Two new high-resolution crystal structures of carboxysome pentamer proteins reveal high structural conservation of CcmL orthologs among distantly related cyanobacterial species

Markus Sutter[a], Steven C. Wilson[b], Samuel Deutsch[a] and Cheryl A. Kerfeld[a,b,c,1]

[a]United States Department of Energy - Joint Genome Institute, Walnut Creek CA 94598, [b]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA and [c]Berkeley Synthetic Biology Institute, Berkeley, CA 94720, USA.

[1]Corresponding author. Phone: (925) 296-5691 E-mail: ckerfeld@lbl.gov

Abstract

Cyanobacteria have evolved a unique carbon fixation organelle known as the carboxysome that compartmentalizes the enzymes RuBisCO and carbonic anhydrase. This effectively increases the local CO_2 concentration at the active site of RuBisCO and decreases its relatively unproductive side reaction with oxygen. Carboxysomes consist of a protein shell composed of hexameric and pentameric proteins arranged in icosahedral symmetry. Facets composed of hexameric proteins are connected at the vertices by pentameric proteins. Structurally homologous pentamers and hexamers are also found in heterotrophic bacteria where they form architecturally related microcompartments such as the Eut and Pdu organelles for the metabolism of ethanolamine and propanediol, respectively. Here we describe two new high-resolution structures of the pentameric shell protein CcmL from the cyanobacteria *Thermosynechococcus elongatus* and *Gloeobacter violaceus* and provide detailed analysis of their characteristics and comparison with related shell proteins.

Keywords: Cyanobacteria, microcompartment, carboxysome, CcmL

Introduction

Bacterial microcompartments are the prokaryotic functional equivalent to eukaryotic organelles, by acting to separate a part of the cell from the cytosolic milieu . This allows for a more organized subcellular environment in which to concentrate metabolites, prevent side reactions or sequester toxic intermediates. The best-characterized bacterial microcompartment is the carboxysome which enhances CO_2 fixation by encapsulating two key enzymes into a cellular microcompartment. Carboxysomes contain carbonic anhydrase (CA) which can convert bicarbonate to CO_2 and ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) which catalyzes the reaction of ribulose-1,5-bisphosphate (RuBP) and CO_2 to form 3-phosphoglycerate (3PG), the key step of the Calvin Cycle. The encapsulation of the enzymes together elevates the local CO_2 concentration and minimizes the competing reaction of RuBisCO with oxygen because the shell is thought to provide a barrier to O_2. Cyanobacterial carboxysomes are divided into α- and β-types, based on the form of RuBisCO they contain . α-carboxysomes contain Form 1A RuBisCO and β-carboxysomes, found for example in *Synechococcus elongatus* 7942, contain Form 1B RuBisCO, the same form as observed in land plants. In carboxysomes, RuBisCO is packed into the interior of the carboxysome in a paracrystalline array .

The carboxysome shell, which could potentially have some self-assembling properties, typically has a diameter of 100-200 nm and is composed of several thousand copies of shell proteins organized with apparent icosahedral symmetry. Structural studies using x-ray crystallography have led to the determination of structures of several components of the carboxysome shell . The flat surfaces are thought to be formed by the hexameric proteins (named CcmK in β-carboxysomes and BMC domains (bacterial microcompartment protein; pfam00936 domain) in general) and the vertices by the pentameric proteins (CcmL in β-carboxysomes, CsoS4A/B in α-carboxysomes , Fig. 1a). While there have been numerous reports of structures of hexameric shell proteins, there is thus far only one structure of the CcmL pentamer from β-carboxysomes determined at 2.4 Å resolution from *Synechocystis* sp. PCC 6803 . The presence of pentamer proteins in all carboxysome operons and their high sequence conservation suggest a critical role in β-carboxysome function. CcmL mutants were characterized in early carboxysome studies and found to have a high CO_2 requiring phenotype and form elongated shapes consistent with the idea that they provide the caps of icosahedral particles.

In addition to carboxysome operons, the same combination of hexameric (pfam00936 domain) and pentameric (pfam03319 domain) proteins are found in other bacteria and have been characterized based on the encapsulated enzymes as ethanolamine utilization (Eut) and propanediol utilization (Pdu) compartments . The shell proteins from these microcompartments are homologous to their carboxysomal counterparts. It was therefore surprising that the structure for the CcmL homolog from Eut compartments (EutN) was found to form hexamers in the

crystals in two independent structural studies . However, recent biochemical characterization of EutN indicates that EutN is a pentamer in solution and the observed hexameric form of EutN is most likely an artifact of crystallization in both cases. The same paper also describes a pentamer form of a pfam03319 domain protein from a glycyl radical-enzyme-containing propanediol utilizing microcompartment.

Here we present two new crystal structures of CcmL proteins from the cyanobacteria *Thermosynechococcus elongatus* and *Gloeobacter violaceus.* Our results confirm their pentameric assembly as opposed to the two known hexameric structures from an *E. coli* microcompartment. Moreover, these structures are from very diverse organisms, with *G. violaceus* being one of the most ancient cyanobacteria and *T. elongatus* is one of the few thermophilic cyanobacteria with a sequenced genome. The high resolution (1.7 Å) crystal structure from *G. violaceus* provides an ideal basis for modeling studies.

Results

Two new high-resolution pentameric CcmL structures

We have determined the crystal structures of CcmL from *Gloeobacter violaceus* and *Thermosynechococcus elongatus* at resolutions of 1.7 and 2.0 Å, respectively, solved by molecular replacement. An atomic model was built into the electron density and resulted in R/R_{free} values of 16.1/19.1 % for *G. violaceus* and 24.5/29.3 % for *T. elongatus* CcmL. Detailed statistics concerning data collection and refinement are found in Table 1. Both structures contain a full pentamer in the asymmetric unit. The final model for *G. violaceus* CcmL contains 95-97 of the 100 residues, the C-terminal 3-5 residues were disordered as well as the C-terminal 6xhistidine tag which was used for purification. In the *T. elongatus* model several loop residues for two chains are disordered as well as the C-terminal 2-5 residues.

Both proteins have the expected OB (oligonucleotide/oligosaccharide binding) fold structure with a five stranded curved beta sheet forming a closed beta-barrel with a short helix located between strands four and five on the inside of the pentamer. The site where structurally related proteins have a oligonucleotide/oligosaccharide binding site is the protein-protein interface between two CcmL monomers. The C-terminal beta strands six and seven add to the beta sheet of the neighboring subunit on the outside (Fig. 1b). The five subunits arrange in a pentamer in the shape of a truncated pyramid (Fig 1b).

All of the structures have a central pore (Fig. 1a) with a similar minimum diameter of about 4 Å (Fig. 2a,b) which opens into a wide funnel on the base of the truncated pyramid. The *H. neopolitanus* and *S.* 6803 structures have an additional constriction towards the base (Fig. 2b). In both structures reported here we modeled a sulfate ion in the middle of the narrowest part of the pore since it is present in solution and

matches the electron density present at this location (Fig. 2c). Refinement with the sulfate in this position does not generate additional difference density, indicating that it is the correct ion for this position. When we refined occupancy of the sulfate it converged at 0.98 for the *G. violaceus* and 0.93 for the *T. elongatus* CcmL structure. In the CsoS4A structure a chloride ion is found in the middle of the pore, while in *S.* 6803 CcmL a glycerol molecule is found at this position. In all cases the coordinating residue is a serine hydroxyl group.

Structure and sequence comparison

The structural homology among the new and previously known structures is high (Supplementary Table S1) despite only modest sequence identity (33-63%) and the very distant phylogenetic relationship of the organisms within the cyanobacterial phylum. The individual subunits superimpose very well (Fig. 1c) and root mean square deviation (rmsd) values range from 0.75 and 1.6 Å between the homologs. Proteins from the carboxysomal microcompartments are, as expected, more closely related to each other than to the EutN protein from the functionally different Eut microcompartment.

Genes encoding the pentameric proteins of α-carboxysomes occur in pairs (CsoS4A and CsoS4B) and are typically shorter by about 10-20 amino acids (Fig. 1d); one of the loops between beta strands two and three is shorter (Fig. 1c, asterisk) and CsoS4A also lacks the two C-terminal beta strands (β5 and β6) which connect to the beta sheet of the neighboring subunit (Fig. 1c). Structurally, the CcmL proteins from β-cyanobacteria have an additional ~5 residues at the C-terminus that are moderately conserved (Fig. 1d, residues 95+); they are located at the periphery of the pentamer base.

We calculated the amino acid conservation of CcmL from 104 cyanobacterial species that contain β-carboxysomes and mapped it onto the structure of the *G. violaceus* CcmL (Fig. 3a, b). The conservation is very high for residues on the perimeter of the truncated pyramid (Fig. 3b, most are absolutely conserved residues). This pattern of conserved amino acids contains many charged and polar residues (Fig. 3b, open pentamers).

In the pore, residues around the smallest constriction are less conserved (79-97% for the residues marked in Fig. 3b) than some of the surface residues (Fig. 3b). The central residue S61, for example, is a serine, glycine or alanine in CcmL (β-carboxysomes), indicating that neither charge nor hydrophobicity matter at this position. The subsequent arginine residue (R65 in *G. violaceus*) however is very conserved among all CcmL homologs (all pfam03319 proteins) and is partly responsible for the positive charge of the pore. The residues at the narrow pore of other microcompartment CcmL orthologs are also not strictly conserved, however there is a high number of serine and glycine residues at this position (see also Supplementary Fig. S2).

Electrostatics

We calculated the electrostatic potential and mapped it on the surface of the available pfam03319 structures (Fig. 4). Note that top refers to the narrower side of the truncated pyramid and bottom to its base. The most striking feature that is conserved among all structures is the positive charge throughout the pore (Fig 3). In addition, some of the structures also have a negatively charged base, although it should be noted that most of the structures are missing several residues at the C-terminus which are located on the periphery of the base.

Discussion

The three structures of CcmL from diverse cyanobacterial species share a high degree of structural homology and confirm their pentameric assembly. Conserved residues are most likely important for forming an interface with other proteins, consistent with the idea of a side-to-side interaction with hexameric or pseudo-hexameric proteins to form icosahedral vertices (Fig. 1a, 2a). Several positions in CcmL are absolutely conserved polar or charged residues (T9, K15, K23, E53, D83 and T84 in *G. violaceus* CcmL) and are located on the perimeter around the middle of the truncated pyramid. The same residues are only partly conserved in CsoS4A of the α-carboxysome (Fig. 1d, Supplementary Fig. S2), indicating that those pentamers might interact differently with their hexameric shell protein counterparts. Another interesting region is the loop connecting beta strands 3 and 4 which is a conserved Gly-Ala-Gly in all of those organisms. The lack of bulky sidechains at this position could be important to provide a flat surface to interact with other shell proteins. β-carboxysomal CcmL proteins also have a short C-terminal extension relative to the pentameric proteins of other types of microcompartments (Fig. 1d); it contains a conserved arginine and could be involved in interactions with other shell proteins.

If CcmL only serves as a pentameric cap then there are only 12 pentamers required for assembly of an icosahedral carboxysome shell compared to several hundred hexamers. This, together with the narrow pore of about 4 Å diameter indicates that the CcmL pentamers are unlikely to play a significant role in the flow of metabolites across the shell. A purely structural role is consistent with the number of genes encoding CcmL in carboxysomes, e.g. β-carboxysomal operons contain only one copy of a gene encoding CcmL while there are always several paralogs of the hexameric CcmK proteins. α-carboxysomes typically have two copies of CsoS4 type genes, annotated as CsoS4A and CsoS4B; they are 40% identical in the model organism *H. neopolitanus*. However homologs to CcmL/CsoS4/B are often present in multiple copies in the genomes of heterotrophic bacteria that encode diverse BMC shell proteins. For example in the Ignavibacterium *Melioribacter roseus*, there are seven genes encoding the pfam03319 protein, all are very different (Supplementary

Fig. S1). This implies that in those organisms this protein could have a role different than merely forming vertices, for example as shell constituents important for the flow of different types of substrates or products across the shell.

Interestingly, CsoS4A/B does not seem to be necessary for determining the shape of carboxysomes in *H. neopolitanus* although some instances of elongated carboxysomes were observed in a Δ*csoS4A/B* mutant. However the strains containing carboxysomes lacking CsoS4A were not able to support growth under low CO_2 conditions, indicating that the shells were "leaky". It is possible for icosahedral particles to form without the pentamer vertex but instead having a hole in that position. These types of structures, called "whiffleballs" have been generated from virus shell particles *in vitro* by removing the pentameric proteins .

The tendency for a negatively charged molecule to be found in the smallest constriction of the pore (Fig. 2c) may be functionally relevant. It is impossible to have a residue on a symmetry axis. A way to nevertheless close the pore would be to bind a readily available ion (like sulfate, Fig. 2c) in the center of the pore to obstruct it. The residue coordinating the central ion is a serine in all available structures; while serine residues at the pore are very common they are not absolutely conserved. In the absence of the serine residue the protein backbone could fulfill the role of complexing of the ion in the center.

Recently, a model of for a double layered carboxysome shell was proposed . The hexameric subunit structure that this proposal is based on is from *T. elongatus*, the source of one of the CcmL pentamer structures described here. There is no evidence for double layering in the crystal packing of the *T. elongatus* pentamer or in any of the other cyanobacterial pentamer structures. Moreover, there has been no report in the literature of pentamers forming higher order oligomers in solution. This implies that if the shell is double layered it is likely single layered around the vertices.

While evidence so far supports the function of the CcmL orthologs as vertices of a microcompartment with icosahedral symmetry, it is still not clear how and with which other shell proteins it interacts. A crystal structure of shell protein complexes or of a full compartment will provide a definitive answer to that question.

Methods

Cloning, protein purification and crystallization

gvip287 (Accession: NP_925040) and tll0945 (Accession: NP_681735) were synthesized at the JGI DNA synthesis facility using standard synthesis approaches . gvip287 was synthesized with the native nucleotide sequence, whereas tll0945 was codon optimized for high expression in *E.coli* using GeneDesign with an empirically derived codon usage table of highly expressed heterologous genes . DNA sequence for tll0945 is provided in Supplementary Table S2. Synthetic genes were cloned into pENTR (Invitrogen, Carlsbad, CA) and sequence verified using Pacific Biosciences

single molecule sequencing technology. Synthesized genes were then amplified with primers (Supplementary Table S3) containing BglBrick restriction sites and a 5' ribosome binding site. The amplified products were digested with EcoRI and BamHI, ligated into a BglBrick shuttle vector, a C-terminal his tag insert was added and the construct was then sequenced and transferred to a BglBrick adapted pET21b vector for expression. The constructs were transformed into BL21 (DE3) cells. Protein expression was induced at an OD_{600} of 0.8 by addition of 0.5 mM IPTG at 37 °C and grown for four hours. Cells were harvested by centrifugation, resuspended in buffer A (20 mM Tris pH 8.0 at RT, 100 mM NaCl) and lysed using a French Press. *T. elongatus* CcmL was purified using a heat step (30 min 60°C) followed by centrifugation and a 30 ml SourceQ column using a NaCl gradient from 100 to 1000 mM NaCl followed by size exclusion chromatography on a HiLoad 26/60 Superdex 75 (GE Healthcare) column equilibrated in 20mM Tris pH 7.4 at RT, 50 mM NaCl for final cleanup. The protein was then concentrated to 5.5 mg/ml for crystallization. In attempt to crystallize another protein of interest from a solution which contained a residual amount of *G. violaceus* CcmL, small crystals were obtained in a condition of 0.2 M Lithium sulfate, 0.1 M Bis-Tris pH 5.0, and 10 % (w/v) PEG 3350. *T. elongatus* CcmL crystallized in a sitting drop condition of 1.4 M MgSO4 and 0.1 M MES pH 6.4 with equal volume of reservoir to protein solution. Crystals were stabilized by adding an 80 % glycerol solution to the drop for a final concentration of 30 % glycerol, frozen in liquid nitrogen and measured at beam lines 5.0.1 and 5.0.2 of the Advanced Light Source (ALS) at Lawrence Berkeley National Lab.

Structure determination and protein sequence and structure analysis

The structure was solved using molecular replacement (Phaser in Phenix) with PDB ID 2QW7. An initial model was built into the density using phenix.autosol which was then manually corrected in Coot in combination with refinement runs using phenix.refine . Structural alignment using secondary structure matching followed by iterative alignment of protein C-α backbone atoms was performed with superpose in CCP4 . The sequence logo was generated at http://weblogo.berkeley.edu/logo.cgi from an alignment of the 104 known CcmL sequences. Alignments were performed with ClustalX . Electrostatics calculations were done with the APBS plugin for PyMOL . Structure figures were prepared with PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.3 Schrödinger, LLC.)

Protein Data Bank Accession Codes: PDB codes are 4JW0 for *G. violaceus* CcmL and 4JVZ for *T. elongatus* CcmL.

Acknowledgements

We would like to thank the entire staff at the Advanced Light Source, Lawrence Berkeley National Laboratory, which is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract

No. DE-AC02-05CH11231. CAK and SCW are supported by the NSF (EF1105897). MS was supported by a Swiss National Science Foundation Postdoctoral Fellowship.

References

Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD (2012) Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr D 68:352-367. doi:Doi 10.1107/S0907444912001308

Anderson JC, Dueber JE, Leguia M, Wu GC, Goler JA, Arkin AP, Keasling JD (2010) BglBricks: A flexible standard for biological part assembly. Journal of biological engineering 4 (1):1

Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. Proceedings of the National Academy of Sciences of the United States of America 98 (18):10037-10041. doi:10.1073/pnas.181342398

Bobik TA, Havemann GD, Busch RJ, Williams DS, Aldrich HC (1999) The propanediol utilization (pdu) operon of *Salmonella enterica* serovar Typhimurium LT2 includes genes necessary for formation of polyhedral organelles involved in coenzyme B_{12}-dependent 1,2-propanediol degradation. Journal of bacteriology 181 (19):5967-5975

Cai F, Menon BB, Cannon GC, Curry KJ, Shively JM, Heinhorst S (2009) The pentameric vertex proteins are necessary for the icosahedral carboxysome shell to function as a CO2 leakage barrier. PloS one 4 (10):e7521

Cai F, Sutter M, Cameron JC, Stanley DN, Kinney JN, Kerfeld CA (2013) The structure of CcmP, a tandem bacterial microcompartment domain protein from the beta-carboxysome, forms a subcompartment within a microcompartment. The Journal of biological chemistry 288 (22):16055-16063. doi:10.1074/jbc.M113.456897

Cot SS, So AK, Espie GS (2008) A multiprotein bicarbonate dehydration complex essential to carboxysome function in cyanobacteria. Journal of bacteriology 190 (3):936-945

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta crystallographica Section D, Biological crystallography 60 (Pt 12 Pt 1):2126-2132. doi:10.1107/S0907444904019158

Forouhar F, Kuzin A, Seetharaman J, Lee I, Zhou W, Abashidze M, Chen Y, Yong W, Janjua H, Fang Y, Wang D, Cunningham K, Xiao R, Acton TB, Pichersky E, Klessig DF, Porter CW, Montelione GT, Tong L (2007) Functional insights from structural genomics. Journal of structural and functional genomics 8 (2-3):37-44. doi:10.1007/s10969-007-9018-3

Hess WR, Rocap G, Ting CS, Larimer F, Stilwagen S, Lamerdin J, Chisholm SW (2001) The photosynthetic apparatus of Prochlorococcus: Insights through comparative genomics. Photosynthesis research 70 (1):53-71. doi:10.1023/A:1013835924610

Iancu CV, Morris DM, Dou Z, Heinhorst S, Cannon GC, Jensen GJ (2010) Organization, structure, and assembly of alpha-carboxysomes determined by electron cryotomography of intact cells. Journal of molecular biology 396 (1):105-117

Kerfeld CA, Heinhorst S, Cannon GC (2010) Bacterial microcompartments. Annu Rev Microbiol 64:391-408

Kerfeld CA, Sawaya MR, Tanaka S, Nguyen CV, Phillips M, Beeby M, Yeates TO (2005) Protein structures forming the shell of primitive bacterial organelles. Science (New York, NY 309 (5736):936-938

Klein MG, Zwart P, Bagby SC, Cai F, Chisholm SW, Heinhorst S, Cannon GC, Kerfeld CA (2009) Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport. Journal of molecular biology 392 (2):319-333

Kofoid E, Rappleye C, Stojiljkovic I, Roth J (1999) The 17-gene ethanolamine (eut) operon of *Salmonella typhimurium* encodes five homologues of carboxysome shell proteins. Journal of bacteriology 181 (17):5317-5329

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23 (21):2947-2948. doi:10.1093/bioinformatics/btm404

Li Y, Conway JF, Cheng N, Steven AC, Hendrix RW, Duda RL (2005) Control of virus assembly: HK97 "Whiffleball" mutant capsids without pentons. Journal of molecular biology 348 (1):167-182. doi:10.1016/j.jmb.2005.02.045

Long BM, Badger MR, Whitney SM, Price GD (2007) Analysis of carboxysomes from *Synechococcus* PCC7942 reveals multiple Rubisco complexes with carboxysomal proteins CcmM and CcaA. The Journal of biological chemistry 282 (40):29323-29335

Price GD, Badger MR (1989) Isolation and Characterization of High CO(2)-Requiring-Mutants of the Cyanobacterium Synechococcus PCC7942 : Two Phenotypes that Accumulate Inorganic Carbon but Are Apparently Unable to Generate CO(2) within the Carboxysome. Plant physiology 91 (2):514-525

Price GD, Howitt SM, Harrison K, Badger MR (1993) Analysis of a genomic DNA region from the cyanobacterium Synechococcus sp. strain PCC7942 involved in carboxysome assembly and function. Journal of bacteriology 175 (10):2871-2879

Quan J, Saaem I, Tang N, Ma S, Negre N, Gong H, White KP, Tian J (2011) Parallel on-chip gene synthesis and application to optimization of protein expression. Nature biotechnology 29 (5):449-452. doi:10.1038/nbt.1847

Richardson SM, Wheelan SJ, Yarrington RM, Boeke JD (2006) GeneDesign: rapid, automated design of multikilobase synthetic genes. Genome research 16 (4):550-556. doi:10.1101/gr.4431306

Samborska B, Kimber MS (2012) A dodecameric CcmK2 structure suggests beta-carboxysomal shell facets have a double-layered organization. Structure 20 (8):1353-1362. doi:10.1016/j.str.2012.05.013

Schneider TD, Stephens RM (1990) Sequence logos: a new way to display consensus sequences. Nucleic acids research 18 (20):6097-6100

Smart OS, Goodfellow JM, Wallace BA (1993) The pore dimensions of gramicidin A. Biophysical journal 65 (6):2455-2460. doi:10.1016/S0006-3495(93)81293-1

Smith HO, Hutchison CA, 3rd, Pfannkoch C, Venter JC (2003) Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. Proceedings of the National Academy of Sciences of the United States of America 100 (26):15440-15445. doi:10.1073/pnas.2237126100

Tanaka S, Kerfeld CA, Sawaya MR, Cai F, Heinhorst S, Cannon GC, Yeates TO (2008) Atomic-level models of the bacterial carboxysome shell. Science (New York, NY 319 (5866):1083-1086

Tsai Y, Sawaya MR, Cannon GC, Cai F, Williams EB, Heinhorst S, Kerfeld CA, Yeates TO (2007) Structural analysis of CsoS1A and the protein shell of the *Halothiobacillus neapolitanus* carboxysome. PLoS Biol 5 (6):e144

Wheatley NM, Gidaniyan SD, Liu Y, Cascio D, Yeates TO (2013) Bacterial Microcompartment Shells of Diverse Functional Types Possess Pentameric Vertex Proteins. Protein Science. doi:10.1002/pro.2246

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS (2011) Overview of the CCP4 suite and current developments. Acta crystallographica Section D, Biological crystallography 67 (Pt 4):235-242. doi:10.1107/S0907444910045749

Yeates TO, Kerfeld CA, Heinhorst S, Cannon GC, Shively JM (2008) Protein-based organelles in bacteria: carboxysomes and related microcompartments. Nat Rev Microbiol 6 (9):681-691

Figures

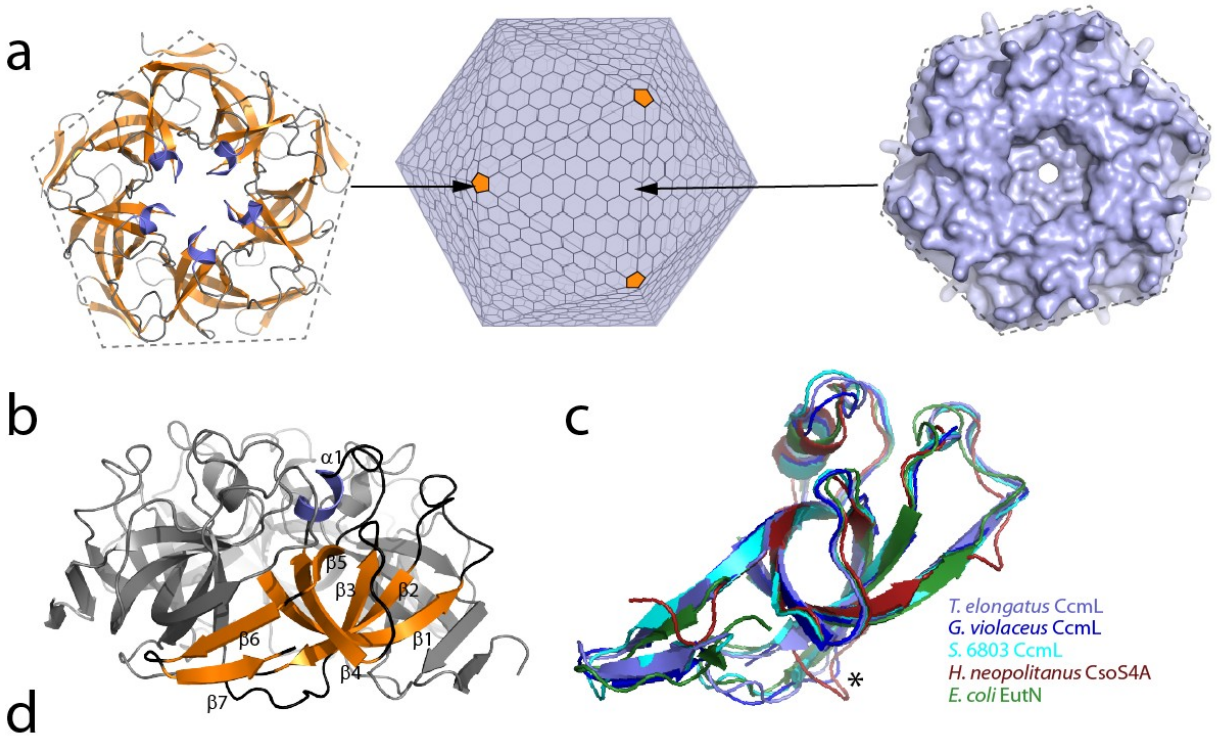


Fig. 1a Icosahedral model of the carboxysome consisting of faces made up by hexameric proteins and pentameric CcmL proteins at the vertices **b** Structure of *G. violaceus* CcmL in cartoon representation; one monomer is colored by secondary structure (orange: beta sheet; slate blue: helix; black: loops) **c** Alignment of the monomers of all available pfam03319 structures. The asterisk marks the shorter CsoS4A loop **d** Amino acid sequence alignment of the structures shown in c, colored according to amino acid properties (green: polar, cyan: hydrophobic, red: basic, magenta: acidic, light brown: glycine, yellow: proline)

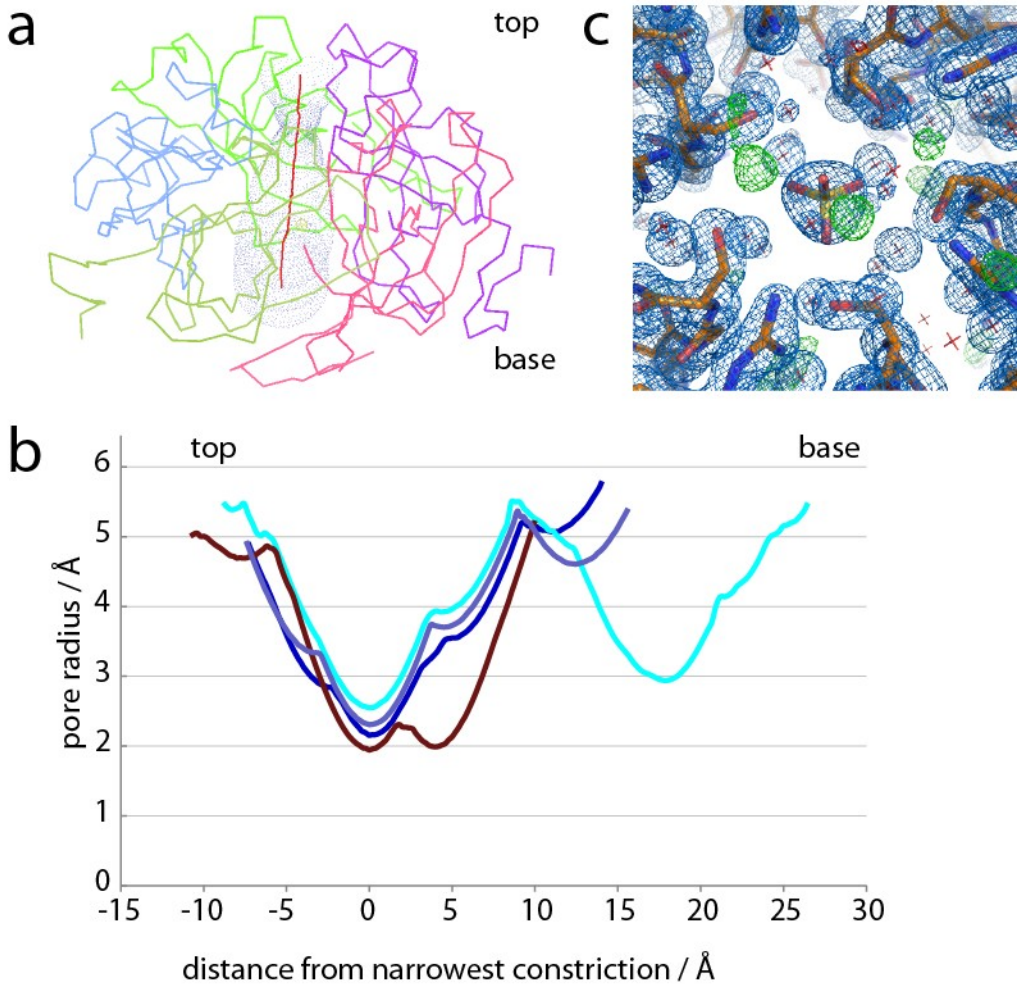


Fig. 2a Visualization of the *G. violaceus* pore using the HOLE implementation in Coot , CcmL chains are shown as C-alpha ribbons and the path along the pore as a red line **b** Graph of the pore radius tracked (as in a) from top to base for the *G. violaceus* (blue), *T. elongatus* (slate blue) *S*. 6803 (cyan) and *H. neopolitanus* (red) structures aligned for the smallest constriction at 0 **c** Electron density and model at the central pore of *G. violaceus* CcmL showing a sulfate molecule as well as several ordered water molecules (crosses). The $2F_o$-F_c density is shown as a blue mesh at 1.8σ and the F_o-F_c density at -4σ (red) and +4σ (green).

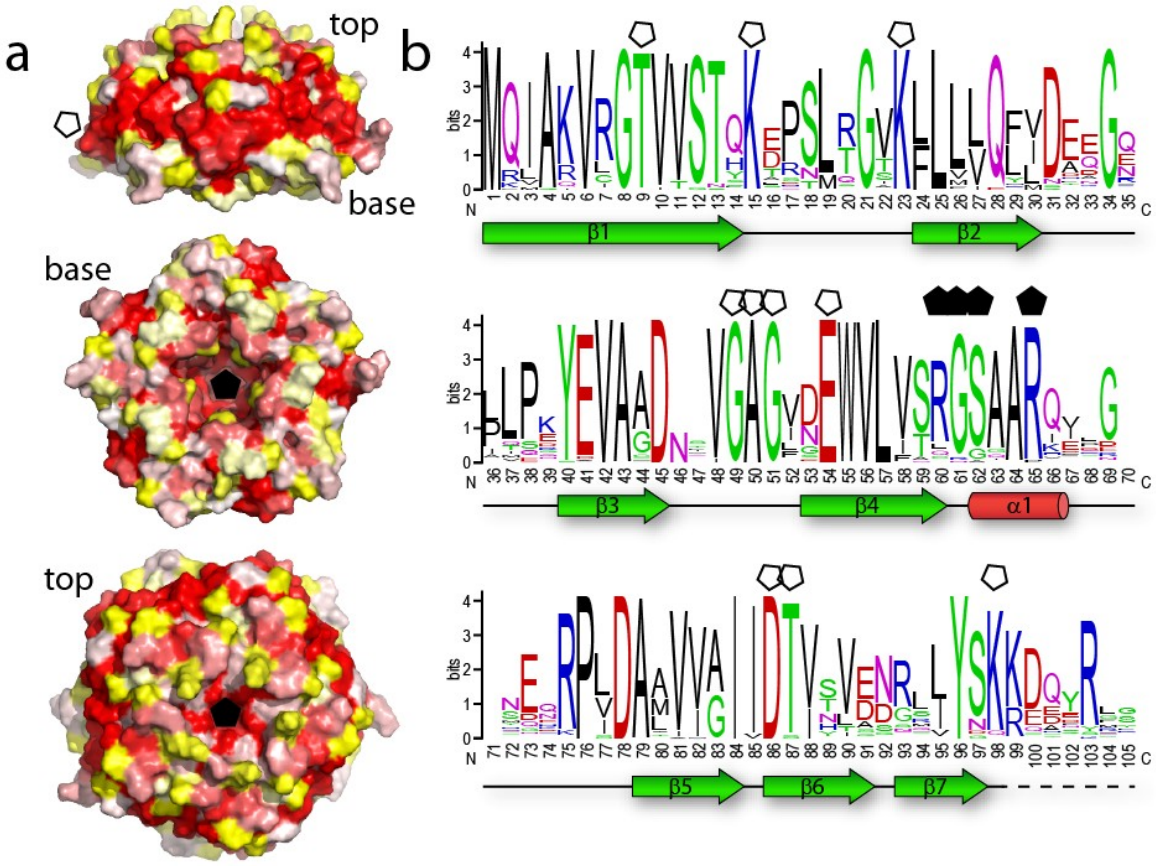


Fig. 3a Mapping of amino acid conservation onto the CcmL surface. (red=high, white=medium, yellow=low) **b** Primary structure conservation logo generated from an alignment of all available CcmL protein sequences with secondary structure denoted as green arrows (beta sheets) and a red tube (helix). The open pentagons refer to conserved residues arranged around the perimeter of the pentamer and the closed pentagons denote conserved residues of the central pore

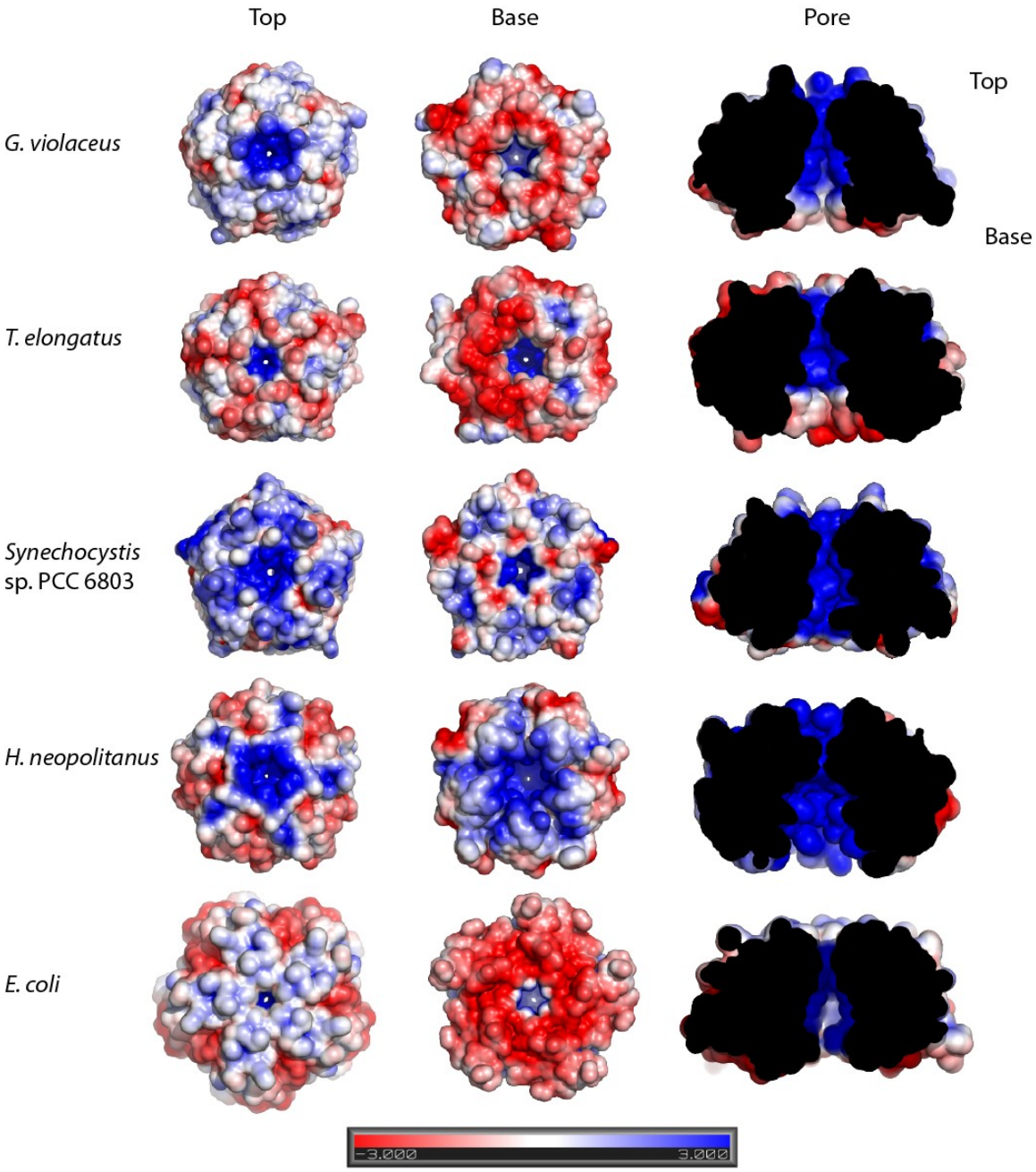
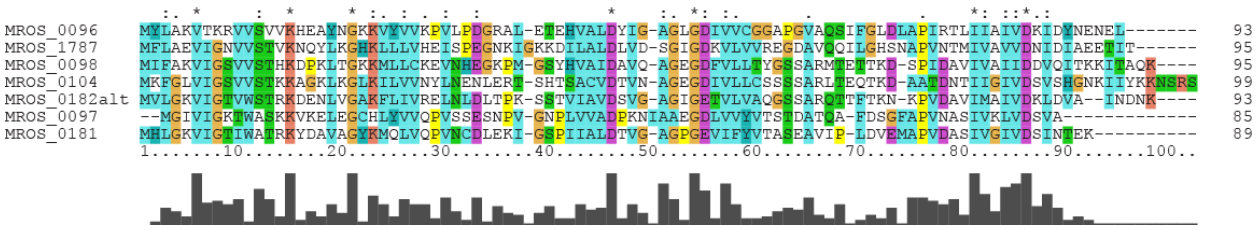


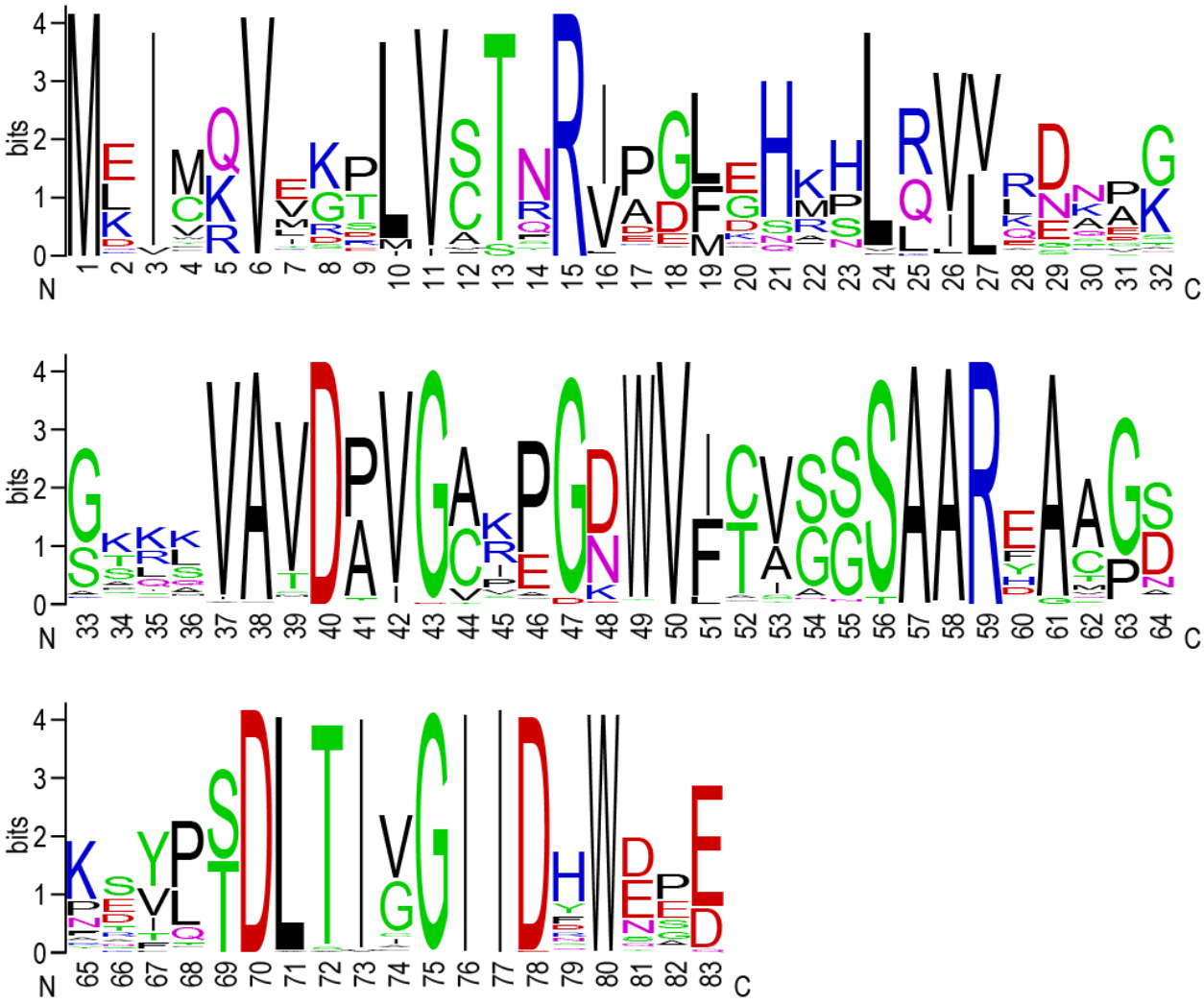
Fig. 4 Electrostatic representation of CcmL, CsoS4A and a EutN model pentamer in top, base and side view, sliced through the central pore. The colors represent potential values, with red at -3, white at 0 and blue at 3 kT/e. The EutN pentamer structure was modeled by aligning the monomers to the most similar structure (CsoS1A from *H. neopolitanus*)

Supplementary Table S1 Primary and tertiary structure alignment. The top line indicates rmsd values from structural alignment of C-alpha atoms in Å with the number of aligned residues in parentheses and the bottom line the amino acid identity / conserved substitution between two proteins.

	CcmL_Gvip	CcmL_Tll	CcmL_S6803	CsoS4A_Hneo	EutN_Ecoli
CcmL_Gvip	-	0.84 (91) 63 / 74%	0.76 (93) 59 / 78%	1.33 (75) 41 / 59%	1.60 (87) 40 / 56%
CcmL_Tll		-	0.96 (91) 56 / 70 %	1.44 (76) 31 / 44%	1.50 (92) 33 / 59%
CcmL_S6803			-	1.34 (75) 37 / 56%	1.60 (93) 39 / 58%
CsoS4A_Hneo				-	1.43 (76) 39 / 54%
EutN_Ecoli					-



Supplementary Fig. S1 Sequence alignment of the seven *Melioribacter roseus* BMV homologs, colored according to amino acid properties (green: polar, cyan: hydrophobic, red: basic, magenta: acidic, light brown: glycine, yellow: proline)



Supplementary Fig. S2 Sequence conservation logo generated from an alignment of CsoS1A and CsoS1B proteins

Supplementary Table S2 DNA Sequence for *T. elongatus* CcmL

ATGAAAATTGCCCGTGTGTGTGGTACCGTGACCAGTACCCAAAAAGAAGATACCCTGACC
GGTGTGAAGTTTCTGGTGCTGCAATACCTGGGTGAAGATGGTGAATTTCTGCCAGATTAC
GAAGTTGCGGCGGACACCGTTGGTGCCGGTCAAGATGAATGGGTGCTGGTGAGTCGCGGT
AGTGCCGCCCGTCACATTATCAATGGCACCGATAAACCAATTGATGCCGCCGTGGTGGCC
ATTATTGATACCGTTAGTCGTGATAATTACCTGCTGTATAGTAAACGTACCCAGTAC

Supplementary Table S3 Primers used for expression vector cloning
gvip287_ccmL_f_rbs

ATAGAATTCATGAGATCTTTTAAGAAGGAGATATACCATGCAGATTGGCAGAGTCCG
gvip287_ccmL_r TTAGGATCCCTAGCGGTACTGGTCCTTTT
tll_ccmL_f_rbs

ATAGAATTCATGAGATCTTTTAAGAAGGAGATATACCATGAAAATTGCCCGTGTGT

tll_ccmL_rs	TTAGGATCCTTAGTACTGGGTACGTTTACTATAC
C_term_his_f	AATTCAGATCTCATCATCATCATCATCATTAAG
C_term_his_r	GATCCTTAATGATGATGATGATGATGAGATCTG