The Carboxysome and other Bacterial Microcompartments

These metabolic modules appear to be adaptive structures for microbes and microbial communities

Cheryl A. Kerfeld, [1,2] William B. Greenleaf[1] and James N. Kinney[1]

Cheryl A. Kerfeld is head of the Structural Genomics and Education Programs at the U.S. Department of Energy-Joint Genome Institute, Walnut Creek, Calif., and an Associate Professor in the Department of Plant and Microbial Biology, University of California, Berkeley; William B. Greenleaf is a Postdoctoral Fellow at the U.S. Department of Energy-Joint Genome Institute, Walnut Creek, Calif.; and James N. Kinney is a Research Associate at the U.S. Department of Energy-Joint Genome Institute, Walnut Creek, Calif.

Summary
● Carboxysomes are part of the carbon concentrating mechanism in cyanobacteria and chemoautotrophs.
● Carboxysomes are a subclass of bacterial microcompartments (BMCs); BMCs can encapsulate a range of metabolic processes.
● Like some viral particles, the carboxysome can be modeled as an icosahedron—in its case, having 4,000-5,000 hexameric shell subunits and 12 surface pentamers to generate curvature.
● The threefold axis of symmetry of the CsoS1D protein in carboxysomes forms a pore that can open and close, allowing for selective diffusion.
● Genetic modules encoding BMC shell proteins and the enzymes that they encapsulate are horizontally transferable, suggesting they enable bacteria to adapt to diverse environments.

Some biologists attribute the evolutionary success of multicellular eukaryotes to the organization and division of tasks within cells in organelles and among cells through differentiation. Yet, this strategy is not reserved solely for eukaryotes. Analyses of microbial metagenomes and metatranscriptomes are revealing comparable divisions of labor among members of these communities. Moreover, individual bacterial cells are much more highly organized than was thought. Examples include membrane-bound compartments such as the magnetosomes of magnetotactic bacteria and the anammoxosomes of the Planctomycetes.
         More than 40 years ago, another type of intracellular compartment was observed in cyanobacteria (Fig. 1). However, because of their resemblance to phage particles, investigators were at first skeptical as to whether these compartments were genuine functional units within bacterial cells. That skepticism began to fade when Jessup Shively at Clemson University and his collaborators purified these inclusions from a chemoautotrophic bacterium and demonstrated that they were packed, not with phage DNA, but with the enzyme, ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO). The Clemson researchers named the intracellular bodies carboxysomes. **[ok?**

them or someone else?ok] We now know that carboxysomes play a significant role in global carbon cycling.

Role of Carboxysomes

Carboxysomes, organelles that fix carbon dioxide (CO_2), are part of the carbon dioxide concentrating mechanism (CCM) found among cyanobacteria and many chemoautotrophs. The CCM involves multiple integrated cellular processes (Fig. 2), worked out by the collective efforts of several groups including that of Aaron Kaplan at the Hebrew Univeristy, Gordon Cannon, Sabine Heinhorst,at the University of Southern Mississippi, Kathleen Scott at the University of South Florida and Murray Badger and Dean Price at the Australian National University.

The CCM involves several types of inorganic carbon transporters in the cell membrane that increase cytosolic concentrations of bicarbonate. The intracellular bicarbonate then diffuses into the carboxysomes, where a carbonic anhydrase converts it into CO_2, the substrate for RuBisCO. RuBisCO combines CO_2 with the 5-carbon sugar, ribulose-1,5-bisphosphate (RuBP), to form two molecules of 3-phosphoglycerate (3PGA). The latter molecules leave the carboxysome for the cytoplasm where they enter the Calvin cycle, yielding metabolic energy and producing molecular building blocks for cell metabolism.

Unlike eukaryotic organelles, carboxysomes are composed entirely of protein. By purifying carboxysomes researchers identified those proteins, which enabled them to determine the gene sequences of the major shell proteins, now known as the bacterial microcompartment (BMC) domain. These efforts led to the discovery of carboxysome-like structures encapsulating other microbial metabolic processes.

Several Types of Bacterial Microcompartments

In some heterotrophic organisms, such as *Salmonella,* genes encoding the BMC domain are found in a operon that also contains genes encoding the enzymes involved in B12-dependent metabolism of 1,2-propanediol (1,2-PD). Growing *Salmonella* on 1,2-PD induced expression of the Pdu (for propanediol utilization) operon. Transmission electron microscopy (TEM) soon confirmed that polyhedral bodies were formed. A portion of the metabolic pathway for another B12-dependent process, ethanolamine degradation (encoded by the ethanolamine utilization, or Eut, operon), is also carried out in carboxysome-like structures (Fig 1C).

Collectively, carboxysomes and these other structures are called bacterial microcompartments. From genetic and biochemical studies on carboxysomes and the Pdu and Eut BMCs, some common themes emerge. Encapsulation in a BMC enhances metabolic function through several interdependent ways: by increasing the local concentration of enzymes and intermediates, and by organizing the enzymes to facilitate substrate transfer between sequential reactions. Many of the enzymes localized in BMCs are oxygen sensitive and catalyze reactions that either produce or consume volatile intermediates.

BMCs function as metabolic modules. Indeed, modularity is apparent across the various levels of BMC structure and function. As semi-autonomous functional modules,

BMCs are integrated with other components of cellular metabolism, including transporters, which supply raw materials, and cytoplasmic enzymes that consume the products of reactions taking place within BMCs.

BMCs also can be viewed as genetic modules. Genes encoding BMCs and associated enzymes typically are grouped in clusters. On a structural level, these gene products can be considered as a set of protein domains, modules that self-assemble to form BMCs. In micrographs, BMCs have striking profiles, with relatively sharp corners and straight edges, both of which are uncommon among biological structures and hint at an unusual level of geometric organization. These features make it particularly promising to investigate BMC function with structural approaches, particularly in the case of shell proteins, in which structure and function are inextricable.

Insights into the Structure of the BMC Shell

In 2004, my collaborator Todd Yeates and I (CAK), then at the University of California, Los Angeles (UCLA), began structural studies of the carboxysome shell proteins of the cyanobacterium *Synechocystis* PCC 6803 (Fig. 1A). We already knew that *Synechocystis* contains four small proteins (each containing about 100 amino acids), CcmK1-CcmK4, each containing a BMC domain, as well as another gene encoding a protein with about 230 amino acids that appeared to include a fusion of two BMC domains.

Those single BMC-domain proteins appear to be the major building blocks of the BMC shell. We overexpressed three of them and tried crystallizing them. Those efforts yielded structures of two of the paralogs, CcmK2 and CcmK4. The fold of the proteins (Fig 3) was the same alpha/beta structure that is observed in numerous unrelated proteins, including ferredoxin.

The six wedge-shaped monomers of the BMC domain are arranged as cyclic hexamers, with distinctive surfaces and relatively straight edges (Figs 3 and 4). **[au: OK?]** The hexamers are arranged in molecular layers suggestive of how the facets of the carboxysome shell might be built. Viewing the BMC proteins as equilateral hexagons, one can readily imagine how they can be tiled in two dimensions to cover a plane without leaving gaps. Other carboxysome and Eut BMC shell proteins have the same quaternary state and crystal packing, substantiating this model for how the BMC shell is built.

This hexameric shape, like the ultrastructure of carboxysomes in micrographs, evokes parallels to key principles of virus capsid assembly. Many virus capsids have icosahedral symmetry, in which a hexameric layer is closed by inserting 12 pentamers (soccer balls are also assembled this way). Thus, we began to search for a protein encoded within bacterial microcompartment gene clusters that forms a pentamer. We determined the crystal structures of two small proteins, CcmL and CsoS4A, from the two types of carboxysome and learned that both form pentamers (Fig 3).

Based on the size of a carboxysome, we estimate that it resembles an icosahedron, having 4,000-5,000 hexameric shell subunits with 60 copies of the pentamer proteins—forming 12 pentamers—inserted to generate curvature (Fig. 3). Pentamers effectively concentrate the curvature. When spaced equidistantly into a layer of hexamers, they can form an icosahedral structure. If pentamers were distributed less evenly, microcompartments would have less regular shapes.

Transport Across the BMC Shell

Like the lipid bilayer of membrane bound-organelles, the bacterial microcompartment shell has conduits through which metabolites selectively pass. In the case of the carboxysome, this is a diffusion process . For example, the relatively high cytoplasmic concentration of bicarbonate (HCO_3^-) established by transporters in cyanobacteria results in a gradient across the carboxysome shell. Once inside the carboxysome, the bicarbonate is quickly converted to CO_2 and consumed by RuBisCO, thereby maintaining that gradient.

In our model of the carboxysome shell, the sixfold axis of symmetry of each hexamer is a compelling pore candidate. While this opening might be important only for structure, several characteristics of the symmetry axis suggest it also is functional. First, there was its size. In hexameric protein oligomers in which the symmetry axis is not functionally important (e.g. glutamine synthetase), the size of the central opening is approximately the same size as the constituent monomers. One can also visualize this by arranging six pennies in a circle; the central vacancy is the same size as a single penny. In contrast, the pronounced wedge shape of the BMC domain results in hexamers with a diameter of about 70 Å and a central opening of only 4 and 7 Å in CcmK4 and CcmK2, respectively (Fig. 4).

Second, the residues that converge at the sixfold axis of symmetry are positively charged. Bringing this many like charges together requires overcoming considerable electrostatic repulsion. Moreover, these residues are absolutely conserved among paralogs. This conservation further implicates the symmetry axis—the putative pore—as functionally relevant because the main metabolites that cross the carboxysome shell are negatively charged.

More Complex Building Blocks: Tandem BMC Domain Proteins

This set of building blocks, single BMC-domain proteins, appears sufficient to build a carboxysome shell that contains perforations large enough to allow passage of bicarbonate. However, this model has a big problem. How could molecules the size of ribulose-1,5-bisphosphate pass through such small pores? Solving this problem was central to understanding how a protein shell provides a selective permeability barrier for ions of different sizes. Determining the structure of a tandem BMC shell protein resolved this paradox for us.

Genes encoding a fusion of two BMC domains are found in the carboxysome, Eut, and Pdu operons. The first crystal structure of one of these, CsoS1D from the marine cyanobacterium *Prochlorococcus*, reveals that trimers of tandem-BMC domain proteins mimic the shape of single BMC domain hexamers (Fig. 3 and 4). Further, they readily fit into our models of the shell.

More importantly, the threefold axis of symmetry of CsoS1D forms a pore of 14Å that is either open or closed in different crystal forms (Fig. 4). Once again, the net charge of the pore is positive due to conserved residues. Moreover, in CsoS1D these side chains control whether the pore is open or closed.

Soon after we published the tandem BMC protein structure from the carboxysome, other research groups reported structures for EutL and EtuB from the BMCs of heterotrophic organisms. These structures contain additional surprises. For instance, instead of a single pore at the symmetry axis, there are three smaller pores, each about 2 Å, between the fused domains. Unlike pores formed at symmetry axes, these pores are asymmetric. Later, Yeates and his group at UCLA determined a second EutL structure in which the trimer contains a single relatively large pore of 11Å at the threefold axis, substantiating our suggestion that some tandem BMC domain proteins have two conformations, corresponding to an open and closed form.

Fusing Domains Provides Structural Versatility

Compounding two BMC domains into one protein increases its potential functional versatility. Moreover, with a sequence identity between the two BMC domains in CsoS1D of only 19%, these two domains perhaps serve distinct functions within the structural module.

In addition to their potential for gated transport, the pseudohexamers formed by tandem BMC proteins also have special characteristics that are likely important for their assembly. In hexamers formed by the same protein, each edge of the hexamer is the same; in contrast, the pseudohexamers have two distinct edges. Subtle differences in edge architecture could govern interactions between hexamers. The potential for differential assembly, coupled with the potential for gating and for forming asymmetric pores, suggests the tandem domain proteins are a relatively important building block. From our survey of the different kinds of BMCs within bacterial genomes, we suspect that most BMC gene clusters will contain at least one tandem BMC domain protein.

While the set of BMC building block structures continues to expand, genetic and biochemical studies continue to challenge simplistic models describing how the shell is organized. Consider the pentameric proteins that are needed to concentrate the curvature in a closed shell of hexamers. A homolog to one of these, EutN, forms hexamers. Moreover, the pentamer proteins can be deleted from the carboxysome operon, and yet carboxysomes of typical shape still form, according to Cannon, Heinhorst, and their collaborators at USM.

These observations suggest that there is a redundancy or pliancy in the set of protein domains that make up the BMC shell. Again, drawing an analogy to viral capsid architecture provides potentially applicable insights. Virus capsids can be assembled from a few proteins that display quasi-equivalence; a single protein may exist in different interchangeable quaternary states—for example, as either a pentamer or hexamer. Comparisons to viruses prompt the question of the evolutionary relatedness between the BMC shell and viral capsids. However, to date, no sequence or structural homology has been found.

We still cannot reconcile another observation with our current understanding of BMC architecture. Some of the hexamers and pseudohexamers tend to interact across their surfaces, as opposed to interacting between edges. For example, CsoS1D forms tight dimers of trimers that resemble proteasomes. Are these incorporated into the shell of the BMC? Or does this arrangement reflect how two carboxysomes physically interact?

Little is known about how BMCs assemble, although recent tomographic evidence suggests that the shell and the interior enzymes assemble concomitantly. Likewise, we do not know how dynamic the carboxysome shell is. An appealing hypothesis is that the composition of the shell changes in response to environmental conditions. Since different shell proteins contain pores of different sizes and possibly different selectivities, their variability would provide one way to alter the flux across the carboxysome shell under different growth conditions. Our analysis of expression data supports this hypothesis.

The Diversity of Bacterial Microcompartments

While the discovery of the first BMCs depended on TEM, we can now rapidly identify organisms with potential BMCs by searching for genes encoding components of the bacterial microcompartment shell. As of December 2009, approximately 20% of microorganisms with genomic sequence data available contain open reading frames (ORFs) for which the presumptive gene products are components of BMC shells, according to our analysis. By examining nearby ORFs (Table 1), it may prove possible to infer what types of metabolic processes those putative BMCs encapsulate.

What is striking is the diversity of phyla and organisms that encode potential BMCs. For example, the Eut BMC operon is found in the *Firmicutes, Fusobacteria, Synergistetes,* and *Gammaproteobacteria*. However, closely related organisms sometimes differ in their number of BMC gene clusters. For example, *Klebsiella pneumoniae* (342) contains three (Eut, Pdu, and one encoding the B12 independent diol dehydratase), whereas *K. pneumoniae* MGH78578 contains only two such operons (Pdu and Eut). Meanwhile, a subset of Mycobacteria contains a BMC gene cluster that appears to have been inserted en bloc and may be involved in converting glutamate to ornithine.

Collectively these observations suggest that genetic modules encoding BMC shell proteins and the enzymes that they encapsulate are horizontally transferred with some frequency. Because these genes potentially encode semi-autonomous functional units, they exemplify a "selfish operon," a concept that John Roth and Jeffrey Lawrence developed in 1996. They also said that newly integrated genes are more likely to persist in genomes if they introduce a new function. The persistence of such operons is predicated on connecting to other functional units, providing the potential for innovation under the pressure of environmental selection.

Sequencing the genomes of organisms from diverse phyla, such as those in the *Genomic Encyclopedia of Bacteria and Archaea* project (GEBA; Table 1) and metagenomic analyses continue to add variations on the BMC theme. Likewise, metatranscriptomic data is revealing their specific roles in microbial communities. For example, various types of BMCs appear to be contributing to the collective metabolic activity in the cow rumen, based on data that we are analyzing with Matthias Hess and Eddy Rubin at the Joint Genome Institute in Walnut Creek, Calif.

Through our growing understanding of BMCs in bacterial isolates and in microbial communities, BMCs are emerging as important functional building blocks of the complex adaptive activities of microbes and microbial communities. Future inquiries

into the structure and function of BMCs should provide insights into how such metabolic modules enable bacteria to adapt to diverse environments.

SUGGESTED READING

Havemann, G. D., and T. A. Bobik. 2003. Protein content of polyhedral organelles involved in coenzyme B12-dependent degradation of 1,2-propanediol in *Salmonella enterica* serovar Typhimurium LT2. J. Bacteriol. **185:**5086-5095.

Iancu, C. V., D. M. Morris, Z. Dou, S. Heinhorst, G. C. Cannon, and G. J. Jensen. Organization, structure, and assembly of [alpha]-carboxysomes determined by electron cryotomography of intact cells. J. Mol. Biol., in press.

Kerfeld, C. A., M. R. Sawaya, S. Tanaka, C. V. Nguyen, M. Phillips, et al. 2005. Protein structures forming the shell of primitive bacterial organelles. Science **309:**936-938.

Klein, M. G., P. Zwart, S. C. Bagby, F. Cai, S. W. Chisholm, et al. 2009. Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport. J. Mol. Biol. **392:**319-333.

Lawrence, J. G., and J. R. Roth. 1996. Selfish operons: horizontal transfer may drive the evolution of gene clusters. Genetics **143:**1843-1860.

Parsons, J. B., S. D. Dinesh, E. Deery, H. K. Leech, A. A. Brindley, et al. 2008. Biochemical and structural insights into bacterial organelle form and biogenesis. J. Biol. Chem. **283:**14366-14375.

Penrod, J. T., and J. R. Roth. 2006. Conserving a volatile metabolite: a role for carboxysome-like organelles in *Salmonella enterica*. J. Bacteriol. **188:**2865-2874.

Price, G. D., and M. R. Badger. 1991. Evidence for the role of carboxysomes in the cyanobacterial CO_2-concentrating mechanism. Can. J. Bot. **69:**963-973.

Tanaka, S., M. R. Sawaya, and T. O. Yeates. 2010. Structure and mechanisms of a protein-based organelle in *Escherichia coli*. Science **327:**81-84.

Yeates, T. O., C. A. Kerfeld, S. Heinhorst, G. C. Cannon, and J. M. Shively. 2008. Protein-based organelles in bacteria: carboxysomes and related microcompartments. Nature Rev. Microbiol. **6:**681-691.

Figure Legends
Figure 1.
Transmission electron micrographs of (A) carboxysomes in the unicellular cyanobacterium *Synechocystis* PCC6803 (Courtesy of Patrick Shih and Cheryl Kerfeld)

and (B) carboxysomes in cells of the filamentous cyanobacterium *Oscillatoria* PCC6407 (Courtesy of Gérard Guglielmi, Nicole Tandeau de Marsac, Muriel Gugger and Rosemary Rippka of the Pasteur Culture Colletcion). (C) Eut BMCs in *Salmonella enterica* LT2 (Courtesy of Rick Webb, Doug Huseby, John Roth.)

Figure 2.
Schematic of carboxysome function in the context of a single cyanobacterial cell. Inorganic carbon transporters are shown in green. CA, carbonic anhydrase; RuBP, ribulose-1,5-bisphosphate, RuBisCO, ribulose-1,5-bisphosphate carboxylase/oxygenase; 3-PGA, 3-phosphoglycerate; G3P, glyceraldehyde 3-phosphate

Figure 3.
The building blocks of the carboxysome shell and a model of overall assembly based on icosahedral symmetry. The single BMC domain monomer and hexamer is shown in blue, the tandem BMC domain monomer and trimer is shown in fuscia, green and gold and the Pf03319 (CcmL-EutN domain) monomer and pentamer are shown in yellow.

Figure 4.
Surface and electrostatic representation of building blocks of the BMC shell with conserved residues surrounding the pore shown as sticks. A) CcmK2 hexamer with Lysine residues. B) CcmK4 hexamer with Arginine residues C) CsoS1D trimer in the open and D) closed conformation with Arginine residues shown as sticks. Electrostatic potential ranges between -10 (red) and +10 (blue).
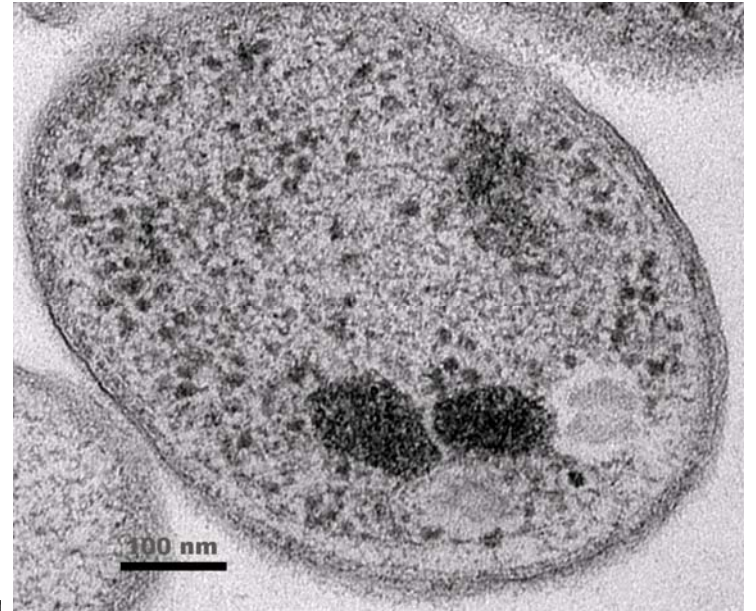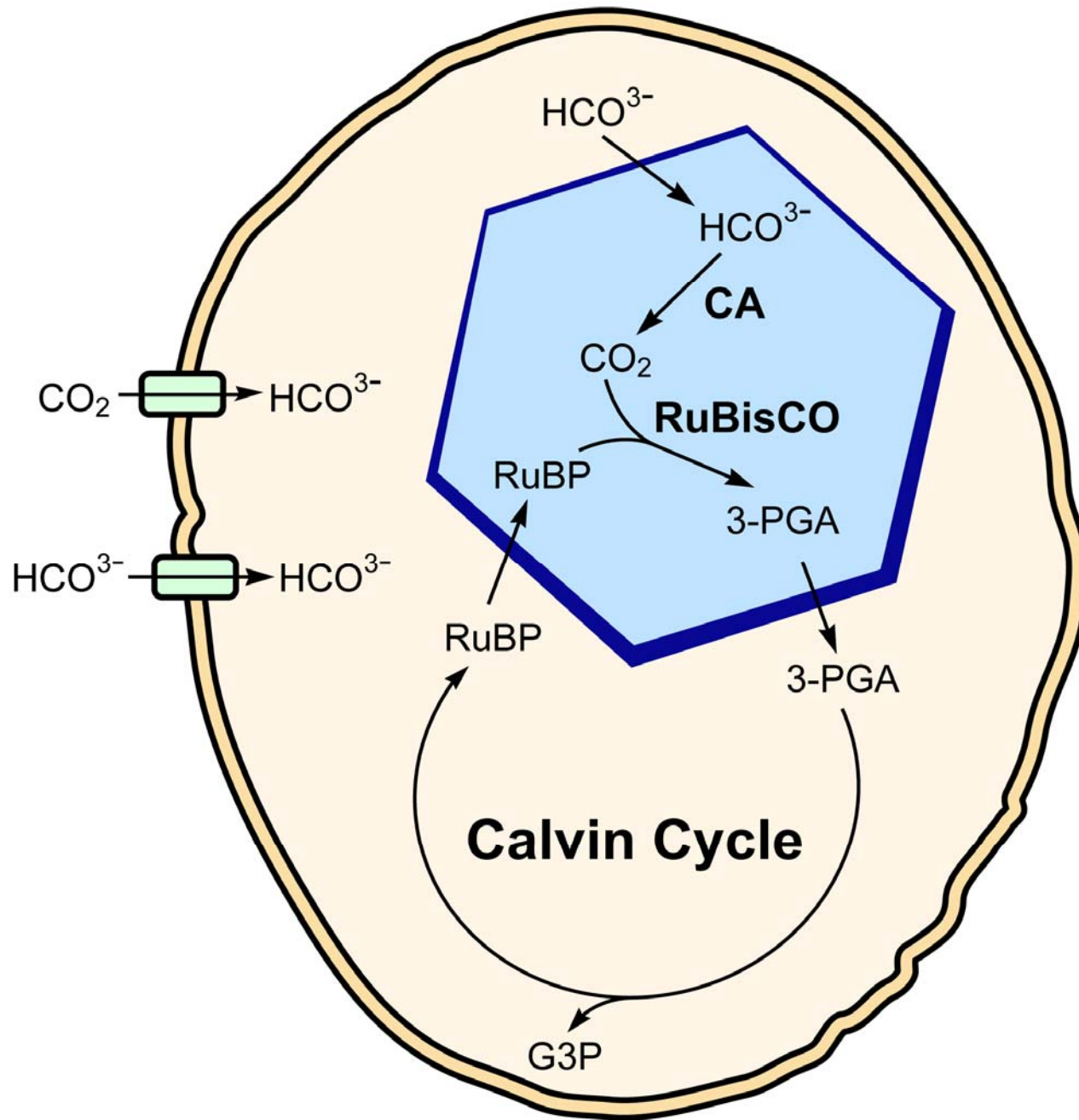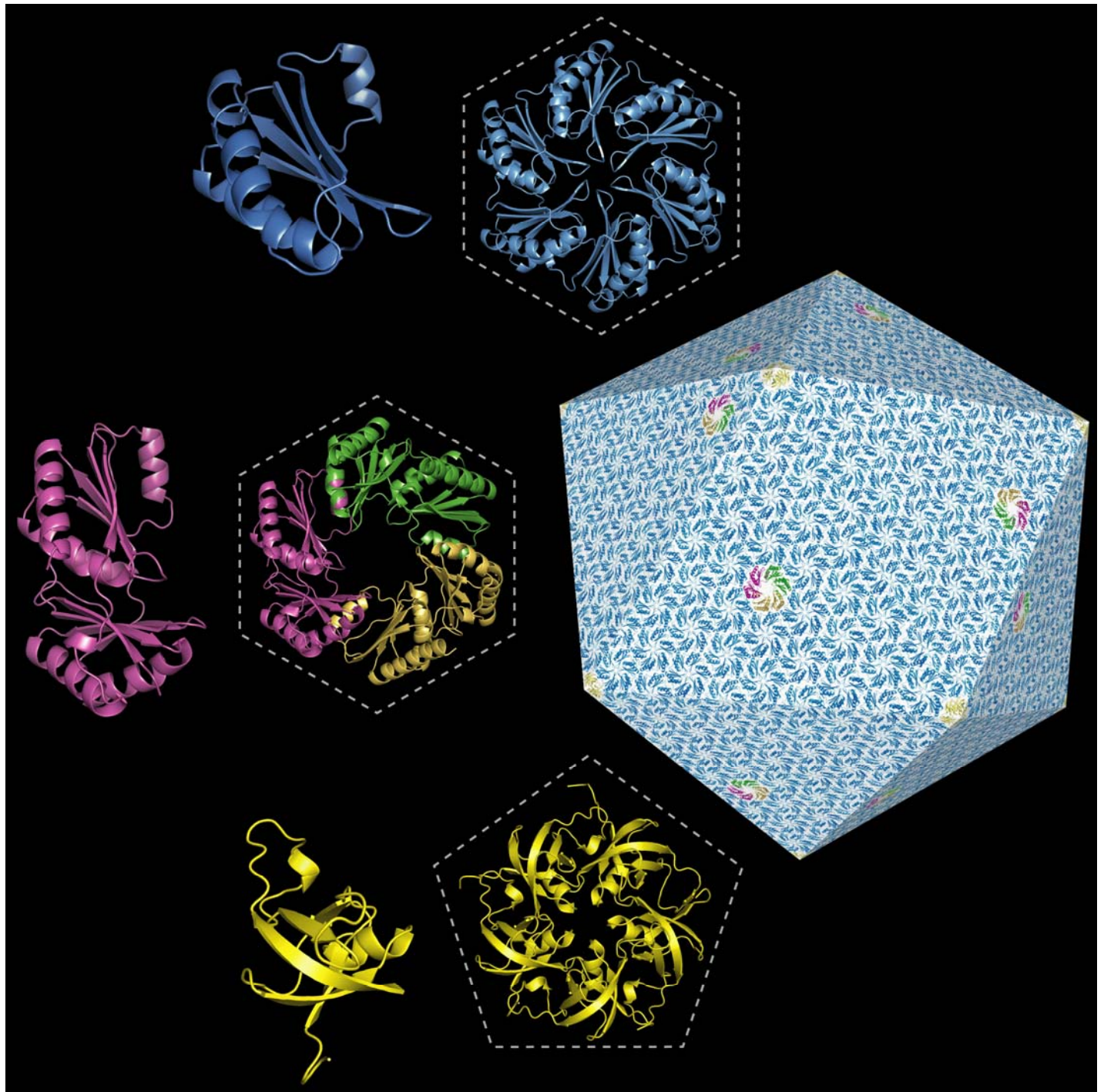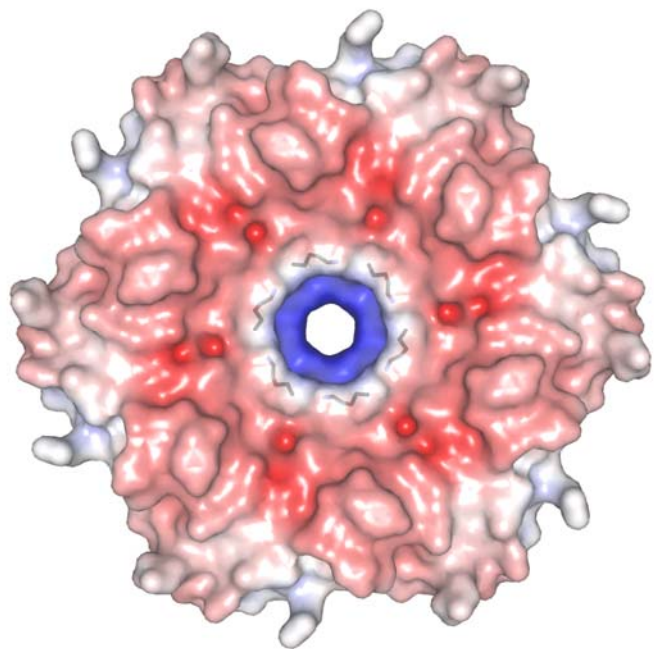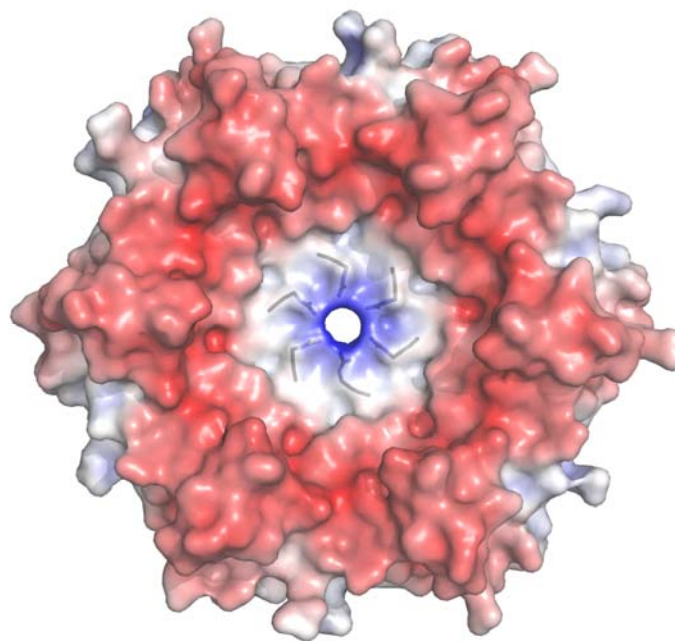
A
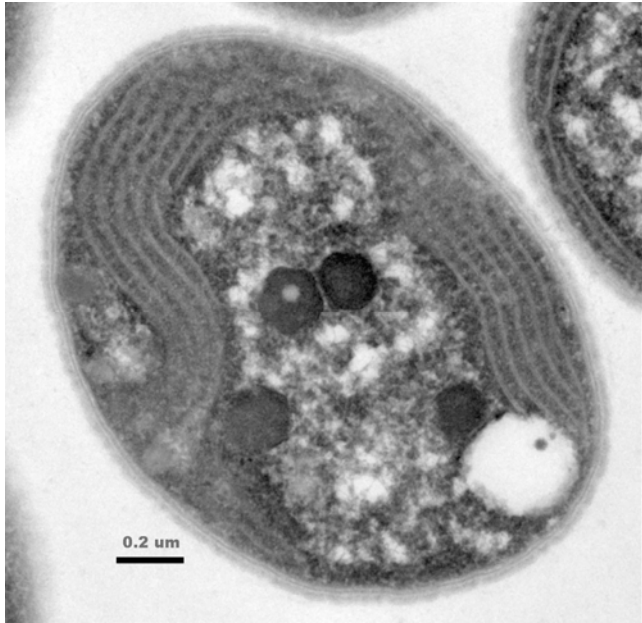
B
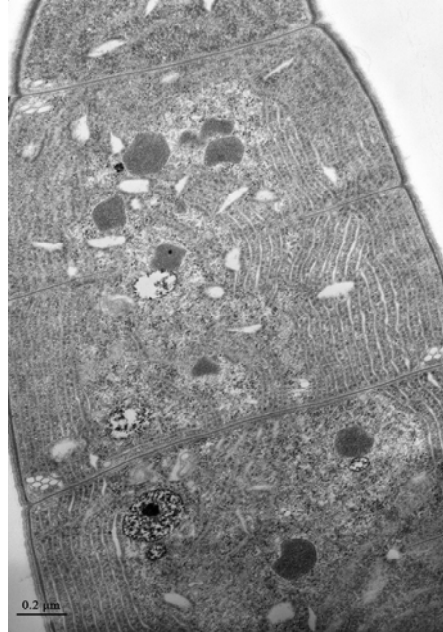
C

Figure 1

Figure 2

Figure 3

A

B

C

D

Figure 4