Supplementary material for the article:

Ljubetič, A.; Lapenta, F.; Gradišar, H.; Drobnak, I.; Aupič, J.; Strmšek, Ž.; Lainšček, D.; Hafner-Bratkovič, I.; Majerle, A.; Krivec, N.; et al. Design of Coiled-Coil Protein-Origami Cages That Self-Assemble in Vitro and in Vivo. *Nature Biotechnology* **2017**, *35* (11), 1094–1101. https://doi.org/10.1038/nbt.3994

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

Ajasja Ljubetič[*], Fabio Lapenta[*], Helena Gradišar, Igor Drobnak, Jana Aupič, Žiga Strmšek, Duško Lainšček, Iva Hafner-Bratkovič, Andreja Majerle, Nuša Krivec, Mojca Benčina, Tomaž Pisanski, Tanja Ćirković Veličković, Adam Round, José María Carazo, Roberto Melero and Roman Jerala[‡]

* Shared equal contribution.
[‡] Correspondence to: roman.jerala@ki.si

This PDF file includes:

Supplementary Discussion
Figs. S1 to S22
Tables S1 to S7
Supplementary Note
References

Other Supplementary Materials for this manuscript includes the following:

Code for the designed platform is available at github.com/NIC-SBI/CC_protein_origami.

The following data is included in figshare 10.6084/m9.figshare.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.

Nature Biotechnology: doi:10.1038/nbt.3994

Supplementary Discussion

Intrinsic flexibility of CC polyhedral protein cages

Convex polyhedra composed of only of triangular faces are structurally rigid[77], even without fixed vertices, given linkers of reasonable size. The tetrahedron fits these criteria, however the pyramid and triangular prism both contain rectangular faces, which are most likely the predominant sources of flexibility.

Even in the tetrahedron the individual CC segments are connected by at least 5 amino acids long linkers, which bestow a large conformational freedom upon designed assemblies. Designing linker with high rigidity might lead to structures with decreased flexibility and is an interesting future direction.

Recently several techniques have been developed to ease the characterization of flexible/dynamic proteins assemblies in solutions[33].

Verification of SAXS sensitivity to the structure of polyhedral CC cages

In order to show that SAXS is sensitive to conformational changes of proteins and therefore a reliable tool for confirming the structure of designed polyhedral proteins solution, we compared the experimental SAXS curve for TET12SN to theoretical SAXS profiles calculated for multiple model systems equal in size to the TET12SN design (Fig. S12A-C). The chosen model systems were an extended coiled-coil (a crystal structure of a natural coiled-coil tropomyosin, PDB ID 1c1g, truncated to contain the same number of amino acid residues as TET12SN), an ideal extended helix with an amino acid sequence matching that of TET12SN (generated using Chimera), and flexibly extended coiled-coil protein (models of the protein cage variant TET12Sscr, serving as a negative control of the presented design strategy). The selected model systems describe the experimental SAXS profile significantly worse than the polyhedral protein cages. Rg values determined for the model systems range from 5-12 nm. Maximum particle dimension (Dmax) for the extended coiled-coil and the extended helix is 37 nm and 70 nm respectively, and ranges from 13 to 22 nm for different TET12Sscr models. The extracted Rg and Dmax values are substantially higher in comparison to those obtained from experimental SAXS curves for tetrahedral protein cages (Table S7). This offers an additional proof that the designed structures have been achieved.

Furthermore, to demonstrate SAXS can also distinguish between different conformations of the tetrahedral protein cage, we constructed a set of morphed structures starting from the largest volume (most geometrically symmetric) structure of TET12SN (Fig 2B, left panel) and ending at highly skewed conformation (Fig. S12D). The morphed trajectory was created using USCF Chimera. This stepwise change in conformation is indeed reflected in SAXS profiles, as a gradual transition of calculated scattering curves, morphing from the curve for the largest volume tetrahedral structure to the profile of the skewed protein cage that has a lower volume cavity. The same procedure was repeated for PYR16SN (Fig. S12E) and TRIP18SN (Fig. S12F).

Determination of SAXS model free parameters

The determined model-free parameters for each design are listed in Table S7. Radii of gyration Rg, obtained through Guinier analysis range from 3.3 to 4.2 nm for larger protein cages and are only slightly larger than what is expected for globular protein

of the same size (~3 nm)[78]. The molecular masses determined from scattering at zero angle were within the expected range. Radius of gyration of the cross-section, R_c, was also determined and values ranged from 1.6-2 nm, which is somewhat larger than R_c of 1.3 nm, determined from an extended coiled-coil of similar size (truncated tropomyosin PDB ID: 1c1g). Normalized Kratky plots confirmed that the proteins were indeed structured, and the plateau that was observed at higher values of the scattering vector *q* indicated the presence of a dynamic component. This is consistent with the analysis of Porod exponents. The latter, with two exceptions, were in the range of 3.6-4, indicating limited flexibility. The two exceptions are TETS and TET12$_{1.6}$S-f$_5$b, for which the determined Porod exponents were 3.2 and 3.3 respectively, indicating more pronounced flexibility.

Agreement of SAXS data to largest volume CC polyhedral protein cages

The conformations of polyhedral protein cages that fit SAXS data best were not ideally geometrically symmetrical but showed slight skewed deviations (Fig. 2B, Fig. 4B). A comparison between largest volume structures and best fitting conformations is shown in Fig. S13A for TET12SN, TET12$_{2.3}$SN-f$_5$, PYR16SN and TRIP18SN. The theoretical SAXS profiles calculated from largest volume models exhibit distinct minima and maxima, which are less pronounced in experimental SAXS profiles. However, the overall trends in theoretical SAXS profiles are recapitulated in experimental SAXS profiles (Fig. S13B and Fig. S13C). The discrepancies are most likely due to intrinsic flexibility present in this type of protein assemblies[33].

Nature Biotechnology: doi:10.1038/nbt.3994

Supplementary Figures

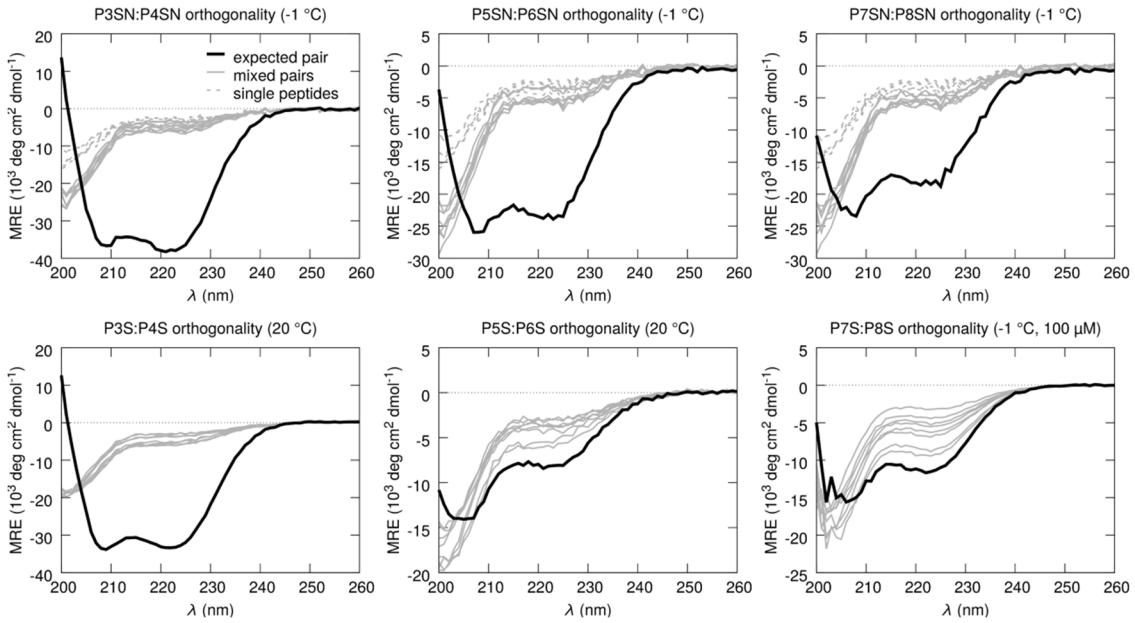
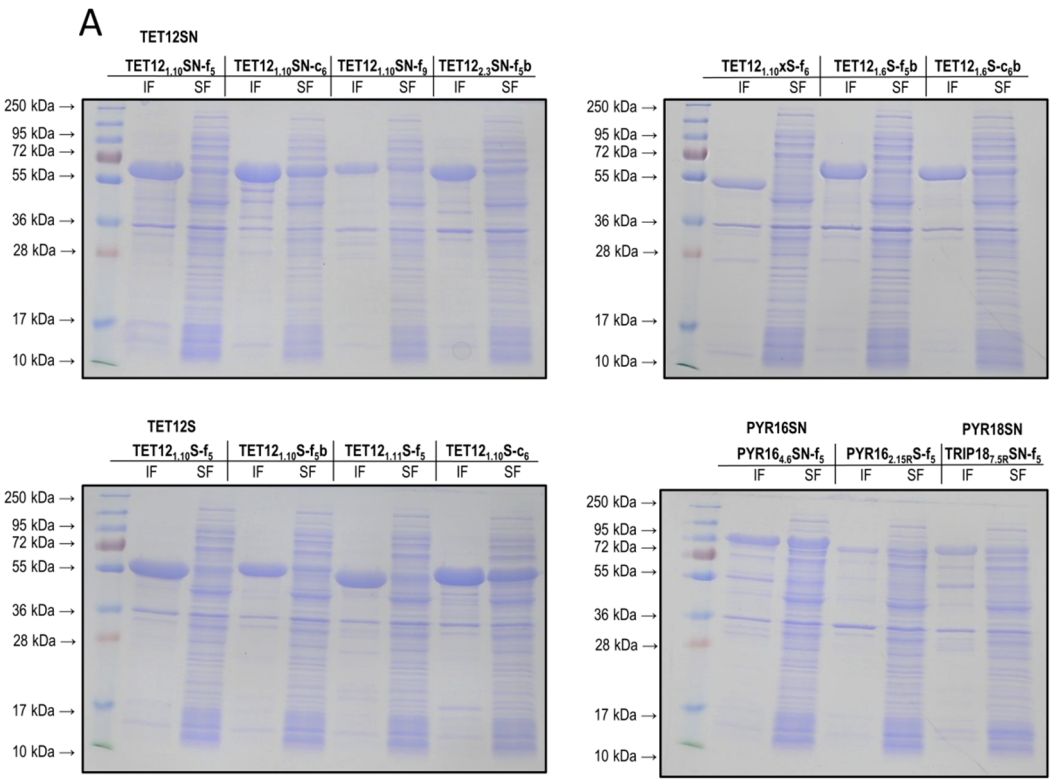


Fig. S1: Orthogonality of the supercharged coiled-coil peptide set.

Circular dichroism spectra for 1:1 molar mixtures of the target pairs are shown in black; all other off-target pairwise combinations of heterodimeric coiled coils are shown in gray; single peptides are shown with a dashed gray line. Both designed sets were tested: P_nSN (P3SN, P4SN, P5SN, P6SN, P7SN, P8SN) and P_nS (P3S, P4S, P5S, P6S, P7S, P8S). The designed peptide partners (shown in black) exhibited a higher helical content than any other combination of peptides (gray), thus indicating an intrinsic preference for peptides to bind to their designed partner. MRE = mean residue ellipticity. Peptide sequences are listed in Table S1.

Nature Biotechnology: doi:10.1038/nbt.3994



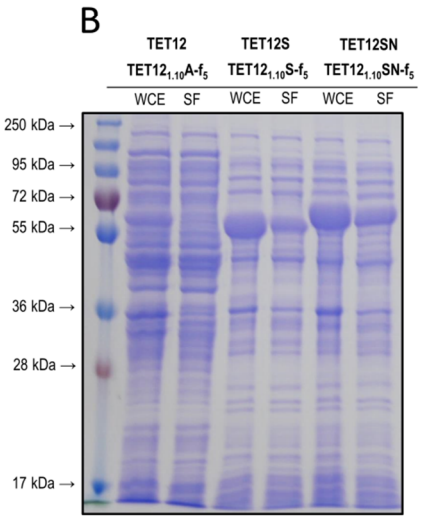


Fig. S2: Production of the soluble designed-protein origami variants in bacteria.

(**A**) SDS-PAGE showing insoluble fractions (IF) and soluble fractions (SF) of different protein variants expressed as described in the Material and Methods section. Soluble fractions of cellular lysate were loaded after sonication and centrifugation; comparable amounts of insoluble fraction were resuspended and loaded in SDS-PAGE. Simplifed names of constructs are given in the topmost row.
(**B**) Comparison of the solubility of TET12$_{1.10}$S-f$_5$, TET12$_{1.10}$A-f$_5$, and TET12$_{1.10}$SN-f$_5$ expressed in bacteria. Whole-cell extracts (WCE) and soluble fractions (SF) were obtained (as previously described), analyzed by SDS-PAGE and stained with Coomassie blue.

Nature Biotechnology: doi:10.1038/nbt.3994

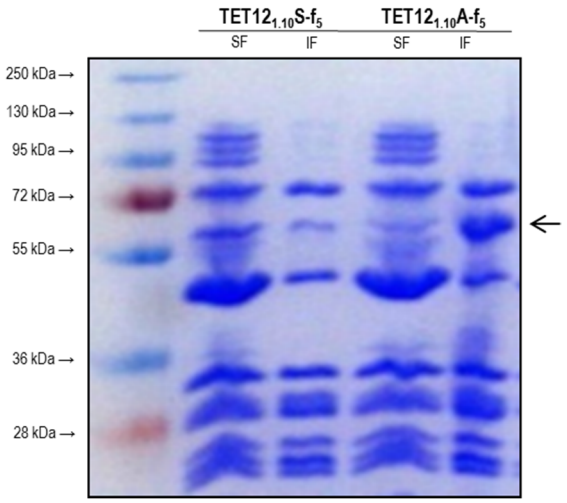
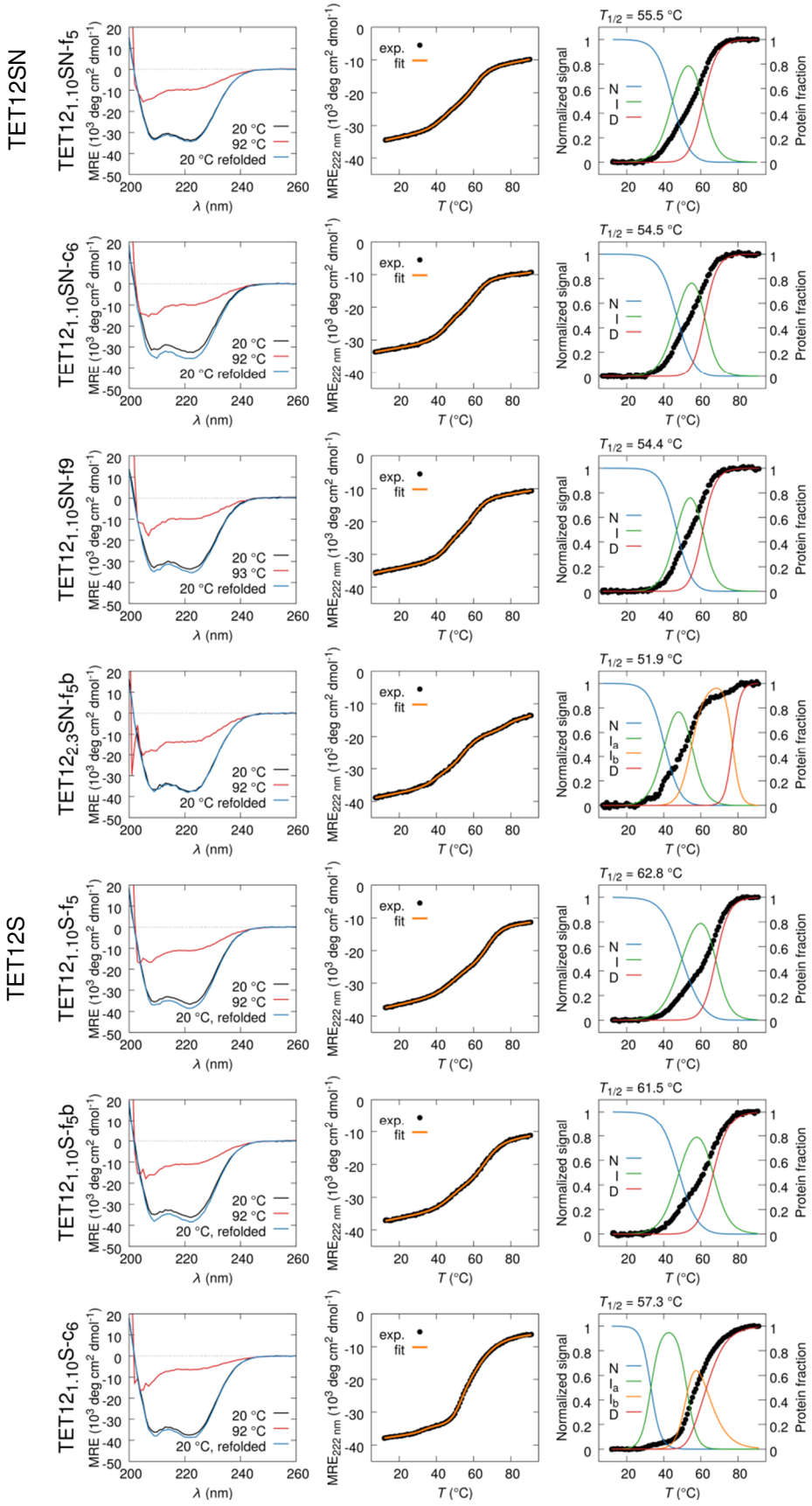


Fig. S3: Comparison of the synthesis of polypeptides TET12$_{1.10}$S-f$_5$ (TET12S) and TET12$_{1.10}$A-f$_5$ (TET12) by *in vitro* translation.

Both polypeptides were synthesized *in vitro* using a cell-free transcription/translation kit, PURExpression (New England Biolabs). TET12$_{1.10}$S-f$_5$ (a variant of a tetrahedron-forming polypeptide composed of second-generation CC peptide segments with an increased number of negatively charged amino acids) was better expressed in soluble fraction (SF), while the first generation TET12$_{1.10}$A-f$_5$ polypeptide was mostly present in the insoluble fraction (IF).

Nature Biotechnology: doi:10.1038/nbt.3994



Nature Biotechnology: doi:10.1038/nbt.3994

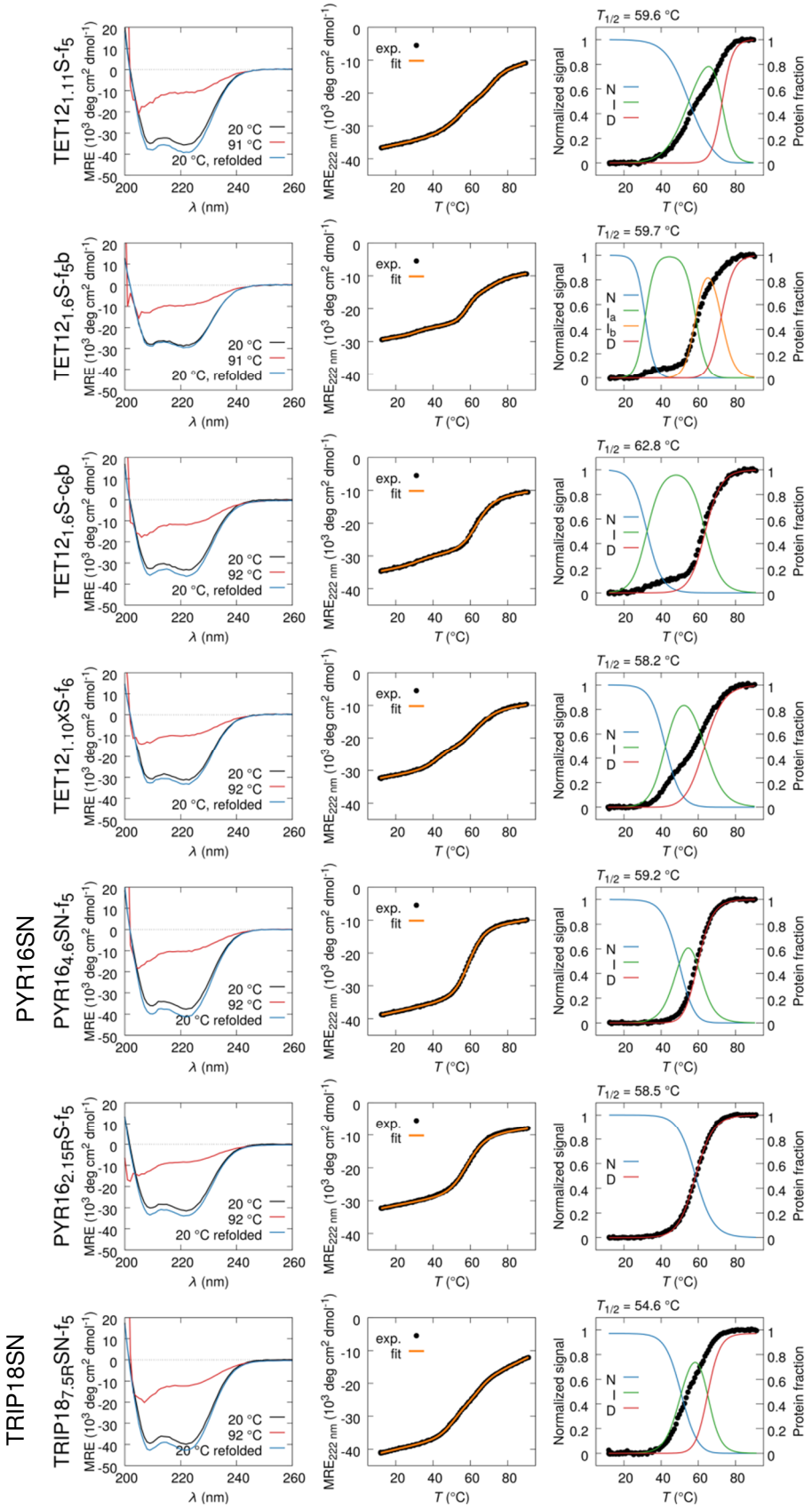


Fig. S4: Secondary structure and thermal stability of the protein polyhedral constructs. Three panels are shown for each polyhedral variant; the left panel shows CD spectra at 20°C, 92°C, and after rapid cooling from 92 to 20°C (refolded). The central panel shows the temperature trace of the CD signal at 222 nm, along with its thermodynamic model fit. The right panel shows the normalized CD temperature trace (0, corresponding to the native state, and 1, corresponding to the final temperature-denatured state) and the fraction of protein found in the native (N), intermediate (I, or I_a and I_b, in case of four-state unfolding), and denatured (D) states at each temperature. These fractions were calculated by fitting an equilibrium thermodynamic model to the experimental data (see Materials and Methods). Most variants exhibited two apparent transitions, while three transitions were required to describe the temperature traces of TET12$_{2.3}$SN-f$_5$, TET12$_{1.10}$SN-c$_6$, and TET12$_{1.6}$S-f$_5$b. The mid-transition temperature ($T_{1/2}$) was determined as the temperature at which the normalized CD signal was halfway between the native and denatured baselines. All experiments were performed at least two times. Simplified names are written for the representative designs.

Nature Biotechnology: doi:10.1038/nbt.3994

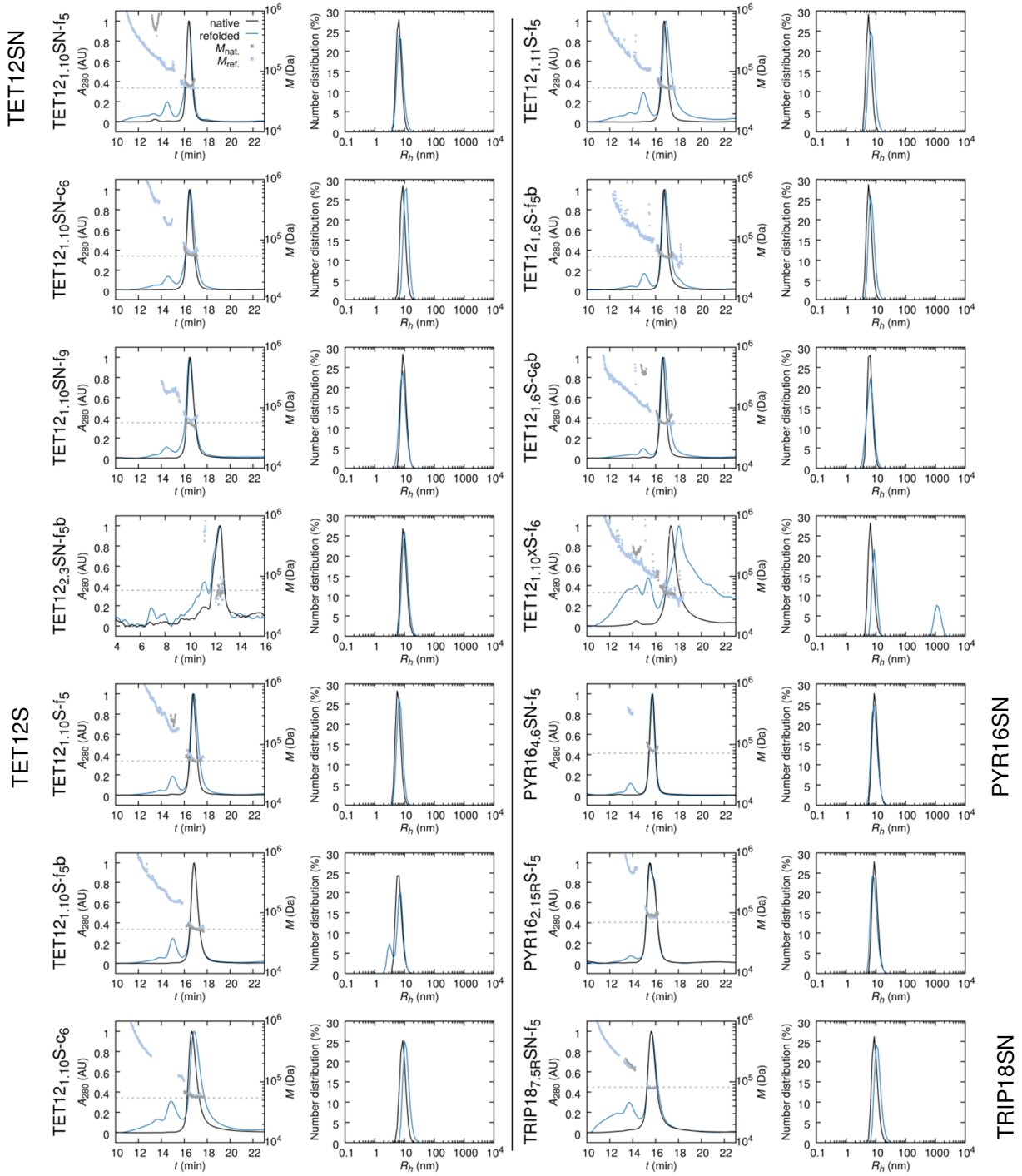


Fig. S5: Multimerization state and size dispersity of isolated designed protein origami particles, as determined by SEC-MALS and DLS.

Two experiments were performed for all polyhedral cages: SEC-MALS and DLS were measured both before (black traces) and after (blue traces) a cycle of thermal denaturation (to 90°C) and rapid cooling back to room temperature. The left panel for each construct represents the size-exclusion chromatogram (normalized so that major peak maximum is set to 1) with an overlay of molar masses calculated from SEC-MALS

Nature Biotechnology: doi:10.1038/nbt.3994

measurements. The dotted line represents the expected molar masses of the monomer. The right panel for each construct shows the numeric distribution of hydrodynamic radii (R_h) calculated from batch DLS measurements. For most constructs (with the notable exception of TET12$_{1.10}$xS-f$_6$), thermal denaturation and rapid cooling did not cause a major amount of oligomerization or aggregation. All experiments were preformed at least two times. Simplified names are written for the representative designs.

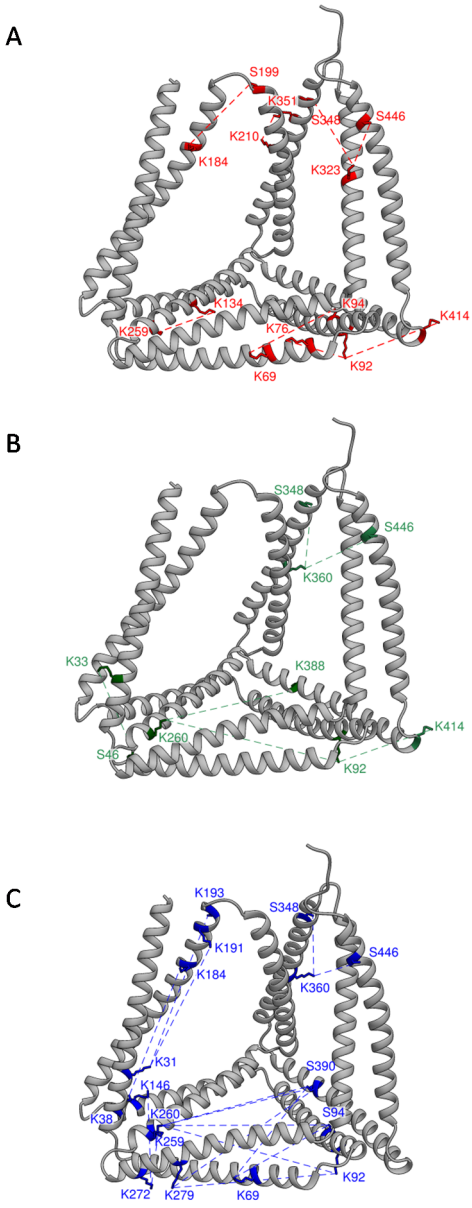


Fig. S6: Identified cross-linked amino acid residues in chemically cross-linked TET12SN.

Cross-links identified in cross-linking experiments with (A) DSS, (B) BS(PEG)$_5$, or (C) BS(PEG)$_9$, analyzed via proteolytic cleavage and mass spectrometry mapped to the model of TET12SN that best fit the SAXS analysis. The most frequently observed informative cross-links between different segments are presented (details in Table S6). Due to high sequence similarity, some of the identified peptide pairs could not be unambiguously ascribed to a specific peptide pair (e.g., the connection between K360 and S446 in panels B and C and a pair with a connection between K360 and S348). Accordingly, we could not differentiate between connections K323-S446 and K323-S348 in panel A. Table S5 presents a summary of all of the identified cross-links.

Nature Biotechnology: doi:10.1038/nbt.3994

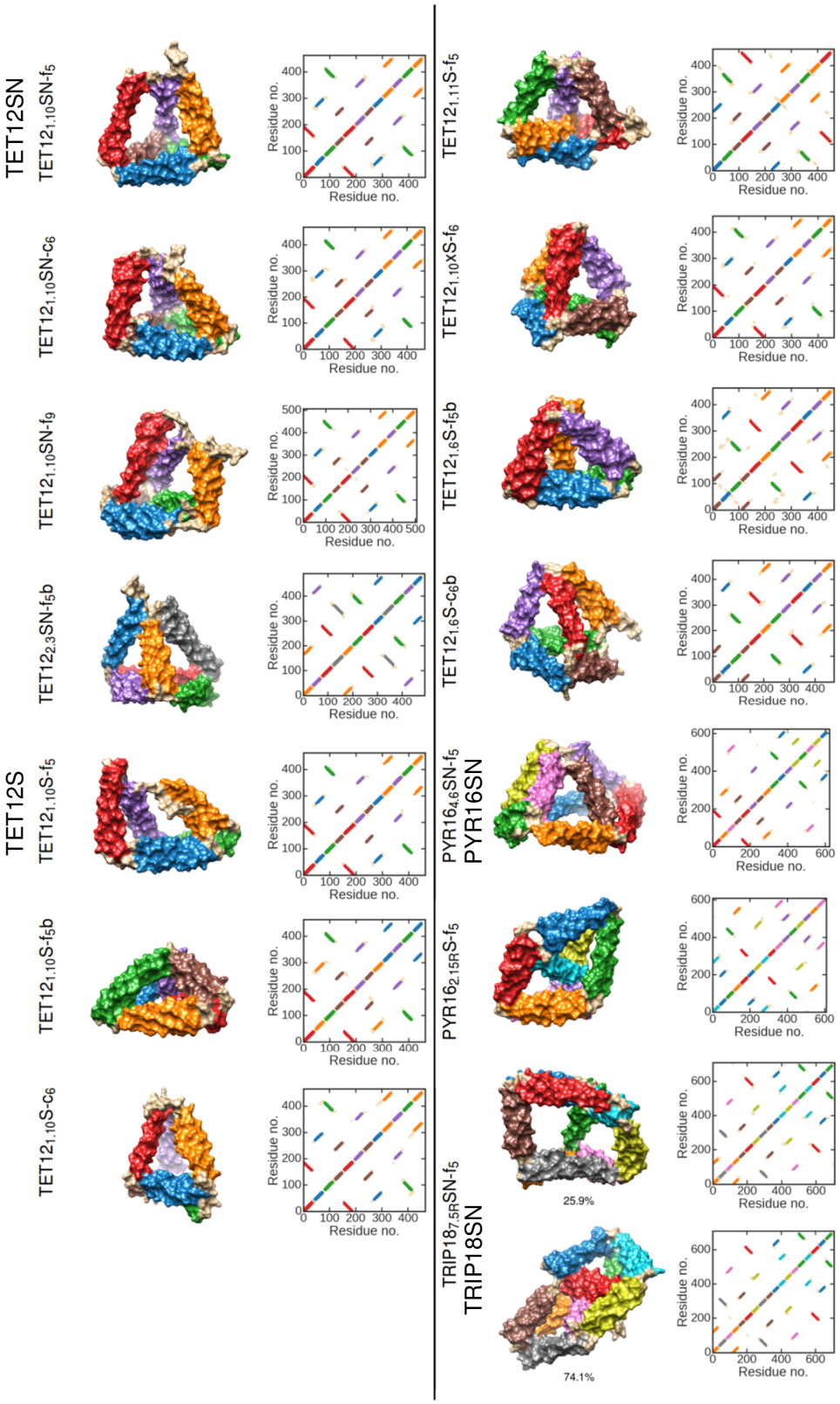


Fig. S7: Molecular models and contact maps of the designed protein origami variants.

For every design, the model with the best agreement to experimental SAXS data (shown in Fig. S9) is presented. Residues were deemed to be in contact if their alpha carbons were less than 10 Å apart. Variants with the same topology have similar contact maps. In the case of the construct TRIP18$_{7.5R}$SN-f$_5$, the orange points correspond to the right, rectangular model, while the black points mark contacts in the left, oblique model. The design nomenclature is presented in Table S2. Simplified names are written for the representative designs.

Nature Biotechnology: doi:10.1038/nbt.3994

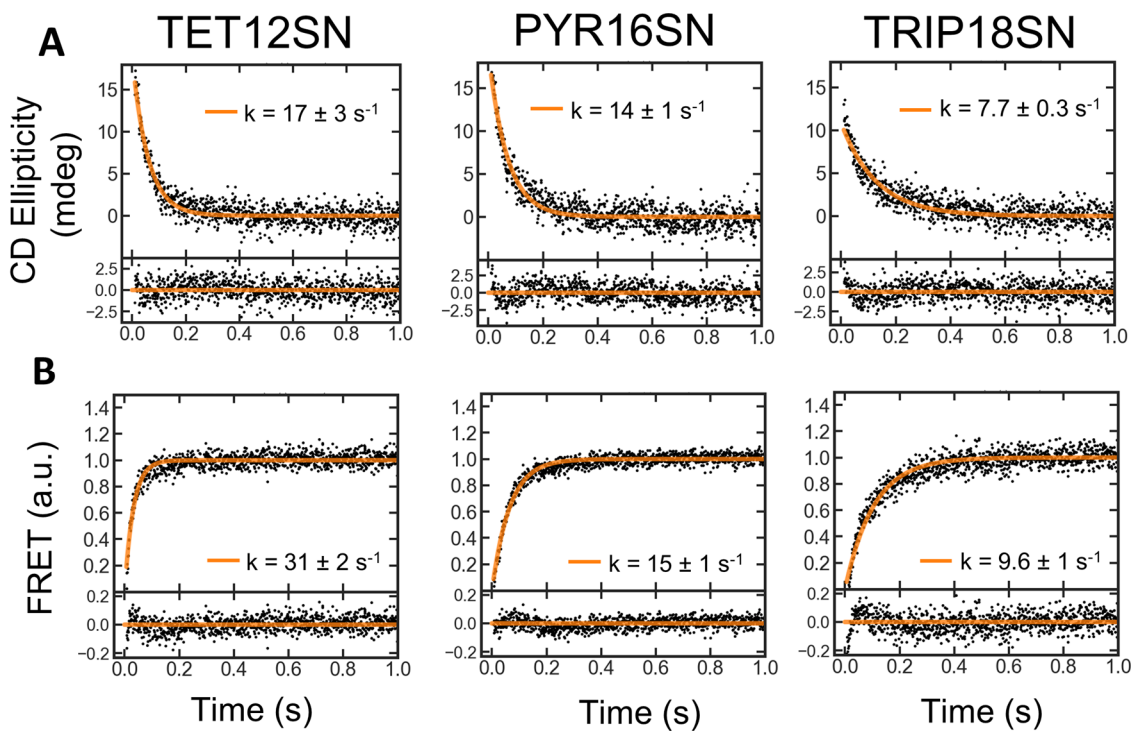
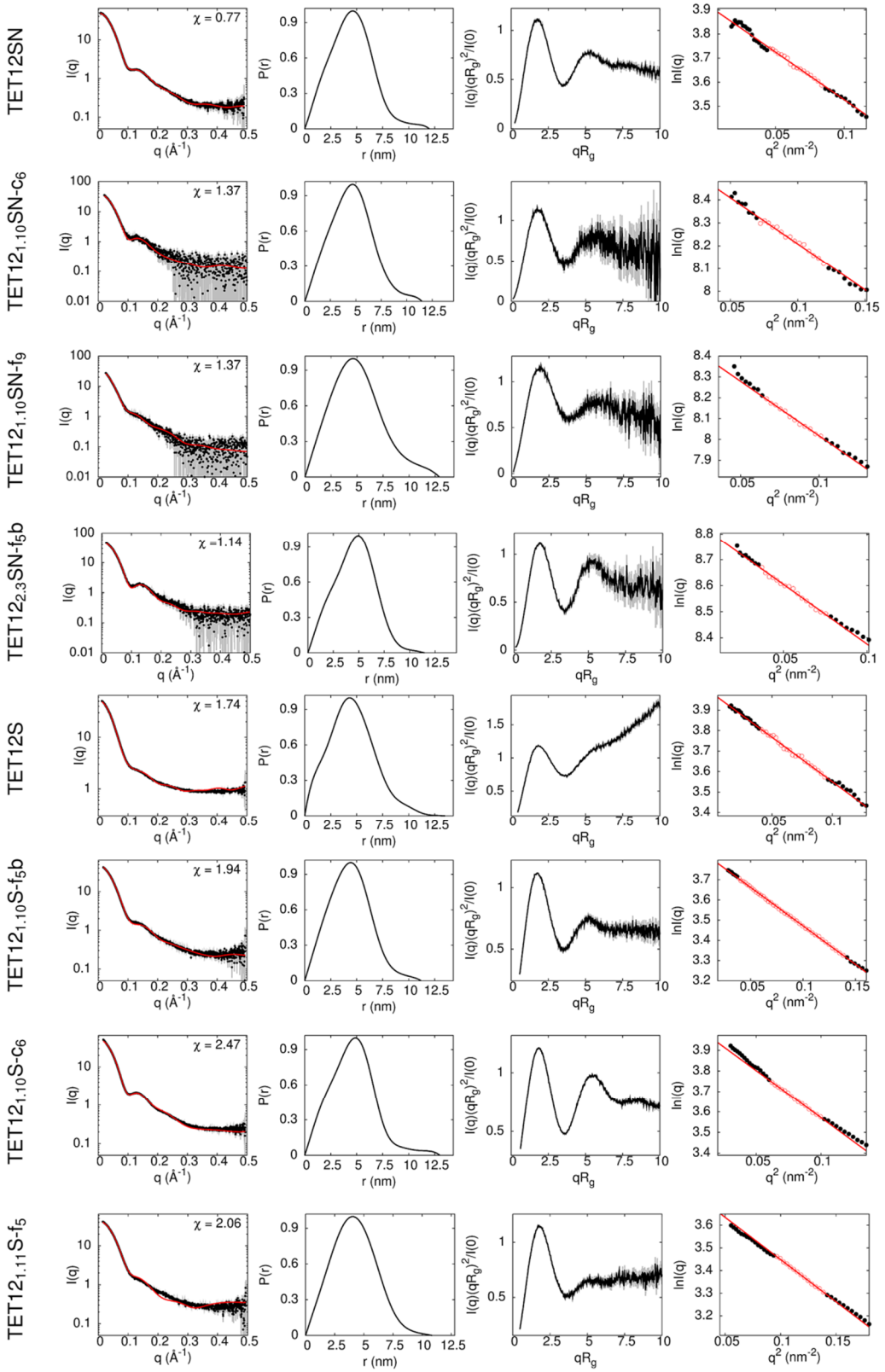


Fig. S8: Folding kinetics of polyhedral designs.

(**A**) Stopped-flow CD kinetics at 225nm of protein refolding in 1 M guanidinium hydrochloride (black). (**B**) Stopped-flow kinetics of polyhedra labeled with Cy3 and Cy5 on the N and C terminals. Cy3 was excited and the emission of Cy5 measured. A one component fit is shown (orange) and the folding rate displayed in inset. An average of three repetitions was used to obtain the experimental trace for CD measurements. For FRET measurements 7, 6 and 9 traces were averaged. The error was obtained by fitting each trace independently and obtaining the standard deviation of the folding rates. The residuals are show at bottom of each panel.

Nature Biotechnology: doi:10.1038/nbt.3994



Nature Biotechnology: doi:10.1038/nbt.3994

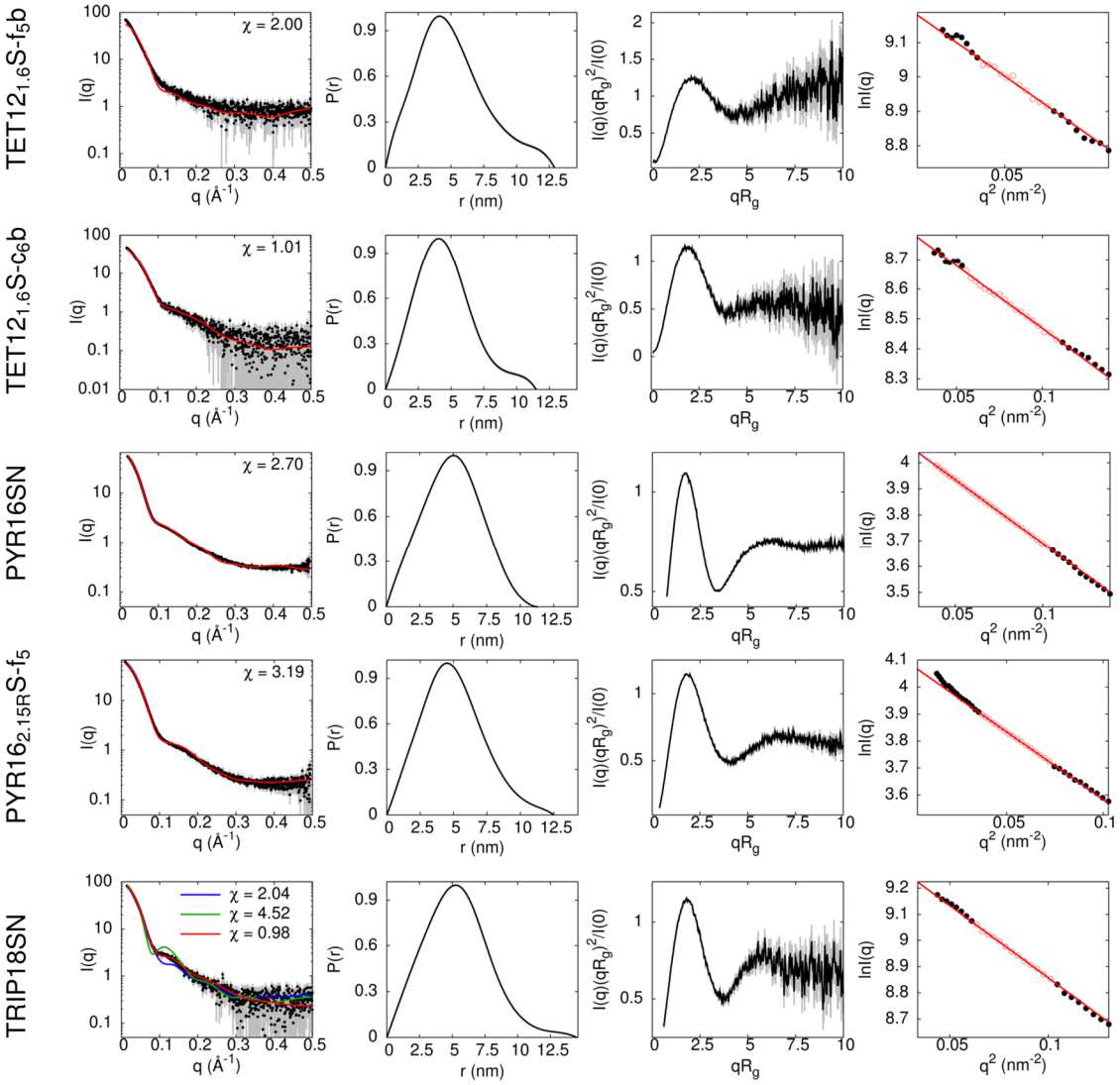


Fig. S9: SAXS analysis of the size and shape of the designed protein origami variants.

For every design, comparisons between the theoretical (red line) and observed (black dots) scattering curves, pair-distance distribution functions, Kratky plots, and Guinier plots are presented. Model-free parameters obtained from Guinier analysis and pair-distance distribution functions are presented in Table S7. For TRIP18SN the calculated SAXS profile for the two-component system is shown in red, while the blue and green line correspond to SAXS profiles calculated for the oblique and rectangular prism, respectively.

Nature Biotechnology: doi:10.1038/nbt.3994

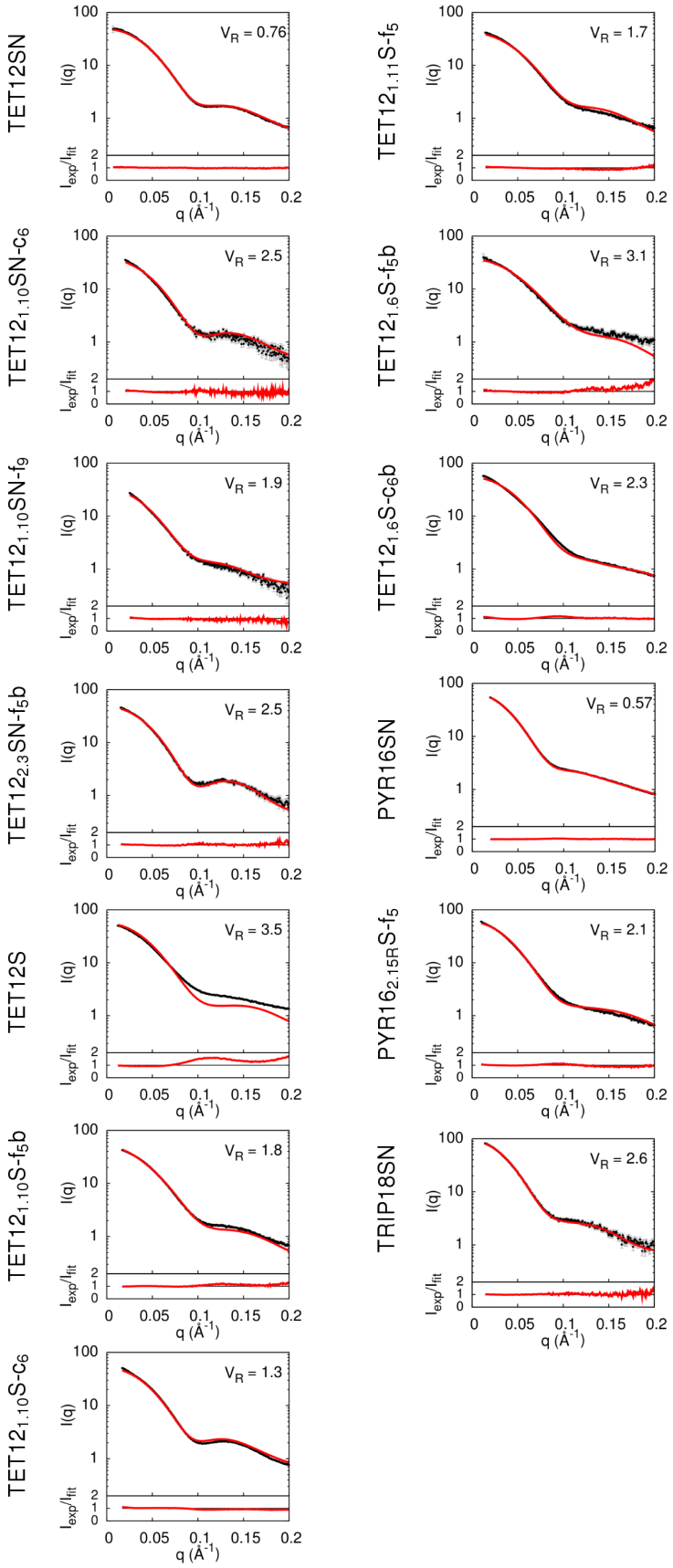


Fig. S10: Fitting of SAXS profiles using the V_R metric.

Theoretical profiles (red line) were calculated from the same structures as in Fig. S9 and compared to experimental SAXS profiles (black dots) using FoXS[66]. Similarly as with the χ metric, model polyhedral protein cages fit well to experimental SAXS data also according to the V_R metric. In case of TRIP18SN fitting with FoXS predicts a slightly different ratio of the two conformations (65%:35% instead of 75%:25% obtained with OLIGOMER in favor of the oblique triangular prism).

Nature Biotechnology: doi:10.1038/nbt.3994

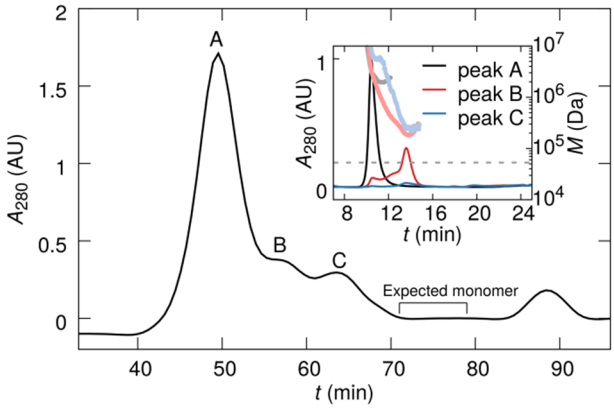


Fig. S11: Characterization of scrambled protein origami (TET12SScr).
TET12SScr was isolated from cell lysate with NiNTA chelating chromatography, after which it was analyzed on an SEC column. Unlike other designed-protein origami variants (which eluted as monomers in the range of 70 to 75 minutes after injection), there was no elution of a TET12SScr monomer in the expected elution range. In order to confirm that the monomer did not elute at an earlier time, SEC-MALS was performed on samples taken from peaks A, B and C. As shown in the SEC-MALS chromatograms (inset), no monomer was found to be present. The expected molecular weight of the monomer is shown as a horizontal dashed line.

Nature Biotechnology: doi:10.1038/nbt.3994

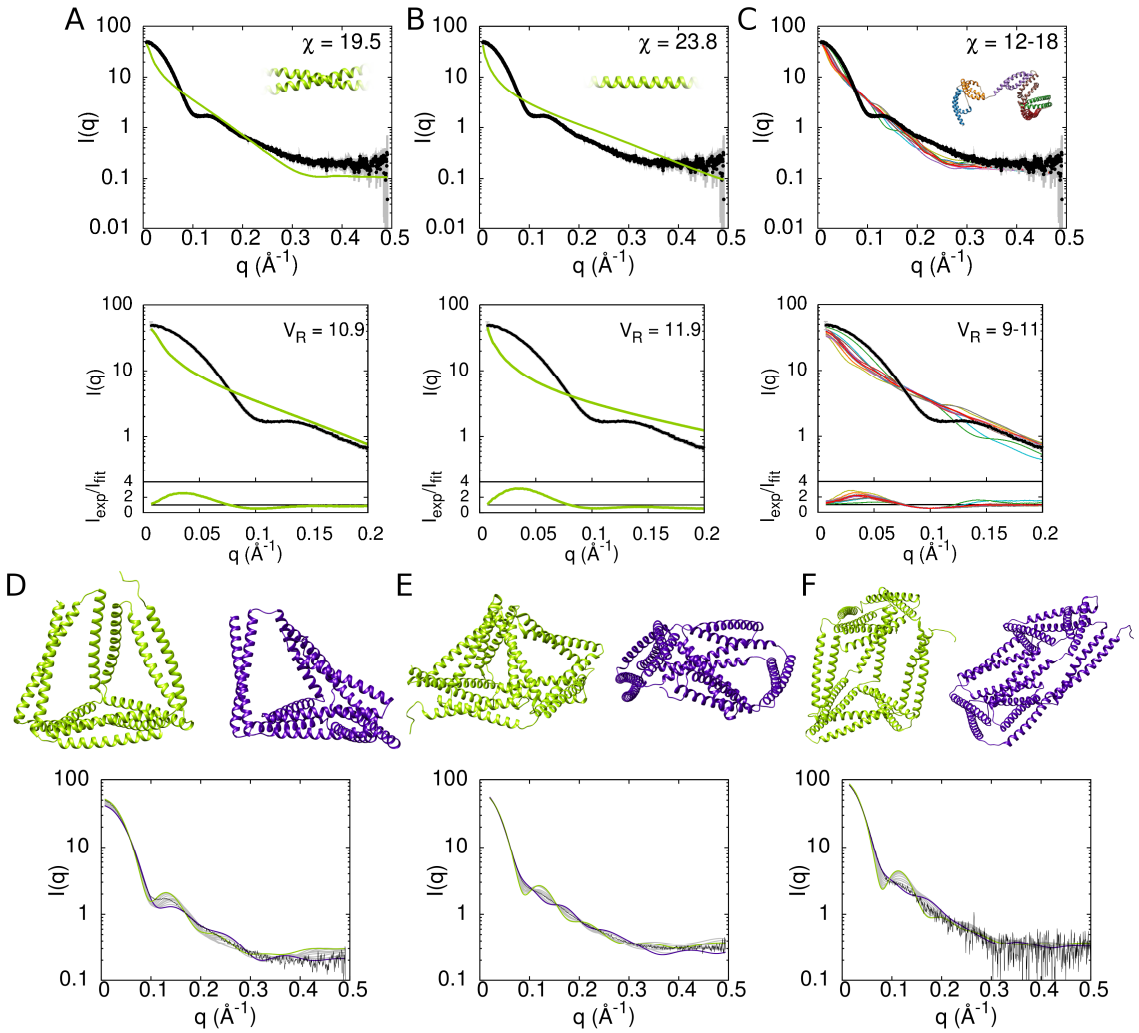


Fig. S12: SAXS as a tool for studying conformation of polyhedral protein cages.
Comparison of SAXS profiles calculated from an (**A**) extended coiled coil, (**B**) extended helix and (**C**) models of negative control variant TET12SScr, to experimental SAXS profile for TET12SN. Model systems are of the same sequence length as TET12SN. The fits are calculated using the χ (left panels) as well as V_R metric (right panels). (**D,E,F**) Calculated SAXS profiles for the largest volume conformations (structure and scattering curve shown in green), and the most skewed protein cage models (shown in violet). Scattering curves for morphed structures are depicted in gray. Panels D, E, and F correspond to TET12SN, PYR16SN, and TRIP18SN respectively.

Nature Biotechnology: doi:10.1038/nbt.3994

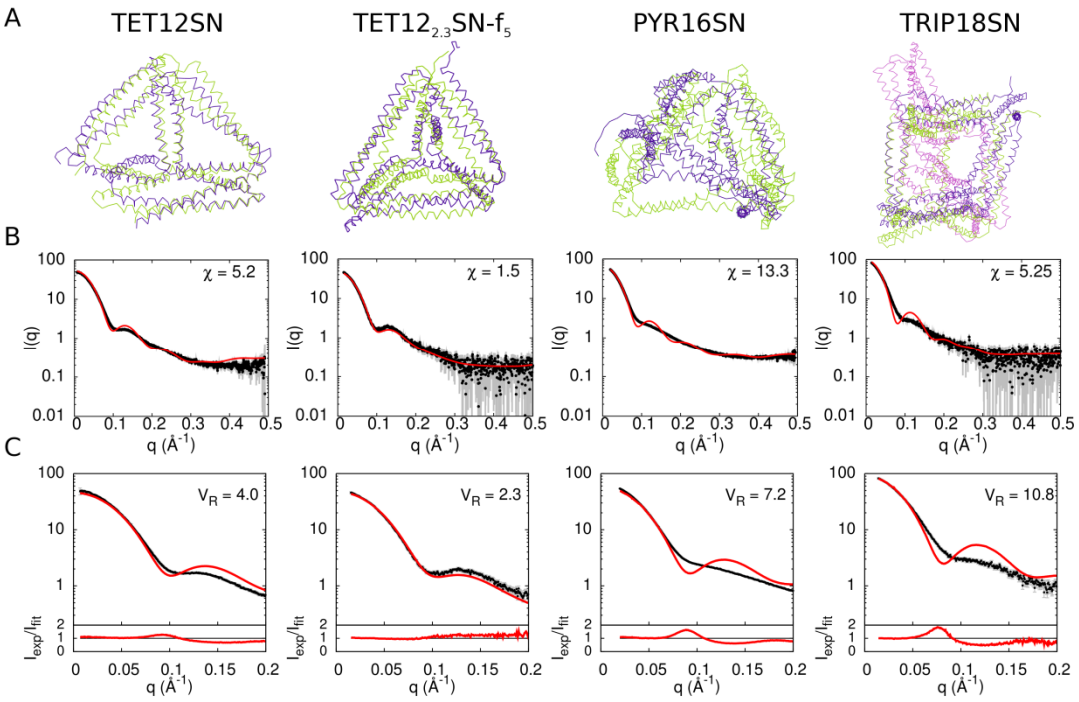


Fig. S13: Comparison of representative largest volume polyhedral protein cages to SAXS data.

(**A**) Superposition of the largest volume polyhedral protein cages (green, displayed also in Fig 2 and Fig 4) and structures displaying the best fit to experimental SAXS profiles (violet, also in Fig 5). (**B**) Comparison of theoretical scattering curves calculated from the largest volume structures to experimental SAXS profiles. The fit is evaluated using the χ metric. (**C**) Comparison of theoretical scattering curves calculated from largest volume structures to experimental SAXS profiles. The fit is evaluated using the V_R metric.

Nature Biotechnology: doi:10.1038/nbt.3994

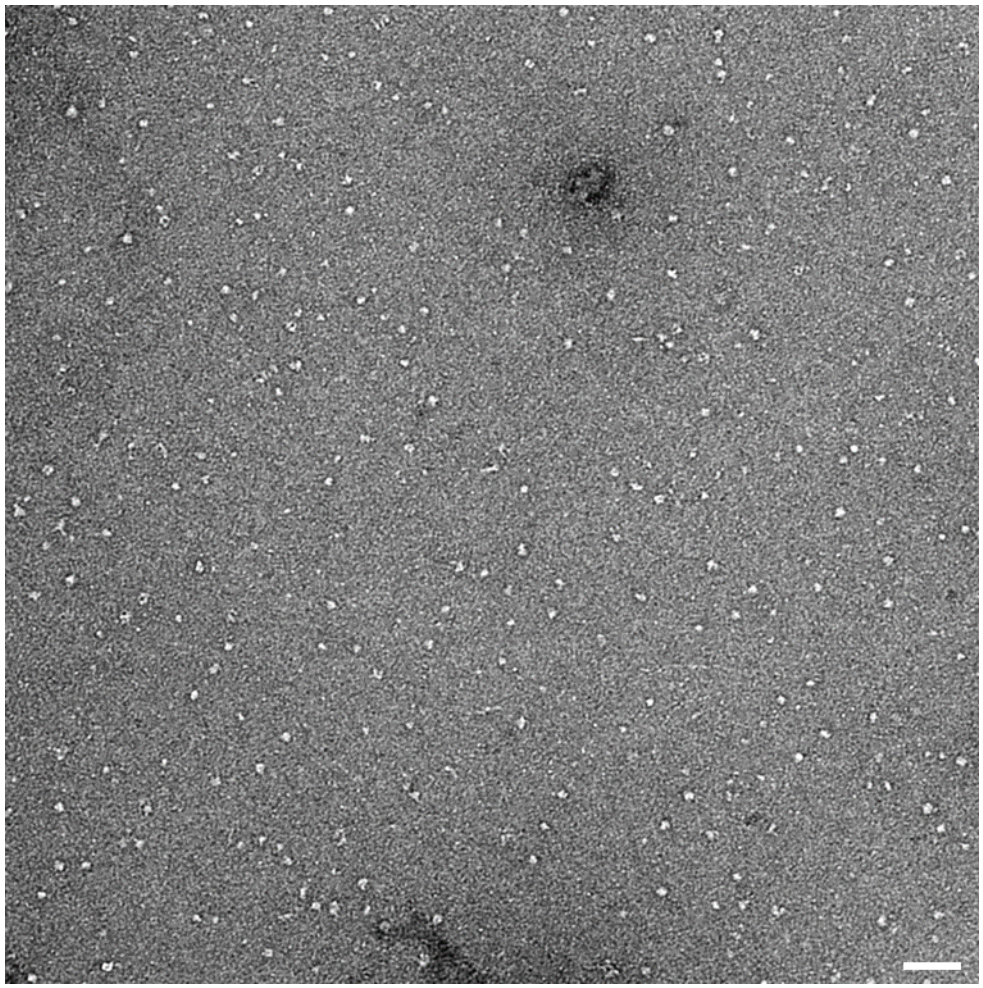


Fig. S14: Representative electron microscopy micrograph of TET12 SN.
One representative micrograph of negative stain EM experiments to show that the images of individual complexes are clearly distinguishable. The scale bar denotes 50 nm.

Nature Biotechnology: doi:10.1038/nbt.3994

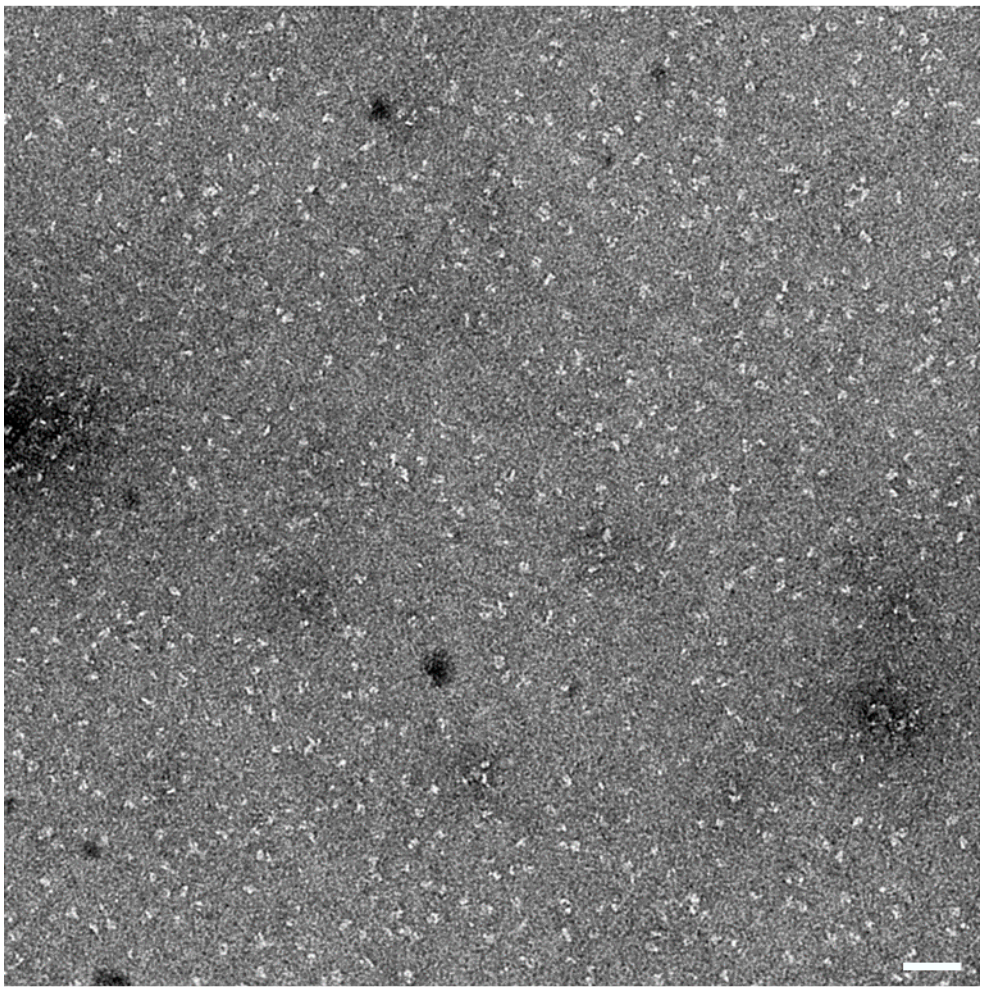


Fig. S15: Representative electron microscopy micrograph of PYR16SN.

One representative micrograph of negative stain EM experiments to show that the images of individual complexes are clearly distinguishable. The scale bar denotes 50 nm.

Nature Biotechnology: doi:10.1038/nbt.3994

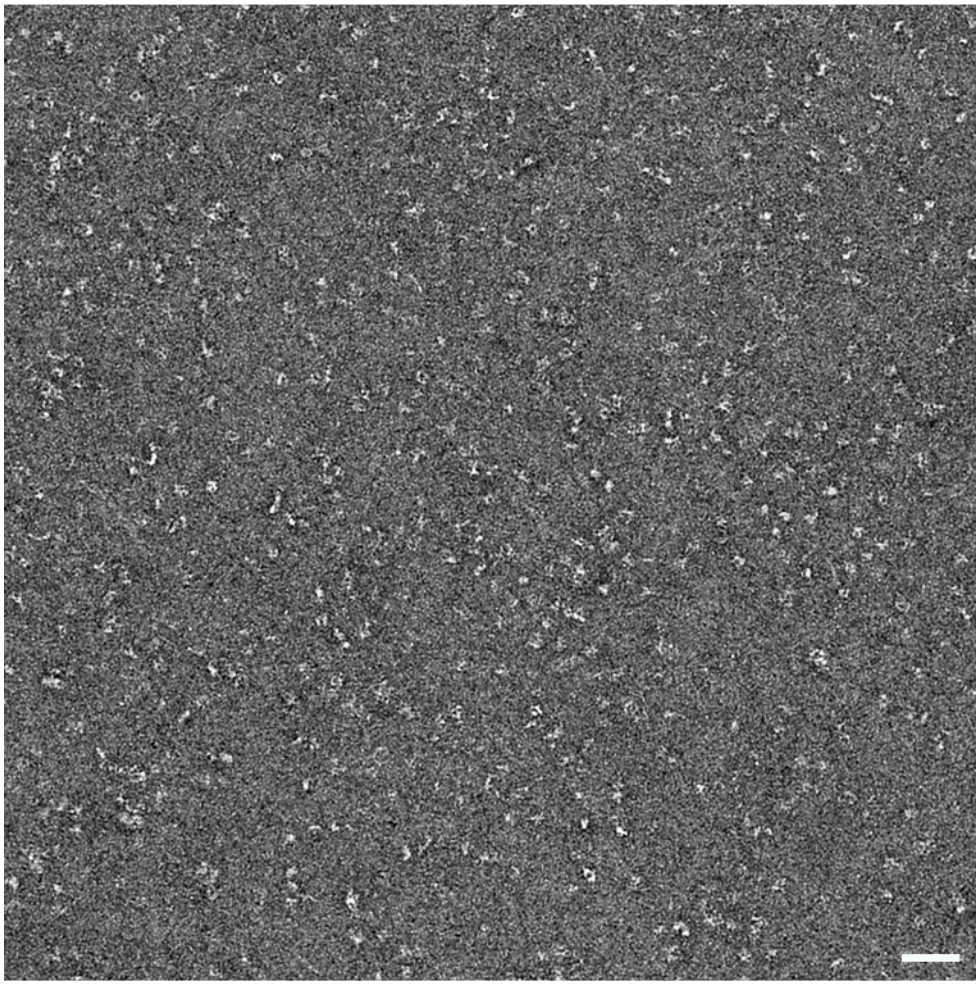


Fig. S16: Representative electron microscopy micrograph of TRIP18SN.
One representative micrograph of negative stain EM experiments to show that the images of individual complexes are clearly distinguishable. The scale bar denotes 50 nm.

Nature Biotechnology: doi:10.1038/nbt.3994

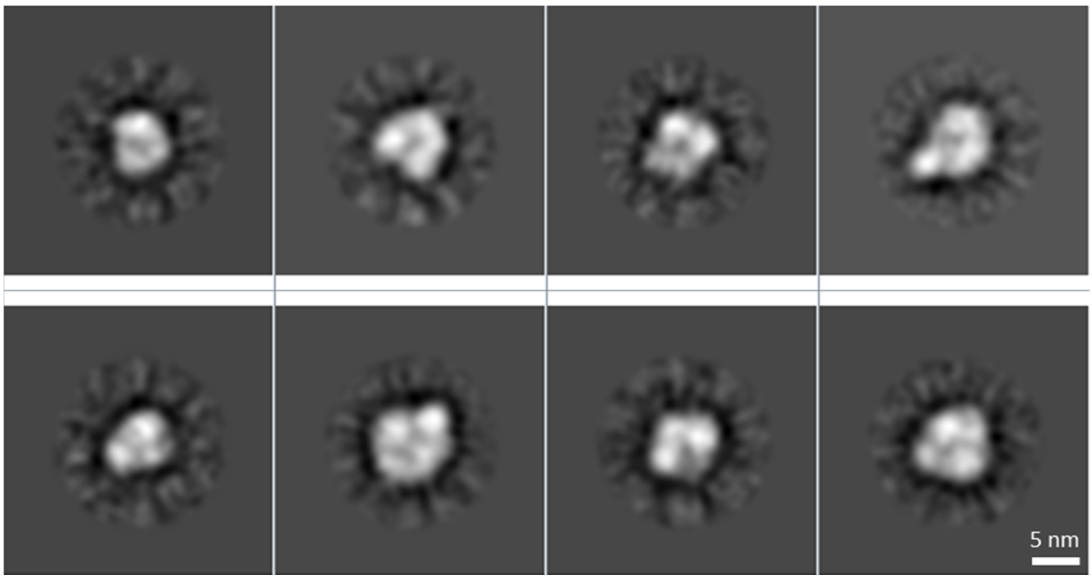


Fig. S17: TET12SN class averages from negative stain TEM micrographs.
Gallery of selected reference-free two-dimensional (2D) averages. The scale bars denotes 5 nm.

Nature Biotechnology: doi:10.1038/nbt.3994

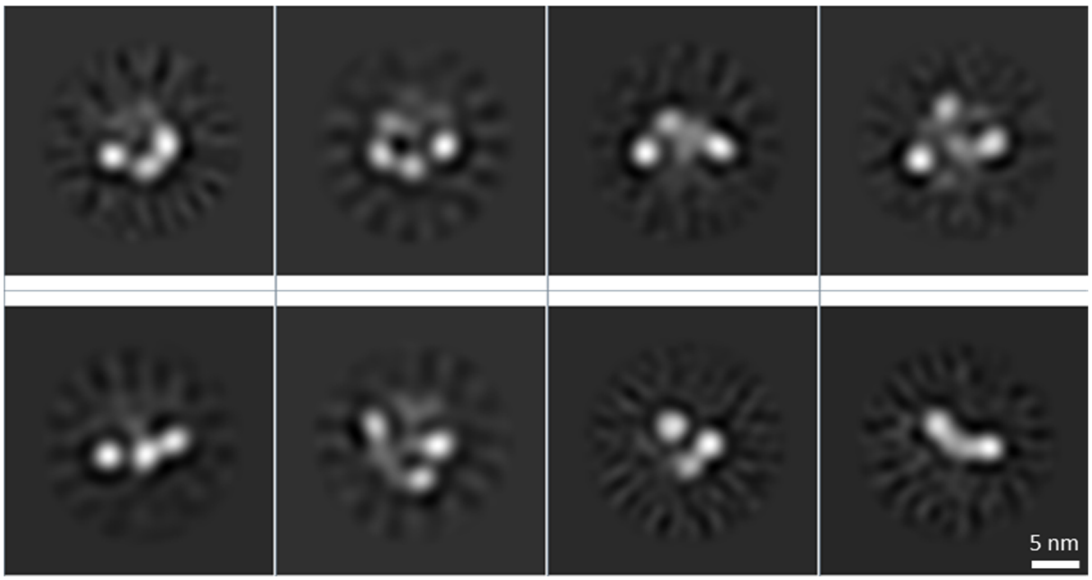


Fig. S18: PYR16SN class averages from negative stain TEM micrographs.
Gallery of selected reference-free two-dimensional (2D) averages. The scale bar denotes 5 nm.

Nature Biotechnology: doi:10.1038/nbt.3994

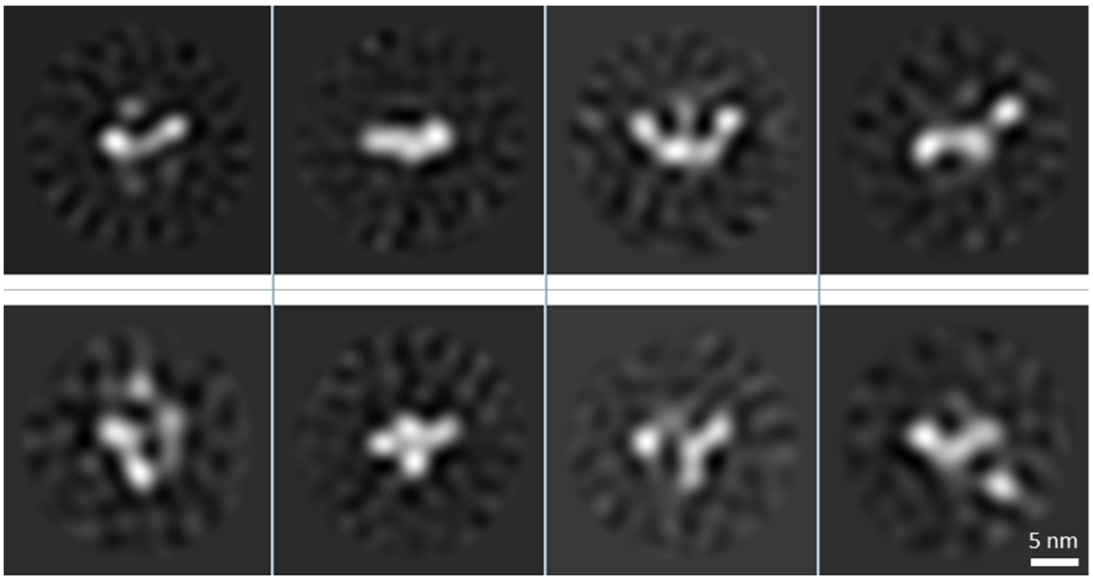


Fig. S19: TRIP18SN class averages from negative stain TEM micrographs.
Gallery of selected reference-free two-dimensional (2D) averages. The scale bar denotes 5 nm.

Nature Biotechnology: doi:10.1038/nbt.3994

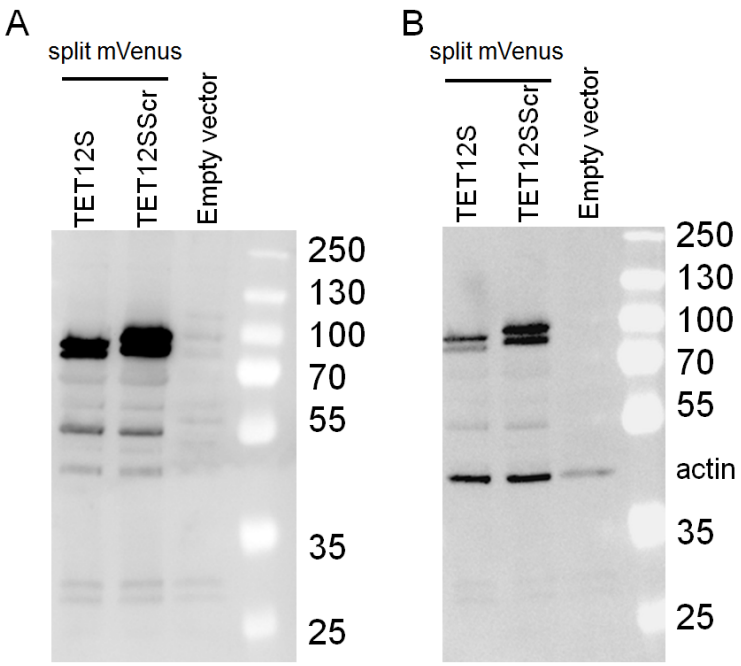


Fig. S20: Control of production of TET12S and TET12SScr in mammalian cell line HEK293.

(A) Western blot analysis of the protein-expression level of TET12S$^{split\text{-}mVenus}$ and TET12SScr$^{split\text{-}mVenus}$ was confirmed by Western blot analysis with antibodies against the fluorescent proteins. (B) After detection of GFP specific bands, membrane was incubated with antibodies against β-actin, which serves as loading control. Representative of 2 Western blots is shown.

Nature Biotechnology: doi:10.1038/nbt.3994

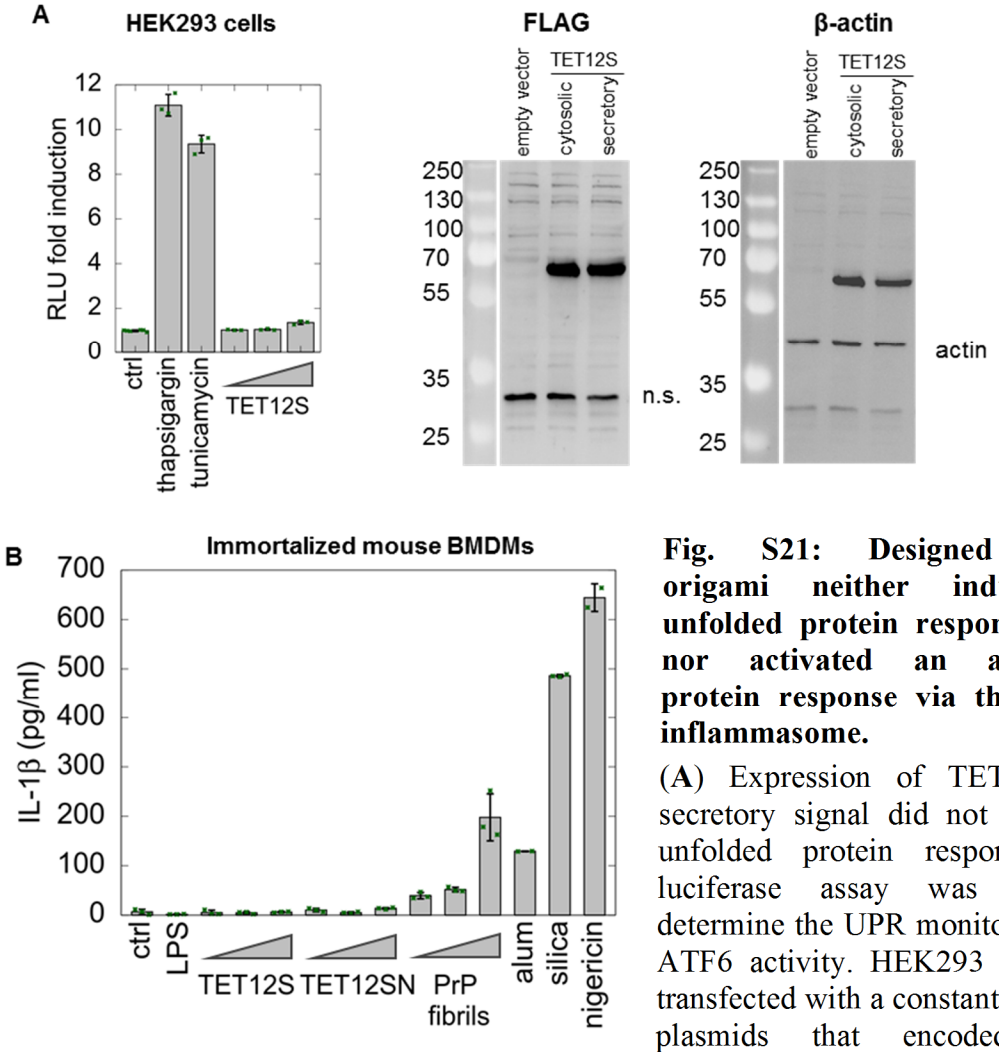


Fig. S21: Designed protein origami neither induced an unfolded protein response (UPR) nor activated an aggregated protein response via the NLRP3 inflammasome.

(**A**) Expression of TET12S with secretory signal did not induce an unfolded protein response. Dual luciferase assay was used to determine the UPR monitored by the ATF6 activity. HEK293 cells were transfected with a constant amount of plasmids that encoded ATF6-responsive firefly luciferase and constitutive *Renilla* luciferase, as well as varying amounts of pCMV3 plasmid-encoding TET12S (1 ng, 10 ng, and 50 ng) and empty plasmid. Thapsigargin (5 μM) and tunicamycin (10 μg/mL) were used as positive controls. Representative experiment (of four independent experiments) is shown. The error bars represent the standard deviation of 6 (ctrl) or 3 (other) cell culture replicates. Individual datapoints are shown in green. Western blot analysis confirmed the expression of both cytosolic and secretory TET12S in cell lysate (FLAG), non-specific band is designated as n.s. After detection of FLAG-tagged proteins, membrane was incubated with antibodies against β-actin, which serves as loading control (+ β-actin). Representative of 2 Western blots is shown.

(**B**) Immortalized bone-marrow-derived macrophages were primed with LPS and stimulated overnight with different concentrations of bacterially expressed and purified TET12S and TET12SN (0.1, 0.15, or 0.25 mg/mL). PrP fibrils were used as the positive control for the induction of NLRP3 inflammasome (0.1, 0.15, or 0.25 mg/mL). Non-protein particulate stimulators alum and silica were used at 0.25 mg/mL overnight and nigericin (10 μM) for 30 min. Representative experiment (of two independent experiments) is shown. The error bars represent the standard deviation of 2 (alum, nigericin) or 3 (other) cell culture replicates.

Nature Biotechnology: doi:10.1038/nbt.3994

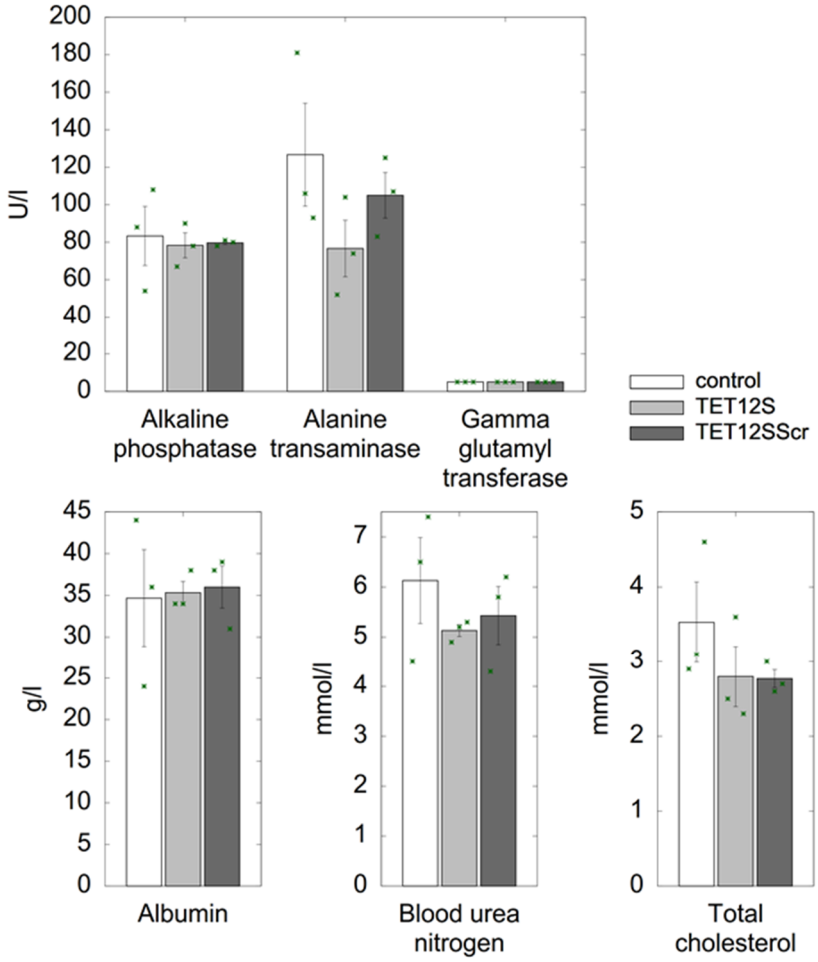


Fig. S22: Analysis of the effect of protein origami expression on the physiological and biochemical parameters of mice.

Two days after the hydrodynamic application of plasmids encoding the expression of TET12S and TET12SScr, biochemistry profiles of the serum and liver were determined. Similar levels of all of the measured parameters in comparison to the control animals (control; injection of sterile saline solution) demonstrated that there was no adverse influence of the liver expression of protein origami on animal physiology. Bars represent average values of measurements from three animals (n=3; error bars represent s.d.); Individual datapoints are shown in green. All pairwise control vs. sample p-values are above 0.05.

Nature Biotechnology: doi:10.1038/nbt.3994

Supplementary Tables

Table S1: Set of coiled-coil forming peptides and linkers used in the protein origami design

Peptide	Polyhedral variants implementing those peptides	Orientation #	Hydrophobic/ electrostatic pattern	No. of amino acids	Net charge	Helical propensity&	Sequence" *gabcdef gabcdef gabcdef gabcdef gabcdef gabcdef gabc*
Original peptides							
P3*	$TET12_{1.10}A$-f_5	he-P	IINN/EEKK	28	-2	3.59	EIQQLEE EIAQLEQ KNAALKE KNQALKY
P4*		he-P	IINN/KKEE	28	-1	1.35	KIAQLKQ KIQALKQ ENQQLEE ENAALEY
P5*		he-P	NINI/EKKE	28	-2	5.45	ENAALEE KIAQLKQ KNAALKE EIQALEY
P6*		he-P	NINI/KEEK	28	-3	4.28	KNAALKE EIQALEE ENQALEE KIAQLKY
P7*		he-P	ININ/EKEK	28	-2	4.01	EIQALEE KNAQLKQ EIAALEE KNQALKY
P8*		he-P	ININ/KEKE	28	-2	3.00	KIAQLKE ENQQLEQ KIQALKE ENAALEY
GCNsh	$TET12_{1.10}A$-f_5	ho-P	/	27	-1	9.21	QLED KVEELLS KNYHLEN EVARLKK LV
BCR		ho-AP	/	36	0	9.62	DIEQ ELERAKA SIRRLEQ EVNQERS RMAYLQT LLAK
APH		ho-AP	/	45	8	26.48	MKQLEK ELKQLEK ELQAIEK QLAQLQW KAQARKK KLAQLKK KLQA
Peptides PnS – introducing of polar amino acids at *b,c* or *f* positions							
P1S	$PYR16_{2.15R}$S-f_5	he-P	INNI/EEEE	28	-8	2.05	EIQALEQ ENAQLEQ ENAALEQ EIAQLEY
P2S		he-P	INNI/KKKK	28	1	2.78	KIDQLKE KNADLKE KNQDLKE KIDALKY
P3S	$TET12_{1.10}$S-f_5 $TET12_{1.10}$xS-f_6 $TET12_{1.10}$S-f_5b $TET12_{1.11}$S-f_5 $TET12_{1.6}$S-f_5b $TET12_{1.6}$S-c_6b $TET12_{Scr}$S-f_5 $TET12_{1.10}$S-c_6 $PYR16_{2.15R}$S-f_5	he-P	IINN/EEKK	28	-2	1.05	EIQQLEE EISQLEQ KNSQLKE KNQQLKY
P4S		he-P	IINN/KKEE	28	-1	1.12	KISQLKQ KIQQLKQ ENQQLEE ENSQLEY
P5S		he-P	NINI/EKKE	28	-2	1.49	ENSQLEE KISQLKQ KNSQLKE EIQQLEY
P6S		he-P	NINI/KEEK	28	-3	1.56	KNSQLKE EIQQLEE ENQQLEE KISQLKY
P7S		he-P	ININ/EKEK	28	-2	1.12	EIQSLEE KNSQLKQ EISQLEE KNQQLKY
P8S		he-P	ININ/KEKE	28	-2	2.15	KISQLKE ENQQLEQ KIQQLKE ENSQLEY
P9S	$PYR16_{2.15R}$S-f_5	he-P	NNII/EKEE	28	-4	1.12	ENQQLEQ KNSQLKQ EISQLEQ EISQLEY
P10S		he-P	NNII/KEKK	28	1	5.02	KNSQLKE ENSQLEE KIQQLKE KIQQLKY
P11S		he-P	NIIN/EEEK	28	-6	1.72	ENQQLEQ EISQLEQ EISQLEQ KNSQLKY
P12S		he-P	NIIN/KKKE	28	-1	1.62	KNSQLKE KISQLKE KIQQLKE ENQQLEY
GCNshS	$TET12_{1.10}$S-f_5 $TET12_{1.10}$xS-f_6 $TET12_{1.10}$S-f_5b $TET12_{1.11}$S-f_5 $TET12_{1.6}$S-f_5b $TET12_{1.6}$S-c_6b $TET12_{Scr}$S-f_5	ho-P	/	27	-1	4.35	QLED KVEELLS KNYHLEN EVSRLKK LV
BCRS	$TET12_{1.10}$S-f_5 $TET12_{1.10}$xS-f_6 $TET12_{1.10}$S-f_5b $TET12_{1.11}$S-f_5	ho-AP	/	36	0	7.41	DIEQ ELERAKQ SIRRLEQ EVNQERS RMQYLQT LLSK

Nature Biotechnology: doi:10.1038/nbt.3994

	TET12$_{1.6}$S-f$_5$b **TET12$_{1.6}$S-c$_6$b** **TET12$_{Scr}$S-f$_5$** **PYR16$_{2.15R}$S-f$_5$**						
APHsh	**PYR16$_{2.15R}$S-f$_5$**	ho-AP	/	34	-1	17.29	ELKQLEE ELQAIEE QLAQLQW KAQARKE KLAQLK
Peptides PnSN – introducing of negatively charged amino acids at *b,c* or *f* positions							
P1SN	**PYR16$_{4.6}$SN-f$_5$**	he-P	INNI/EEEE	28	-4	2.84	EIRQLEQ ENSQLER ENQRLEQ EIYQLER
P2SN		he-P	INNI/KKKK	28	0	3.93	KIEELKE KNSQLKE KNEELKQ KIYELKE
P3SN	**TET12$_{1.10}$SN-f$_5$** **TET12$_{1.10}$SN-c$_6$** **TET12$_{1.10}$SN-f$_9$** **TET12$_{2.3}$SN-f$_5$b** **PYR16$_{4.6}$SN-f$_5$**	he-P	IINN/EEKK	28	-4	1.17	EIQQLEE EISQLEQ KNSELKE KNQELKY
P4SN		he-P	IINN/KKEE	28	-2	1.13	KISQLKE KIQQLKQ ENQQLEE ENSQLEY
P5SN		he-P	NINI/EKKE	28	-3	1.41	ENSQLEE KISQLKQ KNSELKE EIQQLEY
P6SN		he-P	NINI/KEEK	28	-5	1.38	KNSELKE EIQQLEE ENQQLEE KISELKY
P7SN		he-P	ININ/EKEK	28	-3	0.88	EIQQLEE KNSQLKQ EISQLEE KNQELKY
P8SN		he-P	ININ/KEKE	28	-3	2.19	KISELKE ENQQLEQ KIQQLKE ENSQLEY
P9SN	**PYR16$_{4.6}$SN-f$_5$**	he-P	NNII/EKEE	28	-4	1.47	ENQSLEQ KNSQLKQ EISQLEQ EIQQLEY
P10SN		he-P	NNII/KEKK	28	-1	2.74	KNSQLKE ENSQLEE KIEQLKE KIQELKY
GCNshSN	**TET12$_{1.10}$SN-f$_5$** **TET12$_{1.10}$SN-c$_6$** **TET12$_{1.10}$SN-f$_9$**	ho-P	/	27	-2	12.2	QLED KVEELLS KNYHLEN EVERLKK LV
GCNshSNb	**PYR16$_{4.6}$SN-f$_5$**	ho-P	/	31	0	10.94	S RMKQLED KVEELLS KNYHLEN EVERLKK LV
BCRSN	**TET12$_{1.10}$SN-f$_5$** **TET12$_{1.10}$SN-c$_6$** **TET12$_{1.10}$SN-f$_9$** **TET12$_{2.3}$SN-f$_5$b** **PYR16$_{4.6}$SN-f$_5$**	ho-AP	/	36	-2	13.4	DIEQ ELERAKE SIRRLEQ EVNQERS RMQYLQT LLEK
APHshSN	**TET12$_{1.10}$S-f$_5$** **TET12$_{1.10}$xS-f$_6$** **TET12$_{1.10}$SN-f$_5$** **TET12$_{1.10}$SN-c$_6$** **TET12$_{1.10}$SN-f$_9$** **TET12$_{2.3}$SN-f$_5$b** **TET12$_{1.10}$S-f$_5$b** **TET12$_{1.11}$S-f$_5$** **TET12$_{1.6}$S-f$_5$b** **TET12$_{1.6}$S-c$_6$b** **TET12$_{Scr}$S-f$_5$** **PYR16$_{4.6}$SN-f$_5$**	ho-AP	/	40	-3	26.03	LEE ELKQLEE ELQAIEE QLAQLQW KAQARKE KLAQLKE KL
APH4SN	**TET12$_{2.3}$SN-f$_5$b**	ho-AP	/	39	-4	54.8	LEQIEE RLEQIEE RLQAKEW EKAQLRE ELQALRE KLAQL
Peptides PnSH – introducing of polar amino acids with high helical propensity at *b,c* or *f* positions							
GCNH3	**TET12$_{1.10}$S-c$_6$**	ho-P	/	30	-1	8.86	RMKQLED KVEELER KNYHLEN EVERLKK EV
BCRSH		ho-AP	/	36	-3	10.97	DIEQ ELERAKQ SIEELER EVNQERS RMQYLQT LLSK
APHshE		ho-AP	/	35	-7	8.01	ELEELER ELQEIEE QLEQLQW KAQERKE KLEQLKE

Nature Biotechnology: doi:10.1038/nbt.3994

Linker	Polyhedral variants implementing those linkers	No. of amino acids	Sequence
f_5	**$TET12_{1.10}S$-f_5** **$TET12_{1.10}SN$-f_5** **$TET12_{1.10}S$-f_5b** **$TET12_{1.11}S$-f_5** **$TET12_{1.6}S$-f_5b**	5	GSGPG
f_6	**$TET12_{1.10}xS$-f_6**	7	GSGPGSG
f_9	**$TET12_{1.10}SN$-f_9**	9	GSGSGPGSG
c_6	**$TET12_{1.10}SN$-c_6** **$TET12_{1.6}S$-c_6b** **$TET12_{1.6}SH$-c_6**	6	GGDGKG or GGKGDG

[#]Type of interaction between chains: he-P, parallel heterodimer; ho-P, parallel homodimer; ho-AP, antiparallel homodimer.
[&]Helical propensity was estimated by Agadir at 20°C, ionic strength of 0.1 M, pH 7.0 [79].
"Amino acid sequence is written in heptad repeats (*gabcdef*); see also Fig. S1.
*used in Gradišar *et al.*[14].

Nature Biotechnology: doi:10.1038/nbt.3994

Table S2: Sequence and properties of protein polyhedral cage variants

Nomenclature of the designed protein polyhedra: The full names of the polyhedra consist of the type of the polyhedron (TET = tetrahedron; PYR = pyramid; TRIP = trigonal prism) followed by the number of dimeric CC segments. The subscripts denote the topology and circular permutation of each polyhedron. The next labels denote the type of CC segments used (S = soluble; SN = soluble, negatively charged), linker type (f = flexible; c = charged), and, in subscript form, the length of the linker. In cases where the two variants have the same name (e.g., in cases of different ordering of CC modules), the letters b, c, d, etc. are appended. The most extensively characterized polyhedra are abbreviated TET12SN ($TET12_{1.10}SN\text{-}f_5$), TET12S ($TET12_{1.10}S\text{-}f_5$), TET12SScr ($TET12_{Scr}S\text{-}f_5$), PYR16SN ($PYR16_{4.6}SN\text{-}f_5$), and TRIP18SN ($TRIP18_{7.5R}SN\text{-}f_5$).

Polyhedron variant	Segment sequence	Linker[#]	SPED[&]	Aa*	MW*	Net charge*	pI*	Description
$TET12_{1.10}A\text{-}f_5$	APH–P3–BCR–GCNshort–APH–P7–GCNshort–P4–P5–P8–BCR–P6	GSGPG	yes	476	53,391	-9	5.73	Original tetrahedron, developed in [14].
$TET12_{1.10}SN\text{-}f_5$	APHshSN–P3SN–BCRSN–GCNshSN–APHshSN–P7SN–GCNshSN–P4SN–P5SN–P8SN–BCRSN–P6SN	GSGPG	yes	461	53,411	-47	4.70	Tetrahedron with negatively super charged segments (SN).
$TET12_{1.10}SN\text{-}c_6$	APHshSN–P3SN–BCRSN–GCNshSN–APHshSN–P7SN–GCNshSN–P4SN–P5SN–P8SN–BCRSN–P6SN	GGKGDG	yes	466	54,346	-47	4.73	As TET12SN; includes a charged linker to further improve solubility/stability.
$TET12_{1.10}SN\text{-}f_9$	APHshSN–P3SN–BCRSN–GCNshSN–APHshSN–P7SN–GCNshSN–P4SN–P5SN–P8SN–BCRSN–P6SN	GSGSGPGSG	yes	505	56,582	-47	4.70	As TET12SN; includes longer flexible linkers.
$TET12_{1.10}S\text{-}f_5$	APHshSN–P3S–BCRS–GCNshS–APHshSN–P7S–GCNshS–P4S–P5S–P8S–BCRS–P6S	GSGPG	yes	461	53,191	-33	4.89	Tetrahedron from CC modules with increased polarity.
$TET12_{1.10}S\text{-}f_5^{split\text{-}mVenus}$	mVenus(1–84)–P3S–P4S–P5S–P6S–P7S–P8S–GCNshS–GCNshS–APHshSN–BCRS–APHshSN–BCRS–mVenus(85–238)	GSGPG	yes	707	80,610	-41	5.03	Used to test correct folding in vivo. When the N and C terminal ends come in proximity, the fluorescence is restored.
$TET12_{1.10}S\text{-}f_5^{split\text{-}fLuc}$	fLuc(1–490)–APHshSN–P3S–BCRS–GCNshS–APHshSN–P7S–GCNshS–P4S–P5S–P8S–BCRS–P6S–fLuc(492–551)	GSGPG	yes	1002	112,548	-38	5.21	Used to test correct folding in vivo. When the N and C terminal ends come in proximity, luciferase activity is restored.
$TET12_{Scr}S\text{-}f_5$	P3S–P4S–P5S–P6S–P7S–P8S–GCNshS–GCNshS–APHshSN–BCRS–APHshSN–BCRS	GSGPG	yes	461	53,191	-33	4.89	Nonvalid topology of $TET12_{1.10}S\text{-}f_5$.
$TET12_{Scr}S\text{-}f_5^{split\text{-}mVenus}$	mVenus(1–84)–P3S–P4S–P5S–P6S–P7S–P8S–GCNshS–GCNshS–APHshSN–BCRS–APHshSN–BCRS–mVenus(85–238)	GSGPG	yes	707	80,610	-41	5.03	Used as a negative control, since the N and C termini are not in proximity, there should be no fluorescence.
$TET12_{Scr}S\text{-}f_5^{split\text{-}fLuc}$	fLuc(1–490)–P3S–P4S–P5S–P6S–P7S–P8S–GCNshS–GCNshS–APHshSN–BCRS–APHshSN–BCRS–fLuc(492–551)	GSGPG	yes	1002	112,548	-38	5.21	Used as a negative control, since the N and C termini are not in proximity, there should be no luciferase activity.
$TET12_{1.10}S\text{-}f_5b$	APHshSN–P5S–BCRS–GCNshS–APHshSN–P7S–GCNshS–P6S–P3S–P8S–BCRS–P4S	GSGPG	yes	461	53,191	-33	4.89	TET12SN variant where more stable segments have been used at the C-terminal. (P3:P4 and P5:P6 switched in comparison to $TET12_{1.10}S\text{-}f_5b$)

Nature Biotechnology: doi:10.1038/nbt.3994

Polyhedron variant	Segment sequence	Linker[#]	SPED[&]	Aa*	MW*	Net charge*	pI*	Description
TET12$_{1.10}$S-c$_6$	APHshE–P3S–BCRSH–GCNH3–APHshE–P7S–GCNH3–P4S–P5S–P8S–BCRSH–P6S	GGKGDG	yes	464	54,547	-47	4.75	TET12S with charged linkers. Includes polar GCN, BCR, destabilized APH, increased stability of GCN.
TET12$_{1.11}$S-f$_5$	P3S–BCRS–GCNshS–APHshSN–P7mS–GCNshS–P4S–P5S–P8S–BCRS–P6S–APHshSN	GSGPG	yes	461	53,191	-33	4.89	A circular permutation of TET12$_{1.10}$S
TET12$_{1.6}$S-f$_5$b	GCNshS–P3S–BCRS–GCNshS–APHshSN–P5S–BCRS–P7S–APHshSN–P4S–P8S–P6S	GSGPG	yes	461	53,191	-33	4.89	A circular permutation of TET12$_{1.10}$S with more stable segments at the end (weak P7:P8 switched with stronger P5:P6).
TET12$_{1.6}$S-c$_6$b	GCNshS–P3S–BCRS–GCNshS–APHshSN–P5S–BCRS–P7S–APHshSN–P4S–P8S–P6S	GGDGKG	yes	472	54,469	-33	4.93	TET12$_{1.6}$S with charged linkers.
TET12$_{1.10}$xS-f$_6$	APHshSN–P3S–BCRS–GCNshS–APHshSN–P7S–GCNshS–P4S–P5S–P8S–BCRS–P6S	GSGPGSG	no	459	52,207	-21	5.19	A variant of TET12$_{1.10}$S without capping sequences.
TET12$_{2.3}$SN-f$_5$b	P5SN–P7SN–APHshSN–APH4SN–P6SN–BCRSN–APHshSN–P3SN–APH4SN–BCRSN–P8SN–P4SN	GSGPG	yes	489	56,682	-51	4.66	A different topology using three antiparallel and three parallel segments.
PYR16$_{4.6}$SN-f$_5$	APHshSN–P5SN–P1SN–GCNshSNb–APHshSN–P7SN–GCNshSNb–P6SN–BCRSN–P3SN–P8SN–P9SN–BCRSN–P2SN–P10SN–P4SN	GSGPG	yes	621	71,626	-60	4.7	Pyramid with the smallest topological contact order constructed from negatively super charged segments (SN).
PYR16$_{2.15R}$S-f$_5$	P11S–P3S–BCRsh–P6S–APHsh–P4S–P9S–P12S–APHsh–P1S–P7S–BCRsh–P10S–P8S–P5S–P2S	GSGPG	yes	607	69,427	-52	4.67	Topology of a pyramid using soluble segments.
TRIP18$_{7.5R}$SN-f$_5$	P5SN–APH4SN–P1SN–P6SN–GCNshSNb–APHshSN–P9SN–APH4SN–GCNshSNb–P3SN–P11SN–P10SN–P2SN–BCRSN–P12SN–APHshSN–P4SN–BCRSN	GSGPG	yes	708	81,867	-69	4.69	Trigonal prism using negatively supercharged segments and extended GCN.
TET12$_{1.10}$SN-f$_5$-2Cys	APHshSN–P3SN–BCRSN–GCNshSN–APHshSN–P7SN–GCNshSN–P4SN–P5SN–P8SN–BCRSN–P6SN	GSGPG	yes	476	55,003	-47	4.71	A variant of TET12$_{1.10}$SN-f$_5$ with two added cysteines, first on the N-terminal part and second on the C-terminal part of the protein.
PYR16$_{4.6}$SN-f$_5$-2Cys	APHshSN–P5SN–P1SN–GCNshSNb–APHshSN–P7SN–GCNshSNb–P6SN–BCRSN–P3SN–P8SN–P9SN–BCRSN–P2SN–P10SN–P4SN	GSGPG	yes	636	72,795	-60	4.70	A variant of PYR16$_{2.15R}$S-f$_5$ with two added cysteines, first on the N-terminal part and second on the C-terminal part of the protein.
TRIP18$_{7.5R}$SN-f$_5$-2Cys	P5SN–APH4SN–P1SN–P6SN–GCNshSNb–APHshSN–P9SN–APH4SN–GCNshSNb–P3SN–P11SN–P10SN–P2SN–BCRSN–P12SN–APHshSN–P4SN–BCRSN	GSGPG	yes	723	83,009	-69	4.69	A variant of TRIP18$_{7.5R}$SN-f$_5$ with two added cysteines, first on the N-terminal part and second on the C-terminal part of the protein.

[#]Amino acids in the linker between segments.
[&]SPED, capping sequence Ser-Pro-Glu-Asp at the N-terminus important for α-helices stabilization [26,51].
*Number of amino acids (Aa), molecular weight (MW), net charge and isoelectric point (pI) were calculated by ProtParam tool (http://web.expasy.org/protparam/).

Nature Biotechnology: doi:10.1038/nbt.3994

Table S3: Amino acid sequences of polyhedral variants

Name:	**TET12$_{1.10}$A-f$_5$**
Common name:	**TET12**
Segment sequence:	APH-P3-BCR-GCNshort-APH-P7-GCNshort-P4-P5-P8-BCR-P6

```
  1 MYHHHHHHSR AGMKQLEKEL KQLEKELQAI EKQLAQLQWK AQARKKKLAQ LKKKLQASGP GSPEDEIQQL EEEIAQLEQK NAALKEKNQA LKYGSGPGDI EQELERAKAS IRRLEQEVNQ ERSRMAYLQT LLAKSGPGQL EDKVEELLSK 150
151 NYHLENEVAR LKKLVGSGPG MKQLEKELKQ LEKELQAIEK QLAQLQWKAQ ARKKKLAQLK KKLQASGPGS PEDEIQALEE KNAQLKQEIA ALEEKNQALK YGSGPGQLED KVEELLSKNY HLENEVARLK KLVGSGPGSP EDKIAQLKQK 300
301 IQALKQENQQ LEEENAALEY GSGPGSPEDE NAALEEKIAQ LKQKNAALKE EIQALEYGSG PGSPEDKIAQ LKEENQQLEQ KIQALKEENA ALEYGSGPGD IEQELERAKA SIRRLEQEVN QERSRMAYLQ TLLAKSGPGS PEDKNAALKE 450
451 EIQALEEENQ ALEEKIAQLK YGSGTS                                                                                                                                          476
```

Name:	**TET12$_{1.10}$SN-f$_5$**
Common name:	**TET12SN**
Segment sequence:	APHshSN-P3SN-BCRSN-GCNshSN-APHshSN-P7SN-GCNshSN-P4SN-P5SN-P8SN-BCRSN-P6SN

```
  1 MLEEELKQLE EELQAIEEQL AQLQWKAQAR KEKLAQLKEK LSGPGSPEDE IQQLEEEISQ LEQKNSELKE KNQELKYGSG PGDIEQELER AKESIRRLEQ EVNQERSRMQ YLQTLLEKSG PGQLEDKVEE LLSKNYHLEN EVERLKKLVG 150
151 SGPGLEEELK QLEEELQAIE EQLAQLQWKA QARKEKLAQL KEKLSGPGSP EDEIQQLEEK NSQLKQEISQ LEEKNQELKY GSGPGQLEDK VEELLSKNYH LENEVERLKK LVGSGPGSPE DKISQLKEKI QQLKQENQQL EEENSQLEYG 300
301 SGPGSPEDEN SQLEEKISQL KQKNSELKEE IQQLEYGSGP GSPEDKISEL KEENQQLEQK IQQLKEENSQ LEYGSGPGDI EQELERAKES IRRLEQEVNQ ERSRMQYLQT LLEKSGPGSP EDKNSELKEE IQQLEEENQQ LEEKISELKY 450
451 GLEHHHHHHH H                                                                                                                                                          461
```

Name:	**TET12$_{1.10}$SN-c$_6$**
Segment sequence:	APHshSN-P3SN-BCRSN-GCNshSN-APHshSN-P7SN-GCNshSN-P4SN-P5SN-P8SN-BCRSN-P6SN

```
  1 MLEEELKQLE EELQAIEEQL AQLQWKAQAR KEKLAQLKEK LGKGDGSPED EIQQLEEEIS QLEQKNSELK EKNQELKYGK GDGDIEQELE RAKESIRRLE QEVNQERSRM QYLQTLLEKG KGDGQLEDKV EELLSKNYHL ENEVERLKKL 150
151 VGGKGDGLEE ELKQLEEELQ AIEEQLAQLQ WKAQARKEKL AQLKEKLGKG DGSPEDEIQQ LEEKNSQLKQ EISQLEEKNQ ELKYGDGKGQ LEDKVEELLS KNYHLENEVE RLKKLVGGDG KGSPEDKISQ LKEKIQQLKQ ENQQLEEENS 300
301 QLEYGDGKGS PEDENSQLEE KISQLKQKNS ELKEEIQQLE YGDGKGSPED KISELKEENQ QLEQKIQQLK EENSQLEYGK GDGDIEQELE RAKESIRRLE QEVNQERSRM QYLQTLLEKG KGDGSPEDKN SELKEEIQQL EEENQQLEEK 450
451 ISELKYLEHH HHHHHH                                                                                                                                                     466
```

Name:	**TET12$_{1.10}$SN-f$_9$**
Segment sequence:	APHshSN-P3SN-BCRSN-GCNshSN-APHshSN-P7SN-GCNshSN-P4SN-P5SN-P8SN-BCRSN-P6SN

```
  1 MLEEELKQLE EELQAIEEQL AQLQWKAQAR KEKLAQLKEK LSGSGPGSGS PEDEIQQLEE EISQLEQKNS ELKEKNQELK YGSGSGPGSG DIEQELERAK ESIRRLEQEV NQERSRMQYL QTLLEKSGSG PGSGQLEDKV EELLSKNYHL 150
151 ENEVERLKKL VGSGSGPGSG LEEELKQLEE ELQAIEEQLA QLQWKAQARK EKLAQLKEKL SGSGPGSGSP EDEIQQLEEK NSQLKQEISQ LEEKNQELKY GSGSGPGSGQ LEDKVEELLS KNYHLENEVE RLKKLVGSGS GPGSGSPEDK 300
301 ISQLKEKIQQ LKQENQQLEE ENSQLEYGSG SGPGSGSPED ENSQLEEKIS QLKQKNSELK EEIQQLEYGS GSGPGSGSPE DKISELKEEN QQLEQKIQQL KEENSQLEYG SGSGPGSGDI EQELERAKES IRRLEQEVNQ ERSRMQYLQT 450
451 LLEKSGSGPG SGSPEDKNSE LKEEIQQLEE ENQQLEEKIS ELKYGLEHHH HHHHH                                                                                                          505
```

Name:	**TET12$_{1.10}$S-f$_5$**
Synonyms:	**TET12S**
Segment sequence:	APHshSN-P3S-BCRS-GCNshS-APHshSN-P7S-GCNshS-P4S-P5S-P8S-BCRS-P6S

```
  1 MLEEELKQLE EELQAIEEQL AQLQWKAQAR KEKLAQLKEK LSGPGSPEDE IQQLEEEISQ LEQKNSQLKE KNQQLKYGSG PGDIEQELER AKQSIRRLEQ EVNQERSRMQ YLQTLLSKSG PGQLEDKVEE LLSKNYHLEN EVSRLKKLVG 150
151 SGPGLEEELK QLEEELQAIE EQLAQLQWKA QARKEKLAQL KEKLSGPGSP EDEIQSLEEK NSQLKQEISQ LEEKNQELKY GSGPGQLEDK VEELLSKNYH LENEVSRLKK LVGSGPGSPE DKISQLKQKI QQLKQENQQL EEENSQLEYG 300
301 SGPGSPEDEN SQLEEKISQL KQKNSQLKEE IQQLEYGSGP GSPEDKISQL KEENQQLEQK IQQLKEENSQ LEYGSGPGDI EQELERAKQS IRRLEQEVNQ ERSRMQYLQT LLSKSGPGSP EDKNSQLKEE IQQLEEENQQ LEEKISQLKY 450
451 GLEHHHHHHH H                                                                                                                                                          461
```

Name:	**TET12$_{1.10}$S-f$_5^{\text{split mVenus}}$**
Common name:	**TET12S$^{\text{split-mVenus}}$**
Segment sequence:	mVenus(1-84)-APHshSN-P3S-BCRS-GCNshS-APHshSN-P7S-GCNshS-P4S-P5S-P8S-BCRS-P6S-mVenus(85-238)

```
  1 MDKQKNGIKV NFKIRHNIED GSVQLADHYQ QNTPIGDGPV LLPDNHYLSY QSALSKDPNE KRDHMVLLEF VTAAGITLGM DELYKSGPGL EEELKQLEEE LQAIEEQLAQ LQWKAQARKE KLAQLKEKLS GPGSPEDEIQ QLEEEISQLE 150
151 QKNSQLKEKN QQLKYGSGPG DIEQELERAK QSIRRLEQEV NQERSRMQYL QTLLSKSGPG QLEDKVEELL SKNYHLENEV SRLKKLVGSG PGLEEELKQL EEELQAIEEQ LAQLQWKAQA RKEKLAQLKE KLSGPGSPED EIQSLEEKNS 300
301 QLKQEISQLE EKNQQLKYGS GPGQLEDKVE ELLSKNYHLE NEVSRLKKLV GSGPGSPEDK ISQLKQKIQQ LKQENQQLEE ENSQLEYGSG PGSPEDENSQ LEEKISQLKQ KNSQLKEEIQ QLEYGSGPGS PEDKISQLKE ENQQLEQKIQ 450
451 QLKEENSQLE YGSGPGDIEQ ELERAKQSIR RLEQEVNQER SRMQYLQTLL SKSGPGSPED KNSQLKEEIQ QLEEENQQLE EKISQLKYGS GPGVSKGEEL FTGVVPILVE LDGDVNGHKF SVSGEGEGDA TYGKLTLKFI CTTGKLPVPW 600
601 PTLVTTFGYG LQCFARYPDH MKQHDFFKSA MPEGYVQERT IFFKDDGNYK TRAEVKFEGD TLVNRIELKG IDFKEDGNIL GHKLEYNYNS HNVYIMALEH HHHHHHH                                                  707
```

Name:	**TET12$_{1.10}$S-f$_5^{\text{split fLuc}}$**
Common name:	**TET12S$^{\text{split-fLuc}}$**
Segment sequence:	fLuc(1-490)-APHshSN-P3S-BCRS-GCNshS-APHshSN-P7S-GCNshS-P4S-P5S-P8S-BCRS-P6S-fLuc(492-551)

```
  1 MEDAKNIKKG PAPFYPLEDG TAGEQLHKAM KRYALVPGTI AFTDAHIEVD ITYAEYFEMS VRLAEAMKRY GLNTNHRIVV CSENSLQFFM PVLGALFIGV AVAPANDIYN ERELLNSMGI SQPTVVFVSK KGLQKILNVQ KKLPIIQKII 150
151 IMDSKTDYQG FQSMYTFVTS HLPPGFNEYD FVPESFDRDK TIALIMNSSG STGLPKGVAL PHRTACVRFS HARDPIFGNQ IIPDTAILSV VPFHHGFGMF TTLGYLICGF RVVLMYRFEE ELFLRSLQDY KIQSALLVPT LFSFFAKSTL 300
301 IDKYDLSNLH EIASGGAPLS KEVGEAVAKR FHLPGIRQGY GLTETTSAIL ITPEGDDKPG AVGKVVPFFE AKVVDLDTGK TLGVNQRGEL CVRGPMIMSG YVNNPEATNA LIDKDGWLHS GDIAYWDEDE HFFIVDRLKS LIKYKGYQVA 450
451 PAELESILLQ HPNIFDAGVA GLPDDDAGEL PAAVVVLEHG KLEEELKQLE EELQAIEEQL AQLQWKAQAR KEKLAQLKEK LSGPGSPEDE IQQLEEEISQ LEQKNSQLKE KNQQLKYGSG PGDIEQELER AKQSIRRLEQ EVNQERSRMQ 600
601 YLQTLLSKSG PGQLEDKVEE LLSKNYHLEN EVSRLKKLVG SGPGLEEELK QLEEELQAIE EQLAQLQWKA QARKEKLAQL KEKLSGPGSP EDEIQSLEEK NSQLKQEISQ LEEKNQQLKY GSGPGQLEDK VEELLSKNYH LENEVSRLKK 750
751 LVGSGPGSPE DKISQLKQKI QQLKQENQQL EEENSQLEYG SGPGSPEDEN SQLEEKISQL KQKNSQLKEE IQQLEYGSGP GSPEDKISQL KEENQQLEQK IQQLKEENSQ LEYGSGPGDI EQELERAKQS IRRLEQEVNQ ERSRMQYLQT 900
901 LLSKSGPGSP EDKNSQLKEE IQQLEEENQQ LEEKISQLKY GTMTEKEIVD YVASQVTTAK KLRGGVVFVD EVPKGLTGKL DARKIREILI KAKKGGKIAV NS                                                  1002
```


Nature Biotechnology: doi:10.1038/nbt.3994

Name:	**TET12$_{Scr}$S-f$_5$**
Common name:	**TET12SScr**
Segment sequence:	P3S-P4S-P5S-P6S-P7S-P8S-GCNshS-GCNshS-APHshSN-BCRS-APHshSN-BCRS

```
  1 MSPEDEIQQL EEEISQLEQK NSQLKEKNQQ LKYGSGPGSP EDKISQLKQK IQQLKQENQQ LEEENSQLEY GSGPGSPEDE NSQLEEKISQ LKQKNSQLKE EIQQLEYGSG PGSPEDKNSQ LKEEIQQLEE ENQQLEEKIS QLKYGSGPGS 150
151 PEDEIQSLEE KNSQLKQEIS QLEEKNQQLK YGSGPGSPED KISQLKEENQ QLEQKIQQLK EENSQLEYGS GPGQLEDKVE ELLSKNYHLE NEVSRLKKLV GSGPGQLEDK VEELLSKNYH LENEVSRLKK LVGSGPGLEE ELKQLEEELQ 300
301 AIEEQLAQLQ WKAQARKEKL AQLKEKLSGP GDIEQELERA KQSIRRLEQE VNQERSRMQY LQTLLSKSGP GLEEELKQLE EELQAIEEQL AQLQWKAQAR KEKLAQLKEK LSGPGDIEQE LERAKQSIRR LEQEVNQERS RMQYLQTLLS 450
451 KLEHHHHHHH H                                                                                                                                                    461
```

Name:	**TET12$_{Scr}$S-f$_5^{split\text{-}mVenus}$**
Common name:	**TET12SScr$^{split\text{-}mVenus}$**
Segment sequence:	mVenus(1-84)-P3S-P4S-P5S-P6S-P7S-P8S-GCNshS-GCNshS-APHshSN-BCRS-APHshSN-BCRS-mVenus(85-238)

```
  1 MDKQKNGIKV NFKIRHNIED GSVQLADHYQ QNTPIGDGPV LLPDNHYLSY QSALSKDPNE KRDHMVLLEF VTAAGITLGM DELYKSGSGS PEDEIQQLEE EISQLEQKNS QLKEKNQQLK YGSGPGSPED KISQLKQKIQ QLKQENQQLE 150
151 EENSQLEYGS GPGSPEDENS QLEEKISQLK QKNSQLKEEI QQLEYGSGPG SPEDKNSQLK EEIQQLEEEN QQLEEKISQL KYGSGPGSPE DEIQSLEEKN SQLKQEISQL EEKNQQLKYG SGPGSPEDKI SQLKEENQQL EQKIQQLKEE 300
301 NSQLEYGSGP GQLEDKVEEL LSKNYHLENE VSRLKKLVGS GPGQLEDKVE ELLSKNYHLE NEVSRLKKLV GSGPGLEEEL KQLEEELQAI EEQLAQLQWK AQARKEKLAQ LKEKLSGPGD IEQELERAKQ SIRRLEQEVN QERSRMQYLQ 450
451 TLLSKSGPGL EEELKQLEEE LQAIEEQLAQ LQWKAQARKE KLAQLKEKLS GPGDIEQELE RAKQSIRRLE QEVNQERSRM QYLQTLLSKS GSGVSKGEEL FTGVVPILVE LDGDVNGHKF SVSGEGEGDA TYGKLTLKFI CTTGKLPVPW 600
601 PTLVTTFGYG LQCFARYPDH MKQHDFFKSA MPEGYVQERT IFFKDDGNYK TRAEVKFEGD TLVNRIELKG IDFKEDGNIL GHKLEYNYNS HNVYIMALEH HHHHHHH                                                  707
```

Name:	**TET12$_{Scr}$S-f$_5^{split\text{-}fLuc}$**
Common name:	**TET12SScr$^{split\text{-}fLuc}$**
Segment sequence:	fLuc(1-490)-P3S-P4S-P5S-P6S-P7S-P8S-GCNshS-GCNshS-APHshSN-BCRS-APHshSN-BCRS-fLuc(492-551)

```
  1 MEDAKNIKKG PAPFYPLEDG TAGEQLHKAM KRYALVPGTI AFTDAHIEVD ITYAEYFEMS VRLAEAMKRY GLNTNHRIVV CSENSLQFFM PVLGALFIGV AVAPANDIYN ERELLNSMGI SQPTVVFVSK KGLQKILNVQ KKLPIIQKII 150
151 IMDSKTDYQG FQSMYTFVTS HLPPGFNEYD FVPESFDRDK TIALIMNSSG STGLPKGVAL PHRTACVRFS HARDPIFGNQ IIPDTAILSV VPFHHGFGMF TTLGYLICGF RVVLMYRFEE ELFLRSLQDY KIQSALLVPT LFSFFAKSTL 300
301 IDKYDLSNLH EIASGGAPLS KEVGEAVAKR FHLPGIRQGY GLTETTSAIL ITPEGDDKPG AVGKVVPFFE AKVVDLDTGK TLGVNQRGEL CVRGPMIMSG YVNNPEATNA LIDKDGWLHS GDIAYWDEDE HFFIVDRLKS LIKYKGYQVA 450
451 PAELESILLQ HPNIFDAGVA GLPDDDAGEL PAAVVVLEHG KSPDEIQQL EEEISQLEQK NSQLKEKNQQ LKYGSGPGSP EDKISQLKQK IQQLKQENQQ LEEENSQLEY GSGPGSPEDE NSQLEEKISQ LKQKNSQLKE EIQQLEYGSG 600
601 PGSPEDKNSQ LKEEIQQLEE ENQQLEEKIS QLKYGSGPGS PEDEIQSLEE KNSQLKQEIS QLEEKNQQLK YGSGPGSPED KISQLKEENQ QLEQKIQQLK EENSQLEYGS GPGQLEDKVE ELLSKNYHLE NEVSRLKKLV GSGPGQLEDK 750
751 VEELLSKNYH LENEVSRLKK LVGSGPGLEE ELKQLEEELQ AIEEQLAQLQ WKAQARKEKL AQLKEKLSGP GDIEQELERA KQSIRRLEQE VNQERSRMQY LQTLLSKSGP GLEEELKQLE EELQAIEEQL AQLQWKAQAR KEKLAQLKEK 900
901 LSGPGDIEQE LERAKQSIRR LEQEVNQERS RMQYLQTLLS KTMTEKEIVD YVASQVTTAK KLRGGVVFVD EVPKGLTGKL DARKIREILI KAKKGGKIAV NS                                                1002
```

Name:	**TET12$_{1.10}$S-f$_5$b**
Segment sequence:	APHshSN-P5S-BCRS-GCNshS-APHshSN-P7S-GCNshS-P6S-P3S-P8S-BCRS-P4S

```
  1 MLEEELKQLE EELQAIEEQL AQLQWKAQAR KEKLAQLKEK LSGPGSPEDE NSQLEEKISQ LKQKNSQLKE EIQQLEYGSG PGDIEQELER AKQSIRRLEQ EVNQERSRMQ YLQTLLSKSG PGQLEDKVEE LLSKNYHLEN EVSRLKKLVG 150
151 SGPGLEEELK QLEEELQAIE EQLAQLQWKA QARKEKLAQL KEKLSGPGSP EDEIQSLEEK NSQLKQEISQ LEEKNQQLKY GSGPGQLEDK VEELLSKNYH LENEVSRLKK LVGSGPGSPE DKNSQLKEEI QQLEEENQQL EEKISQLKYG 300
301 SGPGSPEDEI QQLEEEISQL EQKNSQLKEK NQQLKYGSGP GSPEDKISQL KEENQQLEQK IQQLKEENSQ LEYGSGPGDI EQELERAKQS IRRLEQEVNQ ERSRMQYLQT LLSKSGPGSP EDKISQLKQK IQQLKQENQQ LEEENSQLEY 450
451 GLEHHHHHHH H                                                                                                                                                    461
```

Name:	**TET12$_{1.10}$S-c$_6$**
Segment sequence:	APHshE-P3S-BCRSH-GCNH3-APHshE-P7S-GCNH3-P4S-P5S-P8S-BCRSH-P6S

```
  1 MELEELEREL QEIEEQLEQL QWKAQERKEK LEQLKEGKGD GSPEDEIQQL EEEISQLEQK NSQLKEKNQQ LKYGGKGDGD IEQELERAKQ SIEELEREVN QERSRMQYLQ TRLSGKGDGR MKQLEDKVEE LERKNYHLEN EVERLKKEVG 150
151 KGDGELEELE RELQEIEEQL EQLQWKAQER KEKLEQLKEG KGDGSPEDEI QSLEEKNSQL KQEISQLEEK NQQLKYGGDG KGRMKQLEDK VEELERKNYH LENEVERLKK EVGDGKGSPE DKISQLKQKI QQLKQENQQL EEENSQLEYG 300
301 GDGKGSPEDE NSQLEEKISQ LKQKNSQLKE EIQQLEYGGD GKGSPEDKIS QLKEENQQLE QKIQQLKEEN SQLEYGGKGD GDIEQELERA KQSIEELERE VNQERSRMQY LQTRLSGKGD GSPEDKNSQL KEEIQQLEEE NQQLEEKISQ 450
451 LKYGLEHHHH HHHH                                                                                                                                                 464
```

Name:	**TET12$_{1.11}$S-f$_5$**
Segment sequence:	P3S-BCRS-GCNshS-APHshSN-P7mS-GCNshS-P4S-P5S-P8S-BCRS-P6S-APHshSN

```
  1 MSPEDEIQQL EEEISQLEQK NSQLKEKNQQ LKYGSGPGDI EQELERAKQS IRRLEQEVNQ ERSRMQYLQT LLSKSGPGQL EDKVEELLSK NYHLENEVSR LKKLVGSGPG LEEELKQLEE ELQAIEEQLA QLQWKAQARK EKLAQLKEKL 150
151 SGPGSPEDEI QSLEEKNSQL KQEISQLEEK NQQLKYGSGP GQLEDKVEEL LSKNYHLENE VSRLKKLVGS GPGSPEDKIS QLKQKIQQLK QENQQLEEEN SQLEYGSGPG SPEDENSQLE EKISQLKQKN SQLKEEIQQL EYGSGPGSPE 300
301 DKISQLKEEN QQLEQKIQQL KEENSQLEYG SGPGDIEQEL ERAKQSIRRL EQEVNQERSR MQYLQTLLSK SGPGSPEDKN SQLKEEIQQL EEENQQLEEK ISQLKYGSGP GLEEELKQLE EELQAIEEQL AQLQWKAQAR KEKLAQLKEK 450
451 LLEHHHHHHH H                                                                                                                                                    461
```

Name:	**TET12$_{1.6}$S-f$_5$b**
Segment sequence:	GCNshS-P3S-BCRS-GCNshS-APHshSN-P5S-BCRS-P7S-APHshSN-P4S-P8S-P6S

```
  1 MQLEDKVEEL LSKNYHLENE VSRLKKLVGS GPGSPEDEIQ QLEEEISQLE QKNSQLKEKN QQLKYGSGPG DIEQELERAK QSIRRLEQEV NQERSRMQYL QTLLSKSGPG QLEDKVEELL SKNYHLENEV SRLKKLVGSG PGLEEELKQL 150
151 EEELQAIEEQ LAQLQWKAQA RKEKLAQLKE KLSGPGSPED ENSQLEEKIS QLKQKNSQLK EEIQQLEYGS GPGDIEQELE RAKQSIRRLE QEVNQERSRM QYLQTLLSKS GPGSPEDEIQ SLEEKNSQLK QEISQLEEKN QQLKYGSGPG 300
301 LEEELKQLEE ELQAIEEQLA QLQWKAQARK EKLAQLKEKL SGPGSPEDKI SQLKQKIQQL KQENQQLEEE NSQLEYGSGP GSPEDKISQL KEENQQLEQK IQQLKEENSQ LEYGSGPGSP EDKNSQLKEE IQQLEEENQQ LEEKISQLKY 450
451 GLEHHHHHHH H                                                                                                                                                    461
```


Nature Biotechnology: doi:10.1038/nbt.3994

Name:	**TET12$_{1.6}$S-c$_6$b**
Segment sequence:	GCNshS-P3S-BCRS-GCNshS-APHshSN-P5S-BCRS-P7S-APHshSN-P4S-P8S-P6S

```
   1 MQLEDKVEEL LSKNYHLENE VSRLKKLVGG DGKGSPEDEI QQLEEEISQL EQKNSQLKEK NQQLKYGGKG DGDIEQELER AKQSIRRLEQ EVNQERSRMQ YLQTLLSKGK GDGQLEDKVE ELLSKNYHLE NEVSRLKKLV GGKGDGLEEE 150
 151 LKQLEEELQA IEEQLAQLQW KAQARKEKLA QLKEKLGKGD GSPEDENSQL EEKISQLKQK NSQLKEEIQQ LEYGGKGDGD IEQELERAKQ SIRRLEQEVN QERSRMQYLQ TLLSKGKGDG SPEDEIQSLE EKNSQLKQEI SQLEEKNQQL 300
 301 KYGGKGDGLE EELKQLEEEL QAIEEQLAQL QWKAQARKEK LAQLKEKLGK GDGSPEDKIS QLKQKIQQLK QENQQLEEEN SQLEYGGDGK GSPEDKISQL KEENQQLEQK IQQLKEENSQ LEYGGDGKGS PEDKNSQLKE EIQQLEEENQ 450
 451 QLEEKISQLK YGLEHHHHHH HH                                                                                                                                          472
```

Name:	**TET12$_{1.10}$xS-f$_6$**
Segment sequence:	APHshSN-P3S-BCRS-GCNshS-APHshSN-P7S-GCNshS-P4S-P5S-P8S-BCRS-P6S

```
   1 MLEEELKQLE EELQAIEEQL AQLQWKAQAR KEKLAQLKEK LSGPGSGEIQ QLEEEISQLE QKNSQLKEKN QQLKYGSGPG SGDIEQELER AKQSIRRLEQ EVNQERSRMQ YLQTLLSKSG PGSGQLEDKV EELLSKNYHL ENEVSRLKKL 150
 151 VGSGPGSGLE EELKQLEEEL QAIEEQLAQL QWKAQARKEK LAQLKEKLSG PGSGEIQSLE EKNSQLKQEI SQLEEKNQQL KYGSGPGSGQ LEDKVEELLS KNYHLENEVS RLKKLVGSGP GSGKISQLKQ KIQQLKQENQ QLEEENSQLE 300
 301 YGSGPGSGEN SQLEEKISQL KQKNSQLKEE IQQLEYGSGP GSGKISQLKE ENQQLEQKIQ QLKEENSQLE YGSGPGSGDI EQELERAKQS IRRLEQEVNQ ERSRMQYLQT LLSKSGPGSG KNSQLKEEIQ QLEEENQQLE EKISQLKYGL 450
 451 EHHHHHHHH                                                                                                                                                          459
```

Name:	**TET12$_{2.3}$SN-f$_5$b**
Segment sequence:	P5SN-P7SN-APHshSN-APH4SN-P6SN-BCRSN-APHshSN-P3SN-APH4SN-BCRSN- P8SN-P4SN

```
   1 MSPEDENSQL EEKISQLKQK NSELKEEIQQ LEYGSGPGSP EDEIQQLEEK NSQLKQEISQ LEEKNQELKY GSGPGLEEEL KQLEEELQAI EEQLAQLQWK AQARKEKLAQ LKEKLGSGPG LEQIEERLEQ IEERLQAKEW EKAQLREELQ 150
 151 ALREKLAQLG SGPGSPEDKN SELKEEIQQL EEENQQLEEK ISELKYGSGP GDIEQELERA KESIRRLEQE VNQERSRMQY LQTLLEKGSG PGLEEELKQL EEELQAIEEQ LAQLQWKAQA RKEKLAQLKE KLGSGPGSPE DEIQQLEEEI 300
 301 SQLEQKNSEL KEKNQELKYG SGPGLEQIEE RLEQIEERLQ AKEWEKAQLR EELQALREKL AQLGSGPGDI EQELERAKES IRRLEQEVNQ ERSRMQYLQT LLEKGSGPGS PEDKISELKE ENQQLEQKIQ QLKEENSQLE YGSGPGSPED 450
 451 KISQLKEKIQ QLKQENQQLE EENSQLEYGL EHHHHHHHH                                                                                                                          489
```

Name:	**PYR16$_{4.6}$SN-f$_5$**
Common name:	**PYR16SN**
Segment sequence:	APHshSN-P5SN-P1SN–GCNshSNb–APHshSN-P7SN–GCNshSNb-P6SN–BCRSN-P3SN-P8SN-P9SN–BCRSN-P2SN-P10SN-P4SN

```
   1 MLEEELKQLE EELQAIEEQL AQLQWKAQAR KEKLAQLKEK LGSGPGSPED ENSQLEEKIS QLKQKNSELK EEIQQLEYGS GPGSPEDEIR QLEQENSQLE RENQRLEQEI YQLERGSGPG SRMKQLEDKV EELLSKNYHL ENEVERLKKL 150
 151 VGSGPGLEEE LKQLEEELQA IEEQLAQLQW KAQARKEKLA QLKEKLGSGP GSPEDEIQQL EEKNSQLKQE ISQLEEKNQE LKYGSGPGSR MKQLEDKVEE LLSKNYHLEN EVERLKKLVG SGPGSPEDKN SELKEEIQQL EEENQQLEEK 300
 301 ISELKYGSGP GDIEQELERA KESIRRLEQE VNQERSRMQY LQTLLEKGSG PGSPEDEIQQ LEEEISQLEQ KNSELKEKNQ ELKYGSGPGS PEDKISELKE ENQQLEQKIQ QLKEENSQLE YGSGPGSPED ENQSLEQKNS QLKQEISQLE 450
 451 QEIQQLEYGS GPGDIEQELE RAKESIRRLE QEVNQERSRM QYLQTLLEKG SGPGSPEDKI EELKEKNSQL KEKNEELKQK IYELKEGSGP GSPEDKNSQL KEENSQLEEK IEQLKEKIQE LKYGSGPGSP EDKISQLKEK IQQLKQENQQ 600
 601 LEEENSQLEY GLEHHHHHHH H                                                                                                                                           621
```

Name:	**PYR16$_{2.15R}$S-f$_5$**
Segment sequence:	P11S-P3S-BCRsh-P6S-APHsh-P4S-P9S-P12S-APHsh-P1S-P7S-BCRsh-P10S-P8S-P5S-P2S

```
   1 MSPEDENQQL EQEISQLEQE ISQLEQKNSQ LKYGSGPGSP EDEIQQLEEE ISQLEQKNSQ LKEKNQQLKY GSGPGDIEQE LERAKQSIRR LEQEVNQERS RMQYLQTLLS KSGPGSPEDK NSQLKEEIQQ LEEENQQLEE KISQLKYGSG 150
 151 PGELKQLEEE LQAIEEQLAQ LQWKAQARKE KLAQLKSGPG SPEDKISQLK QKIQQLKQEN QQLEEENSQL EYGSGPGSPE DENQQLEQKN SQLKQEISQL EQEISQLEYG SGPGSPEDKN SQLKEKISQL KEKIQQLKEE NQQLEYGSGP 300
 301 GELKQLEEEL QAIEEQLAQL QWKAQARKEK LAQLKSGPGS PEDEIQALEQ ENAQLEQENA ALEQEIAQLE YGSGPGSPED EIQSLEEKNS QLKQEISQLE EKNQQLKYGS GPGDIEQELE RAKQSIRRLE QEVNQERSRM QYLQTLLSKS 450
 451 GPGSPEDKNS QLKEENSQLE EKIQQLKEKI QQLKYGSGPG SPEDKISQLK EENQQLEQKI QQLKEENSQL EYGSGPGSPE DENSQLEEKI SQLKQKNSQL KEEIQQLEYG SGPGSPEDKI DQLKEKNADL KEKNQDLKEK IDALKYGLEH 600
 601 HHHHHHH                                                                                                                                                          607
```

Name:	**TRIP18$_{7.5R}$SN-f$_5$**
Common name:	**TRIP18SN**
Segment sequence:	P5SN-APH4SN-P1SN-P6SN-GCNshSNb-APHshSN-P9SN-APH4SN-GCNshSNb-P3SN-P11SN-P10SN-P2SN-BCRSN-P12SN-APHshSN-P4SN-BCRSN

```
   1 MSPEDENSQL EEKISQLKQK NSELKEEIQQ LEYGSGPGLE QIEERLEQIE ERLQAKEWEK AQLREELQAL REKLAQLGSG PGSPEDEIRQ LEQENSQLER ENQRLEQEIY QLERGSGPGS PEDKNSELKE EIQQLEEENQ QLEEKISELK 150
 151 YGSGPGSRMK QLEDKVEELL SKNYHLENEV ERLKKLVGSG PGLEEELKQL EEELQAIEEQ LAQLQWKAQA RKEKLAQLKE KLGSGPGSPE DENQSLEQKN SQLKQEISQL EQEIQQLEYG SGPGLEQIEE RLEQIEERLQ AKEWEKAQLR 300
 301 EELQALREKL AQLGSGPGSR MKQLEDKVEE LLSKNYHLEN EVERLKKLVG SGPGSPEDEI QQLEEEISQL EQKNSELKEK NQELKYGSGP GSPEDENQSL EQEISQLEQE IQQLEQKNSE LKYGSGPGSP EDKNSQLKEE NSQLEEKIEQ 450
 451 LKEKIQELKY GSGPGSPEDK IEELKEKNSQ LKEKNEELKQ KIYELKEGSG PGDIEQELER AKESIRRLEQ EVNQERSRMQ YLQTLLEKGS GPGSPEDKNE QLKEKISELK EKIEQLKEEN QSLEYGSGPG LEEELKQLEE ELQAIEEQLA 600
 601 QLQWKAQARK EKLAQLKEKL GSGPGSPEDK ISQLKEKIQQ LKQENQQLEE ENSQLEYGSG PGDIEQELER AKESIRRLEQ EVNQERSRMQ YLQTLLEKLE HHHHHHHH                                                  708
```

Name:	**TET12$_{1.10}$SN-f$_5$-2Cys**
Common name:	**TET12SN-2Cys**
Segment sequence:	APHshSN-P3SN-BCRSN-GCNshSN-APHshSN-P7SN-GCNshSN-P4SN-P5SN-P8SN-BCRSN-P6SN

```
   1 MCTGLDLEEE LKQLEEELQA IEEQLAQLQW KAQARKEKLA QLKEKLSGPG SPEDEIQQLE EEISQLEQKN SELKEKNQEL KYGSGPGDIE QELERAKESI RRLEQEVNQE RSRMQYLQTL LEKSGPGQLE DKVEELLSKN YHLENEVERL 150
 151 KKLVGSGPGL EEELKQLEEE LQAIEEQLAQ LQWKAQARKE KLAQLKEKLS GPGSPEDEIQ QLEEKNSQLK QEISQLEEKN QELKYGSGPG QLEDKVEELL SKNYHLENEV ERLKKLVGSG PGSPEDKISQ LKEKIQQLKQ ENQQLEEENS 300
 301 QLEYGSGPGS PEDENSQLEE KISQLKQKNS ELKEEIQQLE YGSGPGSPED KISELKEENQ QLEQKIQQLK EENSQLEYGS GPGDIEQELE RAKESIRRLE QEVNQERSRM QYLQTLLEKS GPGSPEDKNS ELKEEIQQLE EENQQLEEKI 450
 451 SELKYGSGEK TKRCDPLEHH HHHHHH                                                                                                                                      476
```


Nature Biotechnology: doi:10.1038/nbt.3994

Name:	**PYR16$_{4.6}$SN-f$_5$-2Cys**
Common name:	**PYR16SN-2Cys**
Segment sequence:	APHshSN-P5SN-P1SN–GCNshSNb–APHshSN-P7SN–GCNshSNb-P6SN–BCRSN-P3SN-P8SN-P9SN–BCRSN-P2SN-P10SN-P4SN

```
  1 MGCGGSGSHM LEEELKQLEE ELQAIEEQLA QLQWKAQARK EKLAQLKEKL GSGPGSPEDE NSQLEEKISQ LKQKNSELKE EIQQLEYGSG PGSPEDEIRQ LEQENSQLER ENQRLEQEIY QLERGSGPGS RMKQLEDKVE ELLSKNYHLE 150
151 NEVERLKKLV GSGPGLEEEL KQLEEELQAI EEQLAQLQWK AQARKEKLAQ LKEKLGSGPG SPEDEIQQLE EKNSQLKQEI SQLEEKNQEL KYGSGPGSRM KQLEDKVEEL LSKNYHLENE VERLKKLVGS GPGSPEDKNS ELKEEIQQLE 300
301 EENQQLEEKI SELKYGSGPG DIEQELERAK ESIRRLEQEVN QERSRMQYL QTLLEKGSGP GSPEDEIQQL EEEISQLEQK NSELKEKNQE LKYGSGPGSP EDKISELKEE NQQLEQKIQQ LKEENSQLEY GSGPGSPEDE NQSLEQKNSQ 450
451 LKQEISQLEQ EIQQLEYGSG PGDIEQELER AKESIRRLEQ EVNQERSRMQ YLQTLLEKGS GPGSPEDKIE ELKEKNSQLK EKNEELKQKI YELKEGSGPG SPEDKNSQLK EENSQLEEKI EQLKEKIQEL KYGSGPGSPE DKISQLKEKI 600
601 QQLKQENQQL EEENSQLEYG GSGGGSLECH HHHHHH                                                                                                                                  636
```

Name:	**TRIP18$_{7.5R}$SN-f$_5$-2Cys**
Common name:	**TRIP18SN-2Cys**
Segment sequence:	P5SN-APH4SN-P1SN-P6SN-GCNshSNb-APHshSN-P9SN-APH4SN-GCNshSNb-P3SN-P11SN-P10SN-P2SN-BCRSN-P12SN-APHshSN-P4SN-BCRSN

```
  1 MGCGGSGSHM SPEDENSQLE EKISQLKQKN SELKEEIQQL EYGSGPGLEQ IEERLEQIEE RLQAKEWEKA QLREELQALR EKLAQLGSGP GSPEDEIRQL EQENSQLERE NQRLEQEIYQ LERGSGPGSP EDKNSELKEE IQQLEEENQQ 150
151 LEEKISELKY GSGPGSRMKQ LEDKVEELLS KNYHLENEVE RLKKLVGSGP GLEEELKQLE EELQAIEEQL AQLQWKAQAR KEKLAQLKEK LGSGPGSPED ENQSLEQKNS QLKQEISQLE QEIQQLEYGS GPGLEQIEER LEQIEERLQA 300
301 KEWEKAQLRE ELQALREKLA QLGSGPGSRM KQLEDKVEEL LSKNYHLENE VERLKKLVGS GPGSPEDEIQ QLEEEISQLE QKNSELKEKN QELKYGSGPG SPEDENQSLE QEISQLEQEI QQLEQKNSEL KYGSGPGSPE DKNSQLKEEN 450
451 SQLEEKIEQL KEKIQELKYG SGPGSPEDKI EELKEKNSQL KEKNEELKQK IYELKEGSGP GDIEQELERA KESIRRLEQE VNQERSRMQY LQTLLEKGSG PGSPEDKNEQ LKEKISELKE KIEQLKEENQ SLEYGSGPGL EEELKQLEEE 600
601 LQAIEEQLAQ LQWKAQARKE KLAQLKEKLG SGPGSPEDKI SQLKEKIQQL KQENQQLEEE NSQLEYGSGP GDIEQELERA KESIRRLEQE VNQERSRMQY LQTLLEKGSG GGSLECHHHH HHH                                723
```

(Data in fasta format also published at figshare with doi: 10.6084/m9.figshare.4003398)

Nature Biotechnology: doi:10.1038/nbt.3994

Table S4: Comparison of model experimental molecular weights and hydrodynamic diameters (D_H)

[*]Hydrodynamic radii were calculated from using HYDROPRO software[80] from molecular models showing the best agreement with experimental SAXS data.

[#]D_H was obtained using Rg (Guinier) reported in Table S7 and the relation $Rg=\sqrt{3/5}R_H$ valid for globular proteins.

[l]The D_H was calculated for the rectangular and oblique triangular prism respectively.

	Molecular weight from sequence	SEC-MALS Native	SEC-MALS Refolded	Model[*]	DLS Native	DLS Refolded	SAXS[#] Native
	Mr (kDa)	Mr (kDa)	Mr (kDa)	D_H (nm)	D_H (nm)	D_H (nm)	D_H (nm)
TET12$_{1.10}$SN-f$_5$	53.4	57.6±1.2%	55.0±1.0%	7.8	6.7±1.5	07.5±2.0	09.0
TET12$_{1.10}$SN-c$_6$	54.3	57.5±1.0%	64.7±1.4%	7.8	9.1±1.9	11.5±2.4	09.3
TET12$_{1.10}$SN-f$_9$	56.6	53.6±0.9%	64.5±1.2%	8.6	9.6±2.0	09.0±2.3	10.1
TET12$_{2.3}$SN-f$_5$b	56.7	52.4±4.1%	51.8±8.7%	8.0	9.8±2.3	10.8±2.7	09.3
TET12$_{1.10}$S-f$_5$	53.2	53.9±0.7%	56.6±1.6%	7.8	6.3±1.3	07.1±1.6	09.3
TET12$_{1.10}$S-f$_5$b	53.2	52.8±0.8%	55.5±2.1%	8.4	6.7±1.7	07.2±1.7	08.7
TET12$_{1.10}$S-c$_6$	54.5	57.5±0.6%	61.7±1.1%	8.0	9.0±2.1	11.8±3.1	08.9
TET12$_{1.11}$S-f$_5$	53.2	54.4±0.8%	54.1±2.3%	7.2	6.0±1.3	07.5±1.9	08.5
TET12$_{1.6}$S-f$_5$b	53.2	53.9±0.2%	53.3±0.2%	7.2	6.1±1.3	07.4±1.9	10.5
TET12$_{1.6}$S-c$_6$b	54.5	58.2±0.8%	55.3±0.9%	7.6	6.4±1.3	06.6±1.8	09.1
TET12$_{1.10}$S-f$_6$	52.2	50.4±0.2%	47.7±9.7%	7.4	7.1±1.6	09.2±1.7	/
PYR16$_{4.6}$SN-f$_5$	71.6	80.9±0.9%	81.1±1.1%	8.8	9.7±2.2	09.3±2.4	09.9
PYR16$_{2.15R}$S-f$_5$	69.4	92.6±1.5%	85.2±0.9%	8.8	9.3±2.2	09.0±2.4	10.0
TRIP18$_{7.5R}$SN-f$_5$	81.9	78.6±1.0%	82.6±0.8%	9.4/9.2[l]	9.6±2.4	12.1±3.4	10.8

Nature Biotechnology: doi:10.1038/nbt.3994

Table S5: Summary of identified DSS, BS(PEG)$_5$ or BS(PEG)$_9$-cross-linked amino acids in the TET12SN.

In the absence of a cross-linker, all lysine-containing peptides were identified and protein coverage more than 95% of the entire sequence was obtained. Rows in white color: cross-linking within the same segment or to neighboring linker. Rows in green color: cross-linked residues within coiled-coil pair. Rows in yellow color: cross-linked residues within consecutive segments. Rows in blue color: long range cross-linking. **Bold** cross-links are presented in Fig. S6 and Table S6.

No. of a cross-link	Crosslinker	Residue 1	Residue 2	Segment 1*	Segment 2*	N[&]	Minimal distance[#]
1	DSS	K31	K38	APHshSN$_1$	APHshSN$_1$	6	9.66
2	DSS	K184	K191	APHshSN$_5$	APHshSN$_5$	5	9.69
3	**DSS**	**K92**	**K414**	**BCRSN$_3$**	**BCRSN$_{11}$**	**5**	**17.37**
4	**DSS**	**L69**	**S94**	**P3SN$_2$**	**BCRSN$_3$**	**3**	**17.78**
5	**DSS**	**K134**	**K259**	**GCNshSN$_4$**	**GCNshSN$_7$**	**3**	**15.01**
6	DSS	K134	K146	GCNshSN$_4$	GCNshSN$_4$	3	17.15
7	**DSS**	**K76**	**K92**	**P3SN$_2$**	**BCRSN$_3$**	**3**	**18.42**
8	DSS	K277	S268	P4SN$_8$	P4SN$_8$	3	12.48
9	DSS	K33	K259	APHshSN$_1$	GCNshSN$_7$	2	14.91
10	DSS	K33	K146	APHshSN$_1$	GCNshSN$_4$	2	13.24
11	DSS	K186	S195	APHshSN$_5$	linker	2	13.33
12	DSS	K69	K92	P3SN$_2$	BCRSN$_3$	2	17.39
13	DSS	K146	Y249	GCNshSN$_4$	GCNshSN$_7$	2	15.89
14	DSS	Y249	K259	GCNshSN$_7$	GCNshSN$_7$	2	14.51
15	DSS	K33	S46	APHshSN$_1$	P3SN$_2$	2	19.75
16	DSS	S425	K444	P6SN$_{12}$	P6SN$_{12}$	2	27.59
17	DSS	K26	K184	APHshSN$_1$	APHshSN$_5$	2	22.46
18	DSS	K26	K31	APHshSN$_1$	APHshSN$_1$	2	7.95
19	DSS	K26	K33	APHshSN$_1$	APHshSN$_1$	2	9.74
20	DSS	K33	S42	APHshSN$_1$	linker	2	13.34
21	DSS	K38	S42	APHshSN$_1$	linker	2	4.93
22	DSS	K26	K38	APHshSN$_1$	APHshSN$_1$	2	17.30
23	DSS	S133	K388	GCNshSN$_4$	BCRSN$_{11}$	2	25.68
24	DSS	K146	S246	GCNshSN$_4$	GCNshSN$_7$	2	18.34
25	DSS	S246	K259	GCNshSN$_7$	GCNshSN$_7$	2	19.22
26	DSS	K423	K444	P6SN$_{12}$	P6SN$_{12}$	2	29.92
27	DSS	K118	S390	BCRSN$_3$	BCRSN$_{11}$	2	21.27
28	DSS	S133	K146	GCNshSN$_4$	GCNshSN$_4$	2	19.07
29	DSS	S133	K259	GCNshSN$_4$	GCNshSN$_7$	2	18.23
30	DSS	K76	S94	P3SN$_2$	BCRSN$_3$	1	20.60
31	DSS	K186	K193	APHshSN$_5$	APHshSN$_5$	1	9.66
32	DSS	S94	K414	BCRSN$_3$	BCRSN$_{11}$	1	21.68
33	DSS	K184	S195	APHshSN$_5$	linker	1	14.55
34	**DSS**	**K184**	**S199**	**APHshSN$_5$**	**P7SN$_6$**	**1**	**18.58**
35	DSS	K184	K193	APHshSN$_5$	APHshSN$_5$	1	13.44
36	DSS	K191	K193	APHshSN$_5$	APHshSN$_5$	1	4.88
37	**DSS**	**K323**	**S348**	**P5SN$_9$**	**P8SN$_{10}$**	**1**	**20.06**
38	**DSS**	**K323**	**S446**	**P5SN$_9$**	**P6SN$_{12}$**	**1**	**13.53**
39	DSS	K31	S46	APHshSN$_1$	P3SN$_2$	1	21.10
40	DSS	S46	K146	P3SN$_2$	GCNshSN$_4$	1	10.56
41	DSS	S46	K259	P3SN$_2$	GCNshSN$_7$	1	6.25
42	DSS	K33	K40	APHshSN$_1$	APHshSN$_1$	1	9.69
43	DSS	K38	S46	APHshSN$_1$	P3SN$_2$	1	11.51
44	DSS	K31	K179	APHshSN$_1$	APHshSN$_5$	1	22.49
45	DSS	K179	K184	APHshSN$_5$	APHshSN$_5$	1	7.67

Nature Biotechnology: doi:10.1038/nbt.3994

46	DSS	K179	K186	$APHshSN_5$	$APHshSN_5$	1	9.57
47	DSS	K179	K191	$APHshSN_5$	$APHshSN_5$	1	17.14
48	DSS	K247	Y249	$GCNshSN_7$	$GCNshSN_7$	1	4.99
49	DSS	Y136	K247	$GCNshSN_4$	$GCNshSN_7$	1	9.07
50	DSS	K240	K388	$GCNshSN_7$	$BCRSN_{11}$	1	21.90
51	**DSS**	**K210**	**K351**	**$P7SN_6$**	**$P8SN_{10}$**	**1**	**8.86**
52	DSS	S415	K444	linker	$P6SN_{12}$	1	41.13
53	DSS	K26	K40	$APHshSN_1$	$APHshSN_1$	1	19.98
54	DSS	K259	K272	$GCNshSN_7$	$P4SN_8$	1	12.15
55	DSS	K146	K272	$GCNshSN_4$	$P4SN_8$	1	19.03
56	DSS	K272	K279	$P4SN_8$	$P4SN_8$	1	9.64
57	DSS	K127	K388	$GCNshSN_4$	$BCRSN_{11}$	1	18.13
58	DSS	K31	S42	$APHshSN_1$	linker	1	14.83
59	DSS	K40	K259	$APHshSN_1$	$GCNshSN_7$	1	12.31
60	DSS	K40	K146	$APHshSN_1$	$GCNshSN_4$	1	10.67
61	DSS	K31	K40	$APHshSN_1$	$APHshSN_1$	1	13.34
62	DSS	S348	K360	$P8SN_{10}$	$P8SN_{10}$	1	17.56
63	DSS	K92	S133	$BCRSN_3$	$GCNshSN_4$	1	36.07
64	DSS	S419	K444	$P6SN_{12}$	$P6SN_{12}$	1	33.74
65	DSS	S419	K428	$P6SN_{12}$	$P6SN_{12}$	1	10.29
66	**$BS(PEG)_5$**	**S348**	**K360**	**$P8SN_{10}$**	**$P8SN_{10}$**	**7**	**17.56**
67	**$BS(PEG)_5$**	**K360**	**S446**	**$P8SN_{10}$**	**$P6SN_{12}$**	**7**	**31.06**
68	$BS(PEG)_5$	K184	K191	$APHshSN_5$	$APHshSN_5$	5	9.69
69	$BS(PEG)_5$	K31	K38	$APHshSN_1$	$APHshSN_1$	5	9.66
70	**$BS(PEG)_5$**	**K92**	**K414**	**$BCRSN_3$**	**$BCRSN_{11}$**	**3**	**17.37**
71	**$BS(PEG)_5$**	**K33**	**S46**	**$APHshSN_1$**	**$P3SN_2$**	**2**	**19.75**
72	$BS(PEG)_5$	K31	K40	$APHshSN_1$	$APHshSN_1$	2	13.34
73	**$BS(PEG)_5$**	**K260**	**K388**	**$GCNshSN_7$**	**$BCRSN_{11}$**	**2**	**31.15**
74	**$BS(PEG)_5$**	**K92**	**K260**	**$BCRSN_3$**	**$GCNshSN_7$**	**2**	**31.60**
75	$BS(PEG)_5$	K31	S42	$APHshSN_1$	linker	2	14.83
76	$BS(PEG)_5$	K184	K193	$APHshSN_5$	$APHshSN_5$	2	13.44
77	$BS(PEG)_5$	S419	K444	$P6SN_{12}$	$P6SN_{12}$	2	33.74
78	$BS(PEG)_5$	K33	S42	$APHshSN_1$	linker	1	13.34
79	$BS(PEG)_5$	K259	K272	$GCNshSN_7$	$P4SN_8$	1	12.15
80	$BS(PEG)_5$	K146	K272	$GCNshSN_4$	$P4SN_8$	1	19.03
81	$BS(PEG)_5$	K184	S195	$APHshSN_5$	linker	1	14.55
82	$BS(PEG)_5$	K69	S94	$P3SN_2$	$BCRSN_3$	1	17.78
83	$BS(PEG)_5$	K69	S390	$P3SN_2$	$BCRSN_{11}$	1	31.44
84	$BS(PEG)_5$	K92	K240	$BCRSN_3$	$GCNshSN_7$	1	34.62
85	$BS(PEG)_5$	K240	K388	$GCNshSN_7$	$BCRSN_{11}$	1	21.90
86	$BS(PEG)_5$	K260	K279	$GCNshSN_7$	$P4SN_8$	1	11.38
87	$BS(PEG)_5$	K423	K449	$P6SN_{12}$	$P6SN_{12}$	1	36.14
88	$BS(PEG)_5$	K346	K360	$P8SN_{10}$	$P8SN_{10}$	1	19.80
89	$BS(PEG)_5$	S212	K224	$P7SN_6$	$P7SN_6$	1	17.58
90	$BS(PEG)_5$	K38	K40	$APHshSN_1$	$APHshSN_1$	1	4.86
91	$BS(PEG)_5$	K127	S390	$GCNshSN_4$	$BCRSN_{11}$	1	18.52
92	$BS(PEG)_5$	S94	K127	$BCRSN_3$	$GCNshSN_4$	1	29.06
93	$BS(PEG)_5$	K92	K127	$BCRSN_3$	$GCNshSN_4$	1	33.20
94	$BS(PEG)_5$	K127	K388	$GCNshSN_4$	$BCRSN_{11}$	1	18.13
95	$BS(PEG)_5$	K186	K193	$APHshSN_5$	$APHshSN_5$	1	9.66
96	$BS(PEG)_5$	K191	S195	$APHshSN_5$	linker	1	4.84
97	$BS(PEG)_9$	S348	K360	$P8SN_{10}$	$P8SN_{10}$	7	17.56
98	**$BS(PEG)_9$**	**K360**	**S446**	**$P8SN_{10}$**	**$P6SN_{12}$**	**7**	**31.06**
99	**$BS(PEG)_9$**	**K38**	**K184**	**$APHshSN_1$**	**$APHshSN_5$**	**6**	**36.48**
100	$BS(PEG)_9$	K184	K191	$APHshSN_5$	$APHshSN_5$	6	9.69
101	$BS(PEG)_9$	K31	K38	$APHshSN_1$	$APHshSN_1$	6	9.66
102	**$BS(PEG)_9$**	**K31**	**K191**	**$APHshSN_1$**	**$APHshSN_5$**	**6**	**36.32**

Nature Biotechnology: doi:10.1038/nbt.3994

103	**BS(PEG)$_9$**	**K92**	**K260**	**BCRSN$_3$**	**GCNshSN$_7$**	**4**	**31.60**
104	**BS(PEG)$_9$**	**K260**	**K388**	**GCNshSN$_7$**	**BCRSN$_{11}$**	**4**	**31.15**
105	**BS(PEG)$_9$**	**K69**	**S390**	**P3SN$_2$**	**BCRSN$_{11}$**	**4**	**31.44**
106	**BS(PEG)$_9$**	**K69**	**S94**	**P3SN$_2$**	**BCRSN$_3$**	**4**	**17.78**
107	**BS(PEG)$_9$**	**K259**	**K272**	**GCNshSN$_7$**	**P4SN$_8$**	**3**	**12.15**
108	**BS(PEG)9**	**K146**	**K272**	**GCNshSN$_4$**	**P4SN$_8$**	**3**	**19.03**
109	BS(PEG)$_9$	K184	K193	APHshSN$_5$	APHshSN$_5$	3	13.44
110	**BS(PEG)$_9$**	**K31**	**K193**	**APHshSN$_1$**	**APHshSN$_5$**	**3**	**40.89**
111	**BS(PEG)$_9$**	**K92**	**K279**	**BCRSN$_3$**	**P4SN$_8$**	**3**	**30.39**
112	**BS(PEG)$_9$**	**K279**	**K388**	**P4SN$_8$**	**BCRSN$_{11}$**	**3**	**39.39**
113	**BS(PEG)$_9$**	**S94**	**K260**	**BCRSN$_3$**	**GCNshSN$_7$**	**3**	**27.23**
114	**BS(PEG)$_9$**	**K260**	**S390**	**GCNshSN$_7$**	**BCRSN$_{11}$**	**3**	**26.42**
115	BS(PEG)$_9$	K92	S274	BCRSN$_3$	P4SN$_8$	2	38.09
116	BS(PEG)$_9$	S274	K388	P4SN$_8$	BCRSN$_{11}$	2	43.93
117	BS(PEG)$_9$	K186	S195	APHshSN$_5$	linker	2	13.33
118	BS(PEG)$_9$	K92	S415	BCRSN$_3$	linker	2	17.43
119	BS(PEG)$_9$	K388	S415	BCRSN$_{11}$	linker	2	38.88
120	BS(PEG)$_9$	S390	K414	BCRSN$_{11}$	BCRSN$_{11}$	2	34.52
121	BS(PEG)$_9$	S94	K414	BCRSN$_3$	BCRSN$_{11}$	2	21.68
122	BS(PEG)$_9$	K388	K414	BCRSN$_{11}$	BCRSN$_{11}$	2	37.19
123	BS(PEG)$_9$	K92	K414	BCRSN$_3$	BCRSN$_{11}$	2	17.37
124	BS(PEG)$_9$	K260	K279	GCNshSN$_7$	P4SN$_8$	2	11.38
125	BS(PEG)$_9$	K31	S195	APHshSN$_1$	linker	2	40.73
126	BS(PEG)$_9$	K184	S195	APHshSN$_5$	linker	2	14.55
127	BS(PEG)$_9$	K146	K147	GCNshSN$_4$	GCNshSN$_4$	1	3.68
128	BS(PEG)$_9$	K147	K259	GCNshSN$_4$	GCNshSN$_7$	1	10.51
129	BS(PEG)$_9$	K33	K193	APHshSN$_1$	APHshSN$_5$	1	45.46
130	BS(PEG)$_9$	K186	K193	APHshSN$_5$	APHshSN$_5$	1	9.66
131	BS(PEG)$_9$	S94	K118	BCRSN$_3$	BCRSN$_3$	1	33.68
132	BS(PEG)$_9$	K118	S390	BCRSN$_3$	BCRSN$_{11}$	1	21.27
133	BS(PEG)$_9$	K69	K388	P3SN$_2$	BCRSN$_{11}$	1	36.17
134	BS(PEG)$_9$	K69	K92	P3SN$_2$	BCRSN$_3$	1	17.39
135	BS(PEG)$_9$	K184	S264	APHshSN$_5$	linker	1	43.10
136	BS(PEG)$_9$	K31	S264	APHshSN$_1$	linker	1	15.02
137	BS(PEG)$_9$	K279	Y407	P4SN$_8$	BCRSN$_{11}$	1	41.44
138	BS(PEG)$_9$	Y111	K279	BCRSN$_{11}$	P4SN$_8$	1	37.65
139	BS(PEG)$_9$	K31	K69	APHshSN$_1$	P3SN$_2$	1	33.44
140	BS(PEG)$_9$	K69	K259	P3SN$_2$	GCNshSN$_7$	1	27.14
141	BS(PEG)$_9$	K127	K184	GCNshSN$_4$	APHshSN$_5$	1	36.80
142	BS(PEG)$_9$	K31	K127	APHshSN$_1$	GCNshSN$_4$	1	30.23
143	BS(PEG)$_9$	K31	S66	APHshSN$_1$	P3SN$_2$	1	31.77
144	BS(PEG)$_9$	K33	S42	APHshSN$_1$	linker	1	13.34
145	BS(PEG)$_9$	K38	S42	APHshSN$_1$	linker	1	4.93
146	BS(PEG)$_9$	S425	K444	P6SN$_{12}$	P6SN$_{12}$	1	27.59
147	BS(PEG)$_9$	K71	S94	P3SN$_2$	BCRSN$_3$	1	14.15
148	BS(PEG)$_9$	K71	S390	P3SN$_2$	BCRSN$_{11}$	1	29.35
149	BS(PEG)$_9$	K146	K260	GCNshSN$_4$	GCNshSN$_7$	1	10.79
150	BS(PEG)$_9$	K259	K260	GCNshSN$_7$	GCNshSN$_7$	1	3.66
151	BS(PEG)$_9$	S419	K444	P6SN$_{12}$	P6SN$_{12}$	1	33.74
152	BS(PEG)$_9$	K423	K449	P6SN$_{12}$	P6SN$_{12}$	1	36.14
153	BS(PEG)$_9$	K428	S446	P6SN$_{12}$	P6SN$_{12}$	1	25.67
154	BS(PEG)$_9$	S348	K428	P8SN$_{10}$	P6SN$_{12}$	1	37.06

* Serial number of the CC segment in subscript is provided in order to present the information on the crosslinked segments. Therefore the polypeptide TET12SN is presented as APHshSN$_1$-P3SN$_2$-BCRSN$_3$-GCNshSN$_4$-APHshSN$_5$-P7SN$_6$-GCNshSN$_7$-P4SN$_8$-P5SN$_9$-P8SN$_{10}$-BCRSN$_{11}$-P6SN$_{12}$.

Nature Biotechnology: doi:10.1038/nbt.3994

[&] Number of occurrences of identified cross-linked peptides in independent experiments.
[#] Minimal distance between Cα atoms of cross-linked lysine, serine or tyrosine residues in Angstroms identified during molecular dynamics simulations of TET12SN (NPT ensemble, 300 K, 1 bar, 50 ns, 2 ps time step, Gromos54a7 force field) using GROMACS molecular dynamics package [56].

Nature Biotechnology: doi:10.1038/nbt.3994

Table S6: Identified DSS, BS(PEG)$_5$ or BS(PEG)$_9$-cross-linked amino acids of TET12SN presented in Fig. S6

The table shows only the highest scoring results for a cross-linked pair. Cross-linked amino acid residues are written in **bold**.

Crosslinker	Score	m/z [~]	Charge	M+H+ [$]	Calculated mass	Deviation (ppm)	No. of a cross-link [&]	Sequence 1 *	Sequence 2 *	Cross-linked residues
DSS	124	879.95	4	3516.759	3516.759	0.07	38	[Q**K**NSELKEEIQQLEYGSGPGSPEDK]	[I**S**ELK]	K323-S446
							37	[Q**K**NSELKEEIQQLEYGSGPGSPEDK]	[I**S**ELK]	K323-S348
DSS	101	741.14	4	2961.525	2961.524	0.35	3 [#]	[MQYLQTLLE**K**SGPGSPEDK]	[A**K**ESIR]	K92-K414
DSS	80	563.31	3	1687.923	1687.922	0.59	4 [#]	[NSEL**K**EK]	[AKE**S**IR]	K69-S94
DSS	73	708.37	5	3537.841	3537.844	-0.71	5 [#]	[SGPGQLEDKVEELLS**K**NYHLENEVER]	[L**K**K]	K134-K259
DSS	73	642.35	5	3207.702	3207.699	0.92	34	[**K**EKLAQLKEK]	[LSGPG**S**PEDEIQQLEEK]	K184-S199
DSS	72	751.38	4	3002.512	3002.507	1.93	7 [#]	[NQEL**K**YGSGPGDIEQELER]	[A**K**ESIR]	K76-K92
DSS	64	1070.5	4	4279.149	4279.146	0.52	51	[LSGPGSPEDEIQQLEE**K**NSQLK]	[ISEL**K**EENQQLEQK]	K210-K351
BS(PEG)$_5$	165	882.81	3	2646.41	2646.408	0.45	66 [#]	[I**S**ELK]	[EENQQLEQ**K**IQQLK]	S348-K360
							67 [#]	[EENQQLEQ**K**IQQLK]	[I**S**ELK]	K360-S446
BS(PEG)$_5$	132	1050.5	4	4199.176	4199.17	1.31	71 [#]	[KE**K**LAQLKEK]	[LSGPG**S**PEDEIQQLEEEISQLEQK]	K33-S46
BS(PEG)$_5$	85	1042.5	3	3125.597	3125.592	1.57	70 [#]	[MQYLQTLLE**K**SGPGSPEDK]	[A**K**ESIR]	K92-K414
BS(PEG)$_5$	75	569.56	4	2275.204	2275.203	0.45	73	[**K**LVGSGPGSPEDK]	[A**K**ESIR]	K260-K388
							74	[A**K**ESIR]	[**K**LVGSGPGSPEDK]	K92-K260
BS(PEG)$_9$	175	645.12	4	2577.447	2577.448	-0.41	107	[L**K**K]	[LVGSGPGSPED**K**ISQLK]	K259-K272
							108	[L**K**K]	[LVGSGPGSPED**K**ISQLK]	K146-K272
BS(PEG)$_9$	145	706.39	4	2822.517	2822.513	1.42	98 [#]	[EENQQLEQ**K**IQQLK]	[I**S**ELK]	K360-S446
BS(PEG)$_9$	110	613.58	4	2451.31	2451.308	0.87	104 [#]	[**K**LVGSGPGSPEDK]	[A**K**ESIR]	K260-K388
							103 [#]	[A**K**ESIR]	[**K**LVGSGPGSPEDK]	K92-K260
BS(PEG)$_9$	106	571	3	1710.999	1710.999	-0.03	102 [#]	[**K**EK]	[LAQL**K**EK]	K31-K191
							99 [#]	[LAQL**K**EK]	[**K**EK]	K38-K184
BS(PEG)$_9$	103	517.55	4	2067.181	2067.179	0.75	112	[E**K**IQQLK]	[A**K**ESIR]	K279-K388
							111	[A**K**ESIR]	[E**K**IQQLK]	K92-K279
BS(PEG)$_9$	98	676.7	3	2028.097	2028.096	0.42	106 [#]	[NSEL**K**EK]	[AKE**S**IR]	K69-S94
							105 [#]	[NSEL**K**EK]	[AKE**S**IR]	K69-S390
BS(PEG)$_9$	91	710.38	5	3547.874	3547.873	0.33	110 [#]	[**K**EK]	[LAQLKE**K**LSGPGSPEDEIQQLEEK]	K31-K193
BS(PEG)$_9$	90	613.58	4	2451.307	2451.308	-0.13	114	[**K**LVGSGPGSPEDK]	[AKE**S**IR]	K260-S390
							113	[AKE**S**IR]	[**K**LVGSGPGSPEDK]	S94-K260

[~] m/z, mass to charge ratio. [$] M+H+, measured mass.

[&] Depicted number of a cross-link is identical as in Table S5.

[*] White label depicts cross-linking within the same segment or to neighboring linker, green depicts cross-linked residues within coiled-coil pair, yellow depicts cross-linked residues within consecutive segments, and blue depicts long range cross-linking.

[#] Cross-linked peptide was identified in two independent experiments.

Nature Biotechnology: doi:10.1038/nbt.3994

Table S7: Analysis of SAXS data

Model free parameters obtained from collected SAXS curves.

	I(0) from Guinier	Rg from Guinier (nm)	Rg of the cross-section	I(0) from P(r)	Rg from P(r) (nm)	Dmax (nm)	Porod volume estimate (nm^3)	Porod exponent
TET12$_{1.10}$SN-f$_5$	51.2±0.2	3.50±0.05	2.00±0.05	49.5	3.41	10.5	107	4.0
TET12$_{1.10}$SN-c$_6$	41.6±4.0	3.60±0.50	2.00±0.10	41.5	3.56	11.5	145	4.0
TET12$_{1.10}$SN-f$_9$	37.0±7.0	3.90±0.10	2.00±0.05	37.3	3.94	12.7	153	3.6
TET12$_{2.3}$SN-f$_5$b	50.0±1.5	3.60±0.20	2.05±0.15	49.8	3.52	11.5	105	4.0
TET12$_{1.10}$S-f$_5$	53.7±0.5	3.60±0.10	1.95±0.05	41.4	3.72	13.5	210	3.2
TET12$_{1.10}$S-f$_5$b	47.2±0.2	3.38±0.03	1.80±0.20	47.1	3.38	11.2	127	3.9
TET12$_{1.10}$S-c$_6$	52.9±0.5	3.45±0.04	1.80±0.20	55.5	3.62	13.0	107	4.0
TET12$_{1.11}$S-f$_5$	45.0±1.0	3.30±0.10	1.65±0.15	44.3	3.20	10.9	105	3.9
TET12$_{1.6}$S-f$_5$b	79.0±1.0	4.08±0.10	1.80±0.10	77.7	4.05	12.9	200	3.3
TET12$_{1.6}$S-c$_6$b	53.2±0.4	3.51±0.05	1.80±0.10	53.0	3.55	11.5	114	3.8
PYR16$_{4.6}$SN-f$_5$	65.4±0.5	3.83±0.03	1.70±0.20	65.3	3.83	11.7	136	4.0
PYR16$_{2.15R}$S-f$_5$	59.4±0.5	3.87±0.05	1.80±0.10	64.8	3.91	14.4	151	3.8
TRIP18$_{7.5R}$SN-f$_5$	92.0±1.0	4.20±0.10	2.10±0.20	92.0	4.20	14.5	210	3.8

Nature Biotechnology: doi:10.1038/nbt.3994

Supplementary Note

Our approach to polyhedra comes from topological graph theory. A polyhedron is defined as a *polygonal complex*, i.e. as collection of polygons together with information how to glue pairwise their edges in order to form a map on a surface. For essentials of graphs on surfaces, see Pisanski et al. [81]. Each (rooted and oriented) polygon is given by a sequence of vertices or oriented edges.

Some of the commercial programs provide faces as lists of vertices of the skeleton graph. We call such a representation: a *vertex-based description of faces*. For us it is more convenient to describe the same polyhedron in terms of directed edges. We describe the algorithm to convert the vertex-based description in to an edge-based later in the text.

For simplicity we use upper and lower case letters to distinguish between edges that are oriented coherently with the orientation of the polygon and edges that are oriented in the opposite direction. We describe this for the case of four-sided pyramid (Fig. S1.1).

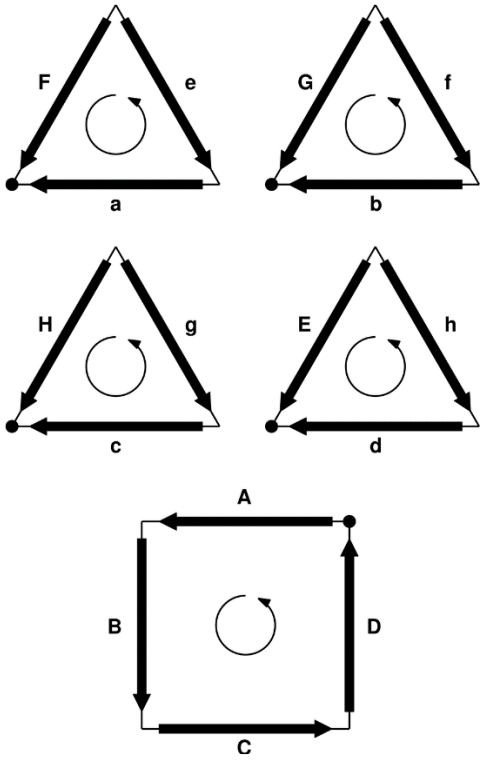


Fig. S1.1. Interpretation of the polygonal complex or scheme for the four-sided pyramid.

The polygons are oriented and rooted. This means that one vertex - the root - is selected and the string describing the edges along the polygon starts at the root and follows the boundary of the polygon along the orientation of the polygon. The upper case

letters signify that the direction of the oriented edge is compatible with the orientation of the polygon. Lower case letters signify that the directions are antiparallel (opposite). The collection of strings, called a scheme by Ringel [82] uniquely defines the skeleton, i.e. the graph of the polyhedron. Here we call such a scheme an *edge-based description of faces*. In order to facilitate the input of various polyhedra, we provided an algorithm that converts vertex-based description of faces to an edge-based description of faces.

The algorithm proceeds as follows:

1. For each face **f** described by a cyclic sequence of vertices
f = $[v_1,v_2,..., v_k]$ we identify the collection of ordered pairs of consecutive vertices:
$s_i = (v_i,v_{i+1})$, for i = 1,2,...,(k-1) and $s_k = (v_k,v_1)$. Denote $s(f) = [s_1,s_2,...,s_k]$
2. If s = (a,b) is any ordered pair above, then define t = (a,b) if a< b and t = (b,a) otherwise. Let T = $[t_1,t_2, ...]$ denote lexicographically ordered set of such pairs.
3. Assign capital letters to the elements of T as follows: to t_1 we assign "A", to t_2 we assign "B", etc. Let L denote such an assignment: $L(t_1)$ = "A", $L(t_2)$ = "B", etc. We extend it to the pairs in the reverse order: if L(a,b) = "A" then let L(b,a) = "a".
4. For each face f we apply the assignment L to s(f) and we obtain
$M(f) = [L(s_1),L(s_2), ..., L(s_k)]$, which is the edge-based description of the polyhedron.

Note that the order in which the pairs are listed is irrelevant. Below we present the results of the intermediate steps of the algorithm for the four-sided pyramid.

```
f1 = [1 2 3 0]
f2 = [0 1 4]
f3 = [0 4 3]
f4 = [3 4 2]
f5 = [1 2 4]


s(f1) = [(1 2) (2 3) (3 0) (0 1)]
s(f2) = [(0 1) (1 4) (4 0)]
s(f3) = [(0 4) (4 3) (3 0)]
s(f4) = [(3 4) (4 2) (2 3)]
s(f5) = [(1 2) (2 4) (4 1)]

T = [(0 1) (0 3) (0 4) (1 2) (1 4) (2 3) (2 4) (3 4)]

L(0 1) = "A"
L(0 3) = "B"
L(0 4) = "C"
L(1 2) = "D"
L(1 4) = "E"
L(2 3) = "F"
L(2 4) = "G"
L(3 4) = "H"

M(f1) = "DFbA"
M(f2) = "AEc"
M(f3) = "Chb"
M(f4) = "HgF"
M(f5) = "DGe"
```

By changing the root vertex in a k-sided polygon we obtain k distinct cyclic permutations. By changing the orientation of the polygon we obtain k additional reversals. The following table represents all possibilities for the top left triangle of a square pyramid.

string	reversal
aeF	**fEA**
eFa	**AfE**
Fae	**EAf**

Table S1.1. Cyclic shifts of the triangle "aeF" and their reversals.

In our previous work [24,27] we have shown how one can obtain all stable double traces by successive gluing of two faces with a common edge. In that approach only topologies with the same vertex-figure as the original polyhedron are constructed. Here we extend this method to construct topologies with the same stable skeleton but allowing self-crossings at polyhedral vertices.

Let P be a polyhedron with *n* vertices and *m* edges. A stable double trace can be described by a string of 2*m* symbols. Since this is only a template and not the actual polypeptide the choice of symbols is arbitrary. Hence we may use what we call a *standard encodin*g as follows:

> We scan the double trace from left to right.
> 1. For the first occurrence of a symbol we choose the first unused letter in the upper case.
> 2. For the second occurrence we use the same symbol that has been used in the first occurrence. We use the upper case if the cases of the original symbols matched otherwise we use lower case, noting the antiparallel gluing.

In generating all possible strands that give rise to the same skeleton we exploit the fact that each strand admits a unique standard encoding. For each strand of length 2m there exist in principle 2m cyclic shifts that we call *oriented equivalents*. By reversing the order of each traversal we obtain 4m strings that we call *all-equivalents*. In case of triangle we start with "ABCABC". The 6 oriented equivalents are »ABCABC", "BCABCA", "CABCAB", "ABCABC", "BCABCA", "CABCAB" and their reverses: "CBACBA", "ACBACB", "BACBAC", "CBACBA", "ACBACB", "BACBAC". However, only 3 oriented equivalents are distinct "ABCABC", "BCABCA", "CABCAB" and they differ from their reverses: "CBACBA", "ACBACB", "BACBAC".

But all six distinct strings give the same standard encoding: "ABCABC". In this case there exists only one *reflexive topology*. In case of tetrahedron, there exist three distinct topologies and each of them is reflexive. A topology is reflexive if a string is equivalent to its reverse. In other words, in a reflexive topology we are unable to distinguish the direction in which the string is traversed. It is well-known that tetrahedron admits one topology with two antiparallel dimers and two topologies with three antiparallel dimers.

The topology having two antiparallel dimers has 12 oriented equivalents

```
'ABCADECFEbDf',  'BCADECFEbDfA',  'CADECFEbDfAB',  'ADECFEbDfABC',
'DECFEbDfABCA',  'ECFEbDfABCAD',  'CFEbDfABCADE',  'FEbDfABCADEC',
'EbDfABCADECF',  'bDfABCADECFE',  'DfABCADECFEb',  'fABCADECFEbD'
```

and 24 all-equivalents:

```
'ABCADECFEbDf',  'fDbEFCEDACBA',  'BCADECFEbDfA',  'DbEFCEDACBAf',
'CADECFEbDfAB',  'bEFCEDACBAfD',  'ADECFEbDfABC',  'EFCEDACBAfDb',
'DECFEbDfABCA',  'FCEDACBAfDbE',  'ECFEbDfABCAD',  'CEDACBAfDbEF',
'CFEbDfABCADE',  'EDACBAfDbEFC',  'FEbDfABCADEC',  'DACBAfDbEFCE',
'EbDfABCADECF',  'ACBAfDbEFCED',  'bDfABCADECFE',  'CBAfDbEFCEDA',
'DfABCADECFEb',  'BAfDbEFCEDAC',  'fABCADECFEbD',  'AfDbEFCEDACB'.
```

However, when applied standard encoding to the 12 oriented equivalents the following sorted list is obtained:

```
'ABCADECFEbDf',  'ABCADEcFdBFE',  'ABCDAEDFCeBf',  'ABCDBEAdFeCF',
'ABCDBEFDaFcE',  'ABCDEBFEaDfC',  'ABCDECAFEbFd',  'ABCDECFBeAfD',
'ABCDEbFECAFd',  'ABCDEbFdAFEC',  'ABCDaEDBFEcF',  'ABCDaEcFEDBF'
```

Note that the reverses do not produce any new standard from. The string `'ABCADECFEbDf'` is lexicographically minimal among all oriented equivalents. It is called the *oriented canonical form.* In general, the reverse string may produce a different oriented canonical form. The minimum of the two is called a *canonical form* of the topology. The string is reflexive if its oriented canonical form is the same as the canonical form of its reverse. In the case of tetrahedron all three topologies are reflexive. Each topology is uniquely determined by its canonical form. The other two topologies of the tetrahedron have the following canonical forms: `'ABCADECFdBef'` and `'ABCADEbDFceF'`. However, one can easily distinguish between the two. While the former has 12 oriented standard forms the latter only has four of them, due to the symmetry of the strand.

Tetrahedron does not exhibit all possible situations that may occur while exploring all topologies of stable self-assembly by dimers. The square pyramid already shows two features that do not appear in tetrahedron.

In general, generation of all possible topologies for a given polyhedron P runs in two phases. In the first phase we generate all possible embeddings of the skeleton of P, i.e. its graph G(P) in different closed surfaces up to equivalence of local rotation (vertex figure) at each vertex of P.

Let v be a vertex of P of valence d, d >2. A *vertex figure* at v is determined by a cyclic permutation of the neighbors of v in P. There are (d-1)! cyclic permutations. Since we are not interested in the orientation, only half of this number counts. There are (d-1)!/2 distinct vertex-figures at v. The total number of non-equivalent embeddings is given by the formula:

$$NE(P) = (d_1\text{-}1)!/2\ (d_2\text{-}1)!/2\ \ldots\ (d_n\text{-}1)!/2$$

where d_i is the valence of the i-th vertex v_i. For a vertex of valence 3 the contribution to the product is 1. Hence for a tetrahedron NE = 1. A vertex of valence 4 contributes a factor of 3 to the product. For a square pyramid NE = 3.
For each polyhedron we choose one embedding as the *basic embedding*. For a convex polyhedron the basic embedding is the planar non-crossing embedding. For other embeddings we may count the *number of crossings*, i.e. the number of vertices of P where the vertex-figure differs from the basic one.

For each non-equivalent embedding we generate all possible topologies by gluing faces of the starting embedding in all possible ways as described in our previous work, until a single face is obtained. This is performed in such a way that the original vertex-figures are never changed. Among the canonical forms of the vertex-figures equivalent topologies we may select the one that is lexicographically minimal and call it super-canonical form. For each of the non-equivalent embeddings (NE) we compute the super-canonical form. If two such embeddings produce the same super-canonical forms they are, in fact, isomorphic, due to some symmetry of the graph of the polyhedron. And this happens in the case of square pyramid. Instead of NE = 3 distinct cases we only obtain 2.

For a square pyramid we have 19 topologies with no crossings. In this case there are 10 reflexive topologies and 9 irreflexive pairs of topologies.

10 reflexive topologies:
ABCDEAFGCHefBgdH,
ABCDEFGaceHfbghd,
ABCADEFCGdHFbHeg,
ABCADEFcGFHEgbDh,
ABCDEAFGeHCfBhdG,
ABCDEAFcGEHfBGdh,
ABCADEFCGFHdBHEg,
ABCDEFbGeHfgachd,
ABCADEFGdBegHcFH,
ABCADEbDFGeFHcgH,

9 irreflexive pairs of topologies:
ABCADEFBGFDHcGeH, ABCADEFGCFHdBgHE
ABCADEFGbDHfcghE, ABCADEFbGeHcfgDH
ABCADEFGcFHEbDhG, ABCADEFbGFCHEgDh
ABCADEFBGeHcGFDH, ABCADEFGbDHGCFhE
ABCADEFCGeHbDHfg, ABCADEFGcFHdBehG
ABCDEAFGeHbFchdG, ABCDEAFcGdHegbFH
ABCADEFcGeHfgbDH, ABCADEcFGDHfbghE
ABCADEFGEHCfHbDg, ABCADEFGEbDgHcFH
ABCADEFCGeHGFbDH, ABCADEcFGDHGBFhE

In addition there are 33 topologies with a crossing.

10 reflexive topologies:

ABCADECFGEHGdBHf,
ABCADECFGbDHGeHf,
ABCDAEFDGCfHGbEH,
ABCDAEFCGEHdGBfH,
ABCDAEFbGDHeGcfh,
ABCADEFbDGfHcgEH,
ABCDEFacGeHfgbhd,
ABCADEFCGdBfHGEH,
ABCADECFGEbHGdHf,
ABCADEFbDGHeGcfh,

23 irreflexive pairs of topologies:

ABCADECFGBeHGDHf, ABCADECFGdBHGEHf
ABCADEFGECHfbDgh, ABCADEFGEHbDgchF
ABCADEFGdBFHcegH, ABCADEFbDGCHeGfh
ABCADEFGceHfbDHG, ABCADEFcGEHbDghF
ABCADECFGdHGEbHf, ABCADEFcGEHGdBhF
ABCDAEFCGBfHdGEH, ABCDAEFCGeBHdGFH
ABCDAEFcGdfHgbEH, ABCDAEFdGEHbgchF
ABCDAEFGbEDHfcgh, ABCDAEFGbEHcgdhF
ABCADEFGdBHegcHF, ABCADEbFGECHfDgh
ABCDEFGadHfbghce, ABCDEFGadfHbgche
ABCADEFcGdHegbHF, ABCADEcFGEHfbDgh
ABCDAEFGdHFbEhcG, ABCDAEFcGHFDgbEh
ABCDAEFGCHeBgdHF, ABCDAEFbGDHFCgEh
ABCADEFGdBfHcGEH, ABCADEFbDGCHFgEh
ABCADEFcGEHDgbhF, ABCADEcFGeHfbDHG
ABCDAEFcGHdfgbEH, ABCDAEFdGHbEgchF
ABCADEFGdBHGEcHf, ABCADEbFGECHGdFh
ABCDAEFGdHeBfhcG, ABCDAEcFGDHfbEgh
ABCADEFGCeHGbDHf, ABCADEFcGEHBGdhF
ABCADECFGDHGBeHf, ABCADEFCGEHGdBfH
ABCADEFDGCHeGbfh, ABCADEFDGbfHcgEH
ABCADEFBGEHcGdfH, ABCADEFGbDHGCeHf
ABCADEFBGdfHcGEH, ABCADEFbDGHFCgEh

By computations of super-canonical forms the actual number of vertex-figure non-equivalent embedding has been reduced to NA = 6.

	square pyramid			
# antiparallel dimers	0 crossings	1 crossing	# topologies	# directed topologies
2	4	7	11	17
3	2	9	11	20
4	8	8	16	24
5	3	7	10	18
8	2	2	4	5
Total	19	33	52	84

Table S1.2. Topologies and directed topologies of a four sided pyramid according to the number of antiparallel dimers.

For a square pyramid there exist 52 topologies and 84 directed topologies. In *directed topologies* we distinguish the irreflexive pairs and thus count each irreflexive topology twice. In general, the following is true:

directed topologies - # topologies = #irreflexive pairs

Nature Biotechnology: doi:10.1038/nbt.3994

There are no totally parallel realizations of square pyramid. At least two antiparallel pairs must be used. There are 11 topologies and 17 directed topologies with two antiparallel dimers. On the other end, there exist 4 totally anti-parallel topologies. One of them is irreflexive, which brings up the total number of directed topologies to 5 (Fig S1.2).

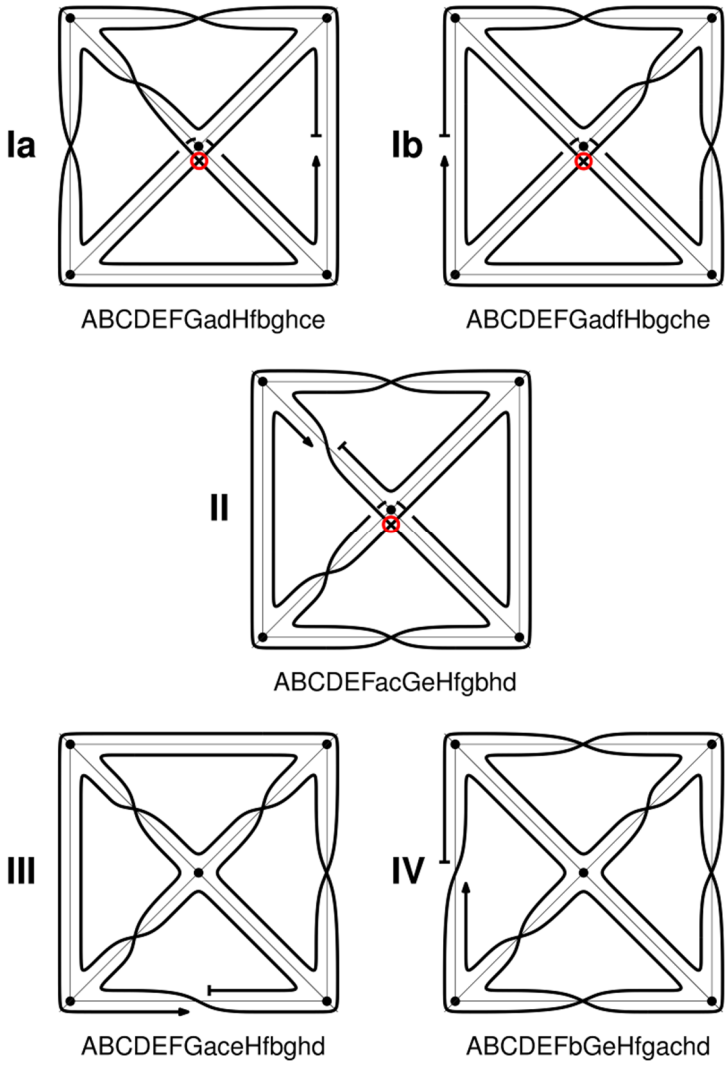


Fig. S1.2. Four topologies and five directed topologies of the four-sided pyramid admitting all antiparallel dimers. Each one is presented by its canonical description. Note that I(a) and I(b) have distinct *directed* canonical descriptions, however only I(a) represents the canonical description of the corresponding topology. The crossing at the vertex of degree four is marked with a red circle.

It is important to note that the edge directions and labels change when we change the topology or directed topology of a polyhedron. For instance, if we keep the topology but change the direction of the double trace, the labels may change. In particular, in

Nature Biotechnology: doi:10.1038/nbt.3994

passing from I(a) to I(b) one has to make the following changes: A -> A, B -> g, C -> f, D -> e, E -> d, F -> c, G -> b.

polyhedron	NE	NA	# topologies	# reflexive	# irreflexive pairs	# directed topologies
tetrahedron	1	1	3	3	0	3
square pyramid	3	2	52	20	32	84
trigonal prism	1	1	25	10	15	40
triangular bipyramid	27	6	470	60	410	880
cube	1	1	40	12	28	68
octahedron	729	38	21479	516	20963	42442

Table S1.3. There is no closed form formula known that would give the number of topologies for an arbitrary polyhedron. Our algorithm determines all NE embeddings up to vertex-figure. This may be exponential in the number of vertices for many classes of polyhedra. For instance for a quartic polyhedron on n vertices (polyhedron with vertices of valence 4), such as the octahedron, the number NE is given by 3n. The actual number NA is much smaller and depends on the symmetry of polyhedron.

Polyhedron	No Topologies	Reflexive	Irreflexive pairs
tetrahedron	3	3	0
square pyramid	19	10	9
	33	10	23
triangular bipyramid	92	8	84
	114	8	106
	106	16	90
	45	8	37
	92	8	84
	21	12	9

Table S1.4. Number of topologies according to different vertex-figures.

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 (DARPins): 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).
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-1β and neuronal toxicity. *Cell. Mol. Life Sci.* **69,** 4215–28 (2012).
77. Cromwell, P. R. *Polyhedra*. (Cambridge University Press, 1997).
78. Smilgies, D. M. & Folta-Stogniew, E. Molecular weight-gyration radius relation of globular proteins: A comparison of light scattering, small-angle X-ray scattering and structure-based data. *Journal of Applied Crystallography* **48,** 1604–1606 (2015).
79. Lacroix, E., Viguera, A. R. & Serrano, L. Elucidating the folding problem of α-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters1. *J. Mol. Biol.* **284,** 173–191 (1998).
80. Ortega, A., Amorós, D. & García De La Torre, J. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys. J.* **101,** 892–898 (2011).
81. Pisanski, T. & Potočnik, P. in *Handbook of Graph Theory* (eds. Gross, J. L., Yellen, J. & Zhang, P.) 730–44 (CRC press, 2013).
82. Ringel, G. *Map Color Theorem*. (Springer Berlin Heidelberg, 1974). doi:10.1007/978-3-642-65759-7