Supplementary Materials for

Assembly principles and structure of a 6.5 MDa bacterial microcompartment shell

Markus Sutter, Basil Greber, Clement Aussignargues, Cheryl A. Kerfeld*
*correspondence to: ckerfeld@lbl.gov

This PDF file includes:

Materials and Methods

Protein expression, purification and crystallization

A plasmid construct co-expressing BMC-H, BMC-T1, BMC-T2 and BMC-T3 (Hoch-5815, Hoch-5812, Hoch-5816 and Hoch-3341, respectively) was cloned into *E. coli* BL21(DE3), cells were grown at 37 °C until on $OD_{600\ nm}$ of 0.8 at which point the temperature was lowered to 22 °C and protein expression was induced by adding 0.05 mM IPTG and the cells were incubated O/N. Cell pellets were resuspended in Buffer A (20 mM Tris pH 7.4, 50 mM NaCl); small amounts of powdered DNase was added and cells were lysed using a French Press at >1000 psi. The lysate was cleared using a 30' 27,000 xg SS-34 spin at 4°C. The supernatant was treated with RNAseA (100 µg per 1 ml lysate) and the sample was incubated on a shaker at RT for 1 h. The sample was then applied on top of a 6 ml sucrose cushion (30% sucrose in Buffer A) and centrifuged in a Ti-70 rotor for 16 h at 257,000 xg. The supernatant was discarded and the pellet gently resuspended in 2 ml Buffer A. Insolubilized matter was removed with a short (2 min) spin in an Eppendorf minifuge at 16,000 xg at 4°C. The supernatant was applied on top of a continuous 10-50% sucrose gradient in Buffer A and centrifuged in a SW-28 swinging bucket rotor for 16 h at 70,000xg at 4°C. Gradient fractions were analyzed on SDS-PAGE and shell containing fractions were pooled and applied on a 5 ml MonoQ column equilibrated in Buffer A. The shells were then eluted by applying a 0-40% gradient of Buffer B (20 mM Tris pH 7.4, 1 M NaCl) over 20 column volumes.

Hoch-5814 BMC-P pentamer was expressed in *E. coli* from a pET vector based on pET21b, no extra sequences or purification tags were added. Expression and lysis were performed the same as for the shells; the cleared lysate was then applied on a column packed with TOYOPEARL SuperQ-650M resin and the pentamer was eluted by applying a 0-40% gradient of Buffer B. Fractions containing pentamer were pooled, concentrated using a Millipore Amicon Ultra-15 concentrator with a 10 kDa molecular weight cut-off and applied on a GE HiLoad 26/60 Superdex 75 size exclusion column.

Purified *H. ochraceum* shells without pentamer were then mixed with a >10x excess (from expected pentamer occupancy) of separately purified pentamer, incubated with slow shaking at room temperature for 1h and applied on MonoQ which separated the excess pentamer from the complete shells. Shell fractions were then pooled and concentrated/buffer exchanged to Buffer A using a Millipore Amicon Ultra-15 concentrator with a 100 kDa molecular weight cut-off. Shells were concentrated to A280 absorption of 2-4 and used to screen for crystallization using an Art Robbins Crystal Phoenix robot. An initial hit was improved in 24-well trays and gave rhombohedral shaped crystals up to 0.2 mm in length in conditions of 5-7 % PEG-20000, 0.1 M Bicine pH 8.7-9.1 with 4 µl of protein mixed with 2 µl of reservoir in a sitting drop plate. Crystals were stabilized by addition of reservoir solution containing 30% ethylene glycol, looped and flash frozen in liquid nitrogen.

The pET11-based plasmids containing the sequence coding for BMC-T2 (Hoch-5816) or BMC-T3 (Hoch-3341) were transformed into *E. coli* BL21 (DE3), and cells were grown in LB broth (Miller) with 100 mg/L ampicillin at 37 ˚C and 160 rpm until the $OD_{600\ nm}$ reached 0.8. IPTG was added to the culture (0.45 mM) and cells were grown for 4 h at the same temperature. Inclusion bodies containing BMC-T2 or BMC-T3 were purified and solubilized, and proteins were refolded using a protocol adapted from

Burgess (*18*). Cells were resuspended in 50 mM Tris pH 7.8, 150 mM NaCl, 10 mM $MgCl_2$ and 5 mM β-mercaptoethanol (buffer C) and lysed in presence of DNase using a French Press at >1000 psi. Triton X-100 was added to a final concentration of 1% (v/v), cell lysate was incubated for 15 min at RT with gentle shaking and centrifuged at 20,000 xg for 10 min. The pellets containing the inclusion bodies were repeatedly washed with buffer C containing 1% Triton X-100, and a final wash without Triton X-100 was performed to remove all traces of detergent. BMC-T2 and BMC-T3 inclusion bodies were solubilized in buffer C containing 8 M urea. Protein concentration was adjusted to 1 mg/ml and refolding was performed by adding the protein quickly (60-fold dilution) to 50 mM Tris pH 7.8, 150 mM NaCl, 5% (v/v) glycerol. DTT was added to 0.5 mM in the case of BMC-T3. The diluted proteins were concentrated using a Millipore Amicon stirred cell (30 kDa molecular weight cut-off) and centrifuged to remove aggregates. For protein crystallization, BMC-T2 and BMC-T3 were buffer-exchanged using GE Healthcare PD-10 columns to 10 mM Tris pH 7.8, 50 mM NaCl and 10 mM Tris pH 7.8, respectively. BMC-T2 crystals were obtained in sitting drop trays by mixing 3 µl of protein (1.5 mg/ml) and 1 µl of a reservoir condition containing 0.1 M Na Acetate pH 5.1, 1.6 M $MgSO_4$. Cryoprotection was achieved using 25% ethylene glycol in reservoir solution before crystal looping and flash freezing in liquid nitrogen. BMC-T3 was crystallized using 0.1 M Na cacodylate pH 6.5, 1.25 M Na citrate (protein concentration of 2 mg/ml, protein/reservoir ratio of 3:1) and cryo stabilized by addition of glycerol to 20 %.

Data collection, analysis and structure determination

Data for *H. ochraceum* shells was collected at SSRL beam line 12.2 (100K, wavelength of data collection 1.03317 Å) diffracting to about 3.5 Å. The crystals belong to space group $C222_1$ and the unit cell dimensions are 394x638x642 Å. Diffraction data were integrated with XDS (*19*) and scaled with SCALA (CCP4) (*20*). Self-rotation functions calculated from the data confirmed the icosahedral nature and from Matthew's coefficient calculations we expected half of a particle in the asymmetric unit. A low resolution (8.7 Å) cryo-electron microscopy density of the whole particle allowed constructing a model by placing the known structures of the hexameric and pentameric subunits which was of suitable quality for molecular replacement. A single solution had slightly higher Z scores and density averaging using icosahedral symmetry operators confirmed the correctness of this solution and enabled exact placement of the subunits for refinement and model building. Manual rebuilding/refinement cycles using COOT(*21*) and phenix.refine (*22*) led to a model with good geometry with regards to the resolution, 91.3 % are in the favored, 7.2 % allowed and 1.5% in the outlier region of the Ramachandran plot. Data for BMC-T2 and BMC-T3 was collected at the ALS beam line 5.0.2 (100K, wavelength of data collection 1.000 Å). BMC-T2 and BMC-T3 structures were solved using molecular replacement with CcmP (4HT5) as a search model and manually rebuilt/refined with COOT/phenix.refine. 98/97 % are in the favored and 2/3 % in the allowed region of the Ramachandran plot for BMC-T3 and BMC-T2, respectively with 1% rotamer outliers.

Cryo-EM specimen preparation, data acquisition and processing

For cryo-EM specimen preparation, C-flat 1.2/1.3 holey carbon grids (Protochips, Morrisville, NC, USA) were covered with a continuous carbon film and plasma cleaned using a SOLARUS plasma cleaner (Gatan, Pleasanton, CA, USA). 4 µl of a 3 mg/ml solution of *H. ochraceum* shells were incubated on the plasma-cleaned grids for 5-7 s in a Vitrobot Mk. IV (FEI Company, Hillsboro, OR, USA) before plunge-freezing in liquid ethane at liquid N_2 temperature. Cryo-EM data were collected using the LEGINON package(*23*) on a Tecnai F20 transmission cryo-electron microscope (FEI Company) operated at 120 kV acceleration voltage and equipped with a US4000 CCD camera (Gatan). Image acquisition was performed using a dose of 25 electrons/$Å^2$, defocus values of -1.5 to -3.0 µm, and a magnification of 107,142 x, resulting in a pixel size of 1.4 Å on the object scale.

Images and power spectra were inspected using the APPION package(*24*), and images showing excessive ice contamination or drift were rejected. Defocus values were estimated using CTFFIND4 (*25*) from within RELION (*26*). Initially, particles were picked using DOG PICKER (*27*) in APPION (*24*) and subjected to 2D classification in RELION (*26*). These initial 2D classes were used as templates for automated particle selection in RELION 1.4 (*26*). Subsequent processing steps were performed using RELION 1.4 (*26*) and are detailed in Fig. S9. A total of approximately 3750 particles were extracted from the micrographs, 2x binned (resulting in a pixel size of 2.8 Å on the object scale), and subjected to reference-free 2D classification. 3450 particles contained in the 2D classes showing near-spherical particles were refined imposing icosahedral symmetry, using as an initial reference a low-pass filtered shell of an icosahedral virus (*28*) (EMD-3351) rescaled to the approximate diameter of the *H. ochraceum* shell. The resulting cryo-EM map no longer resembles the initial reference, indicating that this reconstruction represents the structure of the *H. ochraceum* shell and is free from model bias. The angular assignments obtained in this initial refinement were used for 3D classification of the aligned particles using local angular searches. Two of the four classes showed only low-resolution features and the particles assigned to these classes were discarded. The remaining 2600 selected particles were split into fully independent half-sets (Gold-Standard refinement) and refined to 8.7 Å resolution according to the Fourier shell correlation (FSC) = 0.143 criterion (*29, 30*).

Figure generation and bioinformatics

Molecular structure figures were prepared with pymol (The PyMOL Molecular Graphics System, Version 1.7 Schrödinger, LLC). The Cryo-EM *H. ochraceum* shell figure was rendered in UCSF Chimera (*31*) and Persistence of Vision Ray Tracer (Persistence of Vision Pty. Ltd. (2004), Persistence of Vision Raytracer (Version 3.6), Retrieved from http://www.povray.org/download/). Sequences were aligned with ClustalX (*32*), phylogenetic trees were generated with PhyML(*33*) with default parameters and visualized with Archaeopteryx (*34*).

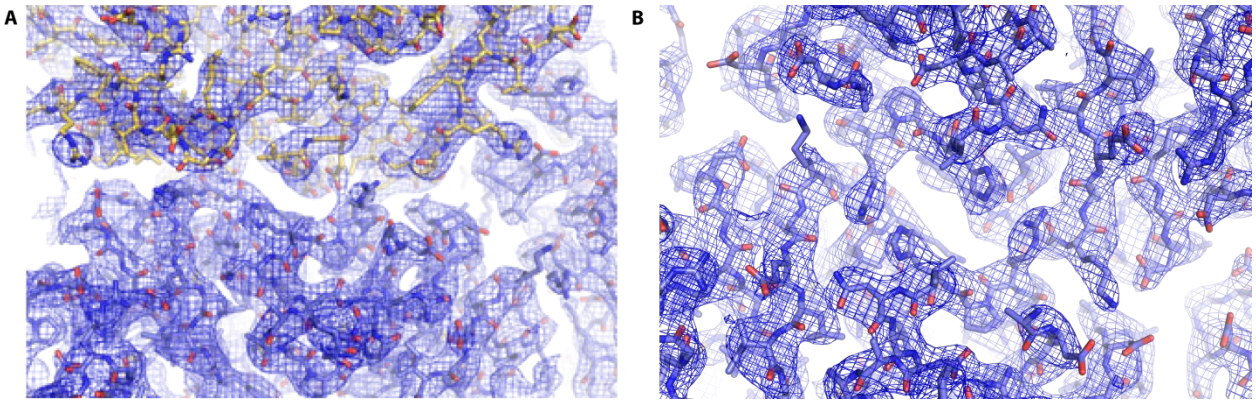

Fig. S1. Electron density figures.

(**A**) Sample 2Fo-Fc density contoured at 1.0 sigma of the hexamer-pentamer interface. (**B**) 2Fo-Fc electron density of “inserting” arginine of PRPH motif contoured at 1.2 sigma.

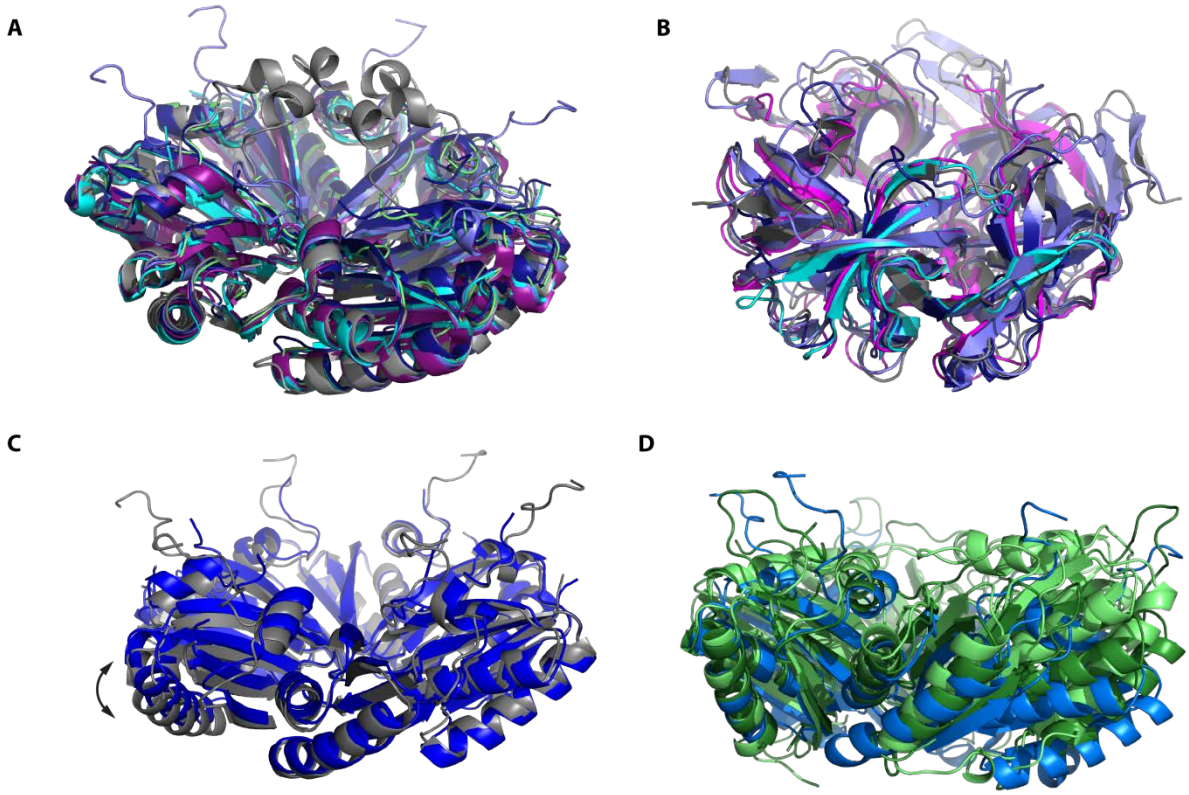

Fig. S2. Structural alignment figures.

(**A**) Alignment of BMC-H representatives show a high structural conservation (*Synechocystis* sp 6803 CcmK2, PDB ID 2A1B, grey; *H. neapolitanus* CsoS1A, PDB ID 2EWH, dark blue; *E. coli* EutM, PDB ID 3I6P, cyan; *S. enterica* PduA, PDB ID 3NGK, green; *M. smegmatis* MSM_0272, PDB ID 5L38, magenta; *H. ochraceum* BMC-H, PDB ID 5DJB, light blue. (**B**) Structural alignment of BMC-P representatives, *Synechocystis* sp 6803 CcmL, PDB ID 2QW7, grey; *H. neapolitanus* CsoS4A, PDB ID 2RCF, dark blue; *E. coli* EutN, PDB ID 2Z9H, cyan (only one chain shown, forms hexamers in crystal); *M. smegmatis* MSM_0273, PDB ID 5L37, magenta; *H. ochraceum* BMC-P, from shell structure, light blue. (**C**) Alignment of the *H. ochraceum* BMC-H in the shell structure (blue) with the crystal structure of the isolated protein (grey, PDB ID 5DJB). (**D**) Structural alignment of the *H. ochraceum* hexamer (blue) with the BMC-T proteins (BMC-T1 (light green) and BMC-T2 (dark green). For the superposition, only one chain of the BMC-T was chosen to minimize deviation instead of an overall alignment.

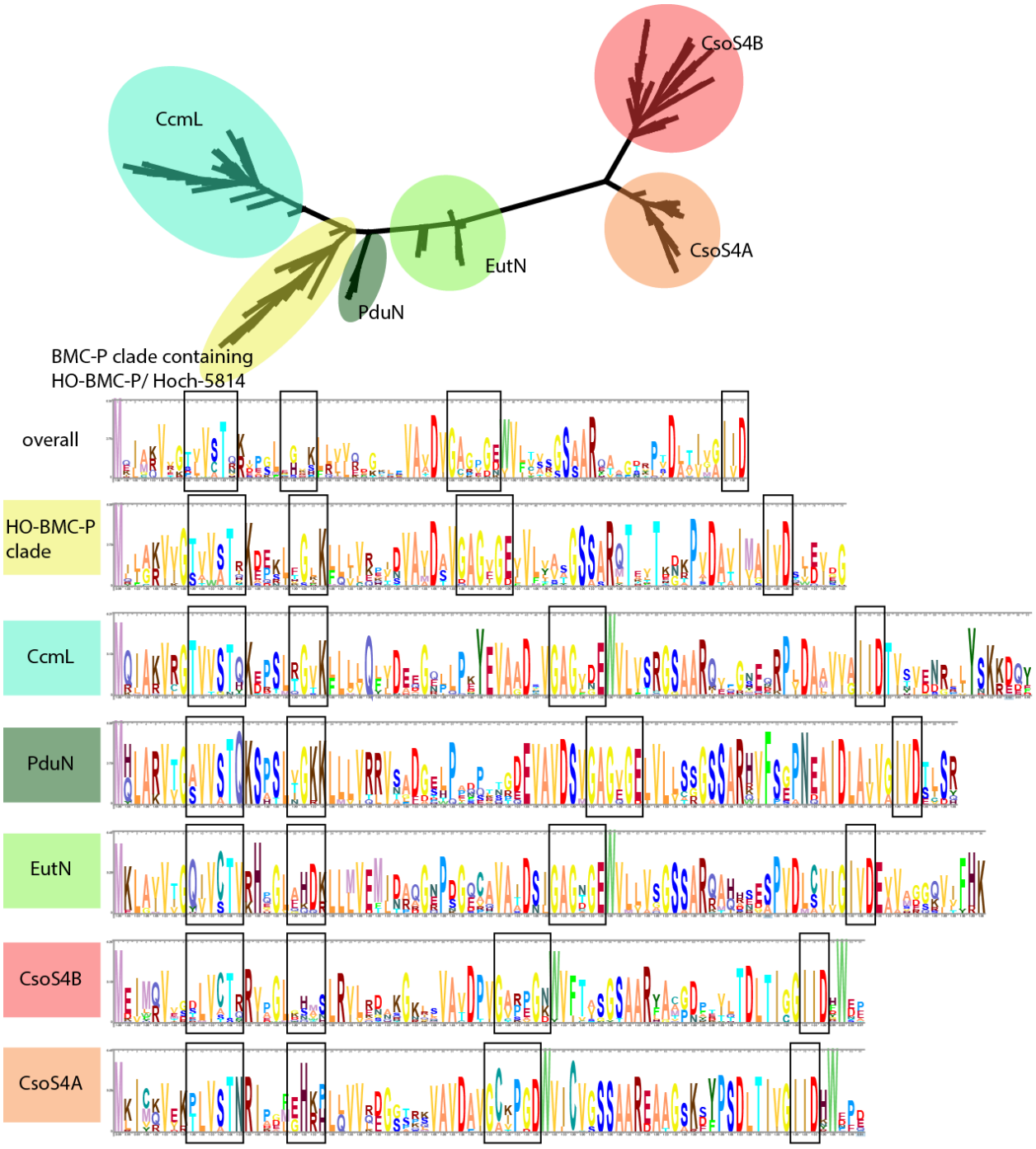

Fig. S3. BMC-P sequence analysis.

Phylogenetic tree of BMC-P proteins (pfam03319) and corresponding sequence logos for overall alignment and the different types. Main regions interacting with the BMC-H are highlighted with boxes.

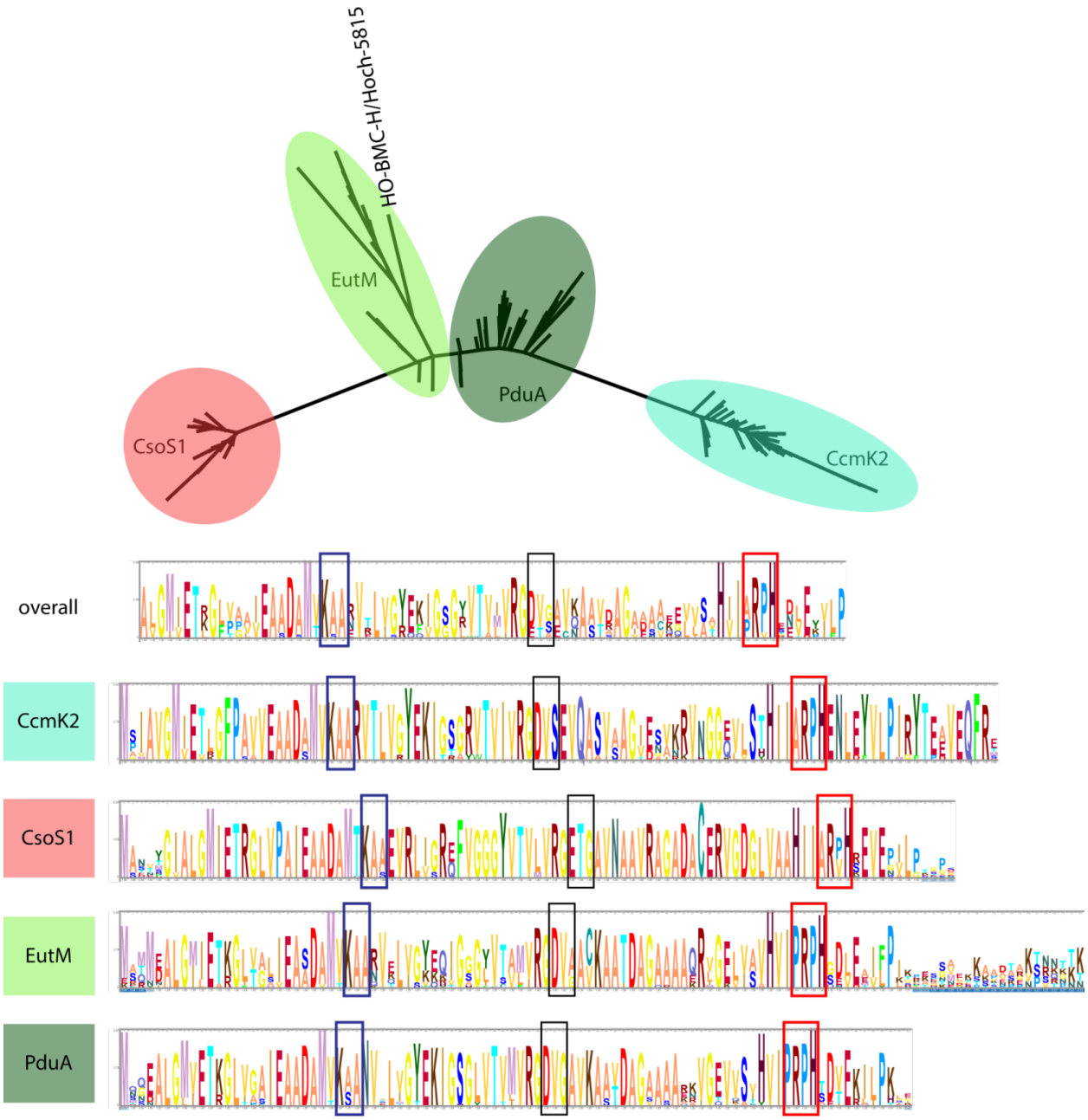

Fig. S4. BMC-H sequence analysis.

Phylogenetic tree of BMC-H proteins from experimentally characterized BMCs and corresponding sequence logos for overall alignment and the different types. Main regions interacting with the BMC-P and other BMC-H are highlighted with boxes (blue for the KAA motif and red for PRPH). The *H. ochraceum* BMC-H protein is a member the clade of sequences containing EutM.

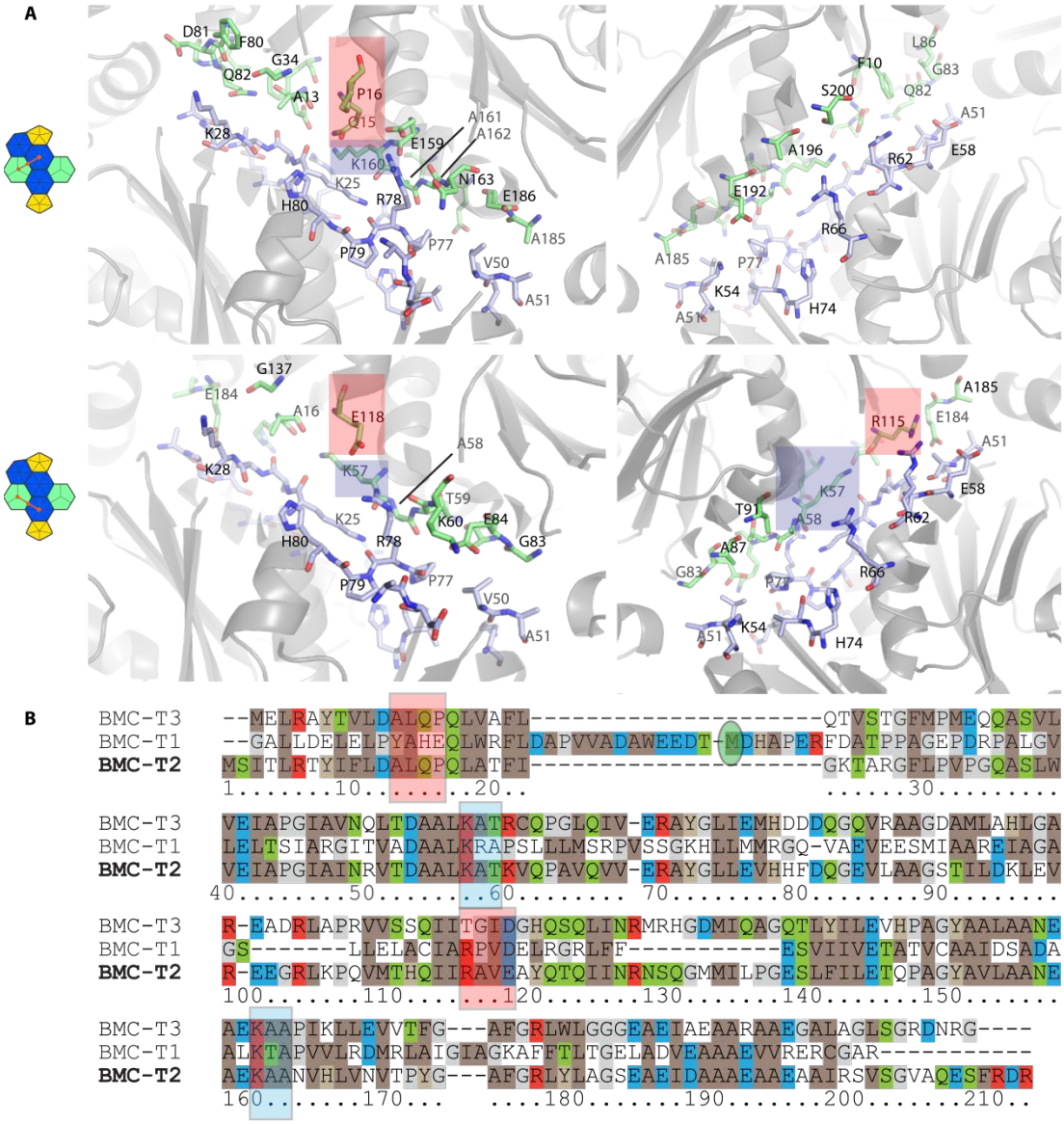

Fig. S5. Detailed view of BMC-T:BMC-H interactions.

(**A**) Detailed view of the two different interfaces of the BMC-T2:BMC-H contacts. (**B**) Structure based sequence alignment of the three BMC-T proteins present in the *H. ochraceum* shell. KAA and PRPH motif equivalent positions are highlighted in (**A**) and (**B**) with blue and red shading, respectively. Note that BMC-T2 and BMC-T3 are circularly permuted with regards to the standard BMC motif so that the C-terminal part of each BMC-T1 is moved to the N-terminus; the initiating methionine is highlighted with a green circle. Numbering follows BMC-T2 and corresponds to the structure in (**A**).

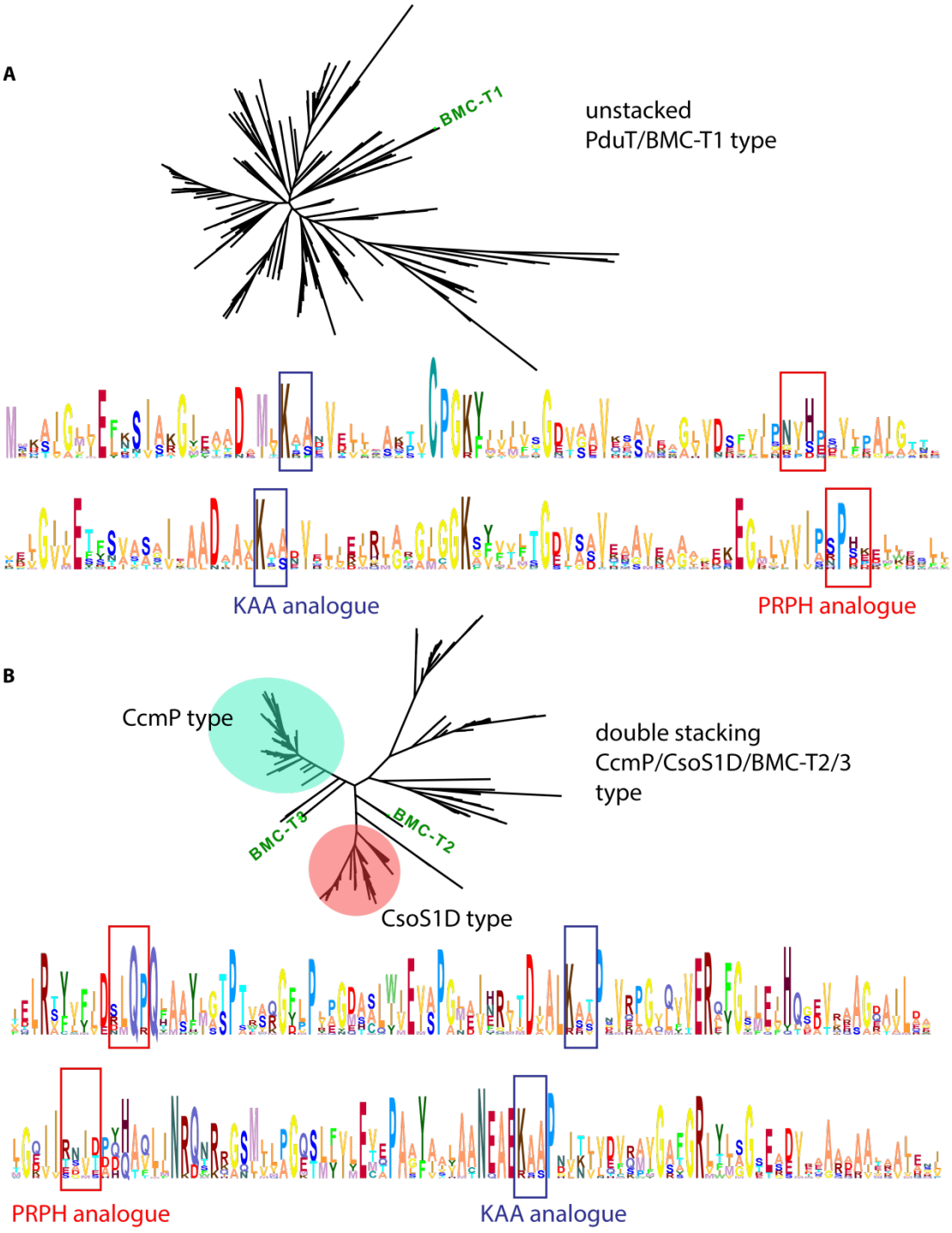

Fig. S6. Phylogenetic trees and conservation logo of the BMC-T type proteins.
(**A**) Single stacking PduT/BMC-T1 type. (**B**) double stacking CcmP/CsoS1D/BMC-T2 or BMC-T3 type. Regions corresponding to the PRPH and KAA motifs are highlighted with boxes. Positions of the *H. ochraceum* shell proteins BMC-T1, BMC-T2 and BMC-T3 on the trees are indicated.

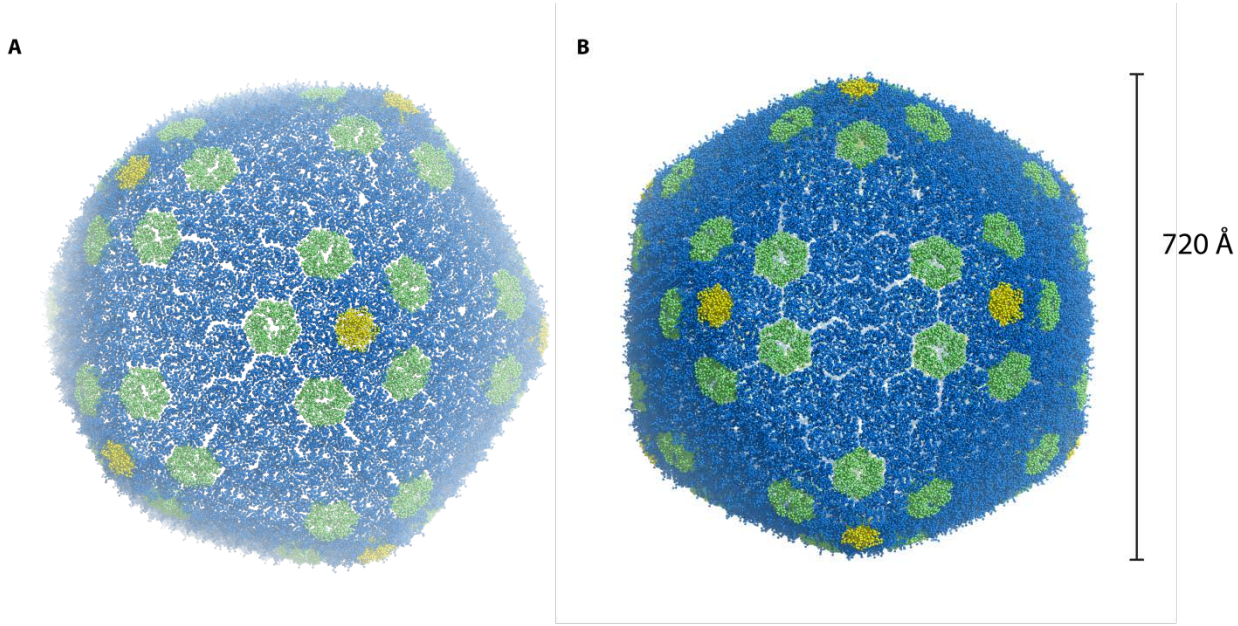

Fig. S7. T=36 BMC shell model.

(**A, B**) Two views of a model of a T=36 icosahedral BMC shell consisting of 290 BMC-H hexamers, 60 BMC-T pseudohexamers and 12 BMC-P pentamers constructed from the proteins and interactions observed in the *H. ochraceum* model. The resulting shell is 720 Å in diameter and approximately 22 MDa in mass.

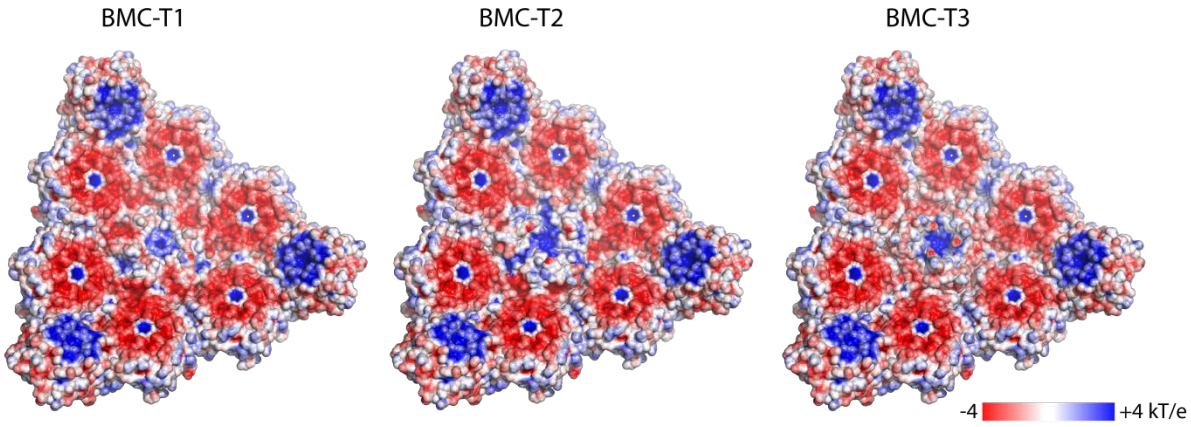

Fig. S8. Influence of BMC-T type on inside surface properties.

Electrostatic surface view of the inside facets with different BMC-T proteins modelled in the center. The three different BMC-T proteins show distinct surface shape and charge distribution. Color gradient from -4 kT/e (red) to +4 kT/e (blue).

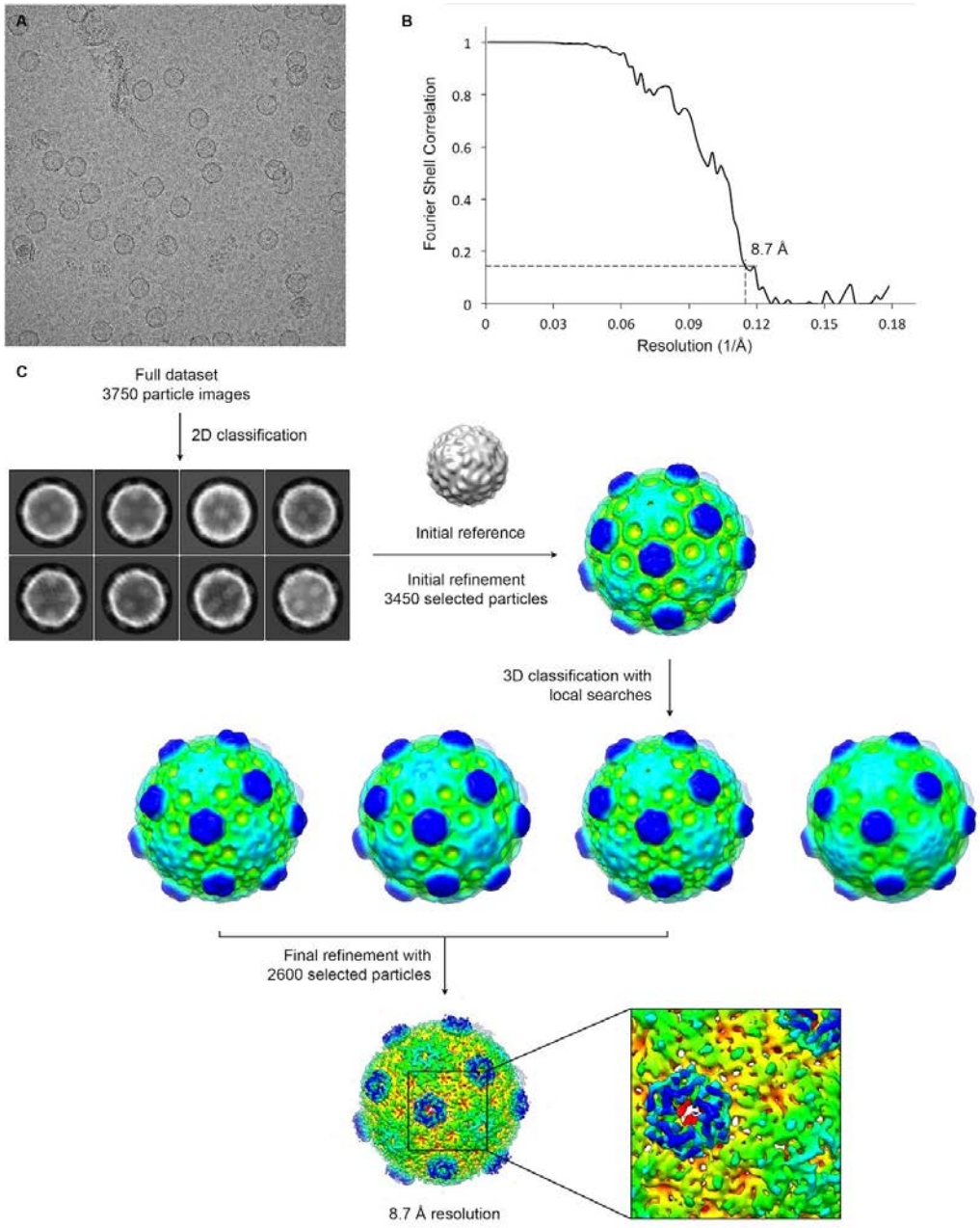

Fig. S9. EM structure determination overview.

(**A**) Sample electron micrograph. The *H. ochraceum* shells are clearly visible as near-spherical particles. (**B**) FSC curve of the final reconstruction of the shell, indicating a resolution of 8.7 Å according to the FSC = 0.143 criterion. (**C**) Data processing and classification strategy used for reconstruction of the *H. ochraceum* shell structure. The reconstructions are coloured using a gradient from red to blue by radius, revealing that the concave surface of the BMC-H subunits is exposed and highlighting the protruding BMC-T subunits (blue). The final reconstruction was sharpened for visualization using a b-factor of -1256, as determined by the post-processing function of RELION. See methods for details.

Table S1. X-ray crystallography data collection and refinement statistics.

X-ray crystallography data collection and refinement statistics for the *H. ochraceum* shell, BMC-T2 and BMC-T3 structures. Statistics for the highest-resolution shell are shown in parentheses.

	H. ochraceum shell	*H. ochraceum* BMC-T2	*H. ochraceum* BMC-T3
Data collection			
Space group	C 2 2 2_1	P 2_1 2_1 2	P 2_1 3
Unit cell dimensions			
a, *b*, *c* (Å)	394.3, 638.1, 642.2	68.1, 126.1, 139.8	110.6, 110.6, 110.6
Resolution (Å)	40-3.51 (3.69-3.51)	39-1.70 (1.79-1.70)	39-1.55 (1.63-1.55)
Unique Reflections	880,323 (135,274)	131,926 (18,813)	65,376 (9,459)
R_{merge}	1.32 (11.6)	0.122 (1.61)	0.103 (1.80)
R_{pim}	0.33 (2.9)	0.033 (0.43)	0.016 (0.27)
Wilson B (Å^2)	85	19.8	22.3
$CC_{1/2}$	0.94 (0.26)	0.999 (0.70)	1.000 (0.87)
I / σI	3.6 (0.4)	21.0 (2.0)	26.9 (2.6)
Completeness (%)	98.2 (92.7)	99.4 (97.9)	100.0 (100.0)
Redundancy	15.8 (13.9)	14.8 (14.8)	42.6 (44.0)
Refinement			
Resolution (Å)	40-3.51	39-1.70	39-1.55
Number of reflections	878,895	131,795	65,297
R_{work} / R_{free}	27.9 / 32.4	22.5 / 26.2	15.7 / 17.1
Number of atoms			
Protein	215,283	9,164	3,504
Ligand/ion	n/a	n/a	6
Water	n/a	670	386
B-factors			
Protein	102	26.3	23.5
Ligand/ion	n/a	n/a	34.4
Water	n/a	29.9	38.4
R.m.s. deviations			
Bond lengths (Å)	0.005	0.003	0.009
Bond angles (°)	0.86	0.57	1.09