CHEMBIOCHEM

Supporting Information

Symmetry-Directed Self-Assembly of a Tetrahedral Protein Cage Mediated by de Novo-Designed Coiled Coils

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

cbic_201700406_sm_miscellaneous_information.pdf

Materials and Methods

Construction of genes, protein expression and purification: A codon-optimized gene encoding the trimeric esterase and ligated into the expression vector pET28b was synthesized commercially and the NEBuilder® HiFi DNA Assembly kit was used to insert the various sequences encoding the Gly linker and trimeric coiled coil at 3'end of the esterase gene. The sequences of the fusion proteins are shown in Figure S2. Protein expression and purification were performed essentially as described previously.[1] Unless otherwise noted, all experiments were performed with protein samples that had been dialyzed against 25 mM HEPES, containing 100 mM NaCl and 1mM EDTA, pH 7.5 (dialysis buffer).

Enzyme Activity: The catalytic activity of the parent esterase and the protein cages assembled form it was measured using *p*-nitrophenyl acetate (*p*NPA) at 45°C, with the absorbance of p-nitrophenol monitored at 405 nm, as described previously (Figure S6). [1]

Thermal stability of protein cages: Thermal unfolding curves of the trimeric esterase and the assembled tetrahedral cage were determined by monitoring the change in ellipticity at 222 nm. Measurements were made in dialysis buffer using a J815 CD spectrometer, Jasco, United Kingdom, between 25 °C and 100 °C with a heating rate of 2 °C /min (Figure S7A). The protein concentration for all samples was 0.1 mg/mL. The thermal stability of the enzyme activity was also analyzed (Figure S7B). Samples of the trimeric esterase and the assembled tetrahedral cage were incubated at various temperatures for 10 min followed by cooling on ice for 5 min. The residual enzyme activity was then measured as described above.[2]

Sedimentation Velocity Analytical Ultracentrifugation: Sedimentation velocity (SV) experiments were conducted using a Beckman Proteome Lab XL-I analytical ultracentrifuge (Beckman Coulter) as described previously.[1] Briefly, samples of Tet-8-H5 ~ 0.5 mg/mL in dialysis buffer were loaded into pre-cooled a cell containing standard sector shaped two-channel Epon centerpieces with 1.2 cm path-length (Beckman Coulter). Cell and rotor were temperature-equilibrated at 6 ºC in the centrifuge for at least 2 h prior to sedimentation. The sample was spun at 39,000 rpm and the sedimentation of the construct was monitored continuously by absorbance

at 280 nm. Data analysis was conducted via 2-dimensional sedimentation spectrum analysis (2DSA) using the finite element modeling module provided with the Ultrascan III software (http:/www.ultrascan.uthscsa.edu).[3] The partial specific volume of the construct was calculated based on the sequence of the Tet-8-H5 monomer. Confidence levels for statistics were derived from 2-DSA data refinement using a genetic algorithm followed by 100 Monte Carlo simulations[4] reaching a final RMSD of 0.0036. Calculations to analyze analytical ultracentrifugation data were performed on the UltraScan LIMS cluster at the Bioinformatics Core Facility at the University of Texas Health Science Center at San Antonio.

Native mass spectrometry: Native MS were performed on a Synapt G2 ion mobility-mass spectrometry platform (Waters Corp. Manchester, UK). Samples were buffer exchanged into 200 mM ammonium acetate, pH 7.5, using a Micro Biospin column (Bio-Rad, Inc. Hercules, California). Buffer-exchanged samples were diluted to ~10 μM and electrosprayed from gold-coated borosilicate capillaries prepared in-house. The instrument parameters were similar to those described previously,[5] with a capillary potential of 1.5 kV, and sampling cone potential of 30 V. The trap collision energy was set to 20 V for efficient transmission of ions with no observable unfolding or dissociation. Spectra were acquired from m/z = 1000 to 13000 for 1 minute. The resulting datasets were processed using Masslynx (Waters Corp. Manchester, UK) and charge-state assignments and masses were calculated using ESIprot.[6]

Protein Crosslinking: The protein cages were crosslinked using bis(sulfosuccinimidyl)suberate (BS3). BS3 concentrations were varied between 1.6 mM and 2.4 mM whereas the protein concentrations were varied between 0.25 and 0.5 mg/mL. The reaction was performed in dialysis buffer pH 7.5 for 30 min at room temperature and quenched immediately by adding Tris-Cl buffer pH 7.5 to a final molar concentration 20-fold higher than the BS3 concentration. The extent of cross-linking was analyzed by SDS-PAGE (20% gel) and native PAGE (3-8% gels) as shown in Figure S5.

EM sample preparation and image acquisition: Protein samples for negative stain EM were diluted to ~0.02 mg/mL and fixed on a grid using conventional negative staining procedures.[7] Imaging was performed at room temperature with a Morgagni 268(D) transmission electron

microscope (FEI Co.) equipped with a tungsten filament operated at an acceleration voltage of 100 kV and a mounted Orius SC200W CCD camera.[8]

Protein samples for cryoEM were concentrated to ~0.1 mg/mL and loaded on a glow-discharged Quantifoil grid (R2/2 200 mesh) and vitrified using a Vitrobot (FEI Mark IV). The sample was imaged on an Arctica transmission electron microscope (FEI Co.) equipped with a field emission electron gun operated at 200 kV. Images were recorded at a magnification of 43,817× on a Gatan K2 Summit camera, resulting in a pixel size of 1.14 Å on the specimen level. All of the images were acquired using a low-dose procedure to minimize radiation damage to the samples, with a defocus value of 2–4 μm.

Cryo-EM image processing and 3D reconstructions: A total of 307 Cryo-EM images were recorded, and dose fractionated image stacks were binned 2 × 2 and subject to whole-frame drift correction using Unblur.[8] A sum of the total 6 sec frames in each image stack was integrated with dose filter applied using Unblur and the integrated images were used for CTF parameters determination by CTFFIND3.[9]

Particles were selected automatically using RELION,[10] then subjected to reference-free alignment and classification using ISAC.[11] Fully assembled and well-defined class average images were selected to generate the initial mode using program e2initialmodel.py.[12] Then, 35,217 particles were extracted from those selected classes for 3D reconstruction using RELION. The initial model was filtered to 60-Å resolution, and then subjected to 3D auto-refinement. Tetrahedral (T) symmetry was enforced during reconstruction, and the final map of the protein cage was produced with an indicated resolution of 13 Å at the 0.143 level of Fourier shell correlation (Figure S6). The crystal structure of the esterase (PDB 1ZOI) was first manually docked in the map with the C terminus in close proximity to the threefold axis. Subsequently, the crystal structure of the de novo protein (4DZL) was manually placed into the protruding density attribute to the coiled-coil with its N terminus facing the C terminus of esterase. The fitting was then refined using the “fit in map” routine in CHIMERA.[13] Map visualization, rendering, and figure generation were performed using CHIMERA.

References:

[1] A. Sciore, M. Su, P. Koldewey, J. D. Eschweiler, K. A. Diffley, B. M. Linhares, B. T. Ruotolo, J. C. Bardwell, G. Skiniotis, E. N. Marsh, *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 8681-8686.

[2] S. Y. Shaw, Y. J. Chen, J. J. Ou, L. Ho, *J Mol Catal B-Enzym* **2006**, *38*, 163-170.

[3] B. Demeler, *Modern Analytical Ultracentrifugation: Techniques and Methods. D. J. Scott, S.E. Harding and A.J. Rowe. Eds. Royal Society of Chemistry (UK)* **2005**, *210-229*.

[4] E. Brookes, B. Demeler, *Prog. Colloid Polym. Sci.* **2006**, *131*, 78-82.

[5] H. Hernandez, C. V. Robinson, *Nat. Protoc.* **2007**, *2*, 715-726.

[6] R. Winkler, *Rapid Commun. Mass Spectrom.* **2010**, *24*, 285-294.

[7] A. Peisley, G. Skiniotis, *Methods Mol. Biol.* **2015**, *1335*, 29-38.

[8] T. Grant, N. Grigorieff, *eLife* **2015**, *4*, e06980.

[9] J. A. Mindell, N. Grigorieff, *J. Struct. Biol.* **2003**, *142*, 334-347.

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

[11] Z. Yang, J. Fang, J. Chittuluru, F. J. Asturias, P. A. Penczek, *Structure* **2012**, *20*, 237-247.

[12] G. Tang, L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees, S. J. Ludtke, *J. Struct. Biol.* **2007**, *157*, 38-46.

[13] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605-1612.

[14] R. Das, D. Baker, in *Ann. Rev. Biochem. Vol. 77*, **2008**, pp. 363-382.

Table S1: Hydrodynamic parameters for protein assemblies formed by Tet8-5H as determined by sedimentation velocity analytical ultracentrifugation[a].

	Solute 1
Molecular weight *(kDa)*	429.5 ± 6.6
Sedimentation coefficient *(S)*	14[b]
Diffusion coefficient *(cm^2/S)*	2.93 (± 0.05)e-07
Frictional ratio *(f/f_0)*	1.47 ± 0.02
Partial specific volume*(Constant)*	0.73
Partial concentration (OD)	0.158
Relative percentage	100 %

[a]Data fitting was conducted with the program Ultrascan (see methods above).

[b]The sedimentation velocity data revealed some minor concentration dependent non-ideality at the concentration the experiment was conducted, due to the size of the oligomer. The non-ideality was expressed in slight boundary sharpening and hence an artificially precise fit of the sedimentation coefficient. Therefore, no standard deviation could be obtained for this parameter.

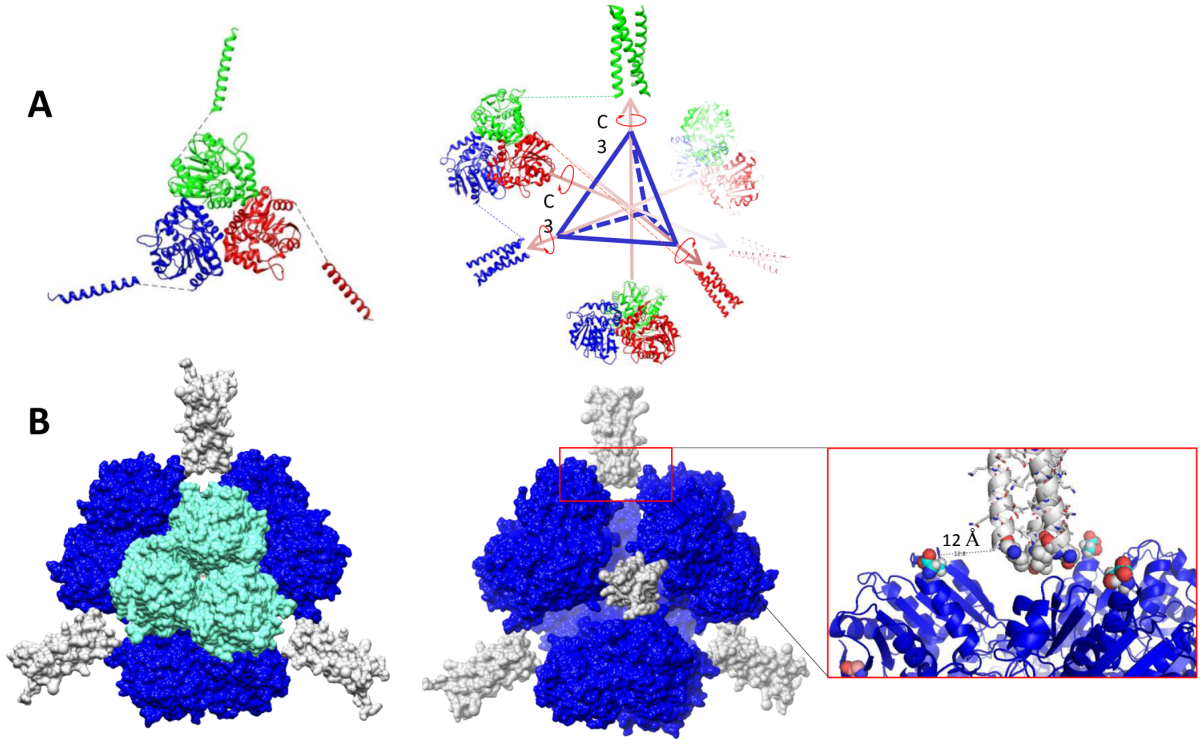

Figure S1: Estimation of the minimal length of the flexible linker sequence needed to connect the esterase and coiled-coil domains. **A)** The C_3 axes of the esterase and coiled-coil were computationally aligned along the C_3 axes defining the tetrahedral point group. Using an algorithm implemented in Rosetta,[14] the distance from the origin of each protein and its angle of rotation about its symmetry axis were allowed to vary whilst maintaining tetrahedral symmetry. Configurations with *inter*-subunit backbone atom distances shorter than 4 Å, representing steric clashes, were discarded. **B)** Views along the 2 C_3 axes of the distance-minimized structure. The between the N-terminus of the coiled-coil and C-terminus of the esterase was estimated to be ~ 12 Å (see inset)

MGSSHHHHHHSSGLVPRGSHMSYVTTKDGVQIFYKDWGPRDAPVIHFHHGWPLSADD
WDAQLLFFLAHGYRVVAHDRRGHGRSSQVWDGHDMDHYADDVAAVVAHLGIQGAV
HVGHSTGGGEVVRYMARHPEDKVAKAVLIAAVPPLMVQTPGNPGGLPKSVFDGFQAQ
VASNRAQFYRDVPAGPFYGYNRPGVEASEGIIGNWWRQGMIGSAKAHYDGIVAFSQTD
FTEDLKGIQQPVLVMHGDDDQIVPYENSGVLSAKLLPNGALKTYKGYPHGMPTTHADV
INADLLAFIRS$(G)_n$(**I**AA**I**KQE)$_x$G

Figure S2. Sequences of the fusion proteins generated in this study. The protein sequence of the trimeric esterase in shown in black font, with the leader sequence encompassing an N-terminal His-tag shown in orange. The position of the glycine linker is indicated in red (n= 4, 6, 8, 10) and the trimeric coiled-coil heptad sequence is shown in blue (x = 4, 5).

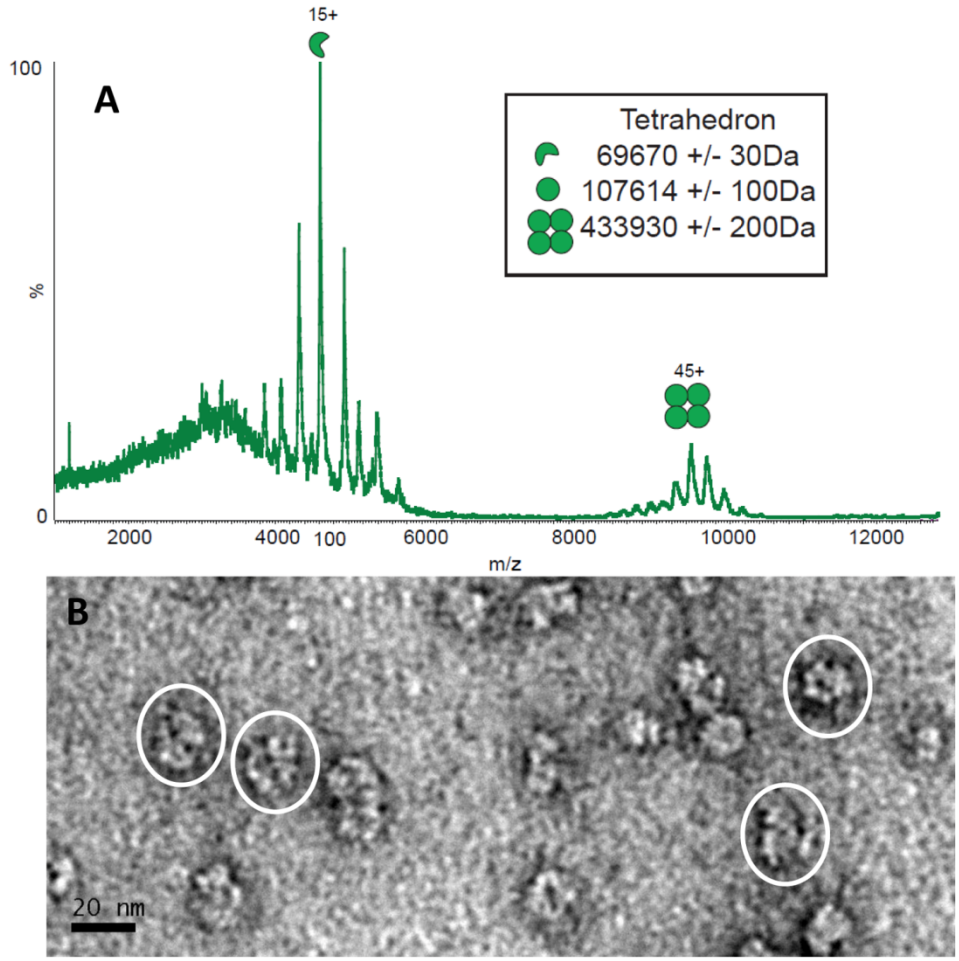

Figure S3: A) Native mass spectrometry of SEC purified Tet8 in ammonium acetate buffer, pH 7.5. The spectrum shows the formation of a species with M_r ~ 434,000 Da consistent with the formation of tetrahedral structure. Other peaks with M_r ~ 107,000 and 70,000 Da are consistent with the presence of unassembled trimer and dimeric forms of the esterase. Dissociation of these species may have occurred during electrospray ionization because no species with molecular weights in this range were detected by native PAGE or SEC (Figure 2 main text). **B)** Negative stain EM images of SEC-purified Tet8. Particles with morphology that could represent correctly formed cages are circled.

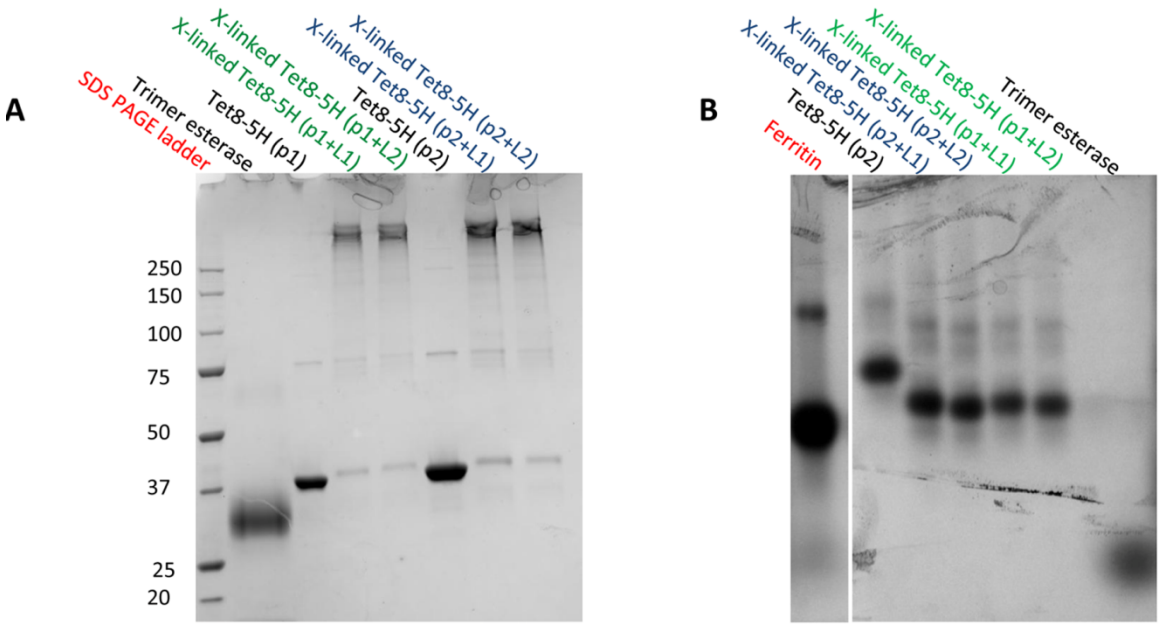

Figure S4. A) SDS-PAGE of cross-linked and uncross-linked Tet8-5H treated at different protein and BS3 cross-linker concentrations. p1: protein concentration = 0.25 mg/mL; p2: protein concentration = 0.5 mg/mL; L1: BS3 concentration = 1.6 mM; L2: BS3 concentration = 2.4 mM. SDS-PAGE analysis of cross-linked protein indicated that the majority of protein is cross-linked to form complexes of M_r > 250 kDa. **B)** Native PAGE of cross-linked and uncross-linked Tet8-5H treated at different protein and cross-linker concentrations. The analysis confirms that crosslinking occurs primarily between subunits of individual cages. The greater electrophoretic mobility of crossed-linked Tet8-5H is presumably due to the neutralization of positively charged surface lysine residues upon cross-linking.

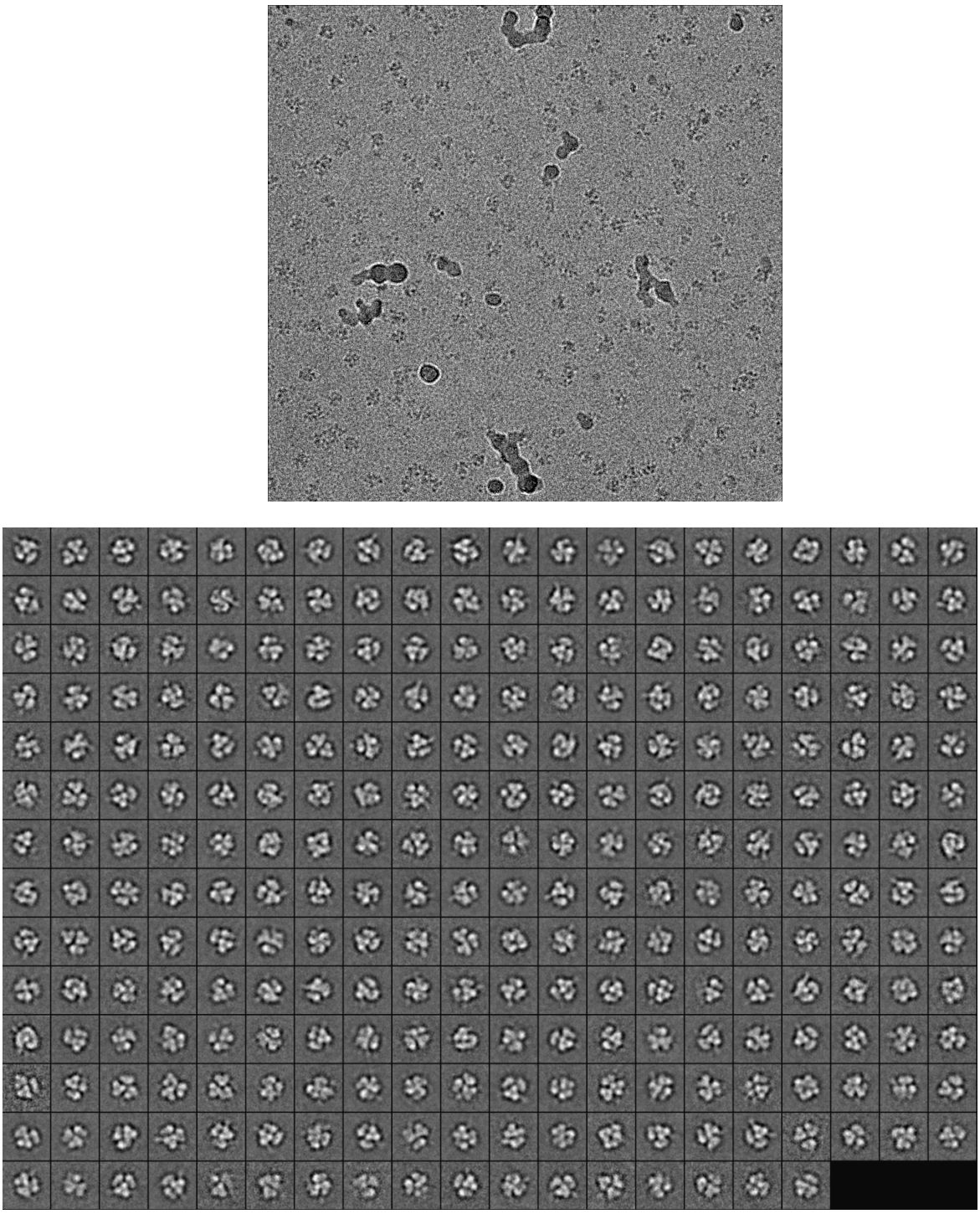

Figure S5. Top: representative cryo-EM field for Tet-8-5H particles; Bottom Reference-free 2D class averages for Tet-8-5H determined from cryo-EM. A total of 35,217 particle images were classified into 277 classes. For details, see the materials and methods section.

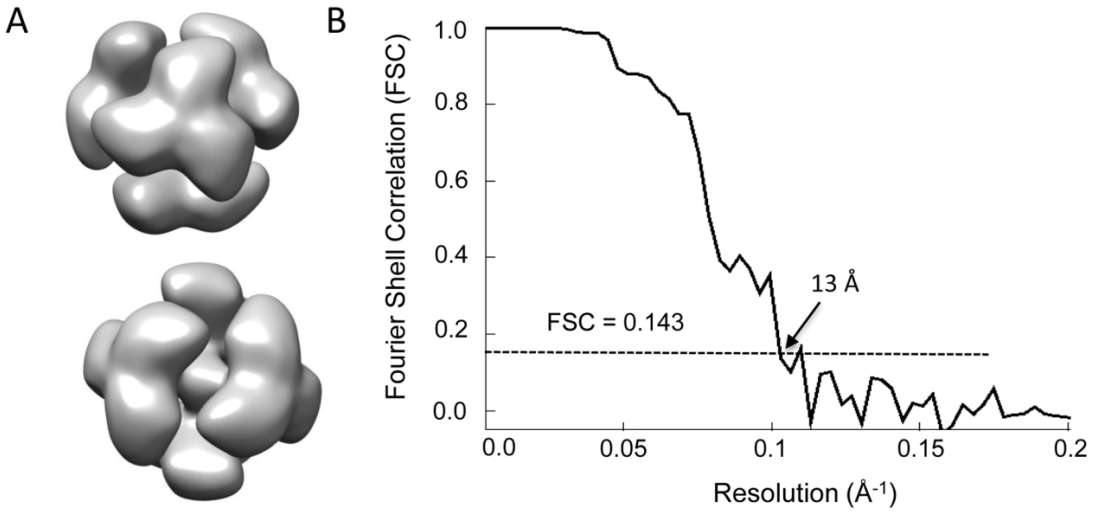

Figure S6. **A)** Initial electron density model used for 3D refinement. Model is shown viewed along threefold and twofold symmetry axes. **B)** Gold standard Fourier Shell Correlation (FSC) curves for the final 3D maps of Tet8-5H indicate resolution of 13 Å at Fourier Shell Correction of 0.143.

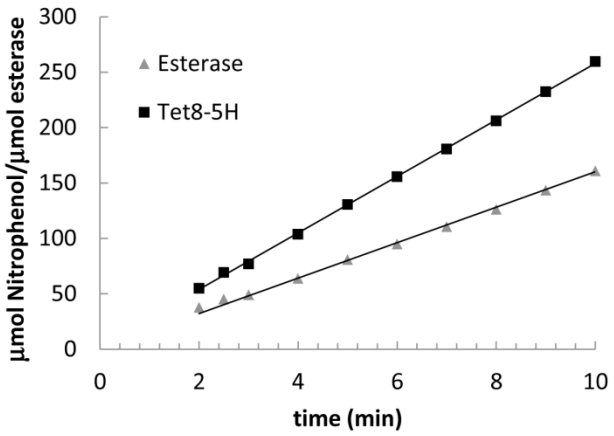

Figure S7. Comparison of the specific activities of the parent esterase and Tet5-8H with p-nitrophenol acetate as substrate (pNPA). Under the conditions of the assay the specific activities of the parent esterase and Tet8-5H were 16.0 ± 0.1 and 25.5 ± 0.8 min^{-1} respectively.

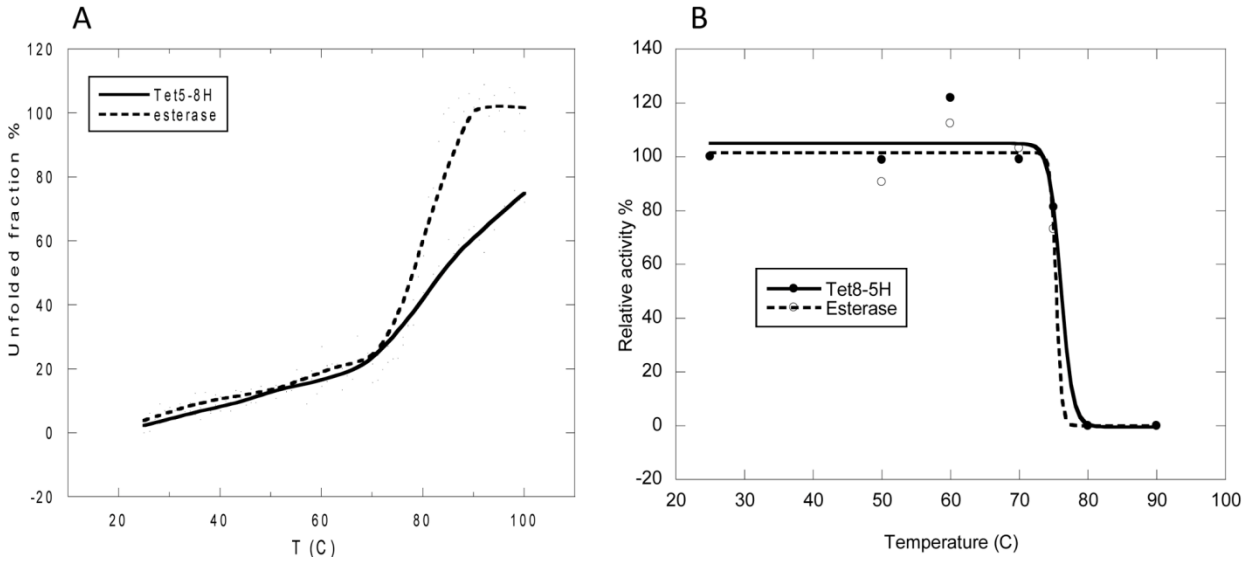

Figure S8 A) Thermal denaturation curves for the trimeric esterase and Tet8-5H protein cage followed by CD spectroscopy at 222 nm. **B)** Thermal stability curves for the trimeric esterase and Tet5-8H protein cage; for details see the Materials and Methods section.