the expressed structures with the designs. We also demonstrated self-assembly of a tetrahedral structure in bacteria, mammalian cells, and mice without evidence of inflammation. A semi-automated computational design platform and a toolbox of CC building modules are provided to enable the design of protein cages in any polyhedral shape.

Modular engineering of DNA has demonstrated the construction of complex *de novo* two- and three-dimensional assemblies¹⁻³ and DNA polyhedral meshes^{4,5}. However, proteins, which self-assemble into complex folds and perform a wide range of functions, could provide more versatile programmable materials. Proteins provide the diverse chemical functionalities of amino acid side-chains, self-assemble rapidly under normal physiological conditions⁶, and can be efficiently and sustainably produced *in vivo*⁷. Only a fraction of all possible protein folds are found in nature⁸, and *de novo* design could provide many new folds but remains challenging. Computational approaches have produced increasingly complex assemblies, predominantly based on engineering of the interaction surfaces of natural protein

domains, remodeling of protein hydrophobic core^{9–12} and on the design of repetitive folds¹¹.

We sought to combine the modular building strategy of DNA nanotechnology with the chemical versatility and functionality of amino acids. Among the various known protein structural motifs, CC dimers provide the best approximation of DNA building blocks because of similar elongated shape and pairwise complementarity that can be designed. The pairing specificity of CC dimers is defined by the combination of hydrophobic and electrostatic interactions between residues at the heptad repeat positions *a:d* and *e:g*, respectively (Fig. 1A). Unlike DNA duplex, CC dimers can interact in either an antiparallel or a parallel orientation (Fig. 1B)¹³. This property is advantageous as it allows construction of more complex folds¹⁴. CC dimers can be used as modular building blocks to form the edges of polyhedral cages¹⁴. According to this strategy, CC-dimer-forming segments are concatenated into a single polypeptide chain via short flexible peptide linkers that act as hinges. This strategy is a molecular embodiment of a mathematical double Eulerian tour: an oriented graph that traverses each edge exactly twice¹⁴. The folding of modular CC protein origami (CCPO) structures is defined by long-range interactions—that is, inter-module contacts that are far apart in the primary structure. In contrast, engineered repeat proteins are based on short- and medium-range interactions between neighboring modules^{11,15–17}. The possibility for long-range interactions suggests that CCPO enables the design of complex shapes not achievable by repeat proteins. However, self-assembly of CCPO structures under physiological conditions may be challenging. Our first-generation CC tetrahedron had to be slowly refolded *in vitro* by dialysis through slow annealing at low concentrations and was poorly soluble¹⁴. Similarly, DNA nanostructures are typically assembled slowly *in vitro* from a large number of

chains. The requirement for *in vitro* refolding significantly limits the potential applications of bionanotechnology¹⁸. CCPO structures are absent in natural proteins, and whether they can self-assemble *in vivo* has been unknown. Our recent finding that a single-chain DNA molecule can rapidly fold into a highly knotted pyramid¹⁹ suggested the possibility of designing modular proteins that self-assemble under physiological conditions.

Here we present second-generation CCPO structures that can be designed in any polyhedral shape, are soluble, and self-assemble *in vitro* and *in vivo*. First, we provide supercharged CC building-modules that have increased number of charged residues at non-interacting positions. (Supplementary Table 1) We investigated the design of a range of modules and linkers in tetrahedral variants with alternative topologies and the design and characterization of new protein polyhedra that contain a four-branched vertex (rectangular pyramid) and a triangular prism. The agreement of the designed folds with experiment was confirmed by solution small-angle X-ray scattering (SAXS) analysis, electron microscopy, and crosslinking analysis. The stability and folding kinetics of the second-generation CCPO structures were found to be similar to those of natural proteins although they lack a compact hydrophobic core. Extensive characterization of a tetrahedral structure expressed in living animals demonstrated biocompatibility as it folded correctly and did not trigger a stress response. Our analysis of different CCPO structures provides rules for a Coiled-coil protein origami design platform (CoCoPOD; see github.com/NIC-SBI/CC_protein_origami and **Supplementary Source Code**) which uses a toolbox of pre-characterized CC modules to assemble arbitrary polyhedral cages.

Results

We recently showed that the stability of CC dimers can be modified without affecting pairing specificity by varying surface residues²⁰. As high stability of CC dimer modules might lead to folding bottlenecks, we sought to decrease their stability by introducing acidic residues at positions *b*, *c*, and *f* of the CC heptad repeats (Fig. 1A), while leaving the interacting residues unchanged (Fig. 1B; Table S1).

Supercharged CC modules, with negatively charged non-interacting residues at positions *b*, *c* and *f* of the heptad repeat conserved the pairing specificity and orthogonality of the CC modules while decreasing their stability (Fig. S1). The nomenclature used to designate the different polyhedral cages generated in this study is shown in **Fig 1C**.

Tetrahedra with various topologies from supercharged modules

We tested whether supercharged CC modules facilitate *in vivo* production of soluble monomeric CCPO tetrahedra. We evaluated 12 tetrahedral designs with different (i) topologies, (ii) circular permutations, (iii) length and sequence of linkers, and (iv) CC modules (Tables S2, S3). Eight designs used soluble (S), negatively charged CC modules and four designs used soluble, negatively supercharged (SN) CC modules (Table S2), differing in the number of charged residues at non-interacting CC positions. A representative negatively charged version (TET12S) had a –33 net negative charge, and a representative negatively supercharged variant (TET12SN) had a –47 net negative charge (Fig. 2), both with five-amino-acid-long flexible linkers. Expression of 12 tetrahedral variants in bacteria demonstrated that all designs were produced as soluble proteins (Fig. S2A), in contrast to the almost completely insoluble first-generation TET12, having the net charge of -9. (ref.14; Fig. S2B). Supercharged modules enabled the isolation of tetrahedral monomers under native

conditions without the need for refolding (Table S4). In addition to bacterial expression, TET12S was produced in soluble form by an *in vitro* transcription/translation system (Fig. S3).

The purified TET12SN protein displayed helical content of 86% in agreement with the predicted 88% according to the design (Fig. 2D; Fig. S4). The protein's hydrodynamic radius, as determined by dynamic light scattering (Fig. S5) was in agreement with the molecular model (Table S4). Size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) analysis (Fig. S5) showed the sample to be highly monodisperse, with up to 99% of the protein eluting in the monomeric fraction, with a calculated M_r in close agreement with the expected molecular weight (Table S4). The correct pairing of the CC modules in the folded state was verified by cross-linking using chemical cross-linker reagents of different lengths, followed by proteolytic digestion and mass spectrometric analysis of the cross-linked peptides²¹. All digested cross-linked peptide pairs between CC segments that were far apart in the primary structure were consistent with the design (Fig. S6; Tables S5–S6).

Analysis of the thermal unfolding of TET12SN indicated several transitions due to the range of stabilities of the CC modules (Fig. 2E). Despite the use of destabilized supercharged CC modules, TET12SN was stable up to 40 °C. Folding was reversible, and the protein remained predominantly monomeric even after rapid temperature quenching (Fig. S5), similar to several other supercharged proteins²². Even though CCPO cages do not have a compact hydrophobic core and are stabilized by long distance interactions between CC modules, their relative contact order (RCO) is lower than that of globular proteins. While the contacts between residues at the *a* and *d* heptad positions of the interacting CC modules are indeed far apart in the

primary sequence, the RCO is the average distance in the primary sequence between contacting residues. In CCPO structures, the RCO is dominated by the short-range ($i+3$ and $i+4$) contacts of the α -helices (Fig. 2C; Fig. S7).

Next we investigated the folding kinetics of TET12SN to determine the effect of the absence of a compact hydrophobic core. The rate of secondary-structure formation, measured by stopped-flow circular dichroism (CD), was similar to that of natural globular proteins of similar sizes²³ (Fig. S8A). Formation of a defined tertiary-structure contact imposed by the CCPO fold was assessed by the rate of fluorescence resonance energy transfer (FRET) increase between the fluorescently labeled N and C termini of the polypeptide chain. This interaction would be unlikely to occur in the absence of the predicted folding as the first and last CC-forming segment do not interact with each other but only come into proximity due to topological constraints. The observed increase in FRET signal (Fig. S8B) indicates that the correct designed structure is formed by the expressed construct. The rate of secondary structure formation (Fig. S8A) is similar to overall structure formation (Fig. S8B) as the CC segments only form well-defined structure upon forming a pair.

For the other designed tetrahedral variants, we varied the length and stability of the constituent CC modules and the length and sequence of the linkers connecting them (Table S2). The linkers varied from five to nine amino-acid residues and comprised charged and small flexible residues. The lengths of the CC dimer-defined edges differed by up to 60% — the longest, APH, had 45 residues, while the shortest, GCNshSN, had 27 residues. This wide range confirmed the ability to deviate at will from ideal symmetric tetrahedra. Because CCPO designs do not require symmetry, the shape and size of the cage and the cage's internal cavity can be tuned; in addition, stability can be tuned by the selection of the CC modules (Table S1).

Mathematical graph analysis has proved that a single-chain tetrahedron can be constructed either from four parallel and two antiparallel or from three parallel and three antiparallel CC segments²⁴. In addition to the 4+2 (parallel + antiparallel dimer building module) topology we therefore designed a new tetrahedron with the 3+3 topology based on the supercharged versions (Table S1)²⁵. Analysis of the produced and purified 3+3 tetrahedral fold (TET12_{2,3}SN-f₅) revealed that the fold was very similar to the TET12SN variant with respect to helical content and hydrodynamic size (Fig. 2). More than 90% of the tetrahedral variants folded successfully, as ascertained by CD (Fig. S4), SEC-MALS, dynamic light scattering (Fig. S5), and SAXS analysis (Figs. S9-S10). In contrast, the variant with a scrambled order of the same CC modules (TET12SScr), which by design cannot fold into a tetrahedral shape exhibited heterogeneous assemblies with increased hydrodynamic radius (Fig. S11; design of TET12SScr in Table S2). The high success rate demonstrated the robustness of the CCPO design strategy.

Comparative biophysical characterization of the tetrahedral variants revealed several design principles. One is the importance of the N-terminal capping residues of the CC modules. The design without the tetrapeptide capping sequence SPED (TET12_{1,10}XS-f₆) was unable to refold after thermal denaturation, in contrast to the TET12_{1,10}S-f₅ design, where the SPED capping sequence was retained (Fig. S5). The N-terminal capping sequence stabilizes the α -helix through a restricted dihedral angle, side-chain hydrogen bonding, and a favorable dipole moment²⁶ as well as by increasing the helical propensity to stabilize the CC dimers. However, a lower net charge may also contribute to the poor solubility of TET12_{1,10}XS-f₆, since a SPED sequence contributes two charged residues.

Second, variations of the linker sequences, ranging from five to nine amino-acid residues, and variations in polarity and charges of residues in the linker, had a minor effect on the final yield of folded tetrahedra (compare TET12_{1.10}SN-f₅, TET12_{1.10}SN-c₆, and TET12_{1.10}SN-f₉ in Fig. S2A) and no effect on stability (Fig. S4).

Third, placement of the least-stable CC modules at the termini of the polypeptide chain considerably decreased the helical content of those variants, most likely because such structures are more prone to fraying at the termini (TET12_{1.10}S-f₅ versus TET12_{1.6}S-f_{5b} in Fig. S4) and should therefore be avoided.

Computational platform for the design of complex polyhedra

To automate the design of CCPO structures, we encoded the design principles into a design platform (CoCoPOD; Fig. 3). CoCoPOD builds 3D models constructed from a single polypeptide chain and provides amino-acid sequences for any arbitrary polyhedral scaffold, where the polyhedron edges are formed from CC dimers. The design strategy consists of several mostly automated steps (Fig. 3): (i) specifying the polyhedral geometry; (ii) routing the chain according to topological constraints; (iii) selecting the optimal topology and circular permutation; (iv) selecting the modules from the CC toolkit; (v) building the 3D model; and (vi) refining and validating the model through folding simulations.

All topologies (i.e., double Eulerian paths) for the selected geometric scaffold are derived via graph analysis²⁷. The optimal topology and circular permutation are selected based on the lowest topological contact order, the orientation of the modules (parallel and antiparallel CC dimers), and the crossing of chains at vertices. The modules are selected from the CC toolbox (Table S1) based on their orthogonality, stability, and other desired properties defined by the user. Molecular models are built

for the designed polypeptide chain, and any folding bottlenecks are identified by the G \bar{o} -type coarse-grained folding simulation, previously used to identify the cooperative folding units of natural proteins^{28,29}.

Building a four-sided pyramid and triangular prism

To demonstrate that the CoCoPOD platform can be used for the construction of larger and more-complex polyhedra than tetrahedron, we designed, isolated, and characterized a four-sided pyramid (PYR16SN) and a triangular prism (TRIP18SN) (Fig. 4). PYR16SN is composed of two antiparallel and six parallel modules and contains a four-branched junction in the vertex that has never been observed in any polypeptide structure in the nature or in engineered protein, to our knowledge. The triangular prism TRIP18SN, comprising >700 residues, is composed of three antiparallel and six parallel CC dimers defined by 18 concatenated CC-forming segments. Both proteins were successfully produced in the soluble fraction of bacteria (Fig. S2). The helical content (Fig. 4D) and hydrodynamic size of both were in agreement with the design determined by CD and SEC-MALS analysis (Fig. S5; Table S4). The thermal-unfolding profile suggested higher cooperativity than in the tetrahedral designs (Fig. 4E). The folding rates for the larger protein polyhedra were found to be slightly slower (Fig. S8) than for the tetrahedral designs, as expected by these polyhedrons' increased complexity.

Structural analysis of designed polyhedra

The shapes of a tetrahedron, four-sided pyramid and triangular prism were analyzed by SAXS, which confirmed that the structures matched the design (Figs. S9-S10, S12-S13; Table S7; see Supplementary Discussion). The molecular models displayed very good agreement with the experimental data, with χ values of 0.8, 2.7 and 1.0 for

TET12SN, PYR16SN and TRIP18SN, respectively (Fig. 5). After creating the models with CoCoPOD, normal mode analysis was used to extensively sample the molecular model space to identify the best matching models. Normalized Kratky plots from the SAXS measurements confirmed that the proteins were structured, while the Porod exponents indicated a limited flexibility due to flexible linkers, which form the vertices (Figs. S9, S13; Table S7; see Supplementary Discussion).

Ab initio calculation of the molecular envelope from the SAXS data revealed that the shapes of all three designed polyhedra were in agreement with the design, with clearly visible internal cavities (Fig. 5D). The scattering curve for the triangular prism may best be explained by a contribution of a rectangular (26%) and an oblique prism (74%). Triangular prism comprises three four-edged and two three-edged faces. Because the angles between the edges of a polyhedron with a four-sided face are not uniquely defined by the length of its edges, this conformational variability was not unexpected.

To further confirm the size and shape determined by SAXS results, particles were imaged by negative-stain transmission electron microscopy (TEM micrographs on Figs. S14-S16; Single particle class averages on Figs. S17-S19). Reconstructed 3D electron density maps for a tetrahedron, rectangular pyramid and triangular prism were also found to be in good agreement with the models and SAXS results (Fig. 5E), corroborating the success of the design strategy.

CCPO structures fold in mammalian cells and in mice

To investigate the potential of CCPO cages for delivery of drugs and vaccines and other health applications, we examined folding of a tetrahedron in mammalian cells and *in vivo*. Subdomains of a split-fluorescent protein were fused to both termini of

TET12S (TET12Ssplit-mVenus) to enable reconstitution of the fluorescence. Fluorescence should occur only if the protein folds according to the design because the N and C termini must coincide at the same vertex, defined by the concurrent formation of three CC segments. Expression of the transfected TET12Ssplit-mVenus plasmid in the HEK293 cell line exhibited high fluorescence, in contrast to the control, which used a scrambled order of modules (TET12SScrsplit-mVenus, Fig. 6A) where both proteins were expressed at similar levels (Fig. S20).

Because eukaryotic cells recognize and respond to misfolded proteins, intracellular production of non-natural protein folds can trigger adverse physiological responses. However, we observed no activation of the unfolded protein response³⁰ in cells that overexpressed TET12S (Fig. S21A). Furthermore, while abnormal protein aggregates can induce an innate immune response via NLRP3 inflammasome activation in the cytosol, expression of TET12SN or TET12S did not activate the NLRP3 inflammasome in primed immortalized mouse macrophages, in contrast to misfolded prion protein amyloid fibrils (Fig. S21B).

To investigate folding *in vivo*, we introduced a plasmid encoding TET12S into the liver of mice by hydrodynamic delivery (Fig. 6B). Histological analysis of liver tissue demonstrated that the correctly folded TET12S was produced in hepatocytes, using reconstitution of a split-protein fluorescent as a reporter (Fig. 6C). Correct folding of TET12S was confirmed by *in vivo* imaging of a reconstituted split luciferase. Mice that produced TET12Ssplit-fLuc exhibited strong bioluminescence in the liver, in contrast to the mice that expressed a scrambled version (TET12SScrsplit-fLuc, Fig. 6D). No increase in inflammation or liver damage markers was observed in the mice (Fig. S22), suggesting that *in vivo* expression of the protein did not elicit

adverse pathological responses and could be readily integrated into biological systems.

Discussion

The CCPO fold, defined by modular interactions between orthogonal CC building blocks that are far apart in the primary sequence, has not been observed in natural proteins. We demonstrate the production of three different complex, nonsymmetrical polyhedra: a tetrahedron, a four-sided pyramid and a triangular prism. The agreement of the isolated protein particles with the design was confirmed by solution small-angle X-ray scattering (SAXS) analysis, electron microscopy, and crosslinking analysis. The stability and folding kinetics of the second-generation CCPO structures were found to be similar to those of natural proteins even though CCPO structures lack a compact hydrophobic core. Extensive characterization of a tetrahedral structure expressed in living animals demonstrated CCPO structures can be efficiently produced and are compatible with cellular physiology. Designed CCPO comprising >700 amino-acid residues were found to fold efficiently under physiological conditions, thus expanding the boundary of accessible designed-protein origami scaffolds.

Although the pairing specificity of CC dimers is not as straightforward as DNA base pairing, polypeptide modules have several far-reaching advantages over DNA nanotechnology such as DNA origami for biotechnology applications. Functional amino-acid residues can be readily incorporated, whereas DNA requires introduction of functional moieties or domains via chemical conjugation³¹.

Furthermore, unlike cytosolic DNA, cytosolic CCPO structures did not trigger an innate immune response in our tests.

Protein assemblies have been designed through genetic fusion of natural oligomerizing domains^{32,33}, novel protein-protein interfaces^{34–36}, and using viral capsids or bacterial microcompartments³⁷. Repeat proteins³⁸ also comprise compact building blocks with defined curvatures. Unlike these approaches, the CCPO strategy does not rely on symmetry, so any amino-acid position in the structure can in principle be individually addressed and functionalized. Reconstitution of the split-fluorescent protein and luciferase also shows that CCPO structures can be combined with natural globular protein domains.

Novel protein scaffolds represent a new class of versatile tools for applications in molecular therapy^{39,40}. Several approaches, such as virus-like particle engineering⁴¹ and computational epitope grafting⁴², have been applied previously to design scaffolded epitopes; one particularly attractive approach is the *in situ* production of genetically encoded vaccines, to present multiple antigens at the surface of designed protein nanoparticles that improved the immune response against influenza⁴³. We envision a diverse array of applications for CCPO structures, including drug delivery and targeting, epitope scaffolding for vaccination and their combination with agonists of innate immune response, compartmentalization and delivery of proteins including cytotoxic enzymes or therapeutic drugs. The approach should also find applications in non-therapeutic areas, such as rational design of protein molecular machines, sensors and functional materials.

Acknowledgments

This research was supported by the Slovenian Research Agency (program P4-0176, projects N4-0037 and J4-5528 (R.J.), L4-6812 (H.G.), and J3-7034 and BI-US/17-18-051 (M.B.); the ERANET SynBio project Bioorigami (ERASYNBIO1-006 to R.J.); COST actions CM1304 (A.L.) and CM1306 (J.A.); a grant from ICGEB (CRP/SLO14-03) to H.G.; ESRF, for making available its facility for performing SAXS measurements; MSCA ITN Tollerant (R.J.); iNEXT project PID1824 VID3978 - Structural Audit for the use of experimental resources (H.G.); NVIDIA Corporation for the donation of the Quadro GP100 GPU (J.M.C); and FP7 project FCUB ERA (GA No. 256716 to T.Č.V) for the use of the proteomics facility. We thank K. Djinović Carugo for performing and analyzing the initial SAXS experiments. We thank staff of Centre for Laboratory Animals at Biotechnical faculty of the University of Ljubljana, where animal experiments were performed. We would like to thank K. Butina, R. Bremšak, I. Škraba, D. Oven, T. Lončar, S. Božič Abram, T. Doles, S. Grudin, J. Mihailović and E. Žagar for their technical support. We also thank E. Žerovnik for granting access to the stopped-flow CD instrument and C. Wood for building preliminary models of the CC pairs.

Author contributions

A.L., F.L., H.G., I.D., J.A., Ž.S., and R.J. designed the CCPO variants. F.L., H.G., Ž.S and N.K. cloned, purified, and experimentally characterized the proteins. A.L., I.D., J.A., and T.P. wrote the CoCoPOD platform. A.M. and T.Ć.V. performed the cross-linking experiments. J.A. and A.R. performed the SAXS experiments and SAXS data analysis. I.H.B. and D.L. performed the experiments on the cells. M.B. performed confocal microscopy imaging. D.L. performed the animal experiments. R.M.R and J.M.C. performed the EM experiments and data processing. R.J. conceived the study, led the research and wrote the initial manuscript. All authors discussed the results and reviewed and contributed to the manuscript.

Competing financial interests

The authors declare no competing financial interests.

References

1. Douglas, S. M. *et al.* Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414–418 (2009).
2. Seeman, N. C. Nanomaterials based on DNA. *Annu Rev Biochem* **79**, 65–87 (2010).
3. He, Y. *et al.* Hierarchical self-assembly of DNA into symmetric supramolecular polyhedra. *Nature* **452**, 198–201 (2008).
4. Veneziano, R. *et al.* Designer nanoscale DNA assemblies programmed from the top down. *Science (80-.)*. **352**, 1534 (2016).
5. Benson, E. *et al.* DNA rendering of polyhedral meshes at the nanoscale. *Nature* **523**, 441–4 (2015).
6. Lupas, A. N. & Alva, V. Ribosomal proteins as documents of the transition from unstructured (poly)peptides to folded proteins. *J. Struct. Biol.* **198**, 74–81 (2017).
7. Demain, A. L. & Vaishnav, P. Production of recombinant proteins by microbes and higher organisms. *Biotechnol. Adv.* **27**, 297–306 (2009).
8. Taylor, W. R., Chelliah, V., Hollup, S. M., MacDonald, J. T. & Jonassen, I. Probing the ‘dark matter’ of protein fold space. *Structure* **17**, 1244–52 (2009).
9. Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364–8 (2003).
10. King, N. P. *et al.* Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **336**, 1171–4 (2012).
11. Doyle, L. *et al.* Rational design of α -helical tandem repeat proteins with closed architectures. *Nature* **528**, 585–8 (2015).
12. Regan, L. & Degrado, W. F. Characterization of a Helical Protein Designed

- from 1st Principles. *Science* (80-.). **241**, 976–8 (1988).
13. Woolfson, D. N. The Design of Coiled-Coil Structures and Assemblies. *Adv Protein Chem* **70**, 79–112 (2005).
 14. Gradišar, H. *et al.* Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. *Nat. Chem. Biol.* **9**, 362–366 (2013).
 15. Plückthun, A. Designed ankyrin repeat proteins (DARPs): binding proteins for research, diagnostics, and therapy. *Annu. Rev. Pharmacol. Toxicol.* **55**, 489–511 (2015).
 16. Grove, T. Z., Cortajarena, A. L. & Regan, L. Ligand binding by repeat proteins: natural and designed. *Curr. Opin. Struct. Biol.* **18**, 507–15 (2008).
 17. Bella, J., Hindle, K. L., McEwan, P. A. & Lovell, S. C. The leucine-rich repeat structure. *Cell. Mol. Life Sci.* **65**, 2307–33 (2008).
 18. Pinheiro, A. V., Han, D., Shih, W. M. & Yan, H. Challenges and opportunities for structural DNA nanotechnology. *Nat. Nanotechnol.* **6**, 763–72 (2011).
 19. Kočar, V. *et al.* Design principles for rapid folding of knotted DNA nanostructures. *Nat. Commun.* **7**, 10803 (2016).
 20. 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.* jacs.7b01690 (2017). doi:10.1021/jacs.7b01690
 21. Götze, M. *et al.* StavroX--a software for analyzing crosslinked products in protein interaction studies. *J. Am. Soc. Mass Spectrom.* **23**, 76–87 (2012).
 22. Lawrence, M. S., Phillips, K. J. & Liu, D. R. Supercharging proteins can impart unusual resilience. *J. Am. Chem. Soc.* **129**, 10110–2 (2007).
 23. Plaxco, K. W., Simons, K. T. & Baker, D. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**,

- 985–994 (1998).
24. Kočar, V. *et al.* TOPOFOLD, the designed modular biomolecular folds: polypeptide-based molecular origami nanostructures following the footsteps of DNA. *Wiley Interdiscip. Rev. Nanomedicine Nanobiotechnology* **7**, 218–237 (2014).
 25. Negron, C. & Keating, A. E. A Set of Computationally Designed Orthogonal Antiparallel Homodimers that Expands the Synthetic Coiled-Coil Toolkit. *J. Am. Chem. Soc.* **136**, 16544–16556 (2014).
 26. Doig, A. J. & Baldwin, R. L. N- and C-capping preferences for all 20 amino acids in alpha-helical peptides. *Protein Sci.* **4**, 1325–36 (1995).
 27. Fijavž, G., Pisanski, T. & Rus, J. Strong Traces Model of Self-Assembly Polypeptide Structures. *MATCH Commun. Math. Comput. Chem.* **71**, 199–212 (2014).
 28. Noel, J. K. *et al.* SMOG 2: A Versatile Software Package for Generating Structure-Based Models. *PLOS Comput Biol* **12**, e1004794 (2016).
 29. Englander, S. W. & Mayne, L. The nature of protein folding pathways. *Proc. Natl. Acad. Sci.* **111**, 15873–15880 (2014).
 30. Agyemang, A. F., Harrison, S. R., Siegel, R. M. & McDermott, M. F. Protein misfolding and dysregulated protein homeostasis in autoinflammatory diseases and beyond. *Semin. Immunopathol.* **37**, 335–47 (2015).
 31. Jahn, K. *et al.* Functional patterning of DNA origami by parallel enzymatic modification. *Bioconjug. Chem.* **22**, 819–23 (2011).
 32. 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.* **98**, 2217–2221 (2001).

33. Lai, Y.-T. *et al.* Designing and defining dynamic protein cage nanoassemblies in solution. *Sci. Adv.* **2**, e1501855 (2016).
34. King, N. P. *et al.* Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510**, 103–108 (2014).
35. Hsia, Y. *et al.* Design of a hyperstable 60-subunit protein icosahedron. *Nature* **535**, 136–9 (2016).
36. Fletcher, J. M. *et al.* Self-Assembling Cages from Coiled-Coil Peptide Modules. *Science (80-.)*. **340**, 595–599 (2013).
37. Wen, A. M. & Steinmetz, N. F. Design of virus-based nanomaterials for medicine, biotechnology, and energy. *Chem. Soc. Rev.* **45**, 4074–4126 (2016).
38. Brunette, T. *et al.* Exploring the repeat protein universe through computational protein design. *Nature* **528**, (2015).
39. Kanekiyo, M. *et al.* Rational Design of an Epstein-Barr Virus Vaccine Targeting the Receptor-Binding Site. *Cell* **162**, 1090–1100 (2015).
40. López-Sagaseta, J., Malito, E., Rappuoli, R. & Bottomley, M. J. Self-assembling protein nanoparticles in the design of vaccines. *Comput. Struct. Biotechnol. J.* **14**, 58–68 (2016).
41. Kushnir, N., Streatfield, S. J. & Yusibov, V. Virus-like particles as a highly efficient vaccine platform: Diversity of targets and production systems and advances in clinical development. *Vaccine* **31**, 58–83 (2012).
42. Correia, B. E. *et al.* Proof of principle for epitope-focused vaccine design. *Nature* **507**, 201–6 (2014).
43. Kanekiyo, M. *et al.* Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies. *Nature* **499**, 102–6 (2013).

Figure legends

Fig. 1: CC module structure and CCPO nomenclature

(A) Coiled-coils (CC) are defined by a regular repeating unit of seven amino acid residues, labeled *a*, *b*, *c*, *d*, *e*, *f*, or *g* (a heptad repeat). Interaction specificity is encoded by hydrophobic interactions at positions *a* and *d* (yellow dashed lines) and electrostatic interactions between positions *e* and *g* (red dashed lines). DNA complementarity rules are comparatively simple, since only four building elements are present.

(B) CC dimers can pair in a parallel or antiparallel orientation, defined by the sequence. Positions *a* and *d* are shown as spheres. B-strand DNA can pair only in an antiparallel orientation.

(C) Nomenclature of CCPO polyhedra. An example (TET12_{1.10}SN-f₅) is shown in the middle of the panel. The full names of the polyhedra are defined by the type of polyhedron (TET = tetrahedron; PYR = pyramid; TRIP = trigonal prism) followed by the number of CC-forming segments (shown in blue). The subscript (in orange) denotes the topology and circular permutation of each polyhedron, i.e. the double Eulerian path that the polypeptide chain makes over the polyhedron. The next labels (green) denote the type of CC modules used (S = soluble; SN = soluble, negatively charged), linker type (f = flexible; c = charged), and, in subscript form (red), the length of the linker. In cases where two variants have the same name (e.g., different ordering of CC modules), the letters b, c, d, etc. are appended. The most extensively characterized polyhedra are referred to by shorter names: TET12SN (TET12_{1.10}SN-

f₅), TET12S (TET12_{1.10}S-f₅), TET12SScr (TET12_{Scr}S-f₅), PYR16SN (PYR16_{2.15R}S-f₅), and TRIP18SN (TRIP18_{7.5R}SN-f₅).

Fig. 2: Design and biophysical analysis of two tetrahedral structures that self-assemble in bacterial cells.

(A) The topology (a double circular path) of the tetrahedral design. TET12SN consists of two antiparallel and four parallel pairs, while TET12_{2,3}SN-f₅ consists of three antiparallel and three parallel CC modules. Each color represents a different CC pair (Table S2).

(B) Representative all-atom models generated by CoCoPOD. The colors are the same as in panel A. The scale bar denotes 5 nm.

(C) Contact maps displaying long-range (i.e., far apart in the primary sequence) contacts between amino acids. Antiparallel pairs are perpendicular to the main diagonal. Differences in the topology between the two designs are clearly visible. The color code is the same as in panel A.

(D) Circular dichroism (CD) spectra at 20°C (black) and 92°C (red).

(E) Normalized CD signal at 222 nm during temperature unfolding (black). The fitted transition is shown in orange; the transition temperature is given in the insets.

Fig. 3: Coiled-coil protein origami design platform (CoCoPOD)

De novo design of protein origami involves the following steps.

(A) Any polyhedral shape is chosen. The user can prepare the input with a 3D rendering program. The size is limited only by the number of available orthogonal CC modules.

(B) All possible topologies (double Eulerian paths) for a given geometry are computed. Since topologies are circular, all possible circular permutations of each topology are examined. The circular permutation with the lowest topological contact order (average distance between segment pairs) is chosen as being optimal.

(C) CC modules are mapped to the abstract topology identified by graph theory. The modules are taken from a toolkit of orthogonal CC building blocks (Table S1).

(D) All-atom 3D models are built, and properties such as volume, contact order, and solvent-accessible surface are calculated.

(E) Designs are also tested using coarse-grained, structure-based folding simulations.

For a detailed description, see Online Methods and github.com/NIC-SBI/CC_protein_origami.

Fig. 4: Design and biophysical analysis of the four-sided pyramid and triangular prism coiled-coil protein origami folds.

(A) A topology is shown for each polyhedral shape. Each color represents a different CC pair (Table S2).

(B) Representative all-atom models generated by CoCoPOD. The colors are the same as in panel A. The scale bar denotes 5 nm.

(C) Contact maps displaying long-range contacts between CC pairs. Antiparallel pairs are perpendicular and parallel pairs are parallel to the main diagonal. The colors are the same as in panel A.

(D) Circular dichroism (CD) spectra at 20°C (black) and 92°C (red).

(E) Normalized CD signal at 222 nm during temperature unfolding (black). The fitted transition is shown in orange; the transition temperature is given in the insets.

Fig. 5: Structural characterization by solution small-angle X-ray scattering (SAXS) and electron microscopy.

(A) Experimental SAXS scattering curve (black) and scattering calculated from models shown in panel C (orange line). The goodness-of-fit, χ , is shown in the insets.

(B) The pair-distance distribution function, $P(r)$, as obtained by the SAXS experiment (black line) and simulated from the model structure shown in panel C (orange line).

The radius of gyration, R_g , is shown in the insets.

(C) Models that best match the scattering curve in panel A. The models were generated using CoCoPOD. For TET12SN and PYR16SN, further sampling of the conformational space was performed using normal mode analysis. The top and side views are shown. The scattering data for TRIP18SN was fitted with a mixture of two conformations corresponding to the rectangular (top model) and oblique (bottom model) triangular prism. The flexibility is expected, since the length of the edges does not uniquely define the angles in faces with four edges. Colors represent unique CC pairs, as in Fig. 2 and Fig. 4.

(D) Ab initio calculations of the molecular envelopes from the SAXS data overlaid with the 3D models shown in panel C. An internal cavity is present in all three polyhedral shapes. The orientation of the models is similar to the top row in panel C. For TRIP18SN, only the rectangular prism model is shown.

(E) Electron densities calculated from negative-stain transmission electron microscopy images overlaid with the 3D models shown in panel C (with similar orientations). Internal cavities are observable in all three constructs. For TRIP18SN, only the rectangular prism model is shown.

Fig. 6: Protein origami folds in mammalian cells and in living animals.

(A) Correctly folded TET12S reconstitutes the split mVenus fluorescence, since the N and C termini coincide at the same vertex. Fluorescence was determined 48 h after transfection of HEK293 cells with plasmids coding for TET12S^{split-mVenus} or TET12SScr^{split-mVenus}. The anti-GFP AlexaFluor 647 antibody (α -GFP) was used to determine the expression level of mVenus. TET12S was localized in the cytosol of HEK293 cells. The scale bar = 50 μ m. A representative image from two independent transfections is shown.

(B) Schematic representation of the *in vivo* experiment based on the hydrodynamic delivery of plasmids for protein expression to the liver of mice.

(C) TET12S correctly folded in liver cells, in contrast to TET12SScr. Tissue samples from mouse liver were thinly sliced and immunostained. The proper formation of TET12S was monitored through the reconstitution of split-mVenus fluorescence. The expression level of split-mVenus of TET12S^{split-mVenus} and TET12SScr^{split-mVenus} was determined with anti-GFP AlexaFluor 647 antibodies (α -GFP); the nuclei were stained with DAPI. The scale bar = 50 μ m. A representative image of two independent experiments is shown, each with three animals.

(D) Correct folding of TET12S in mice was shown by *in vivo* imaging based on the reconstitution of the split luciferase. Mice that received plasmids coding for TET12S^{split-fLuc} produced bioluminescence within the liver, which was not observed in mice injected with TET12SScr^{split-fLuc} plasmid.

Online Methods

Coiled-coil protein origami design (CoCoPOD) computational methods

CoCoPOD is implemented as a Python module (`cocopod`); it depends on Modeller⁴⁴, Chimera⁴⁵, and MDTraj⁴⁶. The source code is available in **Supplementary Source Code**, at github.com/NIC-SBI/CC_protein_origami, and at figshare 10.6084/m9.figshare.4003398. A complete example is shown using an IPython notebook⁴⁷. The design strategy consists of several mostly automated steps (Fig. 3):

1. *Specifying the polyhedral geometry*

The geometry of the polyhedral shape can be specified either manually as a list of polygon faces or, more conveniently, it can be prepared using a 3D rendering program. We have done the latter, in which the platform parses the widely used and supported Stanford PLY ASCII format, which contains the necessary connectivity information.

In the following text we refer to a pair of CC dimers as a building **module**. A **segment** is a single peptide (half a module). The tetrahedron was built from 6 modules and 12 segments.

2. *Routing the chain*

Stable traces are paths in which consecutive edges share a face (i.e., the path does not loop back on itself). As a result, stable traces hold the polyhedron's shape. For any given polyhedron, many different double Eulerian trails are possible¹⁴. In the context of single-chain polyhedral cages, we refer to double Eulerian trails as a **polyhedral topology**, since the order of the building blocks in a chain is determined

by the double trail, while the order of the segments is preserved under continuous deformations.

All strong double traces were enumerated using the method of 1-face embedding, as described by Fijavž *et al.*²⁷. Briefly, a polyhedron is defined as a polygonal complex of oriented and rooted polygons. All stable double traces can be constructed by the successive gluing of two faces with a common edge. In the present work, we also consider trails with crossings in their vertex figures. We provide more details in the supplementary note, “Topologies of a pyramid with crossings.”

3. *Selecting the optimal topology and circular permutation*

Because double Eulerian trails are circular, while polypeptide sequences are not, the circular tour needs to be linearized. The chain can begin at any vertex, after traversing the polyhedral shape the chain will return this vertex. In other words, the N and C termini of the polypeptide are colocalized in the same vertex. For a polyhedron with N edges there are at most 2N possible linear trails (and at most 2N reflected linear trails; see supplementary note). The selection of the starting vertex defines the linear order of the segments, which is termed a **circular permutation**.

For example, there are three topologies for the tetrahedral shape: one with two antiparallel and four parallel segments and two with three antiparallel and three parallel segments. Each topology can be realized using different circular permutations.

The topologies are represented in edge-face notation as a string of letters, where each letter represents one CC-forming segment. The same letters (A–F) denote a matching pair of segments. Differently cased letters (e.g., Aa, aA) indicate an

antiparallel pair, while same-cased letters (e.g., AA, aa) indicate a parallel pair. The three topologies of the tetrahedron are presented below:

- ABCADECfEbdF
- ABCADECfDbEf
- ABCADEbDFceF

The four-sided pyramid and trigonal prism demonstrated in this report can be realized in 52 and 25 topologies, respectively.

The topologies and circular permutations may have different stabilities or folding properties due to the varying amino-acid distances between segments; this is similar to an earlier demonstration of the differences in circular permutations of the DNA single-chain four-sided pyramid¹⁹.

Analogously to protein-contact order⁴⁸, which has been successfully used to predict proteins' folding rates, we introduce a **topological contact order** (TCO) to choose the optimal topology and circular permutation. The TCO is defined as the average number of segments between a pair:

$$\text{TCO} = \frac{1}{N} \sum_{i=1}^N \sum_{j=i+1}^N \begin{cases} j-i, & \text{if } i \text{ and } j \text{ form a parallel pair} \\ j-i-1, & \text{if } i \text{ and } j \text{ form an antiparallel pair} \\ 0, & \text{if } i \text{ and } j \text{ do not form a pair} \end{cases}$$

Where N is the numbers of segments and i and j are the positions of the segment in the chain. Antiparallel and parallel pairs must be handled differently, since the end-to-end distance is longer for the former than for the latter.

Since several circular permutations can have the same TCO, we also calculate the standard deviation of the distances between pairs (stdTCO):

$$\text{stdTCO} = \sqrt{\frac{1}{N} \sum_{i=1}^N \sum_{j=i+1}^N \begin{cases} (\text{TCO} - j - i)^2, & \text{if } i \text{ and } j \text{ form a parallel pair} \\ (\text{TCO} - j - i - 1)^2, & \text{if } i \text{ and } j \text{ form an antiparallel pair} \\ 0, & \text{if } i \text{ and } j \text{ do not form a pair} \end{cases}}$$

Where N is the numbers of segments, i and j are the positions of the segment in the chain, and TCO is the topological contact order.

As a guiding rule, the algorithm selects the circular permutations with the **smallest TCO and stdTCO**. This rule is subject to other design constraints, such as the selection and availability of the modules.

All of the topologies and circular permutations can be enumerated. The circular permutations are then given a canonical form (as described in section 2) and are sorted by the number of crossings, the number of antiparallel segments, and the lexicographical order of the canonical form. Each topology and circular permutation can therefore be assigned a unique name. See file “Topologies-circular-permutations-TCO.xlsx”

4. *Selecting the modules from the CC toolkit*

Once a circular permutation was selected, CC modules from the available toolbox were selected for each segment in the topology. For example, the pair BB can be realized by any parallel CC dimer in the toolbox (e.g., P3-P4 or P3S-P4S), depending on their required properties, such as their stability or charge or the presence of selected residues at particular positions. Linkers between pairs must also be selected and can be adjusted according to the requirements of flexibility or solubility.

The CC toolbox includes a set of natural and designed CC hetero- and homo-dimers with defined orientations, all of which have been experimentally characterized.

In the PDB database, there are 199 parallel and 266 antiparallel CC dimers that are longer than 21 residues⁴⁹. In addition, several designed CC dimers have been characterized based on natural leucine zippers⁵⁰ or have been designed and their orthogonality tested^{25,51}. We used 22 different CC dimers for the construction of the polyhedra in this report; the dimers' sequences are described in Supplementary Table 1.

CC pairs were divided into orthogonality groups (determined by the combinations *g*, *a*, *d*, and *e* positions; Fig. S1). Modifications of the *b*, *c*, and *f* sites did not significantly affect the orthogonality (Fig. S2) but they can strongly affect properties such as solubility and stability. For example, P3, P3S, and P3SN all belong to the same orthogonality group and differ only in residues at the *b*, *c*, and *f* sites.

This part of the design was performed in a semiautomated interactive manner in which the user is allowed to make appropriate selections based on the desired properties of the polyhedra while taking into account the orthogonality of the modules.

The user can then consecutively select each of the building elements from the available toolbox; the algorithm warns the user if several segments within the same orthogonality group have been used. The remaining segments can then be provided based on their desired physicochemical properties, such as stability, polarity, net charge, and helical propensity.

Based on several key biochemical and experimental requirements in this report (Figs. S5, S6 and S9), guidelines for the optimal assignment of CC modules were derived: the least stable segment pairs should not be selected as the terminal segments. The more stable CC modules should also be chosen for the segment pairs

that are far apart in the amino acid sequence; this should help compensate for the entropic penalty of bringing distant segments into close proximity during folding.

Using different linkers can affect the flexibility, folding kinetics, and protein-origami solubility. As such, linker modifications can be used to adjust the desired properties.

In general, stability was not significantly affected by the choice of linkers, as may be seen from the denaturation of TET12_{1.10}SN-f₅, TET12_{1.10}SN-f₉, and TET12_{1.10}SN-c₆ (Fig. S5), which have linkers of length 5, 9 and 6, respectively. This indicates that various flexible hydrophilic linkers can be accommodated within the polyhedral cage.

5. *Constructing 3D models*

The 3D polyhedra models were built using a combination of Modeller and Chimera. First, the entire sequence was forced into an ideal alpha helix; the CC modules were then made rigid, and a rigid-body molecular dynamics simulation was started. Harmonic restraints were applied to each pair consecutively; the order of the pairing was performed according to the stability of the modules (e.g., APH, BCR, P3-P4, P5-P6, GCN, P7-P8). In order to force the orientation of the modules for each pair, two additional restraints were applied: the first between the first C-alpha atoms of each segment and the second between the last C-alpha atoms of each segment. (In antiparallel segments, the restraints were first-last and last-first C-alpha atoms.) Twenty thousand steps were used for each stage of the simulation.

The generated model was then refined by a homology-modeling step based on the 3D structure of the CC dimers. For some CC pairs (e.g., BCR⁵², GCN⁵³) the

crystal structure was known; for others, the models were constructed using CCBUILDER⁵⁴.

This elaborate protocol is necessary because Modeller homology modeling does not reliably converge unless a good initial structure is provided. Constructing the model only via homology is also possible but is significantly more time consuming and more difficult to automate, since the minimization rarely converges. Our step-wise approach to building an initial model results in a speed-up of several orders of magnitude, thus reducing the time necessary to build one model from several days to less than an hour.

The MD and homology models were then repeated to obtain multiple models. At least 20 models were generated for each of the designed polyhedra variants. The models were evaluated by calculating the volume of the cavity, the contact order, the fraction of buried residues, and the presence of topological knots; they were also filtered manually, since bringing segments together using harmonic restraints may sometimes trap the remainder of the chain between two segments.

6. Refinement/validation of the models via folding simulations

All-atom MD simulations were performed on TET12S and TET12SN; we found that the models remained stable throughout the course of the simulations (60 ns). Earlier work on the folding of single-chain DNA origami¹⁹ raised concerns that some of the protein-origami designs might not fold correctly due to kinetic restraints. To address this problem, structure-based model (SBM) folding simulations (also known as Gō simulations) were performed on all designs. This type of simulation has previously shown good agreement with experimental results on natural proteins' folding units (9). The Gō simulations in our work were set up using the SMOG web server²⁸ and were performed using NAMD⁵⁵ and GROMACS⁵⁶. By analyzing the folding simulations, we did not observe any topological or entropic barriers to folding in any of the designs that are presented in this report.

The CC modules that are presented in this report comprised between four and five heptads and therefore did not include a full super-helical turn. As such, the kinetic restraints identified via the DNA pyramid¹⁹ do not apply in this case. We do expect, however, that topological barriers to folding would become more relevant for longer-CC building blocks, since these can introduce knotting. As observed in single-chain DNA polyhedra, the stability of the building segments could be used to guide the folding pathway¹⁹. In this case, this final step in the design platform should be able to identify the topology with the most favorable folding properties. In this iterative way, the models were built and the SBM simulations were performed until a successful design was identified *in silico*.

Construction of plasmids

All genes were synthesized by Genewiz (USA), except for TET12_{1.10}SN-c₆, TET12_{1.10}SN-f₉ and TRIP18_{7.5R1}SN-f₅, which were synthesized by Gen9 (USA). Genes were obtained or cloned into expression vector pET-41a using restriction sites NdeI and XhoI. TET12SScr (TET12_{scr}S-f₅) was obtained as a PCR product from fused gene of TET12SScr and split-mVenus, and was ligated into pET-41a cut with NdeI and XhoI. For expression of FLAG-tagged TET12S in mammalian cells, ORF encoding TET12S (TET12_{1.10}S-f₅) was amplified by PCR using Phusion HF (Thermo Fischer Scientific) and cloned into pcDNA3.1 via EcoRI and NotI restriction sites or into pFLAG-CMV-3 (Sigma-Aldrich, USA) via NotI and EcoRV restriction sites. All restriction enzymes were from New England Biolabs. pFLAG-CMV-3 includes preprotrypsin leader sequence for secretion.

TET12S and TET12SScr with split mVenus were constructed from PCR amplified N-terminal split mVenus (from 1 to 84 amino acid residue of mVenus), C-terminal split mVenus (from 85 to 238 amino acid residue) and appropriate TET12 construct. Obtained constructs were cloned into pcDNA3.1 and pFLAG-CMV.

TET12S and TET12SScr with split luciferase for mammal expression were constructed from PCR amplified N-terminal split firefly luciferase (from 1 to 491 amino acid residue of luc2P firefly luciferase), C-terminal split firefly luciferase (from 492 to 553 amino acid residue) and appropriate TET12 construct. Obtained constructs were fused with Gibson assembly and inserted into pcDNA3.1 and pFLAG-CMV.

Polypeptide expression and purification

E. coli NiCo21(DE3) strain (New England Biolabs) was transformed with respective constructs and grown at 37°C overnight (160 rpm) in Luria-Bertani (LB)

medium containing kanamycin (50 µg/mL). Bacterial cultures were transferred to LB medium at OD of 0.1, grown at 37°C until OD reached 0.6 and induced with 0.5 mM IPTG. After induction, the cultures were cultured for 4 additional hours at 30°C. The harvested cells were resuspended and lysed on ice with a lysis buffer: 50 mM Tris buffer at pH 8, 150 mM NaCl, 10 mM imidazole, 1 mM MgCl₂, 0.5 mg/mL Lysozyme (Sigma-Aldrich), 15 U/mL Benzonase (Millipore) and CPI protease inhibitor mix (Sigma-Aldrich). Cell lysis was completed by ultrasonication on ice for 15 min, at intervals of 1 s pulse and 3 s pause (50% amplitude). Subsequently, cellular lysates were centrifuged at 16000 g (4°C) for 30 min, respective soluble fractions were filtered through 0.2 µm filter units (Sartorius) and incubated for 1 h at 4°C with Ni-NTA resin (Golden Biotechnology) previously equilibrated with buffer A (50 mM Tris buffer at pH 8.0, 150 mM NaCl, 10 mM Imidazole). After washing with buffer A and B (50 mM Tris buffer at pH 8.0, 150 mM NaCl, 10 mM imidazole and 50 mM Tris buffer at pH 8.0, 150 mM NaCl, 20 mM imidazole respectively) the bound fraction was eluted with buffer C (50 mM Tris buffer at pH 8.0, 150 mM NaCl, 250 mM imidazole). EDTA (1 mM) and Glycerol (10% v/v) were added to the eluted fractions. The samples were then concentrated (Millipore centrifugal unit 10 MWCO), injected to a Size Exclusion column (HiLoad16/600 Superdex pg, GE Healthcare) and separated at 1 mL/min (SEC buffer: 20 mM Tris buffer at pH 7.5, 150 mM NaCl, Glycerol 10% v/v). The respective monomeric peaks were pooled, concentrated, shock frozen in liquid nitrogen and stored at -80°C. The samples were characterized by SDS-PAGE, CD, SEC-MALS and dynamic light scattering.

To produce larger amounts of proteins for structural characterization, 4 liter cultures were produced and samples were isolated through four chromatographic passages. Cellular pellets were resuspended in lysis buffer as described above, but

without imidazole, loaded on a 20 mL Q-Sepharose column (GE Healthcare) and eluted within a gradient of NaCl from 150 to 550 mM. Selected fractions were loaded in a 10 mL Ni-Sepharose-6-FF column (GE Healthcare) and eluted within a gradient of imidazole from 0 mM to 550 mM. Pooled fractions were subsequently loaded on a 10 mL column packed with Source-15Q resin (GE Healthcare) and eluted within a gradient of NaCl from 150 mM to 550 mM. Obtained samples were then separated with a size exclusion chromatography and treated as described above. Additionally, all the constructs having cysteines in the sequence have been isolated in presence of TCEP 1 mM.

Circular dichroism (CD) measurement

CD measurements were performed on a Chirascan CD spectrometer equipped with a Peltier temperature controller (Applied Photophysics, UK). CD spectra of the polypeptides in 20 mM Tris buffer pH 7.5, 150 mM NaCl, were recorded in far-UV in a 1 mm quartz cuvette (Hellma, Germany) at 20°C using 1 nm step, 1 nm bandwidth, and 1 s sampling. Secondary structure of 5 μ M polypeptide solution was analyzed by scan measurement from 200 to 260 nm. The results are the average of three scans. The helical content was estimated from average mean residue ellipticity at 222 nm, $[\theta_{222}]$, using the following equation⁵⁷:

$$\text{Helical content (\%)} = [\theta_{222}] / ([\theta_{222}]^H \times (1 - 2.57/n)),$$

where n is the length of amino acid sequence, and $[\theta_{222}]^H$ is the theoretical mean residue ellipticity of an infinitely long helix. For $[\theta_{222}]^H$ the value of -39,500 deg cm² dmol⁻¹ was used⁵⁷. The expected helicity was calculated using CCPO model structures that fit best to SAXS data and PyMOL⁵⁸.

Thermal stability of the isolated proteins was determined by measuring the ellipticity at 222 nm in a temperature range of 10 - 92°C.

Orthogonality of coiled-coil forming peptides

The heterodimeric CC-forming segments used in protein origami constructs are based on a previously published orthogonal set, including P3, P4, P5, P6, P7 and P8⁵¹. Peptides P3-P8 were used as building blocks in TET12 design¹⁴. A new orthogonal set was designed, comprising peptides PnS and PnSN, respectively, by modifying the non-contacting *b*, *c*, and/or *f* residues of the heptad repeats (sequences in Table S1), in order to retain the orthogonality of the set, increase the net negative charge and decrease stability. Peptides were purchased from Proteogenix (France). Binding specificity of peptides was tested by measuring the CD spectra of all possible pairs of peptides P3S, P4S, P5S, P6S, P7S, P8S, and similarly of P3SN, P4SN, P5SN, P6SN, P7SN and P8SN (Fig. S1). CD spectra were measured using a Chirascan CD spectrometer equipped with a Peltier temperature controller (Applied Photophysics, UK), from 200 to 260 nm, bandwidth 1 nm, scanning rate 2 nm/s, and the optical path length was 1 mm. Final peptide concentrations were 20 μM each in 20 mM Tris buffer pH 7.5, 150 mM NaCl, if not specified differently. The temperature was maintained using the instrument's Peltier element and monitored using a temperature probe. It was normally set to -1°C in order to capture the presence of CC dimers that may not be stable at room temperature and at measured concentrations. The peptides PnS were used in TET12S design, and PnSN were used in TET12SN, PYR16SN and TRIP18SN design.

Protein stability

Stability of protein constructs was determined by fitting a thermodynamic model to experimental the CD temperature denaturation data as described previously⁵⁹. Briefly, thermal denaturation is modeled as an equilibrium between the native, denatured and one or two intermediate states. The equilibrium constants K_i that define the ratio between equilibrium concentrations of each molecular species change with temperature according to the Van't Hoff relation $\frac{d \ln K_i}{dT} = \frac{\Delta H_i^\circ}{RT^2}$, where T is the temperature, R is the gas constant, and ΔH_i° is the standard molar enthalpy of the transition. Using this relation the equilibrium species concentrations and subsequently the experimentally measurable signals can be calculated at any temperature, subject to a few parameters (K_i at a reference temperature and ΔH_i° for each transition i ; ΔH_i° were assumed to be independent of temperature, given the small hydrophobic surface area buried in the native structure). These parameters were adjusted so that the model-calculated signal was fitted to the experimentally measured one. The fitting was carried out by a custom-written C++ program making use of the Nelder-Mead optimization algorithm⁶⁰ and a modified version of Powell's Hybrid method⁶¹ for solving the system of equations that yields equilibrium species concentrations.

Circular Dichroism Stopped flow (SF) measurements

Stopped flow measurements were performed on a MOS-500 spectrometer with an SFM-3000 stopped flow unit (BioLogic, France). Proteins at concentration from 1 to 1.5 mg/mL were prepared in 20 mM Tris buffer pH 7.5, 150 mM NaCl, 10% Glycerol and 6 M Guanidinium chloride (GndCl). For folding experiments proteins were diluted with buffer A (20 mM Tris pH 7.5, 150 mM NaCl) and B (20 mM Tris pH 7.5, 150 mM NaCl, 1 M GndCl). All buffers were filtered through a 0.45 μm PES

Minisart® filter (SN 16533, Sartorius, Germany). The final protein concentration after mixing was 0.2 mg/mL. The final GndCl concentration was 1M. The SFM-3000 has three independent syringes (S1, S2, S3). The protein sample was in S3, Buffer A in S1 and buffer B in S2. Total flow was set to 10.5 mL (estimated dead time 7.1 ms using an HDS mixer). Total volume per measurement was ~400 µL, at least 63 uL of the sample from S3 was used in one shot. Three repetitions were performed and averaged. The cuvette used was 2 mm by 2 mm in size (FC20). CD signal was measured at 227 nm with a 5 nm bandwidth and 300 millidegree dynamic range. The signal was recorded each ms for 2 seconds.

Fluorescent labelling for Stopped flow (SF) measurements

Constructs containing a free cysteine at the N and C terminal (TET12_{1,10}SN-f₅-2cys, PYR16_{4,6}SN-f₅-2cys and TRIP18_{7,5R}SN-f₅-2cys) were isolated and purified as described above. About 0.5 ml of 20 µM protein in 20 mM Tris, pH 7.5, 150 mM NaCl, 10 % Glycerol and 1 mM TCEP was used for one experiment. The protein was labeled with five times molar excess of Sulfo-Cyanine3 maleimide and Sulfo-Cyanine5 maleimide. Each dye was dissolved in DMSO to form a 10mg/ml stock solution. Importantly, the corresponding volumes of stock solution Cy3 and Cy5 were thoroughly mixed before addition to the protein solution. The reaction vessel was wrapped in foil, mildly stirred for two hours at room temperature and then over night at 4 °C.

The reaction mixture was purified on a PD 10 disposable desalting column (Sigma Aldrich, USA). First the column was equilibrated with 25 mL of buffer A (20 mM Tris, pH 7.5, 150 mM NaCl). The sample was added to the column and left to enter the column completely, then further buffer A was added and fractions of 0.5 ml were collected. Usually the 6th and 7th fractions were pulled and used for stop flow measurements without the need for

further concentration. The degree of labeling was estimated from the corresponding UV/Vis spectra. The concentration of labeled protein was approximately 9 μM .

TCEP - Tris(2-carboxyethyl)phosphine hydrochloride - was purchased from Sigma Aldrich, USA. The fluorescent dyes were purchased from LumniProbe, Germany.

Stopped flow (SF) measurements

Guanidine hydrochloride (GndCl) was added in powder form to the labelled protein samples to obtain a final concentration of 5M GndCl and $\sim 5 \mu\text{M}$ concentration of protein (478 mg of GndCl were added to 1 ml of labeled protein sample). The sample was left for one hour at room temperature to ensure complete unfolding.

The measurements were performed on MOS-200 spectrometer with an SFM-3000 stopped flow unit (BioLogic, France). The final protein concentration after mixing was $\sim 1 \mu\text{M}$. The S1 and S2 syringes had a capacity of 10 mL, the S3 syringe had a 1.9 mL barrel. The sample was placed in syringe S3 and buffer A in syringes S1 and S2. Total flow was set to 6 mL (estimated dead time 3.7 ms). Total volume per one measurement was 314 μL and least 62 μL of the sample from S3 was used in one shot, resulting in a five times dilution. At least five repetitions were performed and averaged. The cuvette used was 1.5 mm by 0.5 mm in size (FC15/5). The instrument was set in fluorescent mode with the detector at 90° to the excitation light. The sample was excited at 547 nm using the Xe/Hg lamp at 150 W. Signal was recorded with a longpass filter FGL665 (Thorlabs, USA) which transmits light with wavelength longer than 655 nm. The signal was recorded each ms for 2 seconds. As a control an equimolar mixture of free Cy3 and Cy5 was used instead of the sample. No change in signal was observed.

Protein folding/unfolding kinetics

The kinetics of protein folding and unfolding was determined by fitting an exponential kinetic model to experimental stopped-flow data. The model was fitted to the average of at least three traces. The fitting was performed using lmfit (lmfit.github.io) in python. The fitted equation was $y = a \exp(-k \cdot t) + b \cdot t + c$, where y is the signal, t , the time and k , a , b and c the fitted parameters. The linear drift was subtracted from the data and fit before plotting. The first 3 point had to be discarded, since noise was still present from the mixing events.

Dynamic light scattering measurement

The size of self-assembled structures was measured on a ZetasizerNano (Malvern, UK) at 20 °C using an angle of 173° and 633-nm laser. The protein size distribution of particles was recorded and a hydrodynamic diameter was calculated using the software provided by the manufacturer.

Size-exclusion chromatography, Multi-angle light scattering (SEC-MALS)

Molecular masses of samples were determined with a HPLC system Alliance e2695 (Waters, France) coupled to a Dawn Heleos II MALS recorder (Wyatt Technology Europe, Germany). 20 µL, 50 µL or 100 µL of all of the purified proteins were filtered through a 0.1 µm centrifugal filter units (Millipore) and injected on a BioSep 3000 column (Phenomenex, CA, USA), with the exception of TET12_{2,3}SN-f_{5b}, which has been separated on a BioSep 4000 column (Phenomenex, CA, USA). Calculations of molecular mass and standard deviations were obtained using ASTRA 6 software (Wyatt Technology Europe, Germany).

Small Angle X-ray Scattering (SAXS)

Scattering curves were measured at B29 beamline at European Synchrotron Radiation Facility (ESRF, Grenoble, France) and at P12 beamline at PETRA-III synchrotron (DESY, Hamburg, Germany). In the former case the wavelength of incident X-rays was 1 Å and the Pilatus 1M detector was placed 2.87 m from the sample leading to a range of the scattering vector from 0.025-5 nm⁻¹, while in the latter case the X-ray wavelength was 1.24 Å, the Pilatus 1M detector was positioned 2 m from the sample and the scattering vector ranged from 0.028-6.7 nm⁻¹. To detect any potential concentration dependent effects a series consisting of at least 3 concentrations was measured for every protein sample. High concentrations ranged from 3-13 mg/mL, while low concentrations were in the range of 1-3 mg/mL. For each sample data was collected over 10 frames lasting 1 or 2 s (ESRF) or 20 frames lasting 0.05 s (PETRA-III). Frames were carefully inspected for radiation damage and those not showing any radiation damage were then averaged. Before and after each sample scattering, matching buffer was measured and subtracted from sample scattering. In case the protein exhibited concentration dependent effects, curves corresponding to different concentrations were carefully merged; otherwise the curve showing the smallest signal-to-noise ratio was used for subsequent analysis. Scattering curves were analyzed using PRIMUS⁶² and Scatter⁶³ software. Ab initio modeling was accomplished using DAMMIF⁶⁴ and DAMAVER program⁶⁴. Theoretical scattering curves were calculated and compared to experimental SAXS profiles using PepsiSAXS (team.inria.fr/nano-d/software/pepsi-saxs/). Volatility of ratio⁶⁵ (V_R) scores were calculated using FoXS⁶⁶.

In case none of the initial models, generated using CoCoPOD, fit well with the measured SAXS profile, we performed additional conformational sampling of the best

fitting model through normal mode analysis (NMA). NMA was performed using ProDy⁶⁷, a python package intended for protein structural dynamics analysis. The proteins were represented using Anisotropic Network Model. Spring constants were set according to the secondary structure and connectivity of the residues, assigning higher spring constant to sequential residues and proximal residues in individual helical segments (for spring constants values set as default in GammaStructureBased class were used). First 20 non-trivial normal modes were calculated and along them up to 2000 conformations were sampled. The RMSD of sampled conformations in reference to the starting structure was set to 4 nm. The generated models were relaxed by performing a fast molecular dynamics simulation (100 steps, time step of 4 fs) using python package MODELLER to correct for bond lengths and angles. At this point, models were compared to the corresponding experimental SAXS profile and a handful of models showing the best agreement were selected. To ensure the knobs-into-hole packing, a fundamental feature of CCs, remained present after NMA, the CC segments in the selected models were refined through homology modelling as described above in the section “Constructing 3D models” in the Supplementary Methods. For each of the selected conformations homology based refinement was performed 40 times. Thus obtained models were then fitted to experimental scattering profiles and the model showing the lowest χ value was chosen as the best. Model structures obtained in this way showed good agreement with χ values in the range of 0.7 to 3.

In case of TRIP18SN, none of the models from the initial ensemble of 3D structures constructed with CoCoPOD gave a sufficiently good fit to the experimental SAXS profile. However, multiple conformations showed a comparable agreement with experimental data. Since the TRIP18SN design contains rectangular faces (it is

composed of two triangular faces and three rectangular faces) its shape is not uniquely defined just by the distances of edges, in contrast to the tetrahedron. To reflect the increased flexibility, we attempted to describe the scattering profile using more than one structure. From the initial pool of models we selected 10 models that represented different possible conformations of the triangular prism protein cage. Using OLIGOMER⁶⁸, the contribution of individual conformations was evaluated. Only three conformations exhibited significant contribution (>15%). OLIGOMER was run again, using only these three conformations as well as only the top two contributing conformations, leading to χ values of 0.9 and 1.0. Due to a small difference between the χ values, we decided to describe the SAXS profile with only two conformations. This does not imply that these are the only two conformations found in solution, but rather that due to its intrinsic flexibility the protein is present in solution a range of different conformations that can be represented as a combination of those selected models.

Chemical cross-linking coupled mass spectrometry (XL-MS)

Crosslinking Reactions

Cross-linking experiments with three different homobifunctional cross-linkers were performed: disuccinimidyl suberate (DSS), bis-N-succinimidyl-(pentaethylene glycol) (BS(PEG)₅), and bis-N-succinimidyl-(nonaethylene glycol) ester (BS(PEG)₉), all obtained from Pierce Biotechnology (Rockford, IL, USA). DSS has spacer arm 11.4 Å that bridges C α -C α distances up to 24.2 Å (linking two lysine residues or lysine and tyrosine residue) or up to 20.2 Å (linking lysine and serine residue). BS(PEG)₅ has spacer arm 21.7 Å that bridges C α -C α distances up to 34.5 Å (linking two lysine residues or lysine and tyrosine residue) or up to 30.5 Å (linking lysine and serine residue). BS(PEG)₉ has spacer arm 35.7 Å that bridges C α -C α distances up to 48.4 Å (linking two lysine residues or lysine and tyrosine residue) or up to 44.5 Å (linking lysine and serine residue). Prior to crosslinking purified monomeric TET12SN was dialyzed against a crosslinking buffer (20 mM sodium phosphate, 150 mM NaCl, pH 7.5 or pH 7.9). The concentration of the TET12SN in crosslinking reactions was 1.2, 3.3, 4.7 or 10 μ M. Cross-linkers were freshly dissolved in anhydrous DMSO (Sigma-Aldrich) (10 mM or 20 mM stock) and added in 10, 20, 50 or 100 molar excess to the protein solution. Reactions were performed at 25°C for 1 h with gently stirring and then quenched for 15 min by the addition of Tris buffer (1 M, pH 7.5) to a final concentration of 50 mM. Crosslinking reaction mixtures were boiled in SDS loading buffer, and the products were separated on 10% SDS PAGE.

Band excision and in gel digestion for MS

The SDS-PAGE gel was stained with Coomassie Brilliant Blue (Sigma Aldrich, USA). A gel band around 55 kDa that corresponds to the size of the monomer of TET12SN was excised and in gel digested with trypsin according to the method

described by Shevchenko et al.⁶⁹. Approximately 150 ng of trypsin, proteomics grade from Pierce Biotechnology (Rockford, IL, USA), in 25 mM ammonium bicarbonate was used per gel piece. Trypsin digestion at ratio protein:trypsin (1:20) was performed overnight at 37°C.

Nano-LC-MS/MS

Trypsin digested peptides were chromatographically separated using EASY-nLC II system (Thermo Fisher Scientific, Bremen, Germany) with a 2 column set up: trap column C18-A1, 2 cm (SC001, Thermo Scientific, Germany) and analytical column PepMap C18, 15 cm x 75 µm, 3 µm particles, 100 Å pore size (ES800, Thermo Fisher Scientific). 2 µL of sample was loaded and separated by 5-70% B linear gradient over the course of 70 min with a flow rate of 300 nL/min (A phase = 0.1% formic acid in water, B phase = 0.1% formic acid in 98% acetonitrile (MS grade solvents, Serva)). Peptides were analyzed by LTQ Orbitrap XL hybrid mass spectrometer (Thermo Fisher Scientific). MS data were acquired in data-dependent MS2 mode: each high-resolution full scan (m/z 300 to 2000, resolution 30000) in the Orbitrap was followed by product ion scans in the LTQ for the five most intense signals in the full-scan mass spectrum (isolation window 3 u). Dynamic exclusion (exclusion duration 60 s, exclusion window 20 ppm) allowed detection of less abundant ions. Data acquisition was controlled via XCalibur 2.1 (Thermo Fisher Scientific).

Identification of cross-linked peptides

Thermo Raw MS/MS files were exported as mascot generic files (.mgf) and analyzed using StavroX software²¹ version 3.4.12 with the following settings: protease cleavage sites: K, R; missed cleavages: K:3, R:1; variable modifications:

methionine oxidation (max. 2); cross-linker: DSS, or BS(PEG)₅, or BS(PEG)₉; cross-linked amino acids: 1. K, N-term.; 2. K, S, T, Y, N-term; precursor mass deviation: 3.0 ppm; fragment ion mass deviation: 0.8 Da; lower mass limit: 200 Da; upper mass limit: 5000 Da; S/N ratio: 2.0; ion types: b-and y-ions; neutral loss: only of identified fragments, to identify cross-linked peptides. A score cutoff was selected to control the false discovery rate (FDR) to within 1% (0.01). At FDR < 0.01 false positive assignments were eliminated. Cross-linked candidates were identified by comparing the query mass of the precursor to the mass of theoretical cross-links from the peptides. The identified cross-linked peptides were manually evaluated.

Negative stain Electron Microscopy (EM)

Concentrated protein samples were diluted to a concentration of ~30 µg/mL in TRIS buffer (20 mM Tris buffer pH 7.5, 150 mM NaCl, 10% Glycerol). Samples were applied to glow discharged carbon-coated copper grids, washed quickly with distilled water and negatively stained with 2% (w/v) uranyl acetate and observed using a JEOL-1230 operated at 100 kV. Images of single molecules were obtained automatically using a TVIPS F416 CMOS and a final magnification of 54926.

Data processing and 3D reconstruction

All image processing was performed using the Scipion platform (<http://scipion.cnb.csic.es>), which is an image processing framework that integrates several software packages into a unified interface⁷⁰. For each sample (TET12SN, PYR16SN and TRIP18SN) a total of (28573, 49510, 30653) particles were extracted from (26, 40, 49) micrographs respectively using the particle picking tool of Xmipp⁷¹. The particles were classified in 2D using Xmipp⁷². The density calculated from the representative (largest volume) all-atom model generated by the CoCoPOD framework was filtered to 40 Å to avoid bias. The filtered density was used as

an initial template for the 3D refinement in projection matching⁷⁰. For the refinement steps 13425, 18251 and 15145 particles were used, respectively.

Fitting of atomic models into density maps

UCSF Chimera⁴⁵ software was used to model atomic structures into the 3D EM reconstruction⁷³. The best matching models from SAXS experiments (Fig. S7) were fitted into the densities obtained from EM reconstructions. Global search was used to find the best orientation of the model with the reconstructed density (using the chimera fitmap command). Fifty initial search positions were used and the orientation with the highest volume overlap was chosen as the most optimal.

Experiments on mammalian cells

Expression of tetrahedron variants

In order to investigate expression, folding and cell response HEK293 cell line (see also Life Sciences Reporting Summary) was chosen due to ease of cultivation and high transfection efficiency. This cell line was obtained from ATTC and was tested negative for mycoplasma contamination. This cell line was not authenticated.

For Western blot, HEK293 cells (ATTC CRL-1573) were transfected with pCDNA3.1 or pFLAG-CMV-3 plasmids encoding tetrahedron variants with Lipofectamine 2000 (Invitrogen). Two days after transfection cell lysates and supernatants were analyzed on Western blot using anti-FLAG polyclonal antibodies (1:320, Sigma-Aldrich, F7425) or anti-GFP antibodies (1:1000; Invitrogen, A11122); anti- β actin (8H10D10) antibodies (1:5000; Cell Signalling Technology, 3700) as primary antibodies and HRP-conjugate goat polyclonal to rabbit IgG secondary antibodies (1:3000; Abcam, ab6721), HRP-conjugated goat anti-mouse IgG secondary antibodies (1:3000; Jackson ImmunoResearch, 115-035-003).

For visualization of split-mVenus reconstitution within HEK293 cells, HEK293 cells (1×10^6 cells/well) were plated on 8-well tissue culture chambers (Ibidi, Integrated BioDiagnostics). The following day, cells were transfected with plasmids expressing TET12S^{split-mVenus}, TET12SScr^{split-mVenus} (10 ng to 50 ng) using Lipofectamine 2000 transfection reagent (Invitrogen). Two days after transfection samples were fixed using paraformaldehyde (4%), Tween20 (0.01%) and stained against GFP using anti-GFP [E385] AlexaFluor 647 antibodies (1:200; Abcam, ab190585).

Unfolded protein response

Unfolded protein response was followed by a dual luciferase assay based on the transcription factor 6 (ATF6) dependent reporter assay. Plasmid encoding firefly luciferase under the ATF6 control (p5XATF6-GL3) was a gift from Ron Prywes (Addgene, plasmid #11976)⁷⁴. One day before transfection, 4×10^5 HEK293 cells were seeded per well of 96-well plate. The next day, cells were transfected by a mixture of plasmids (50 ng p5XATF6-GL3, 5 ng phRL-TK (Promega), 100 ng of combination of plasmids encoding TET12S or empty vector) using Lipofectamine 2000 (Invitrogen). A day later, media was exchanged and compounds inducing endoplasmic reticulum (ER) stress were added as positive controls. Two days after transfection cells were lysed in passive lysis buffer (Promega) and analyzed for reporter protein activities on a plate reader Orion II (Berthold Technologies). Relative luciferase activity was calculated by normalizing the firefly luciferase activity to constitutive Renilla luciferase activity measured within the same sample.

NLRP3 inflammasome activation

Immortalized mouse bone-marrow derived macrophages, a gift from Kate Fitzgerald⁷⁵ (see also Life Sciences Reporting Summary), were cultured in DMEM supplemented with 10% FBS. This cell line was tested negative for mycoplasma contamination and was not authenticated. A day prior activation 0.75×10^5 cells were seeded per well of 96 well plate. The next day cells were primed with ultra-pure LPS (100 ng/mL, Invivogen) for 8 h after which medium was removed and different concentrations of potential activators in DMEM were added and left overnight (or 30 min for nigericin). The concentration of secreted IL-1 β was measured by ELISA (eBioscience) according to manufacturer's instructions. Prion protein fibrils were prepared from recombinant murine prion protein as previously described⁷⁶. Silica was from Invivogen, alum from Thermo Scientific and nigericin from Sigma-Aldrich.

Animal experiments

Mice

Balb/c mice were purchased from Harlan (Italy), respectively. 8-12 weeks old male and female mice were used for experiments. No sample size estimation was performed. Three mice were used in two separate experiments for each measurement. All the animal samples were included in the analysis. No blinding or randomization was performed. All animal experiments were performed according to the directives of the EU 2010/63 and were approved by the Administration of the Republic of Slovenia for Food Safety, Veterinary and Plant Protection of the Ministry of Agriculture, Forestry and Foods, Republic of Slovenia (Permit no.U34401-37/2015/5) (see also Life Sciences Reporting Summary).

Hydrodynamic injection

Balb/c mice were hydrodynamically (during 4-7 seconds) injected into the tail vein with a volume of saline solution equivalent to 10% of body weight, containing 60 µg of plasmid of DNA expressing TET12S^{split-mVenus}, TET12SScr^{split-mVenus}, pFLAG- TET12S^{split-mVenus}, pFLAG-TET12SScr^{split-mVenus}, TET12S^{split-fLuc} or TET12SScr^{split-fLuc}, using a 3-mL latex free syringe with 27G needle (Beckton Dickinson).

Immunohistochemistry

After 2 or 5 days mice were sacrificed, blood was collected and serum was prepared (Sarstedt; 3000 rpm/30 min). Liver samples were collected for further analysis. Liver samples for immunostaining were fixated overnight in 10% neutral buffered formalin (Sigma-Aldrich) and then embedded in paraffin (Leica Paraplast). The paraffin blocks were cut 7 µm thick with a rotation microtome RM 2245 (Leica). Tissue sections undergone deparaffinization and rehydration using xylene and different dilutions of ethanol (Sigma Aldrich). Slides were washed twice with wash buffer (PBS+ 0.05% Tween20) (Sigma Aldrich). Blocking was done by incubating the samples in 2% BSA (Sigma Aldrich) in wash buffer for 2 h at room temperature. Primary antibodies (anti-GFP [E385] AlexaFluor 647 antibodies; 1:200; Abcam, ab190585) were incubated overnight in humidified chamber at 4°C. After washing, nuclei were stained with DAPI. Tissue slides were mounted using Leica CV Mount (Leica) and imaged using confocal microscopy.

Liver profile

Mice liver profile was determined in mice sera using VetScan Mammalian Liver Profile reagent rotor and analyzed on biochemistry analyzer VetScan VS2 (Abaxis).

In vivo luciferase imaging

Two days after the hydrodynamic plasmid delivery mice were anesthetized with isoflurane. Each mouse received D-luciferin (Xenogen) intraperitoneally, 150 mg/kg of body weight. After 10-15 min mice were imaged *in vivo* using IVIS Lumina Series III (Perkin Elmer) with CCD Camera (Pixel Dimensions 13x13 μm). Images were acquired using 5 min of exposure, binning 8, f/stop 1. Data were analyzed with Living Image 4.5.2 (Perkin Elmer).

Confocal microscopy

Images of HEK293 cells (fixed and permeabilised) and liver tissue sections were acquired using a Leica TCS SP5 inverted laser-scanning microscope on a Leica DMI 6000 CS module (Leica Microsystems) equipped with a HCX Plane-Apochromat lambda blue 63 \times oil-immersion objective with NA 1.4, with two hybrid detectors HyD and two photomultipliers and with high efficiency spectral photometry detection and optical Acusto Optical Beam Splitter (AOBS).

Images were acquired and processed using Leica Application Suite (LAS X) software, dimensions 512 \times 512 \times 1 pixels (pixel dimension 418.5 nm \times 418.5 nm), 8 bit resolution with no resolution enhancement and linear whole range lookup table (LUT) and 400 Hz acquisition speed.

For excitation of reconstituted split-mVenus, a 515 nm laser line (90% laser power) of the 25 mW-multi ion argon laser (25% laser power) was used, the emission

was detected between 520 and 570 nm. The mVenus stained with anti-GFP antibodies (Alexa Fluor 647) were excited with a 10 mW, 633-nm HeNe laser (90% laser power), and detected between 650 and 700 nm. For DAPI excitation, a 50 mW, 405-nm diode laser (35% laser power) was used. Fluorescence emission was detected from, 450 to 500 nm.

Software and statistics

Graphs were prepared with Gnuplot 5.0 (<http://www.gnuplot.info/>) and Matplotlib 2.0.1 (<https://matplotlib.org/>). Microsoft Excel 2013 was used for statistical purposes. An unpaired two-tailed t-test (equal variance was assessed with the F-test assuming normal data distribution) was used for the statistical comparison of the data. Function T.TEST was used in Excel.

Data availability

Electron microscopy data has been deposited into the electron microscopy database (EMDB - www.ebi.ac.uk/pdbe/emdb) with accession codes EMD-3781 (TET12SN), EMD-3788 (PYR16SN) and EMD-3789 (TRIP18SN).

The following data is included in figshare [10.6084/m9.figshare.4003398](https://www.figshare.com/figure/4003398):

- List of topologies and circular permutations in file.
- List of all the design sequences in fasta format.
- SAXS scattering curves of constructs presented in the main article.
- Representative models generated by CoCoPOD, including models with best fit to SAXS data.
- Negative-stain density map reconstructions.
- Source code for CoCoPOD platform.

Method only references

44. Eswar, N. *et al.* Comparative protein structure modeling using MODELLER. *Curr. Protoc. protein Sci.* **50**, 2.9.1-2.9.31 (2007).
45. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–12 (2004).
46. McGibbon, R. T. *et al.* MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **109**, 1528–32 (2015).
47. Perez, F. & Granger, B. E. IPython: A System for Interactive Scientific Computing. *Comput. Sci. Eng.* **9**, 21–29 (2007).
48. Ivankov, D. N. *et al.* Contact order revisited: Influence of protein size on the folding rate. *Protein Sci.* **12**, 2057–2062 (2003).
49. Testa, O. D., Moutevelis, E. & Woolfson, D. N. CC+: a relational database of coiled-coil structures. *Nucleic Acids Res.* **37**, D315-22 (2009).
50. Grigoryan, G., Reinke, A. W. & Keating, A. E. Design of protein-interaction specificity gives selective bZIP-binding peptides. *Nature* **458**, 859–864 (2009).
51. Gradišar, H. & Jerala, R. De novo design of orthogonal peptide pairs forming parallel coiled-coil heterodimers. *J. Pept. Sci.* **17**, 100–106 (2011).
52. Zhao, X., Ghaffari, S., Lodish, H., Malashkevich, V. N. & Kim, P. S. Structure of the Bcr-Abl oncoprotein oligomerization domain. *Nat. Struct. Mol. Biol.* **9**, 117 (2002).
53. Oshaben, K. M., Salari, R., McCaslin, D. R., Chong, L. T. & Horne, W. S. The native GCN4 leucine-zipper domain does not uniquely specify a dimeric oligomerization state. *Biochemistry* **51**, 9581–9591 (2012).
54. Wood, C. W. *et al.* CCBUILDER: an interactive web-based tool for building, designing and assessing coiled-coil protein assemblies. *Bioinformatics* **30**,

- 3029–3035 (2014).
55. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
 56. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2**, 1–7 (2015).
 57. Chen, Y. H., Yang, J. T. & Chau, K. H. Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. *Biochemistry* **13**, 3350–3359 (1974).
 58. Schrödinger, LLC. *The {PyMOL} Molecular Graphics System, Version~1.8.* (2015).
 59. Drobnak, I., Vesnaver, G. & Lah, J. Model-Based Thermodynamic Analysis of Reversible Unfolding Processes. (2010).
 60. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes in C++: The Art of Scientific Computing.* (Cambridge University Press, 2002).
 61. *GNU Scientific Library Reference Manual.* (Network Theory Ltd., 2009).
 62. Konarev, P., Volkov, V., Sokolova, A., Koch, M. & Svergun, D. PRIMUS - a Windows-PC based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 (2003).
 63. Förster, S., Apostol, L. & Bras, W. Scatter: Software for the analysis of nano- and mesoscale small-angle scattering. *J. Appl. Crystallogr.* **43**, 639–646 (2010).
 64. Franke, D. *et al.* DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346 (2009).

65. Hura, G. L. *et al.* Comprehensive macromolecular conformations mapped by quantitative SAXS analyses. *Nat. Methods* **10**, 453–454 (2013).
66. Schneidman-Duhovny, D., Hammel, M. & Sali, A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res.* **38**, W540–W544 (2010).
67. Bakan, A., Meireles, L. M. & Bahar, I. ProDy: Protein dynamics inferred from theory and experiments. *Bioinformatics* **27**, 1575–1577 (2011).
68. Petoukhov, M. V. *et al.* New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350 (2012).
69. Shevchenko, A., Tomas, H., Havliš, J., Olsen, J. V & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **1**, 2856–2860 (2007).
70. de la Rosa-Trevín, J. M. *et al.* Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *J. Struct. Biol.* **195**, 93–99 (2016).
71. Abrishami, V. *et al.* A pattern matching approach to the automatic selection of particles from low-contrast electron micrographs. *Bioinformatics* **29**, 2460–2468 (2013).
72. Sorzano, C. O. S. *et al.* A clustering approach to multireference alignment of single-particle projections in electron microscopy. *J. Struct. Biol.* **171**, 197–206 (2010).
73. Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* **157**, 281–287 (2007).
74. Wang, Y. *et al.* Activation of ATF6 and an ATF6 DNA binding site by the endoplasmic reticulum stress response. *J. Biol. Chem.* **275**, 27013–20 (2000).

75. Hornung, V. *et al.* Silica crystals and aluminum salts activate the NALP3 inflammasome through phagosomal destabilization. *Nat Immunol* **9**, 847–856 (2008).
76. Hafner-Bratkovič, I., Benčina, M., Fitzgerald, K. A., Golenbock, D. & Jerala, R. NLRP3 inflammasome activation in macrophage cell lines by prion protein fibrils as the source of IL-1 β and neuronal toxicity. *Cell. Mol. Life Sci.* **69**, 4215–28 (2012).

Supplementary materials

Supplementary Discussion

Figs. S1 to S22

Tables S1 to S7

Supplementary Note

References