Supplementary Information

Fusion then fission: splitting and reassembly of an artificial fusion-protein nanocage

Naoya Ohara, [a] Norifumi Kawakami*, [a] Ryoichi Arai, [b,c] Naruhiko Adachi, [d,e] Akihito Ikeda, [d] Toshiya Senda, [d] Kenji Miyamoto* [a]

[a.] Department of Bioscience and informatics, Faculty of Science and Technology, Keio University, Yokohama, Kanagawa 223-8522, Japan

[b.] Department of Biomolecular Innovation, Institute for Biomedical Sciences, Interdisciplinary Cluster for Cutting Edge Research, Shinshu University, Ueda, Nagano 386-8567, Japan

[c.] Department of Applied Biology, Faculty of Textile Science and Technology, Shinshu University, Ueda, Nagano 386-8567, Japan

[d.] Structural Biology Research Center, Institute of Materials Structure Science, High Energy Accelerator Research Organization (KEK), Oho, Tsukuba 305-0801, Japan

[e.] Life Science Center for Survival Dynamics, Tsukuba Advanced Research Alliance (TARA), University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577, Japan

Materials and Methods

Materials. Tris(hydroxymethyl)aminomethane (Tris), 4-(2-hydroxyethyl)piperazine-1-ethane-sulfonic acid (HEPES), ethylenediaminetetraacetate (EDTA), imidazole, calcium chloride, strontium chloride hexahydrate, barium chloride dihydrate, sodium chloride, potassium chloride, zinc chloride, copper(II) chloride dihydrate, nickel(II) chloride hexahydrate, cobalt(II) chloride, LB medium, chloramphenicol, and isopropyl β-D-1-thiogalactopyranoside (IPTG) were purchased from Nacalai Tesque. 2-(N-morpholino)ethanesulfonic acid (MES), guanidinium hydrochloride (GdmCl), Iron(II) chloride tetrahydrate, iron(III) chloride hexahydrate, manganese(II) chloride tetrahydrate, and

Proteinase K were purchased from FUJIFILM Wako. Hrv 3C Protease was purchased from Takara Bio. Bovine serum albumin (BSA) and ammonium molybdate was purchased from Sigma-Aldrich. PCR primers were purchased from Integrated DNA Technology.

Plasmid construction. The previously constructed TIP60 gene[1] was used as a template for site-directed mutagenesis using a PrimeSTAR Mutagenesis Basal Kit (TaKaRa) with appropriate primers (see Table S1 for the nucleotide sequences). The bicistronic co-expression vector was constructed by inserting a stop codon, ribosome binding site, and initiation codon into the split site. GFP-fused CoreC(K26E) was constructed by inserting the gene of GFPuv5 to the appropriate site of CoreC(K26E) using In-fusion HD Cloning kit (Takara Bio) (see Table S1 for the nucleotide sequences). See Table S2 for the amino acid sequences.

Protein expression. Plasmids encoding TIP60 or its mutants were introduced into *E. coli* BL21(DE3) cells for protein expression. The cells were pre-cultured in LB medium containing 100 μg mL^{-1} chloramphenicol for 16 h at 37°C with shaking at 200 rpm. Then, the pre-culture was transferred to a 100-fold volume of fresh LB medium containing 100 μg mL^{-1} chloramphenicol. After shaking at 180 rpm and 37°C for 2 h, protein expression was induced by the addition of 0.1 mM IPTG. After an additional 6 h, the cells were harvested by centrifugation at 6000×*g* for 10 min and stored at −30°C until use.

Protein purification.

CoreN

CoreN fragment was almost found in insoluble form when it was expressed alone. Thus, it was isolated from the inclusion body by purifying under denatured condition and re-folded *in vitro*, as follows. The cells were lysed by ultrasonication in 20 mM Tris-HCl buffer (pH 8.0) then centrifuged at 10,000×*g* for 10 min. The resulting pellet containing the aggregated CoreN was resuspended in 20 mM Tris-HCl buffer (pH 8.0) containing 6 M GdmCl. The suspension was then cleared by centrifugation at 10,000×*g* for 10 min followed by Ni-NTA column chromatography (Qiagen). The Ni-NTA column was equilibrated with 20 mM Tris-HCl buffer (pH 8.0) containing 6 M GdmCl prior to loading then washed with 20 mM Tris-HCl buffer (pH 8.0) containing 20 mM imidazole and 6 M GdmCl. The bound protein was eluted with 20 mM Tris-HCl buffer (pH 8.0) containing 400 mM imidazole and 6 M GdmCl. The isolated CoreN was dialyzed against 20 mM MES buffer (pH 6.0) containing 1 mM EDTA (2 h×3) to induce refolding. After dialysis, the protein appeared to have undergone slight aggregation and the solution was thus cleared by brief centrifugation and subsequent filtration through a membrane filter (0.20 μm, Merck Millipore). The resulting protein was again dialyzed against 20 mM HEPES buffer (pH 8.0). The protein concentration was determined by gel densitometry and adjusted to a monomer

concentration of 1500 μM. Because the protein was susceptible to gradual aggregation over the course of several weeks when stored at 4°C, aliquots were frozen in liquid nitrogen and stored at −80°C.

Others

The cells were lysed by ultrasonication in 20 mM Tris-HCl buffer (pH 8.0). The lysate was cleared by centrifugation at 10,000×*g* for 10 min. The supernatant was loaded onto the Ni-NTA column after equilibration with 20 mM Tris-HCl buffer (pH 8.0). The column was washed with 20 mM Tris-HCl buffer (pH 8.0) containing 20 mM imidazole, and the bound protein was eluted with 20 mM Tris-HCl buffer (pH 8.0) containing 500 mM imidazole. The isolated protein was dialyzed against 20 mM Tris-HCl buffer (pH 8.0) containing 1 mM EDTA (2 h×3). In the case of CoreC(K26E), the protein was again dialyzed against 20 mM HEPES buffer (pH 8.0) to remove EDTA prior to the metal-induced assembly experiments.

Determination of protein concentrations. In general, protein concentrations were determined by UV/vis spectroscopy using the theoretical molar extinction coefficient of TIP60 (8480 M^{-1} cm^{-1} at 280 nm). The concentration of the *in vivo* co-expressed TIP120 was calculated on the assumption that the CoreN and CoreC fragments assembled in a 1:1 molar ratio.

In the case of the independently purified split fragments, the fragment concentrations were determined by band densitometry on a tricine SDS-PAGE gel because CoreN contained no aromatic amino acid residues. BSA solutions of various concentrations were used as a reference.

It should be noted that the molar concentrations of TIP60 and its mutants were calculated as the concentration of the 60-mer protein unless otherwise stated (the monomer concentration is 60 times higher).

Tricine SDS-PAGE analysis. Owing to the relatively low molecular weight of CoreN, tricine SDS-PAGE was adopted on account of its good separation of small peptides[2]. The samples were denatured by the addition of 2% SDS followed by heat treatment at 95°C for 10 min. The sample components were then separated on a 10%/16% polyacrylamide gel under a constant voltage of 150 V. After electrophoresis, the protein bands were visualized using CBB Stain One (Nacalai Tesque).

Native PAGE analysis. Samples were diluted with (4×) sample buffer (Invitrogen) and separated on a native PAGE 4–16% Bis-Tris gel (Invitrogen) at a constant voltage of 150 V. The protein bands were visualized using CBB Stain One, and the relative band intensities were quantified using the ImageLab 6.0 software (BioRad).

TEM observations. Samples were diluted with the appropriate buffer to a protein concentration of

0.05 μM. A carbon-coated copper grid (NEM) was hydrophilized using a plasma ion bombarder (PIB-10, Vacuum Device Inc.). The samples were then mounted on the grid and stained with 2% (w/v) ammonium molybdate. TEM images were acquired on a Tecnai Spirit system (FEI) operating at 120 kV.

Thermal stability assay of TIP120. The original TIP60 and TIP120 (from *in vivo* co-expression) were diluted to 0.5 μM with 20 mM Tris-HCl buffer (pH 8.0) containing 1 mM EDTA. The samples were incubated at various temperatures (40, 50, 55, 60, 70, or 80°C) for 5 min then immediately placed on ice prior to native PAGE analysis.

Proteolytic degradation assay of TIP120. The original TIP60 and TIP120 (from *in vivo* co-expression) were diluted for a final concentration of 1.0 μM with 20 mM Tris-HCl buffer (pH 8.0) containing 1 mM EDTA. The samples were pre-incubated at 30°C for 5 min. The reaction was initiated by adding 20 μg/mL Proteinase K (dissolved in 20 mM Tris-HCl buffer (pH 8.0) containing 1 mM EDTA) at each time. After the reaction, the samples were placed on ice and immediately separated by Native PAGE.

***In vitro* assembly of TIP120.** The separately purified CoreN and CoreC were mixed in a molar ratio of 2:1 (monomer concentrations of 300 and 150 μM, respectively) then incubated at 30°C for 16 h.

Metal-induced assembly of CoreN/CoreC(K26E). The separately purified CoreN and CoreC(K26E) were mixed in a molar ratio of 2:1 (monomer concentrations of 300 and 150 μM, respectively) in the presence or absence of various alkaline-earth metal ions (4 mM $MgCl_2$, $CaCl_2$, $SrCl_2$, or $BaCl_2$). The samples were incubated at 30°C for at least 4 h. The effects of other metals such as monovalent cations and transition-metal ions were also examined by the same procedure.

Trp fluorescence analysis of the metal-induced assembly. The mixture of CoreN and CoreC(K26E) (monomeric concentrations of 300 and 150 μM, respectively) were incubated with 0, 1, 1.5, 2, 3, and 4 mM of the alkaline earth metal ions: $MgCl_2$, $CaCl_2$, $SrCl_2$, and $BaCl_2$. After the incubation at 30°C for at least 4 h, they were diluted 50 times with 20 mM HEPES buffer (pH8.0). Then, Trp fluorescence spectrum was acquired by Duetta (HORIBA). The excitation wavelength was 295 nm.

Analytical SEC. Analytical SEC was performed using a LC-2030 Plus HPLC system (Shimadzu) with a COSMOSIL Diol 1000-IIcolumn (Nacalai Tesque). mTIP120-Ba assembled with 4 mM $BaCl_2$ was injected to the system (100 μL). The flow rate for analysis was 1.0 mL/min. The column oven temperature was 30°C. The eluted samples were detected by UV/vis absorbance at 280 nm and

collected just downstream of the UV cell. They were analyzed by Tricine SDS-PAGE.

Cleavage assay of the His-tag. CoreC(K26E) was diluted to a monomeric concentration of 60 μM with 20 mM HEPES buffer (pH8.0) and pre-incubated at 30°C for 5 min. Hrv 3C Protease (Takara Bio) was diluted to 20 mU/μL with 20 mM HEPES buffer (pH8.0). The cleavage reaction was initiated by adding the diluted Hrv 3C Protease solution for a final concentration of 2 mU/μL. After 5, 10, 15, 20, and 40 min at 30°C, the samples were aliquoted and quenched by SDS-PAGE sample buffer (Nacalai Tesque). The cleavage reaction was also tested for 1 μM mTIP120-Ba (corresponding to a monomeric concentration of 60 μM for CoreC(K26E)) with the same manner. They were analyzed by Tricine SDS-PAGE.

Small-angle X-ray scattering (SAXS).

Sample preparation

TIP120 and mTIP120-Ba were further purified by size-exclusion chromatography (SEC) on Sephacryl S-400 HR (Cytiva) in 25 mM HEPES buffer (pH 8.0) containing 5% (v/v) glycerol, 100 mM NaCl, and 1 mM EDTA or 4 mM $BaCl_2$. Fractions containing the 60-mer were concentrated using Amicon Ultra centrifugal filters with a molecular weight cutoff (MWCO) of 10 kDa then subjected to SAXS analysis. SEC-purified mTIP120-Ba was incubated overnight with 32 mM EDTA at 30°C and also subjected to the analysis.

Data acquisition and processing

SAXS experiments were performed using synchrotron radiation (λ=0.15 nm) with a PILATUS3 2M detector (Dectris) at the Photon Factory BL-10C beamline (High Energy Accelerator Research Organization (KEK), Tsukuba, Japan). Scattering data were acquired for protein concentrations of 1, 2, 3, 4, and 5 mg/mL. The acquired scattering images were analyzed essentially as described previously[1]. Briefly, the scattering intensity curves were calculated for each protein concentration using the SAngler program[3]. The radius of gyration was estimated for each protein concentrations by Guinier analysis. They were extrapolated to zero-concentration and used for latter analyses. The radius of gyration was again estimated by using extrapolated scattering. The pair distance distributions were estimated using the ATSAS program suite[4]. For TIP120 and mTIP120-Ba, low-resolution models composed of dummy atoms were reconstructed using the *ab initio* modeling programs DAMMIF[5] and DAMMIN[6].

Cryo-EM sample preparation and data collection. mTIP120-Ba was further purified by SEC on Sephacryl S-400 HR in 25 mM HEPES buffer (pH 8.0) containing 100 mM NaCl and 4 mM $BaCl_2$. Fractions corresponding to the 60-mer were concentrated using Amicon Ultra centrifugal filters with

an MWCO of 10 kDa and adjusted to a protein concentration of 5 mg/mL. Since high concentrations of $BaCl_2$ cause severe aggregation in the thin ice layer on the EM grid, buffer exchange was performed immediately prior to observation. Specifically, the sample was diluted 10-fold with 25 mM HEPES buffer (pH 8.0) containing 100 mM NaCl, and the protein was concentrated 10-fold using Amicon Ultra centrifugal filters with an MWCO of 10 kDa. This process was repeated twice, such that the Ba concentration was reduced by 100-fold. Then, 3 μL of the sample was immediately applied to a holey carbon grid (Quantifoil, Cu, R1.2/1.3, 300 mesh). The grid was rendered hydrophilic by subjecting it to glow discharge for 30 s in air (11 mA current; PIB-10, Vacuum Device Inc.). The grid was blotted for 5 s (blot force 15) at 18°C and 100% humidity and then flash-frozen in liquid ethane using a Vitrobot Mark IV (Thermo Fisher Scientific). A total of 6,191 movies were acquired on a Talos Arctica cryo-EM system (Thermo Fisher Scientific) operating at 200 kV in the nanoprobe mode using the EPU software for automated data collection. The movies were collected on a Falcon 4 direct electron detector operating in the electron counting mode at a nominal magnification of 120,000 (0.84 Å/pixel). Fifty movie fractions were recorded at an exposure of 1.00 electrons per $Å^2$ per fraction, corresponding to a total exposure of 50 $e^-/Å^2$. The defocus steps used were −0.8, −1.2, −1.6, and −2.0 μm.

Cryo-EM data processing. Datasets were processed in the RELION-4.1 software[7]. See Fig. S14 and Table S4 for details regarding the cryo-EM data processing. The movie fractions were aligned, dose weighted, and averaged using the RELION implementation on 5 × 5 tiled fractions with a B-factor of 300. From the 6,191 micrographs, those whose total accumulated motions were >80 Å were discarded and 5,986 micrographs were selected. The non-weighted movie sums were used for contrast transfer function (CTF) estimation with the Gctf program[8], whereas the dose-weighted sums were used for all subsequent steps of image processing. The 3,801 images whose CTF maximum resolutions were better than 6 Å were selected. Next, 366,131 particles were automatically picked with the crYOLO program[9] and extracted while rescaling to 2.52 Å/pixel with a box size of 180 pixels. The extracted particle images were subjected to reference-free 2D classification for two cycles (200 expected classes, 300 Å mask diameter), then the 106,109 particles corresponding to the best 78 classes displaying secondary structural elements were selected and used for *ab initio* reconstruction (C_1 symmetry, single expected class, 300 Å mask diameter). The generated *ab initio* map was used as an initial 3D reference for the 3D classification (C_1 symmetry, four expected classes, 300 Å mask diameter, and an angular sampling interval of 7.5°). The best two 3D classes, which contained 81,967 particles displaying the highest resolution, were used for the 3D refinements (C_1 symmetry, 300 Å mask diameter, and an initial angular sampling of 7.5°). The refined volume and 81,967 particle images were rescaled to 0.84 Å/pixel with a box size of 540 pixels and used for the 3D refinement (C_1 symmetry, 300 Å mask diameter). Post-processing with a soft-edged 3D mask (15 pixel extension, 30 pixel soft cosine edge) afforded a final map (C_1) with 7.44 Å resolution.

Icosahedral symmetry was imposed on this map, which was used for the 3D refinement (*I* symmetry, 300 Å mask diameter, and an initial angular sampling of 3.7°). The refined map was used as a 3D reference for the no-alignment 3D classification (two expected classes, *T*=8, 300 Å mask diameter). The 54,925 particles of the major 3D class were used for the 3D refinements (*I* symmetry, 300 Å mask diameter). After removing duplicates, two cycles of CTF refinement and Bayesian polishing were performed with 54,685 particles. The 3D refinement (*I* symmetry, 300 Å mask diameter) with a soft-edged 3D mask (15-pixel extension, 30-pixel soft cosine edge) was executed after each CTF refinement and Bayesian polishing step. A final 3D refinement (*I* symmetry, 300 Å mask diameter) with a soft-edged 3D mask (15-pixel extension, 30-pixel soft cosine edge) and post-processing afforded a final map (*I*) with 4.58 Å resolution.

Structural alignment to the cryo-EM map. The model structure of mTIP60-Ba (PDB ID: 7XM1)[10] was fitted to the cryo-EM map (C_1 or *I*) of mTIP120-Ba using UCSF Chimera[11].

Incorporation of GFP. GFPuv5 was genetically fused to the N-terminus of CoreC(K26E) without a linker. It was mixed with CoreC(K26E) at various molar ratio (See Fig. S19, right panel). Then, they were incubated with 120 μM CoreN and 2 mM $BaCl_2$ at 30°C for at least 4 h and analyzed by Native PAGE. After the electrophoresis, the fluorescence band corresponding to the GFP was observed with FAS5 imaging system with a blue/green LED transilluminator (NIPPON Genetics) before CBB staining.

Data availability

The cryo-EM map of mTIP120-Ba (C_1) have been deposited in the Electron Microscopy Data Bank (EMDB ID: EMD-38931). The SAXS data and low-resolution models have been deposited in the Small Angle Scattering Biological Data Bank (SASBDB) under accession codes SASDUC2 (TIP120), SASDUD2 (mTIP120-Ba), and SASDUE2 (mTIP120-Ba with excess EDTA).

References for methods

1 N. Kawakami, H. Kondo, Y. Matsuzawa, K. Hayasaka, E. Nasu, K. Sasahara, R. Arai and K. Miyamoto, *Angew. Chem. Int. Ed.*, 2018, **57**, 12400–12404.

2 H. Schägger, *Nat. Protoc.*, 2006, **1**, 16–22.

3 N. Shimizu, K. Yatabe, Y. Nagatani, S. Saijyo, T. Kosuge and N. Igarashi, *AIP Conf. Proc.*, 2016, **1741**, 050017.

4 K. Manalastas-Cantos, P. V. Konarev, N. R. Hajizadeh, A. G. Kikhney, M. V. Petoukhov, D. S. Molodenskiy, A. Panjkovich, H. D. T. Mertens, A. Gruzinov, C. Borges, C. M. Jeffries, D. I. Sverguna and D. Frankea, *J. Appl. Crystallogr.,* 2021, **54**, 343–355.

5 D. Franke and D. I. Svergun, *J. Appl. Crystallogr.*, 2009, **42**, 342–346.

6 D. I. Svergun, *Biophys. J.*, 1999, **76**, 2879.

7 D. Kimanius, L. Dong, G. Sharov, T. Nakane and S. H. W. Scheres, *Biochem. J.,* 2021, **478**, 4169–4185.

8 K. Zhang, *J. Struct. Biol.*, 2016, **193**, 1–12.

9 T. Wagner, F. Merino, M. Stabrin, T. Moriya, C. Antoni, A. Apelbaum, P. Hagel, O. Sitsel, T. Raisch, D. Prumbaum, D. Quentin, D. Roderer, S. Tacke, B. Siebolds, E. Schubert, T. R. Shaikh, P. Lill, C. Gatsogiannis and S. Raunser, *Commun. Biol., 2019*, **2**, 1–13.

10 N. Ohara, N. Kawakami, R. Arai, N. Adachi, T. Moriya, M. Kawasaki and K. Miyamoto, *J. Am. Chem. Soc.*, 2023, **145**, 216–223.

11 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605–1612.

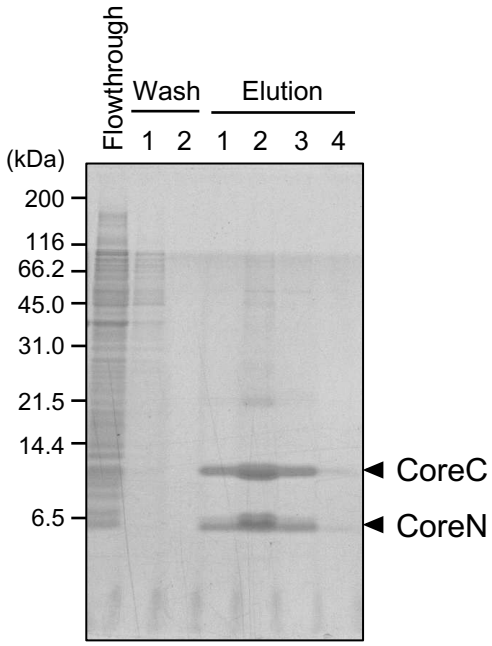


Figure S1. Tricine SDS-PAGE analysis of the fractions from the Ni-NTA column purification of the co-expressed split TIP60 fragments.

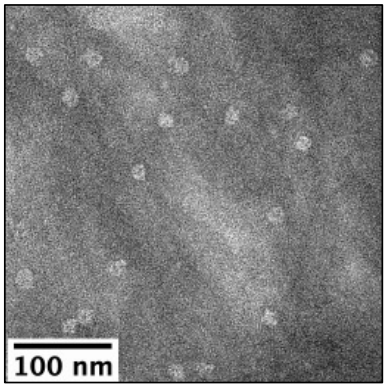


Figure S2. Negative-stained TEM image of TIP120 obtained from the *in vivo* co-expression.

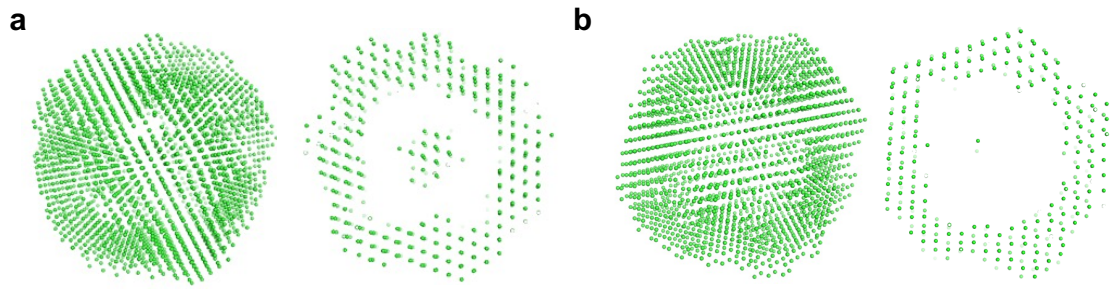


Figure S3. Dummy atom models constructed from the pair distance distributions obtained by SAXS analysis: (**a**) TIP120 and (**b**) mTIP120-Ba.

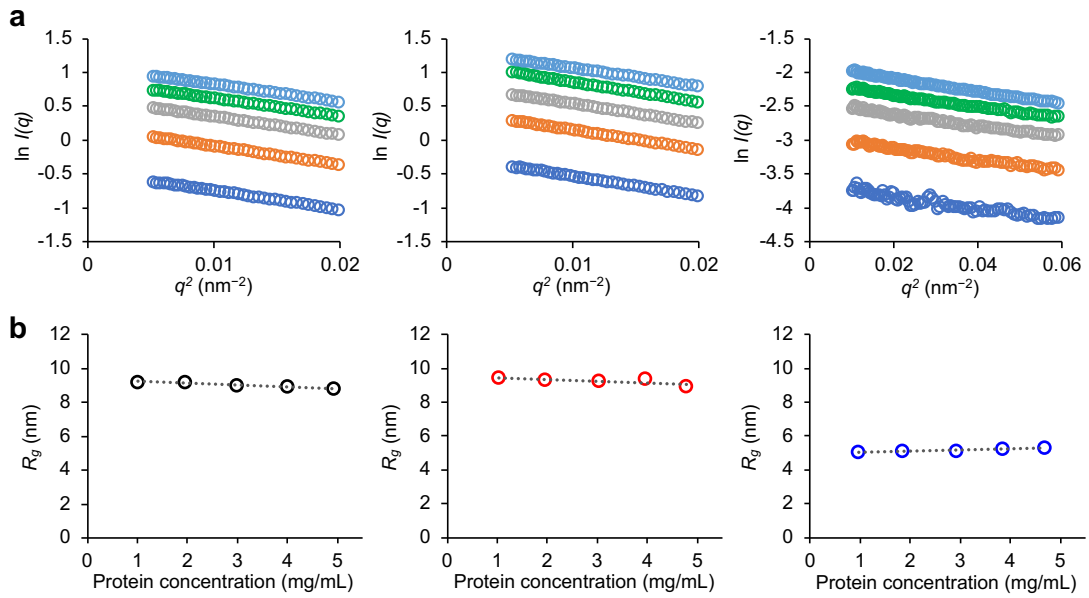


Figure S4. Guinier analysis of TIP120 (left), mTIP120-Ba (middle), and mTIP120-Ba with excess EDTA (right): (**a**) Guinier plots for various protein concentrations (1, 2, 3, 4, and 5 mg/mL). (**b**) Estimated R_g values for various protein concentrations.

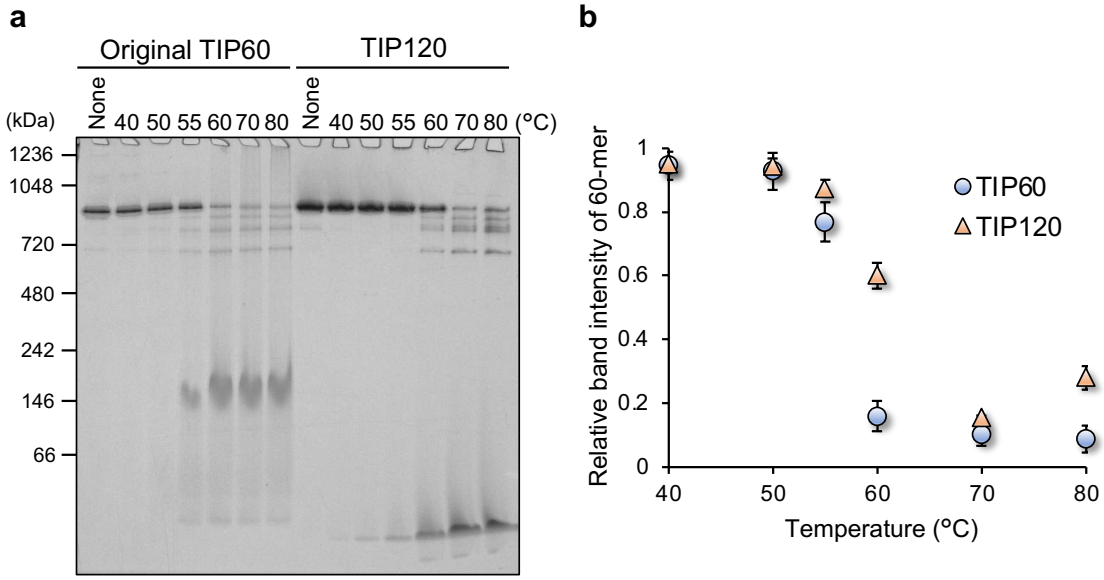


Figure S5. Thermal stability assays of the original TIP60 and TIP120. (**a**) Native PAGE analysis after heat treatment at various temperatures. (**b**) Relative band intensities corresponding to the 60-mer versus the heat treatment temperature. The values are plotted as the mean ± s.d. (n=3).

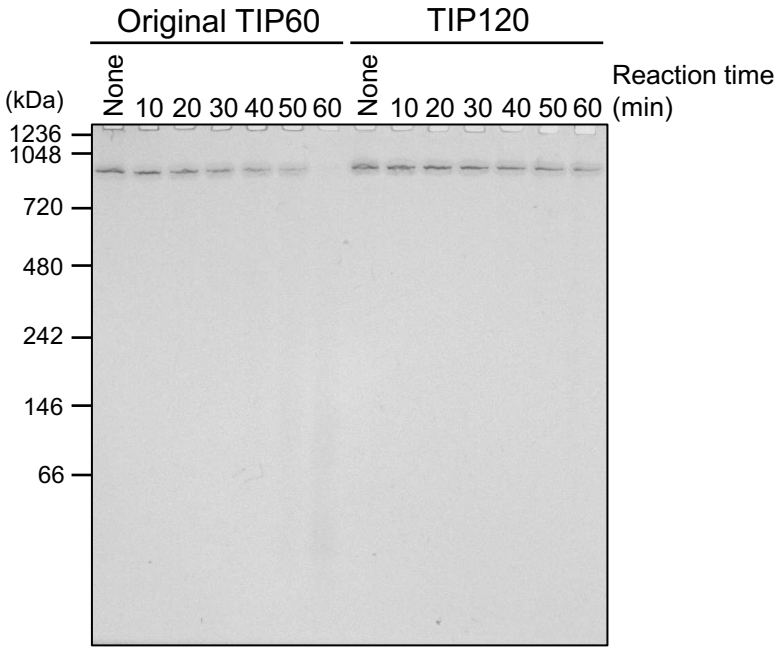


Figure S6. Native PAGE analysis of the original TIP60 and TIP120 (obtained from *in vivo* co-expression) after the digestion with Proteinase K.

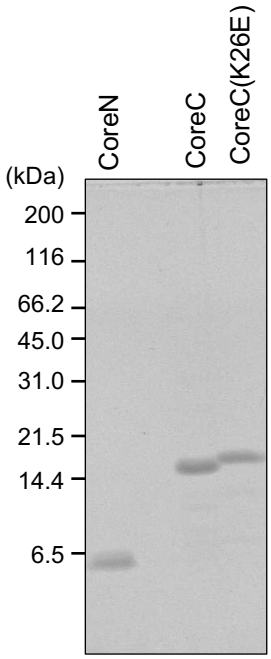


Figure S7. Tricine SDS-PAGE analysis of the independently expressed and purified fragments (CoreN, CoreC, and CoreC(K26E)).

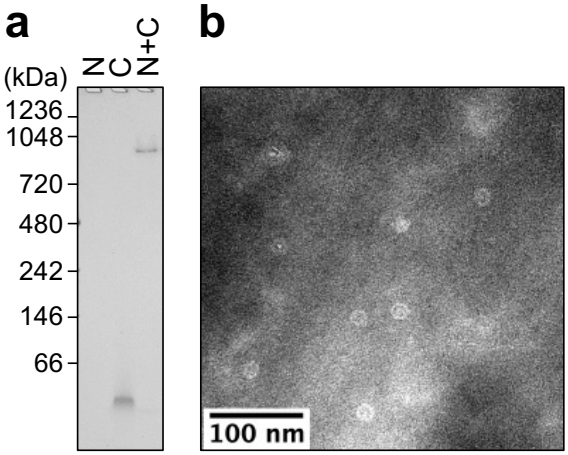


Figure S8. *In vitro* assembly of the split TIP60 fragments. (**a**) Native PAGE analysis of the mixture of independently purified CoreN and CoreC. (**b**) Negative-stained TEM image of the *in vitro* assembled TIP120.

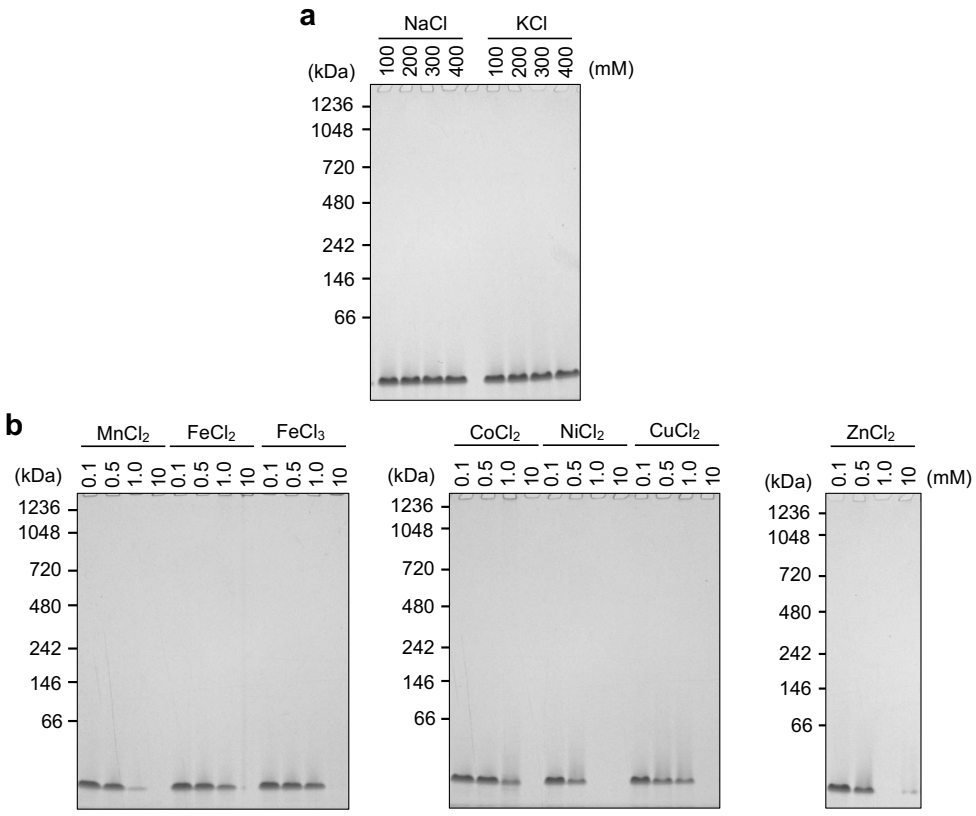


Figure S9. Native PAGE analysis of CoreN/CoreC(K26E) mixtures in the presence of various metal ions: (**a**) monovalent cationic ions (Na^+ and K^+) and (**b**) transition metal ions (Mn^{2+}, Fe^{2+}, Fe^{3+}, Co^{2+}, Ni^{2+}, Cu^{2+}, and Zn^{2+}).

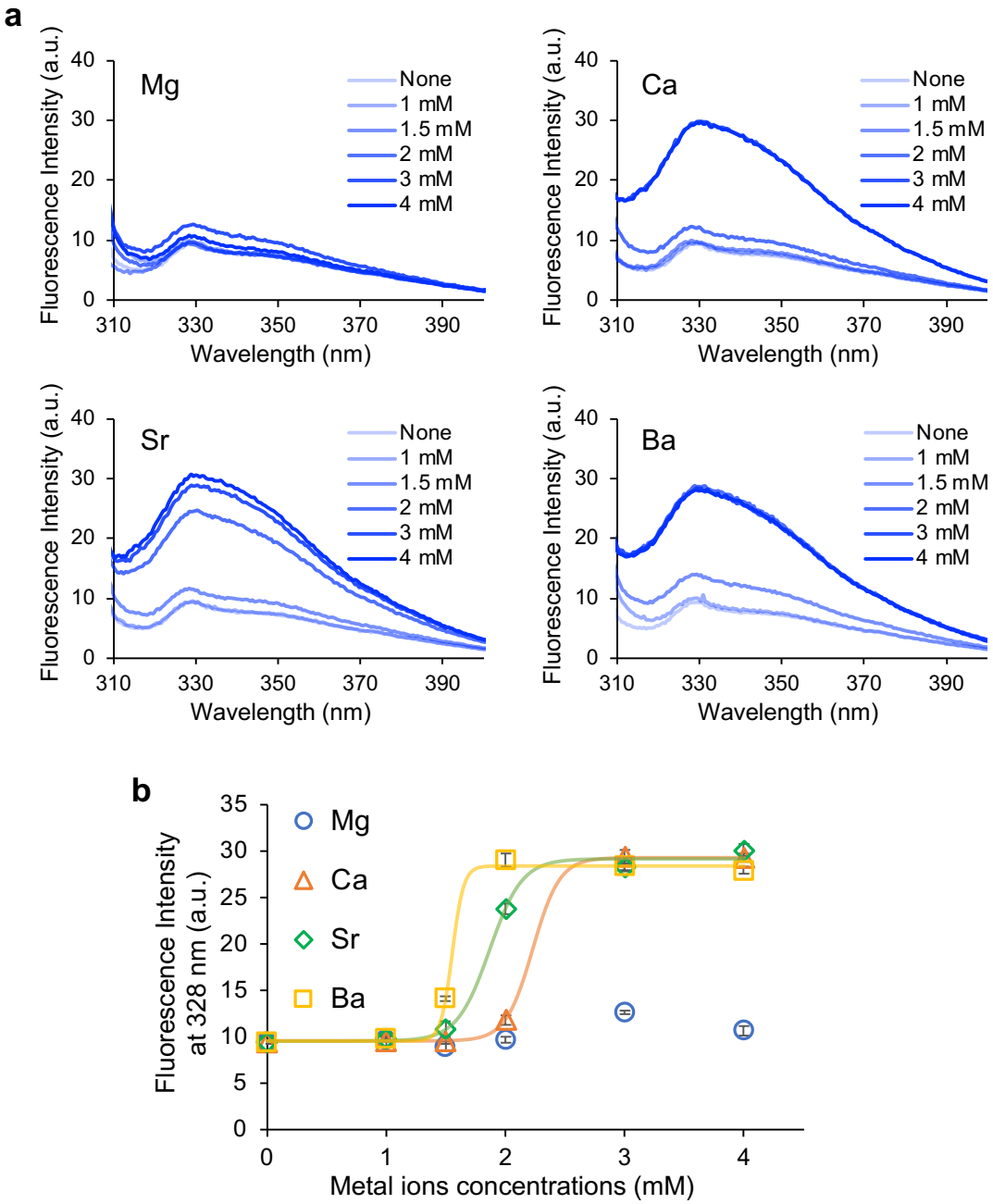


Figure S10. Trp fluorescence analysis of metal-induced assembly of CoreN and CoreC(K26E). (**a**) Trp fluorescence spectrum of the mixture of CoreN and CoreC(K26E) with various concentrations (0, 1, 1.5, 2, 3, and 4 mM) of alkaline metal ions: Mg^{2+}, Ca^{2+}, Sr^{2+}, and Ba^{2+}. λ_{ex}=295 nm. (**b**) The Trp fluorescence intensity at 328 nm against the metal ions concentrations. Each symbol shows the mean ± s.d. (*n*=3). The data were fitted with sigmoid curve and determined the affinity for Ca, Sr, and Ba ions as 2.2, 1.9, and 1.6 mM, respectively, from inflection points.

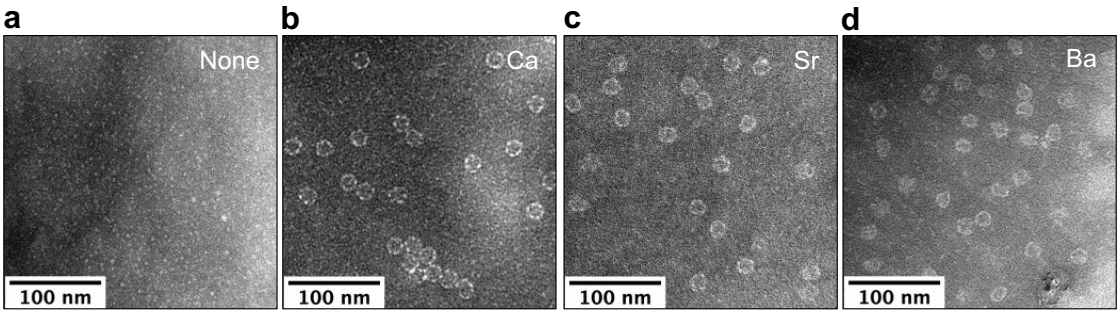


Figure S11. Negative-stained TEM images of (**a**) the mixture of CoreN and CoreC(K26E) without metal ions, (**b**) mTIP120-Ca, (**c**) mTIP120-Sr, and (**d**) mTIP120-Ba.

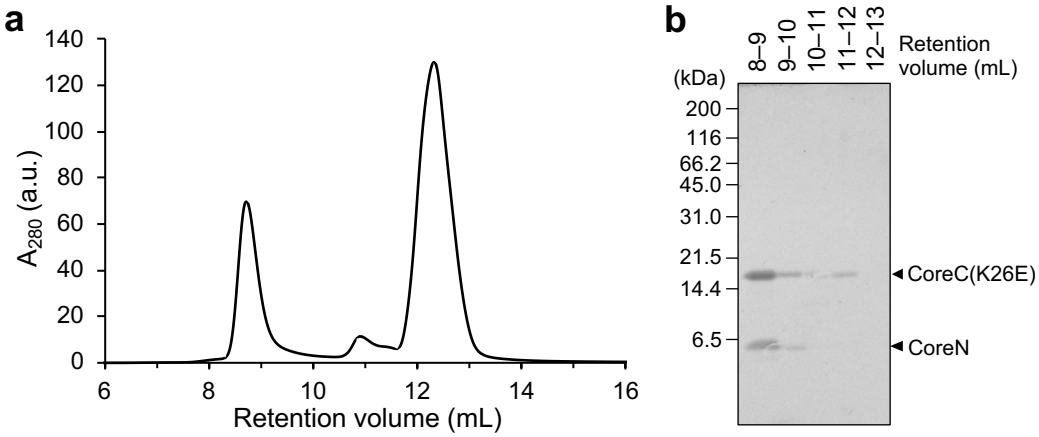


Figure S12. SEC analysis of mTIP120-Ba assembled with 4 mM $BaCl_2$. (**a**) The chromatograph. (**b**) Tricine SDS-PAGE analysis of each fraction from retention volume of 8 to 13 mL. The assembly yield was estimated as 91% by peak area ratio of the first peak (assembled cage) and the second peak (unassembled fragments). The third peak almost reached permeation limit and the corresponding fraction showed no protein bands.

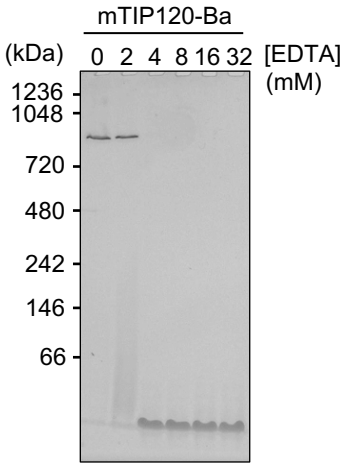


Figure S13. Native PAGE analysis of mTIP120-Ba after incubation with various concentrations of the metal-ion chelator EDTA.

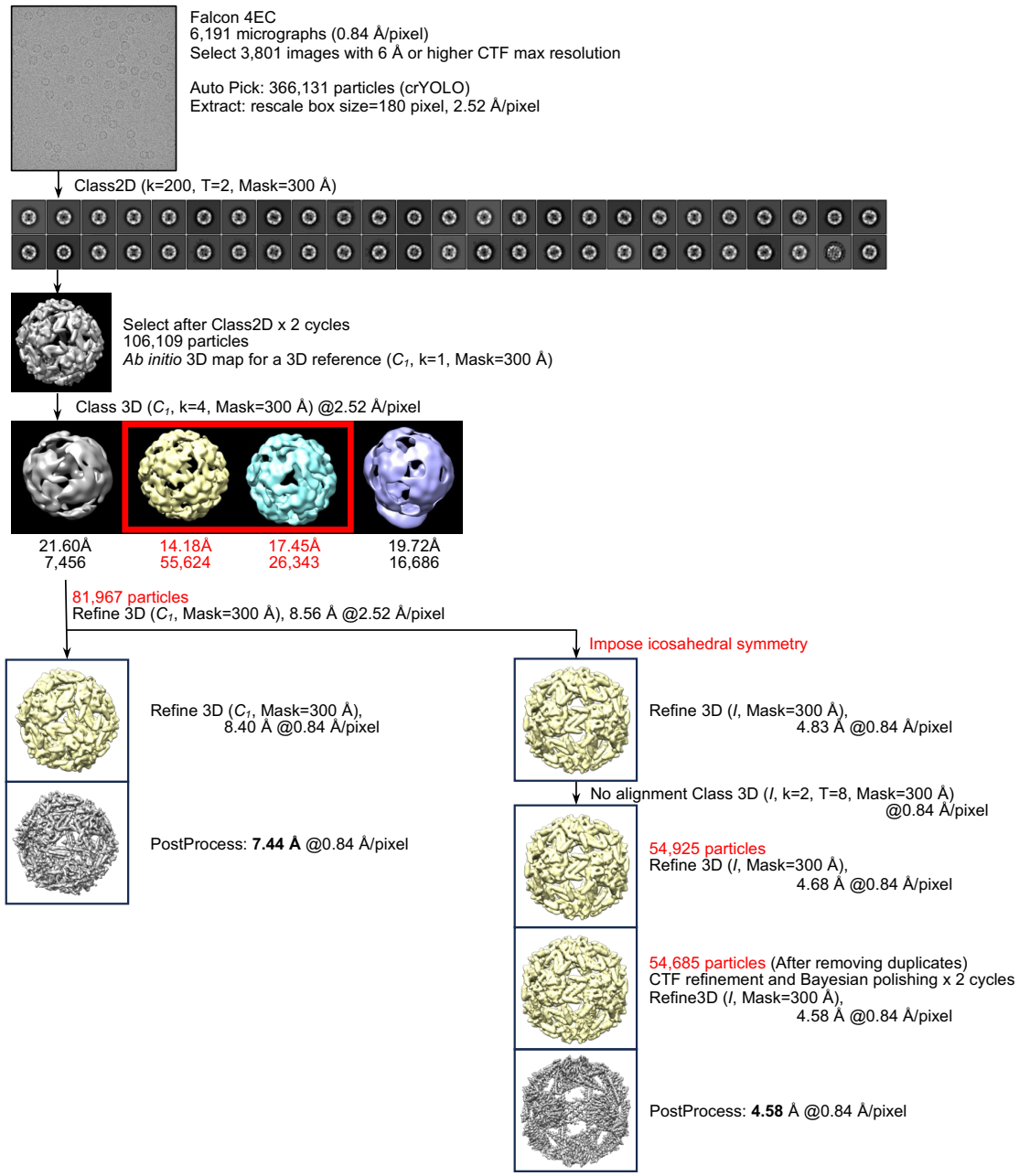


Figure S14. Summary of the cryo-EM single-particle analysis workflow.

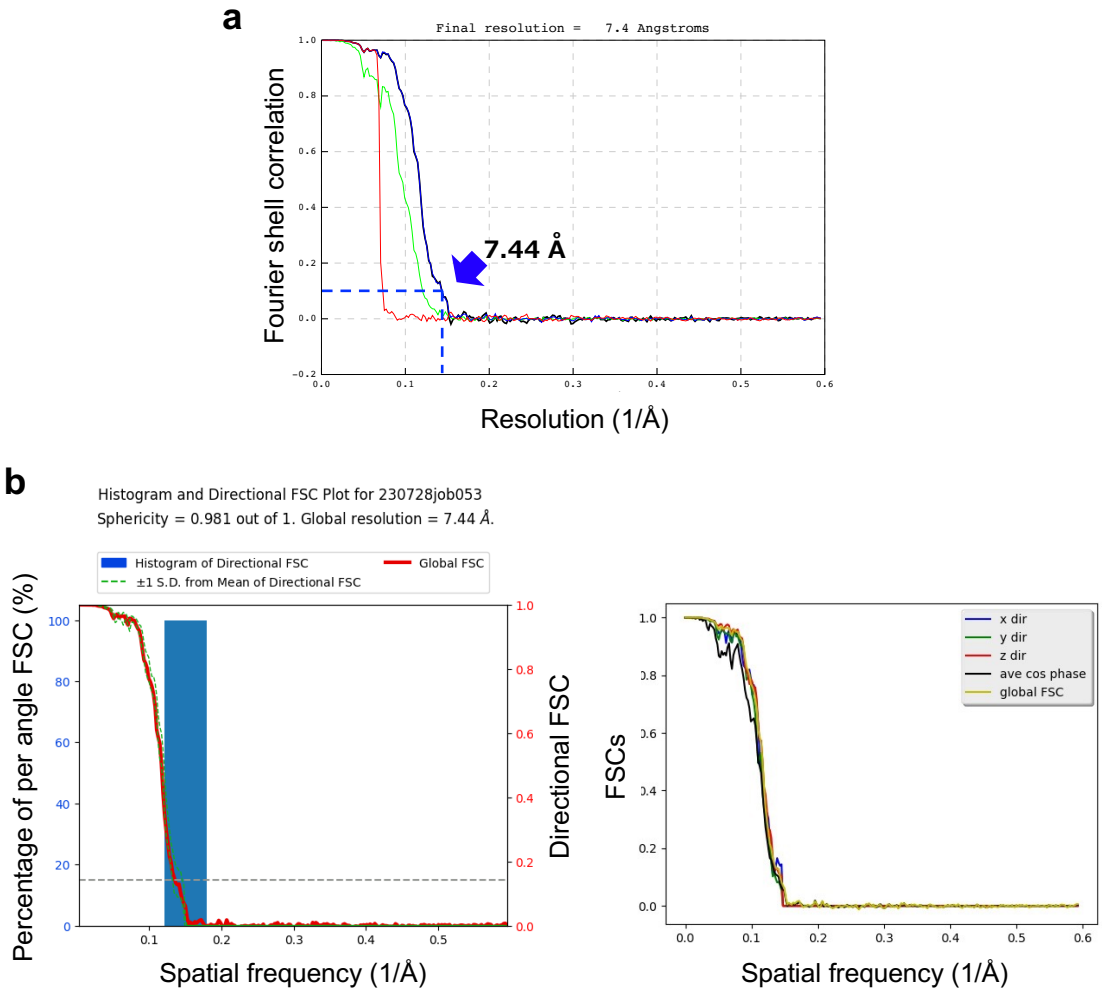


Figure S15. Map resolution of mTIP120-Ba without symmetry constraints. (**a**) Half-map Fourier shell correlation (FSC) curves for mTIP120-Ba (C_1), showing the corrected map (black), masked map (blue), unmasked map (green), and phase-randomized map (red). The resolution was estimated at 7.44 Å with an FSC threshold of 0.143. (**b**) 3D-FSC curves for the final map. The sphericity was 0.981.

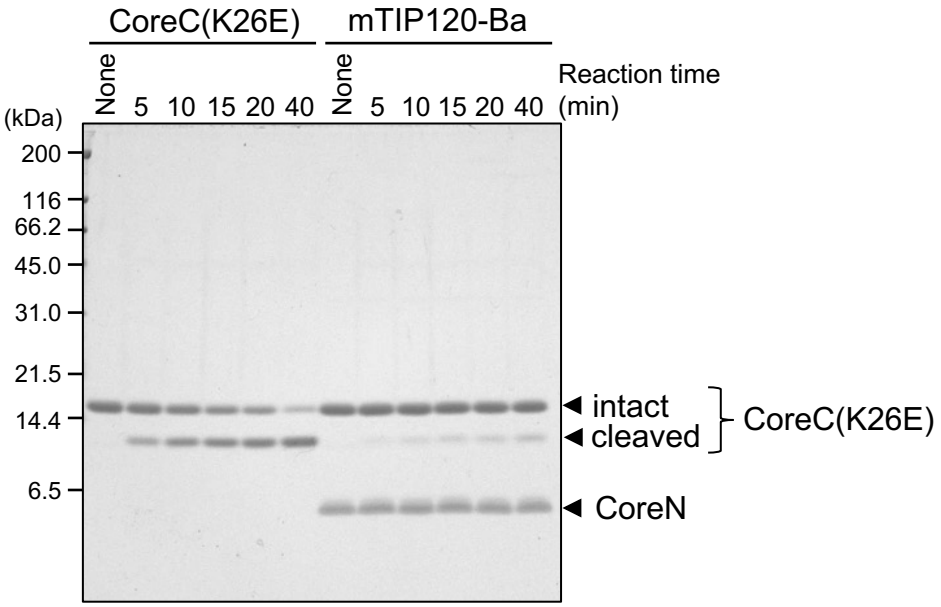


Figure S16. Tricine SDS-PAGE analysis of CoreC(K26E) (one fragment only) and mTIP120-Ba (assembled cage structure) after the reaction with Hrv 3C protease.

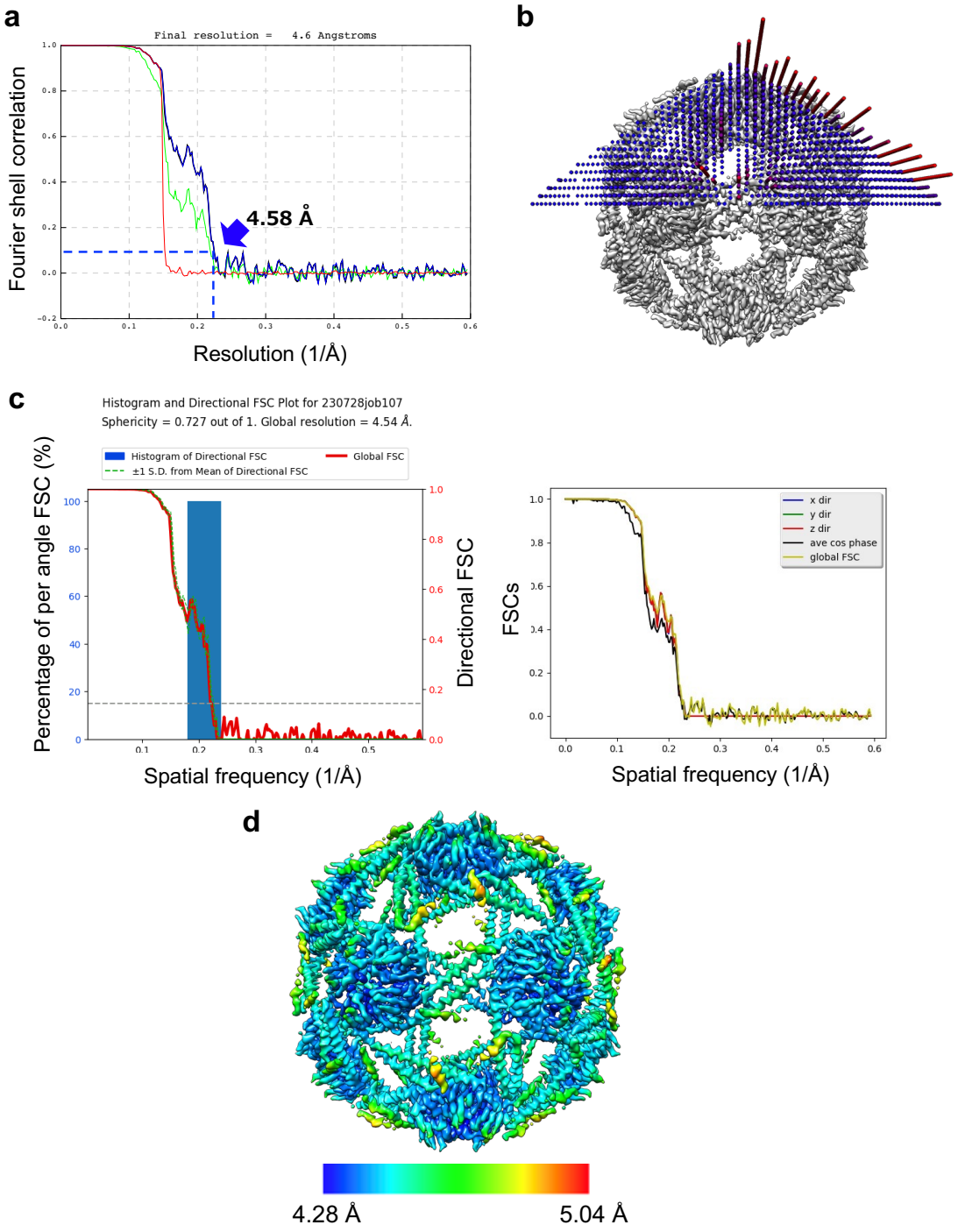


Figure S17. Map resolution of mTIP120-Ba with icosahedral symmetry constraints. (**a**) Half-map FSC curves for the corrected map (black), masked map (blue), unmasked map (green), and phase-randomized map (red). The resolution was estimated at 4.58 Å with an FSC threshold of 0.143. (**b**) Orientation distribution for the final map. (**c**) 3D-FSC curves for the final map. The sphericity was 0.727. (**d**) Local resolution of the final map.

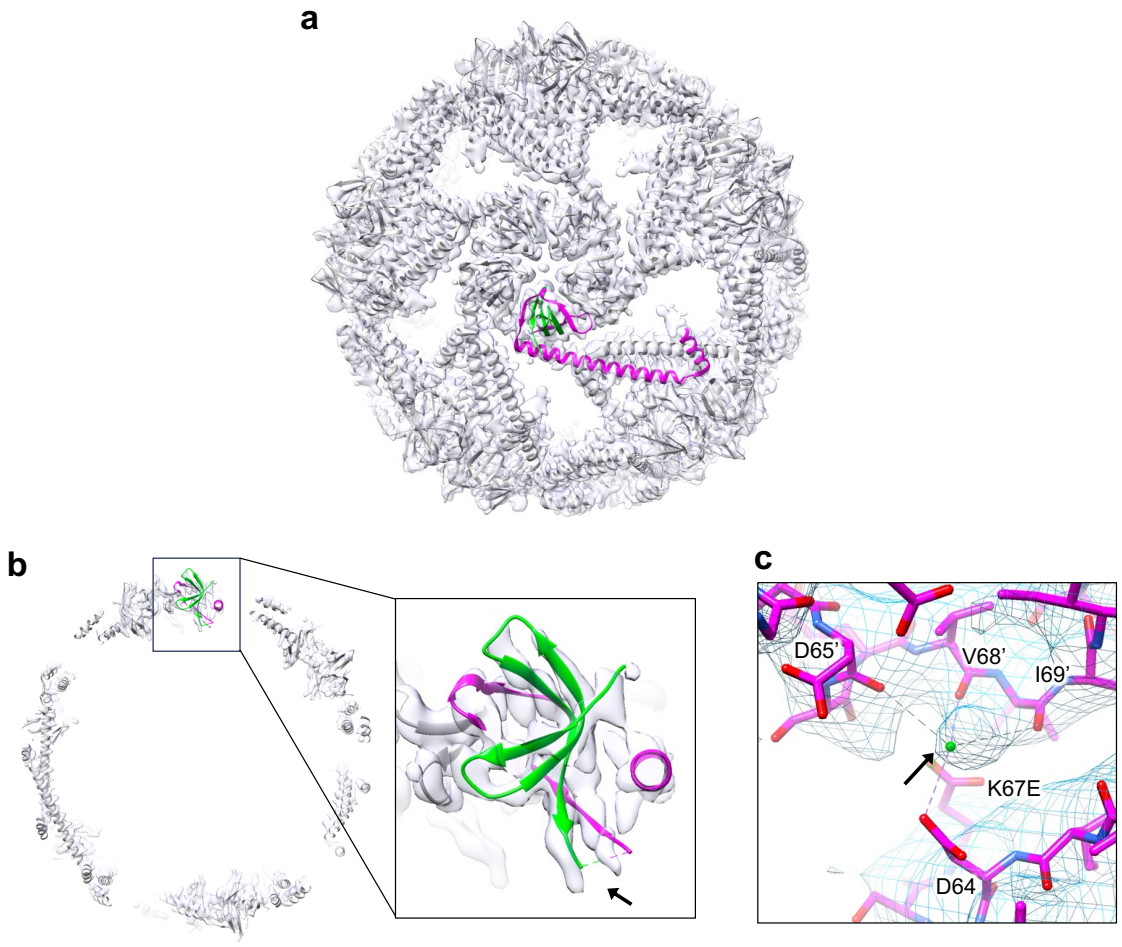
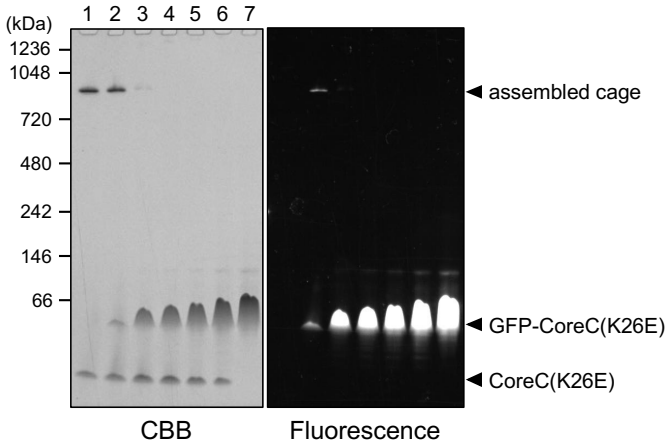


Figure S18. Comparison between the cryo-EM map of mTIP120-Ba with icosahedral symmetry constraints and the model structure of mTIP60-Ba (PDB ID: 7XM1). (**a**) Overall structure. The regions corresponding to CoreN and CoreC of one subunit are highlighted in green and magenta, respectively. (**b**) Cross-sectional view. The arrow indicates the splitting site. (**c**) Expanded view of the region corresponding to the Ba-coordinated site of mTIP60-Ba. The two highlighted adjacent subunits were fitted to the map.



	Protein concentrations (uM)		
Lane	CoreN	CoreC(K26E)	GFP-CoreC(K26E)
1	120	60	0
2	120	58	2
3	120	50	10
4	120	45	15
5	120	40	20
6	120	30	30
7	120	0	60

Figure S19. Native PAGE analysis of mTIP120 assembled with GFP-fused CoreC(K26E). Mixing ratio of the fragments for each lane were summarized in right panel.

Table S1. Nucleotide sequences of the PCR primers used for site-directed mutagenesis.

Mutation	Primer pair (5′→3′)
Insert	/5′Phos/TAGAAGGAGATATACATATGGGTAAACCGGTTCAACTGGTAC
TAG/RBS/ATG	/5′Phos/AGGAGTCGCTTGCATGGG
Delete C-terminus	GACTCCTTAAAAGCTTGCGGCCGCA
	AGCTTTTAAGGAGTCGCTTGCATGGG
Delete N-terminus	/5′Phos/GGTAAACCGGTTCAACTGGTACTG
Insert His-tag	/5′Phos/GGGTCCCTGAAAGAGGACTTCAAGCGGATCCTGGCTGTGGTG
K26E (CoreC)	GACAGCGAAGTGATCACCATTACCAG
	GATCACTTCGCTGTCATCAATCACAA
GFP-CoreC(K26E)	GGTAAACCGGTTCAACTGG
(In-fusion cloning)	GGGTCCCTGAAAGAGGAC
	CTCTTTCAGGGACCCATGAGTAAAGGAGAAGAACTTTTCA
	TTGAACCGGTTTACCTTTGTATAGTTCATCCATGCCA

Table S2. Amino acid sequences of the fragments.

		Amino acid sequence
TIP120 (co-expression)	CoreN	MGSSHHHHHHSQDPKNIKIMRLVTGEDIIGNISESQGLIT IKKAFVIIPMQATP
	CoreC	MGKPVQLVLSPWQPYTDDKEIVIDDSKVITITSPKDDIIK SYESHTRVLENKQVEEILRLEKEIEDLQRMKEQQELSLT EASLQKLQERRDQELRRLEEE
Independently purified	CoreN	MGSSHHHHHHSQDPKNIKIMRLVTGEDIIGNISESQGLIT IKKAFVIIPMQATP
	CoreC	MGSSHHHHHHSQDPLEVLFQGPGKPVQLVLSPWQPYT DDKEIVIDDSKVITITSPKDDIIKSYESHTRVLENKQVEEI LRLEKEIEDLQRMKEQQELSLTEASLQKLQERRDQELR RLEEE (HRV3C protease recognition site)
	CoreC(K26E)	MGSSHHHHHHSQDPLEVLFQGPGKPVQLVLSPWQPYT DDKEIVIDDSEVITITSPKDDIIKSYESHTRVLENKQVEEI LRLEKEIEDLQRMKEQQELSLTEASLQKLQERRDQELR RLEEE (HRV3C protease recognition site)
	GFP-CoreC(K26E)	MGSSHHHHHHSQDPLEVLFQGPMSKGEELFTGVVPILV ELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLP VPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEG YVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNYNSHNVYITADKQKNGIKANFKTR HNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSA LLKDPNEKRDHMVLLEFVTAAGITHGMDELYKGKPVQ LVLSPWQPYTDDKEIVIDDSEVITITSPKDDIIKSYESHTR VLENKQVEEILRLEKEIEDLQRMKEQQELSLTEASLQKL QERRDQELRRLEEE (GFPuv5)

Table S3. Summary of the SAXS analysis results.

	R_g (nm)	D_{max} (nm)	SASBDB ID
Original TIP60[1]	9.5	23.0	SASDDZ8
TIP120	9.2	22.5	SASDUC2
mTIP120 with Ba	9.4	22.4	SASDUD2
mTIP120 with EDTA	4.7	13.9	SASDUE2

Table S4. Cryo-EM data collection, refinement, and validation statistics

	mTIP120-Ba (EMD-38931)
Data collection and processing	
Microscope	Talos Arctica G2
Voltage (kV)	200
Detector	Falcon 4
Magnification	120,000
Pixel size (Å)	0.84
Automated data collection	EPU
Total exposure (e^{-}/Å^2)	50
Exposure rate (e^{-}/Å^2 fraction)	1.00
No. of fractions	50
Defocus (μm)	–0.8, –1.2, –1.6, –2.0
No. of collected micrographs	6,191
No. of selected micrographs	3,801
No. of particles for Class2D	366,131
No. of particles for Class3D	106,109
No. of particles for Refine3D	81,967
Symmetry imposed	C_1
Map resolution (Å)	7.44
FSC threshold	0.143