1 **Title:** Bacterial microcompartments and the modular construction of microbial
2 metabolism

3

4 **Authors:** Cheryl A. Kerfeld^{1, 2, 3, 4} and Onur Erbilgin¹,

5 1) Department of Plant and Microbial Biology, University of California Berkeley,
6 Berkeley, CA USA

7 2) DOE Plant Research Laboratory, Michigan State University, East Lansing,
8 Michigan, USA

9 3) Physical Biosciences Division, Lawrence Berkeley National Laboratory,
10 Berkeley, California, USA

11 4) Berkeley Synthetic Biology Institute, Berkeley, California, USA

12

13 **Corresponding Author:** Kerfeld, C.A. (ckerkfeld@lbl.gov)

14

15 **Keywords:** Bacterial Microcompartments, Carboxysomes, Metabolosomes,
16 Synthetic Biology, Bacterial Organelles

Abstract

Bacterial microcompartments (BMCs) are protein-bound organelles predicted to be present across 23 bacterial phyla. BMCs facilitate carbon fixation as well as the aerobic and anaerobic catabolism of a variety of organic compounds. These functions have been linked to ecological nutrient cycling, symbiosis, pathogenesis, and cardiovascular disease. Within bacterial cells, BMCs are metabolic modules that can be further dissociated into their constituent structural and functional protein domains. Viewing BMCs as genetic, structural, functional, and evolutionary modules provides a framework for understanding both BMC-mediated metabolism, and for adapting their architectures for applications in synthetic biology.

Bacterial microcompartments are specialized metabolic modules

In the past decade, genomics-enabled technologies have firmly established the concept of bacterial communities functioning as multi-cellular organisms consisting of specialist functional guilds [1,2] (Figure 1A). At the same time it has become apparent that eukaryotic-like division of labor likewise extends subcellularly; it is now known that some enzymes catalyzing sequential reactions are encapsulated in organelles, for example in bacterial microcompartments (BMCs) (Figure 1B). BMCs consist of functionally linked enzymes enclosed by a protein shell. They include anabolic carboxysomes (Figure 1C), found in cyanobacteria and some chemoautotrophs, that function in carbon fixation [3,4]. In addition, a variety of catabolic BMCs, termed metabolosomes, (Figure 1D) are found in heterotrophs; metabolosomes have been implicated in the degradation of propanediol (PDU),

ethanolamine (EUT), ethanol (ETU) [3], choline (CUT) [5,6], in the fermentation of plant biomass (GRP) [7,8], and aerobic degradation of algal polysaccharides (PVM) [9]. In all of these examples BMCs play a key role in the organism's success in its particular niche.

Although functionally distinct BMCs vary in their encapsulated enzymes, all are defined by homologous shell proteins. The first crystal structures of each of the three types of proteins constituting BMC shells came from the carboxysome: BMC-H, a hexamer of proteins containing a single PF00936 domain [10], BMC-T, a trimer of proteins containing two PF00936 domains [11], and BMC-P, a pentamer formed by proteins containing one PF03319 domain [12] (Figure 1B, 2A-C). Given the structural similarities between BMC shell proteins (hexagons and pentagons) and those of icosahedral viruses, a model for the BMC shell based on geometric constraints was proposed in which hexagonal BMC-H and BMC-T proteins form the facets and the pentagonal BMC-P proteins cap the vertices [10].

Although first described more than 50 years ago [13], BMCs have only recently become a prominent area of research, with several complementary advances driving the field forward in the last decade (Box 1). The first X-ray crystallographic studies [10] revealed the structure of the basic building block of the shell, how the facets are formed, and provided a hypothesis for the structural basis of the selective permeability. The structures were placed in the context of genomic sequence data, showing that, unexpectedly, many bacterial species have the potential to form

carboxysome-like shells [10]. This anticipated insights from the growth of microbial (meta)genomics and bioinformatics, which amplified and extended these observations [3,14–17]. To date, evidence for BMCs has been found in 23 bacterial phyla, with 23 distinct loci identified [14] (Table 1). A third recent major advance was the demonstration that BMC proteins could be fluorescently labeled without disrupting function [18–20]; this led to studies of the cell biology of BMCs, including their spatial organization in the cell and their assembly. Collectively, the insights from these advances span a wide range of the levels of biological organization, from the genes that encode BMC-associated proteins to their organization into loci that are frequently horizontally transferred as genetic modules, to the self-assembly of various proteins into organelles that play key roles in microbial ecology and evolution. Because of their attributes as self-assembling metabolic modules, they have motivated efforts by synthetic biologists and bioengineers to find ways to utilize these features with the aim of building designed organelles [21]. In this review, we will focus on these recent advances in understanding various aspects of BMC structure, function and assembly in the context of modularity – from the protein domain to the role of the organism in the environment (Figure 1) – and how this can be applied to biotechnology.

BMCs within host-microbe and microbe-environment interactions

The relevance of BMCs in microbial metabolism in terms of host-microbe communication and microbe-environment colonization has recently gained significant attention. Bioinformatic surveys indicate that a diversity of pathogens,

commensals, and other environmentally important microbes encode metabolosomes [10,14–17]. The well-characterized PDU and EUT metabolosomes have been linked to the virulence of intracellular pathogens, including *Salmonella enterica* and *Listeria monocytogenes* [22–26]. BMCs are also found in the human gut microbiome and are relevant to human health: for example, the microbial production of trimethylamine from choline is catalyzed by a glycyl radical enzyme [5] that is found in a BMC locus [6,14]; the oxidation of microbially-derived trimethylamine in the liver has been linked to cardiovascular disease [27,28]. The importance of BMCs in the metabolism of environmentally important microbes is likewise becoming apparent. For example, several lines of evidence indicate that members of the enigmatic phylum *Planctomycetes* are robust symbionts of macroalgae (reviewed in [29]), and many planctomycetes species encode a BMC to aerobically degrade fucose, rhamnose, and fucoidan (a sulfated polysaccharide found predominantly in algae) indicating that the organelle may be involved in this symbiosis [9]. In addition, a *Clostridium* species has been shown to use a BMC for the fermentation of fucose [7], likely derived from plant sources.

The metabolic cores of BMCs

Despite the apparent functional diversity among BMCs, some overarching themes in the biochemical transformations they catalyze can be identified. Recently, two bioinformatically predicted BMC loci were experimentally shown to encode organelles that are involved in the degradation of carbon compounds via an aldehyde [7,9], much like the well characterized PDU and EUT BMCs. While these

four types of metabolosomes differ in their aldehyde-generating enzyme (which we term the 'signature enzyme'), they all use the same three kinds of enzymes to oxidize the aldehyde to its cognate alcohol and acyl-phosphate: (i) a coenzyme A (CoA)-dependent aldehyde dehydrogenase (AldDH), (ii) an alcohol dehydrogenase (AlcDH), and (iii) a phosphotransacylase (PTAC) (Figure 1D). The prevalence of this shared functional core has been recently quantified: of all non-carboxysome BMC loci, 94% encode AldDH, 76% encode AlcDH, and 66% encode PTAC, indicating that majority of metabolosomes share a common core biochemistry involving degradation of an aldehyde, while differing in the signature enzyme and ancillary genes [14].

While the majority of bioinformatically identified BMC loci fit this model, there are others that lack these core enzymes and do not encode any obvious aldehyde-generating enzyme. Among these are the anabolic carboxysomes that encapsulate carbonic anhydrase (CA), which generates CO_2 , and the enzyme ribulose biphosphate carboxylase/oxygenase (RuBisCO), which fixes CO_2 (Figure 1C). Two other BMC types, predicted to be catabolic, also lack some or all of the metabolosome core genes [BMC of unknown function (BUF) and metabolosome with incomplete core (MIC), Table 1], suggesting that they encapsulate different reactions not involved in aldehyde degradation or carbon fixation [14]. These presumably new types of BMCs await experimental characterization.

The BMC shell: sequestration and permeability

The enzymatic cores of all BMCs are surrounded by a selectively permeable membrane – a protein shell (Table 1). Interestingly, substrates known to cross BMC shells are all highly polar (e.g., phosphorylated compounds), while those known to be retained are less so (e.g., aldehydes), or nonpolar (e.g., CO₂). Effectively, the shell has permeability properties inverse to those of a lipid bilayer, and structural studies of shell proteins have shown how this might be achieved.

The structures of many BMC-H shell proteins from both carboxysomes and metabolosomes are now known (Figure 2) [10,12,30–41]. The majority of the observed hexamers contain a central pore (coincident with the cyclic symmetry axis) that is assumed to be constitutively open (Figure 2A). The typical size of these pores (4-7Å) is large enough to permit the passage of smaller metabolites (e.g., bicarbonate into carboxysomes, alcohols and carboxylic acids across metabolosome shells; Table 1). Supporting this hypothesis, residues lining the pore tend to be polar and well conserved among orthologous shell proteins [10,41–45]. Mutation of some of these residues leads to permeability defects [31], while others appear to be tolerated [45,46]. Crystal structures of BMC-T proteins have revealed that residues surrounding the pores may have alternate conformations that can change the diameter of the pore by several angstroms or occlude it completely [11,37,43,47,48]. The open conformation of the pores in these shell proteins is typically larger than those in hexameric shell proteins, presumably to allow bulkier metabolites to pass (e.g., 5-carbon sugar phosphate into carboxysomes; Table 1). Closure of the larger pores is likely required to maintain the diffusive barrier and selectivity of the shell.

A distinct clade of BMC-T proteins form stacked trimers [43] resulting in a central cavity accessible via gated pores on each side (Figure 2B); these BMC-T proteins are proposed to function like an airlock, with alternating open and closed pore conformations [11,43]. A BMC-H from the β -carboxysome likewise stacked face-to-face in crystals, albeit loosely [49]. Some measurements of BMC shell thickness support the hypothesis that the shell is double-layered [49,50], which would be expected to increase the stringency of permeability and the robustness of the shells [49], while making the inner and outer surface of the shell physicochemically the same. On the other hand, shell proteins with C-terminal fusions, such as a RhoN nucleotide binding domain, which would presumably interfere with stacking, can be found in some loci (e.g., the *eut* operon of *Brevibacillus* species); this suggests that at least some shell subunits are not likely to stack.

In addition to small molecules, recent studies indicate that protons and electrons can cross the BMC shell. By localizing pH-sensitive GFP to the α -carboxysome, Menon *et al.* showed that the luminal pH is similar to that of the cytoplasm [19]. Studies on PduT, a BMC-T shell protein, show that it contains an iron-sulfur cluster [51], which is likely bound in the central pore formed by the three-fold symmetry axis [34]. Interestingly, PduS, a presumed luminal protein that also contains an iron-sulfur cluster and binds a flavin cofactor [52,53], has been shown to interact with PduT; the pair is proposed to be a conduit for transfer of electrons across the shell [52]. Finally, a recent comprehensive survey of BMC loci noted that genes encoding several different types of flavin-binding proteins (in addition to PduS)

occur in many BMC loci [14]; given that flavins frequently act as electron carriers between iron-sulfur clusters, it appears that electron transfer into and within BMCs may be widespread.

Although necessarily permissive to some metabolites, the shell is also known to be a barrier to other molecules. Deletion of specific shell proteins results in loss of retention of CO₂ by carboxysomes [54], of propionaldehyde from PDU BMCs [55], and of acetaldehyde from EUT BMCs [56]. Furthermore, when either the alcohol dehydrogenase (AlcDH) or phosphotransacylase (PTAC) are deleted in the PDU or EUT metabolosomes, growth on propanediol or ethanolamine, respectively, is severely hampered – even if a non-BMC-associated enzyme homolog is heterologously expressed [57,58]. When the integrity of the BMC shell is subsequently compromised in these mutants, growth is restored [57,58]. These observations indicate that the PDU and EUT shells are impermeable to NAD⁺/H and CoA, with AlcDH and PTAC necessarily recycling the cofactors used by the AldDH [57,58] (Figure 1C). In addition to sequestering cofactors, CO₂, and volatile aldehydes, there are several indications that the shell also provides insulation from oxygen. The EUT AlcDH is oxygen-sensitive [57], yet the EUT BMC is functional in aerobic conditions, suggesting that the lumen-localized enzyme is protected from cytosolic oxygen. Glycyl radical enzymes [signature enzymes in the glycyl radical enzyme-containing metabolosomes (GRMs); Table 1] are also oxygen sensitive, and are presumably likewise protected by encapsulation within the BMC shell. Similarly, protection of RuBisCO from oxygen has long been proposed to be the

function of the carboxysome shell. Furthermore, the redox state of the carboxysome lumen can also be independent from that of the cytosol [59], a proposed requirement for the enzymatic function of one of the two encapsulated carbonic anhydrases in the beta-carboxysome [60].

Despite these structural and functional studies, our understanding of the permeability properties of BMC shells is still far from complete. For example, glycyl radical enzymes are S-adenosylmethionine (SAM)-dependent; SAM is cleaved to provide the radical and therefore must be replenished. Therefore, SAM must somehow cross the shell and, for those GRM BMCs that appear to encapsulate methionine adenosyl transferase, ATP and methionine must cross the shell. Even for a well characterized BMC like the carboxysome, genomic and bioinformatic analysis has raised new questions about shell permeability. For example, some cyanobacterial genomes that encode β -carboxysomes (containing form 1B RuBisCO) also encode a RuBisCO activase fused to a domain known to be involved in the assembly of the carboxysome [61]. Many α -carboxysome (which encapsulate Form 1A RuBisCO) loci also contain a gene for RuBisCO activase [14,62]. RuBisCO activase requires ATP hydrolysis, suggesting the possibility of ATP/ADP flux across carboxysome shells. The presence of ATP:cobalamin adenosyltransferase (PduO) in the PDU [63] and EUT metabolosomes [64] suggests that ATP may cross these shells as well. To-date quantitative data for metabolite flux across any BMC shell has not been obtained; these data are essential to modeling the impact of heterologously expressed BMCs, for example carboxysomes in plants [65].

224

225 Assuming icosahedral geometry for the BMC shell, only two types of building blocks
226 are required: one hexagonal (BMC-H or BMC-T), and one pentagonal (BMC-P).
227 However, all BMC loci encode multiple paralogs containing the PF00936 domain
228 (the canonical domain of BMC-H and BMC-T proteins) and many organisms encode
229 additional shell proteins in satellite loci [14]. The observation that shell protein
230 paralogs have distinct conserved pore residues and, consequently, pore sizes led to
231 the proposal that the different shell proteins within a given BMC have different
232 permeability properties [10]. This may explain the redundancy of shell protein
233 genes. A corollary of this assumption is that the composition of the shell could be
234 (evolutionarily or synthetically) tuned in response to the environment. A second
235 mystery regarding a multiplicity of shell protein genes in BMC loci has emerged
236 from the increasing number of loci identified in genomic sequence data. Many
237 contain an apparent surfeit of genes (up to 5) encoding BMC-P proteins [9,14,44].
238 This is odd for a presumably minor component of the shell: only 60 copies (12
239 pentamers) are required to cap the 12 vertices of an icosahedron. The multiple
240 BMC-P proteins in these loci are apparently not recent gene duplications of one
241 another, and are conserved for a given type of BMC across phyla [9,14]; these
242 observations suggest that the redundancy has been selected for and is
243 physiologically relevant. Perhaps these BMC-P proteins assemble into hetero-
244 pentamers or have a role other than vertex formation, such as forming a building
245 block of the shell facet.

246

Assembly of BMCs

BMC shell proteins self-assemble

The shell is the interface between the metabolic core of the BMC and the cytosol, and this protein membrane displays a remarkable tendency to self-assemble. The first structure of a BMC-H protein indicated that the hexamers tiled into uniformly oriented layers in the crystals, immediately suggesting a model for facet formation [10]. In modeling the first structure of a BMC-T protein into facets composed of BMC-H proteins, Klein *et al.* identified a motif (D/N-X-X-X-K), the most highly conserved among all PF00936 domains, and suggested that the antiparallel association of these lysine residues between adjacent (pseudo)hexamers was critical for shell assembly [11] (Figure 2). Recently, the crucial role of this lysine was confirmed experimentally [30,31]. In some crystal structures, BMC-H and BMC-T proteins tile into strips or layers [10,12,30,35,38,41,43,45,47] and form nanotubes, lattices, and ‘swiss rolls’ *in vivo* when overexpressed heterologously [18,30,33,37,51]. Ordered arrays have even been captured *in vitro*, in electron micrographs of purified proteins [43,50]; taken together, these observations highlight the propensity of BMC shell proteins to self-assemble into higher-order structures.

Efforts have recently focused on manipulating BMC shells for bioengineering applications; these include assembling chimeric shells *in vivo*, composed of BMC-H proteins from different sources [45], and assembling empty shells. Empty BMC shells (α -carboxysome shells) were first observed *in vivo* in *Halothiobacillus*

neapolitanus mutants lacking RuBisCO [66,67]. In heterologous systems, metabolosome shells have been produced by expressing shell proteins from the PDU [18], EUT [68], or an uncharacterized BMC [50] (Figure 3D). Surprisingly, shells are also formed in the absence of BMC-P proteins [50,54,68], although these shells are presumably highly porous [54]. These observations have parallels in the studies of viral capsids; some do not require vertex pentons to form icosahedra [69], and the lack of pentons leads to loss of packaged material [70].

A peptide extension mediates enzyme encapsulation

Studies to understand and control shell assembly are motivated by the goal of producing bespoke organelles for synthetic biology. To this end, understanding how thousands of proteins, constituting the cargo and the shell, assemble and organize into functional organelles provides crucial information for bio-inspired designs. Initially, efforts were directed to identifying BMC targeting signals on the encapsulated enzymes. The first experimentally verified example was an N-terminal sequence extension on PduP (AldDH) that is not present in non-metabolosome homologs [71]. The next, found at the C-terminus of the carboxysomal protein CcmN, was necessarily identified entirely by experimental methods because there are no unencapsulated orthologs to CcmN [15]. This study also marked the first demonstration that the peptide interacted with the shell. Based on this observation and considering that shell proteins are the unifying features of all BMCs, similar peptides mediating the association of proteins with shells were predicted for 10 different types of BMC loci through the identification of their shared sequence

features: namely a region of 15-20 amino acids predicted to form an amphipathic alpha helix separated from the functional domain(s) of the protein by a poorly conserved linker [15]. To date, the predictions have been functionally verified for the AldDH of the *Haliangeum ochraceum* BMC [50] and of PduD [71–74]. The predicted helical conformation has also been confirmed [74].

From protein domain interactions to fully functional organelle

These peptides were first referred to as ‘targeting peptides,’ seemingly analogous to those that direct proteins to eukaryotic organelles, with the implicit assumption that BMC shells form first, with subsequent import of the cargo enzymes (Figure 3A). Indeed, the observations that shells can be formed by expression of shell protein in the absence of cargo would seem to support this. However, viral capsids provide a precedent for proteinaceous shell systems that can self-assemble in the absence of cargo [75]. The question of BMC assembly has recently been addressed through advances in methods to visualize BMCs; these have opened up the study of the cell biology of the organelles [18–20].

The ability to fluorescently label carboxysomes *in vivo* led to studies culminating in the elucidation of the β -carboxysome assembly pathway and the demonstration that, unlike eukaryotic organelles, carboxysomes assemble from the inside out (Figure 3B). Fluorescence labeling of CcmN was used to show that this protein, of previously unknown function, links the shell to the enzymatic core of the carboxysome [15]. Cameron *et al.* put the β -carboxysome (*ccm*) operon under

control of an inducible promoter and used fluorescently labeled RuBisCO to visualize *de novo* carboxysome assembly, while monitoring cell physiology to ensure the structures observed were fully functional organelles [76]. Through systematic deletion of the genes in the *ccm* operon and fluorescence labeling, the step-by-step assembly pathway was elucidated: procarboxysomes are initiated, typically at one pole of the cell, by the aggregation of RuBisCO and the RuBisCO small subunit-like domains of CcmM [76]; CcmM and CcmN form a complex [15,77], and the peptide on the C-terminus of CcmN (therefore renamed the ‘encapsulation peptide’) facilitates interaction with the shell [15,76] (Figure 3B). Chen *et al.* likewise put fluorescently labeled RuBisCO under the control of an inducible promoter and, using pulse-chase assays, determined that new carboxysomes can be born from pre-existing ones [59]. Fluorescence recovery experiments also showed that encapsulated RuBisCO is not fluid in the carboxysome [59], and in conjunction with electron microscopy observations [76] support earlier hypotheses that the encapsulated enzyme is paracrystalline [78]. This implies that, within carboxysomes, a high concentration of active sites is present in an ordered array and most likely served by substrate channeling mediated through specific protein–protein interactions among the luminal components.

The observed sequence of events in carboxysome biogenesis is consistent with hierarchical assembly of the constituent protein domains. This could indicate cooperativity within the assembly process [79–81], where the interaction of two protein domains creates a high-affinity binding site for a third, with the ternary

339 complex binding a fourth. Several lines of evidence suggest that metabolosomes may
340 likewise follow an ordered assembly pathway, with the formation of a pro-
341 metabolosome core followed by encapsulation by the shell. There is evidence both
342 experimentally and bioinformatically suggesting that the core proteins of
343 metabolosomes also form an ordered network of protein-protein interactions. The
344 signature enzymes of both the PDU and EUT metabolosome contain subunits with
345 encapsulation peptides that render the enzymes insoluble [82,83]. The enzymes are
346 soluble and active after removal of the peptides, suggesting that the full-length
347 proteins require a binding partner *in vivo*. *Salmonella enterica* PduP (AldDH) co-
348 purifies with propanediol dehydratase (PduCDE) [84]; likewise PduQ (AlcDH) is
349 known to physically interact with PduP (AldDH) [57]. When expressed alone, *S.*
350 *enterica* PduP (AldDH) forms insoluble aggregates [85], suggesting that the
351 presence of binding partners may be important for PduP folding. Moreover, some
352 PDU Metabolosome components need to be encapsulated due to their supporting
353 role in maintenance of the signature enzyme and would be ideally positioned near
354 these enzymes. The vitamin B₁₂ cofactor of PduCDE undergoes radical damage and
355 is recycled by PduGH, PduS, and PduO [53]; considering that PduCDE is
356 encapsulated, the vitamin B₁₂ recycling enzymes are presumably also encapsulated.
357 Supporting this hypothesis, PduS binds the shell protein PduT [52], and PduS and
358 PduO interact with one another [53]. Notably, in the metabolosome loci of several
359 genomes (e.g., *Desulfosporosinus* species), PduL and PduO are fused; given the
360 observation that fusions of otherwise isolated gene products is considered a reliable
361 indicator of both functional and physical protein-protein interactions [86,87] and

can be reflective of sequential protein–protein interactions in macromolecular assemblies [81], this would suggest a complex of PduS–PduO–PduL that has at least one shell protein connection through the PduS–PduT interaction. Organizing such a network of protein domain interactions after targeting to the confines of a pre-formed shell is difficult to envision.

The signature enzyme of the EUT metabolosome is B₁₂-dependent but, as-yet, a complete set of B₁₂ recycling enzymes has not been identified in *eut* loci (only EutT, a PduO homolog is known). This has led to the suggestion that the signature enzyme, ethanolamine ammonia lyase (EutBC) is associated with the exterior of the metabolosome and the acetaldehyde formed crosses the shell [58]. However, in some *eut* operons, *eutB* and *eutG* (encoding the large subunit of ethanolamine ammonia lyase and alcohol dehydrogenase, respectively) are fused (e.g., in *Shigella boydii* and many strains of *Shigella flexneri*), indicating that the core AlcDH would be outside the BMC and unable to recycle NAD⁺/H. There are several genes of unknown function that could play roles in maintenance (structural or physiological) and/or biogenesis of both EUT and PDU organelles such as EutQ, EutJ, PduV/EutP, PduM, and PduX. Some of these are also prevalent among diverse BMCs; for example, PduV/EutP is found in 68% of all non-carboxysome loci [14], and localizes to synthetic BMC shells [18]. Finally, many proteins presumably encapsulated in diverse BMC cores contain counterparts to the peptide that associate the pro-carboxysome with the shell [15,71–74].

In contrast to the β -carboxysome and the majority of bioinformatically identified BMC loci, encapsulation peptides have not been identified for any α -carboxysome proteins. In addition, the shell proteins of α -carboxysomes are less closely related to β -carboxysome and metabolosome shell proteins than the latter types are to each other [45], perhaps reflective of some distinctive differences in shell-cargo association relative to the majority of BMCs. Another rather stark difference between α - and β -carboxysomes is the absence of RuBisCO small subunit-like domains on any of the α -carboxysome components; these are critical in the early steps of β -carboxysome assembly. Indeed, the observation of partially formed *H. neapolitanus* α -carboxysomes in a cryo-electron tomography study suggested that they form by simultaneous coalescence of enzymes and shell proteins [88], and may represent a distinct pathway for assembly (Figure 3C).

Post-assembly cellular organization

Once assembled, BMCs are organized within the cytosol by interactions with the cytoskeleton. PDU BMCs were observed to be associated with intracellular (presumably cytoskeletal) filaments, mediated by PduV, a protein of unknown function [18] that is widespread among BMC loci [14]. Similarly, the cytoskeletal protein ParA was shown to be involved in the spatial organization of β -carboxysomes, playing a role in their equal segregation between daughter cells [20]. ParA homologs are frequently found in α -carboxysome loci [62], and genes for cytoskeletal elements are found in several different BMC loci types [14]. Spatial

positioning within the cell is likely an important consideration for optimal function of the BMC within the context of cellular metabolism.

Implications of modular assembly of BMCs for synthetic biology

BMC operons can be viewed as genetic modules that encode the necessary parts (protein domains) to construct a functional organelle (Figure 1E, F) [14]. They were among the first loci described as ‘selfish operons’ [89] and, indeed, recent bioinformatic analyses provides evidence of rampant horizontal gene transfer of BMC loci [16,89,90]. As metabolic modules, BMCs are semi-autonomous functional devices nested within the context of the host metabolism. Interestingly, some bacterial genomes (e.g., *Clostridium phytofermentans*) encode up to three BMC loci [7,14,16], presumably for the catabolism of different organic compounds [14]. This insulation of catabolic reactions by compartmentalization likely facilitates the simultaneous use of pathways that would otherwise compete and allows partitioning of metabolites to distinct fates [91].

The engineering of new functions into microbes via BMCs has recently begun to be explored through heterologous expression studies. The first showed that the *pdu* operon from *Citrobacter freundii* could be expressed in *Escherichia coli* where it conferred the ability to grow on propanediol [51]. Subsequently, the α -carboxysome operon from *H. neapolitanus* was expressed in *E. coli*; carboxysome-like structures were formed and the RuBisCO was functional [92]. However, in these examples the operon ‘donor’ species all belong to the γ -Proteobacteria, and the *E. coli* (also a γ -

proteobacterium) expression strains contain the *eut* operon [14]; the presence of host genes could have contributed to heterologous BMC formation. Recently, the first interkingdom heterologous expression and assembly of BMC proteins was reported. When several β -carboxysome genes (CcmK2, L, M, N, O) were expressed in the plant *Nicotiana benthamiana*, carboxysome-like structures formed in the chloroplasts [93], hinting that carboxysomes may be able to self-assemble within the eukaryotic organelle.

The first steps toward development of designed organelles have involved associating foreign proteins (e.g., GFP) with heterologously expressed BMC shells by fusing the proteins to full length luminal proteins or to encapsulation peptides [50,68,72,74]. Evidence thus far suggests encapsulation peptides are important for interaction with the shell, and not in organization of the contents of the BMC lumen. However, the advantages of compartmentalization are fully realized by co-localizing multiple enzymes in the optimal orientations for substrate channeling, suggesting the need for design principles for constructing BMC cores as well as for associating proteins with the shell.

We propose an approach based on knowledge of the structures of the component protein domains in conjunction with an understanding of their interactions in assembly (Figure 1F, G). Indeed, metabolosome loci can be viewed as being composed of several functional modules composed of one or more gene products (Figure 1E), that can be further deconstructed into their constituent structural and

functional domains [14] (Figure 1F). For example, as noted above, BMC shells can form when one or more shell proteins are heterologously expressed [18,45,50,68]; the PF00936 and PF03319 domains of the BMC-H, BMC-T and BMC-P gene(s) can, accordingly, be viewed as composing a ‘shell’ module. Furthermore, carboxysomes and metabolosomes contain conserved biochemical cores and signature enzymes (Figure 1C-E, Table 1). These likewise can be further deconstructed into their structural domains (Figure 1F). Given the important roles of specific domains of multidomain proteins in carboxysome assembly (e.g., the RuBisCO small subunit-like domains of CcmM interacting with RuBisCO [94,95], and the encapsulation peptide of CcmN interacting with the shell[15]), approaching the design of new types of BMCs from a structural point of view, using knowledge of BMC-related domain structures, architectures and interdomain interactions may be more expedient for assembling multienzyme organelles than *de novo* designs.

Indeed, some bacteria encode fused and streamlined (non-redundant) PDU-EUT or GRM-PDU loci [14], which may represent bi-functional BMCs and underscores the interchangeability of protein domains both structurally and functionally among BMCs. Alternatively, basing designs on the folds encapsulated naturally constituting BMC cores can circumvent the numerous challenges associated with building up a BMC core from scratch. A major advance in the efforts to express carboxysomes in plants exemplifies the potential of this approach; when cyanobacterial RuBisCO and the SSLDs of CcmM were expressed in the plants, RuBisCO aggregated into structures similar to pro-carboxysomes [96].

476

477 Likewise knowledge of the domain structures of encapsulated enzymes provides a
478 powerful approach to the goal of creating new organelles with non-native
479 biochemical activities. The active sites of signature and core enzymes could be
480 modified to allow for more substrate ambiguity. Such substrate ambiguity is
481 already observed in extant BMCs, where the core AldDH can utilize a variety of
482 short-chain aldehydes [97], this may have played a role in the evolution of the
483 functional diversity of BMCs. Alternatively, the signature enzyme could be
484 replaced with another member of its superfamily. Because diverse catalytic
485 activities are frequently supported by the same structural fold [98], viewing the
486 encapsulated proteins domain as a set of protein folds and altering their catalytic
487 residues to function as a different member of the same structural superfamily offers
488 a relatively simple approach to modifying the encapsulated biochemistry (Figure
489 1G). This would minimize the structural differences in order to preserve protein-
490 protein interactions required for assembly.

491

492 In conjunction with manipulation of BMC-encapsulated enzymes, metabolite and
493 cofactor transport across the shell would also need to be tailored; customizing
494 metabolite flux may be approached through the engineering of residues
495 surrounding the pores or through the construction of chimeric shells; both
496 approaches have recently been shown to be feasible [45]. The selective permeability
497 and self-assembly of BMC shell proteins also have implications for the co-option of
498 their repetitive nanoscale ordering to build nanomaterials, such as scaffolds or

selectively permeable protein membranes. Empty BMC shells with designed permeability properties could also be used as transport or storage capsules, especially structurally robust ones that can be predictably assembled and purified in high yield [50]. Because these can form without pentamers, the gaps formed at the vertices could serve as conduits to introduce compounds into the empty shells, and pentamers could theoretically be added later, sealing the capsule. Finally, in addition to the structural elements of the BMC, many BMC loci encode genes for other required functions such as cofactor assembly and recycling (e.g., PduGH/S/O), as well as ancillary modules for functions acting upstream or downstream of the organelle(e.g., transporters) [Figure 1C (green arrows), Figure 1E] to integrate the BMC function with other cellular processes or the environment (Figure 1A).

Concluding remarks

Viewing BMCs as simultaneously structural, functional, genetic, and evolutionary modules embedded within microbes, within the environment, underscores their central role in microbial evolution and ecology. This perspective also can inform their applications in synthetic biology. Parsing BMCs into their constituent structural and functional domains and deducing their order of assembly provides the prerequisite design criteria for engineering bacterial organelles as well as other bio-inspired designs for nanoscale materials. As metabolic modules that contribute to an organism's fitness in specific environments, BMCs also have potential uses in synthetic ecology: mixtures of species, some perhaps differentiated by the role of their specific BMC, could be integrated as a microbial consortium. Accordingly,

BMCs are situated centrally on the continuum between top-down and bottom-up approaches to engineering microbial metabolism. In this context, the concept of modularity provides a framework for understanding the fundamental biological principles governing bacterial microcompartment structure, function, assembly, and, prospectively, design (Box 2).

Acknowledgements

The authors would like to thank members of the Kerfeld laboratory for valuable discussions, and to thank Cristina Iancu and Grant Jensen for providing the micrograph used in Figure 3. This material is based upon work supported by the U.S. Department of Energy, Office of Basic Energy Sciences under contract no. DE-AC02 05CH11231 and by grants from the National Science Foundation (MCB1160614 and EF1105892).

References

- 1 Klitgord, N. and Segrè, D. (2011) Ecosystems biology of microbial metabolism. *Curr. Opin. Biotechnol.* 22, 541–6
- 2 Shapiro, J.A. (1998) Thinking about bacterial populations as multicellular organisms. *Annu. Rev. Microbiol.* 52, 81–104
- 3 Kerfeld, C.A. *et al.* (2010) Bacterial microcompartments. *Annu. Rev. Microbiol.* 64, 391–408
- 4 Rae, B.D. *et al.* (2013) Functions, compositions, and evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate CO₂ fixation in cyanobacteria and some proteobacteria. *Microbiol. Mol. Biol. Rev.* 77, 357–79

547 5 Craciun, S. and Balskus, E.P. (2012) Microbial conversion of choline to
548 trimethylamine requires a glycyl radical enzyme. *Proc. Natl. Acad. Sci. U. S. A.*
549 109, 21307–12

550 6 Kuehl, J. V *et al.* (2014) Functional Genomics with a Comprehensive Library of
551 Transposon Mutants for the Sulfate-Reducing Bacterium *Desulfovibrio*
552 *alaskensis*. *MBio* 5, e01041–14

553 7 Petit, E. *et al.* (2013) Involvement of a Bacterial Microcompartment in the
554 Metabolism of Fucose and Rhamnose by *Clostridium phytofermentans*. *PLoS*
555 *One* 8, e54337

556 8 Scott, K.P. *et al.* (2006) Whole-genome transcription profiling reveals genes
557 up-regulated by growth on fucose in the human gut bacterium “*Roseburia*
558 *inulinivorans*.” *J. Bacteriol.* 188, 4340–9

559 9 Erbilgin, O. *et al.* (2014) Characterization of a planctomycetal organelle: a
560 novel bacterial microcompartment for the aerobic degradation of plant
561 saccharides. *Appl. Environ. Microbiol.* 80, 2193–205

562 10 Kerfeld, C.A. *et al.* (2005) Protein structures forming the shell of primitive
563 bacterial organelles. *Science* 309, 936–8

564 11 Klein, M.G. *et al.* (2009) Identification and structural analysis of a novel
565 carboxysome shell protein with implications for metabolite transport. *J. Mol.*
566 *Biol.* 392, 319–33

567 12 Tanaka, S. *et al.* (2008) Atomic-level models of the bacterial carboxysome
568 shell. *Science* 319, 1083–6

569 13 Drews, G. and Niklowitz, W. (1956) Beiträge zur Cytologie der Blaualgen II.
570 Mitteilung Zentroplasma und granuläre Einschlüsse von *Phormidium*
571 *uneinatum*. *Arch. für Mikrobiol.* 24, 147–162

572 14 Axen, S.D. *et al.* (2014) A taxonomy of bacterial microcompartment loci
573 constructed by a novel scoring method. *PLoS Comput. Biol.*

574 15 Kinney, J.N. *et al.* (2012) Elucidating the essential role of the conserved
575 carboxysomal protein CcmN reveals a common feature of bacterial
576 microcompartment assembly. *J. Biol. Chem.* 287, 17729–17736

577 16 Abdul-Rahman, F. *et al.* (2013) The Distribution of Polyhedral Bacterial
578 Microcompartments Suggests Frequent Horizontal Transfer and Operon
579 Reassembly. *Phylogenetics Evol. Biol.* 1, 1000118

580 17 Jorda, J. *et al.* (2012) Using comparative genomics to uncover new kinds of
581 protein-based metabolic organelles in bacteria. *Protein Sci.* 22, 179–195

582 18 Parsons, J.B. *et al.* (2010) Synthesis of empty bacterial microcompartments,
583 directed organelle protein incorporation, and evidence of filament-associated
584 organelle movement. *Mol. Cell* 38, 305–15

585 19 Menon, B.B. *et al.* (2010) The carboxysome shell is permeable to protons. *J.*
586 *Bacteriol.* 192, 5881–6

587 20 Savage, D.F. *et al.* (2010) Spatially ordered dynamics of the bacterial carbon
588 fixation machinery. *Science* 327, 1258–61

589 21 Frank, S. *et al.* (2013) Bacterial microcompartments moving into a synthetic
590 biological world. *J. Biotechnol.* 163, 273–9

591 22 Joseph, B. *et al.* (2006) Identification of *Listeria monocytogenes* genes
592 contributing to intracellular replication by expression profiling and mutant
593 screening. *J. Bacteriol.* 188, 556–568

594 23 Klumpp, J. and Fuchs, T.M. (2007) Identification of novel genes in genomic
595 islands that contribute to *Salmonella typhimurium* replication in
596 macrophages. *Microbiology* 2, 1207–1220

597 24 Kendall, M.M. *et al.* (2012) Ethanolamine controls expression of genes
598 encoding components involved in interkingdom signaling and virulence in
599 enterohemorrhagic *Escherichia coli* O157:H7. *MBio* 3, 1–10

600 25 Harvey, P.C. *et al.* (2011) *Salmonella enterica* typhimurium colonizing the
601 lumen of the chicken intestine grows slowly and upregulates a unique set of
602 virulence and metabolism genes. *Infect. Immun.* 79, 4105–21

603 26 Maadani, A. *et al.* (2007) *Enterococcus faecalis* mutations affecting virulence in
604 the *Caenorhabditis elegans* model host. *Infect. Immun.* 75, 2634–7

605 27 Tang, W.H.W. *et al.* (2013) Intestinal microbial metabolism of
606 phosphatidylcholine and cardiovascular risk. *N. Engl. J. Med.* 368, 1575–84

607 28 Wang, Z. *et al.* (2011) Gut flora metabolism of phosphatidylcholine promotes
608 cardiovascular disease. *Nature* 472, 57–63

609 29 Lage, O.M. and Bondoso, J. (2014) Planctomycetes and macroalgae, a striking
610 association. *Front. Microbiol.* 5, 1–9

- 611 30 Pang, A. *et al.* (2014) Structural insights into higher-order assembly and
612 function of the bacterial microcompartment protein PduA. *J. Biol. Chem.* DOI:
613 10.1074/jbc.M114.569285
- 614 31 Sinha, S. *et al.* (2014) Alanine Scanning Mutagenesis Identifies an Asparagine-
615 Arginine-Lysine Triad Essential to Assembly of the Shell of the Pdu
616 Microcompartment. *J. Mol. Biol.* DOI: 10.1016/j.jmb.2014.04.012
- 617 32 Thompson, M.C. and Yeates, T.O. (2014) A challenging interpretation of a
618 hexagonally layered protein structure. *Acta Crystallogr. D. Biol. Crystallogr.* 70,
619 203–8
- 620 33 Pitts, A.C. *et al.* (2012) Structural insight into the *Clostridium difficile*
621 ethanolamine utilisation microcompartment. *PLoS One* 7, e48360
- 622 34 Pang, A. *et al.* (2011) Structure of PduT, a trimeric bacterial
623 microcompartment protein with a 4Fe-4S cluster-binding site. *Acta*
624 *Crystallogr. D. Biol. Crystallogr.* 67, 91–6
- 625 35 Crowley, C.S. *et al.* (2010) Structural insight into the mechanisms of transport
626 across the *Salmonella enterica* Pdu microcompartment shell. *J. Biol. Chem.* 285,
627 37838–46
- 628 36 Tanaka, S. *et al.* (2010) Structure and mechanisms of a protein-based
629 organelle in *Escherichia coli*. *Science (80-.)*. 327, 81–4
- 630 37 Heldt, D. *et al.* (2009) Structure of a trimeric bacterial microcompartment
631 shell protein, EtuB, associated with ethanol utilization in *Clostridium kluyveri*.
632 *Biochem. J.* 423, 199–207
- 633 38 Sagermann, M. *et al.* (2009) Crystal structure of the EutL shell protein of the
634 ethanolamine ammonia lyase microcompartment. *Proc. Natl. Acad. Sci. U. S. A.*
635 106, 8883–7
- 636 39 Tanaka, S. *et al.* (2009) Insights from multiple structures of the shell proteins
637 from the beta-carboxysome. *Protein Sci.* 18, 108–20
- 638 40 Crowley, C.S. *et al.* (2008) Structure of the PduU shell protein from the Pdu
639 microcompartment of *Salmonella*. *Structure* 16, 1324–32
- 640 41 Tsai, Y. *et al.* (2007) Structural analysis of CsoS1A and the protein shell of the
641 *Halothiobacillus neapolitanus* carboxysome. *PLoS Biol.* 5, e144
- 642 42 Kinney, J.N. *et al.* (2011) Comparative analysis of carboxysome shell proteins.
643 *Photosynth. Res.* 109, 21–32

- 644 43 Cai, F. *et al.* (2013) The Structure of CcmP, a Tandem Bacterial
645 Microcompartment Domain Protein from the β -Carboxysome, Forms a
646 Subcompartment Within a Microcompartment. *J. Biol. Chem.* 288, 16055–63
- 647 44 Sutter, M. *et al.* (2013) Two new high-resolution crystal structures of
648 carboxysome pentamer proteins reveal high structural conservation of CcmL
649 orthologs among distantly related cyanobacterial species. *Photosynth. Res.* doi,
650 10.1007/s11120-013-9909-z
- 651 45 Cai, F. *et al.* (2014) Engineering bacterial microcompartment shells: chimeric
652 shell proteins and chimeric carboxysome shells. *ACS Synth. Biol.* DOI:
653 10.1021/sb500226j
- 654 46 Rae, B.D. *et al.* (2012) Structural determinants of the outer shell of β -
655 carboxysomes in *Synechococcus elongatus* PCC 7942: roles for CcmK2, K3-K4,
656 CcmO, and CcmL. *PLoS One* 7, e43871
- 657 47 Takenoya, M. *et al.* (2010) Crystallographic insights into the pore structures
658 and mechanisms of the EutL and EutM shell proteins of the ethanolamine-
659 utilizing microcompartment of *Escherichia coli*. *J. Bacteriol.* 192, 6056–63
- 660 48 Pang, A. *et al.* (2012) Substrate channels revealed in the trimeric *Lactobacillus*
661 *reuteri* bacterial microcompartment shell protein PduB. *Acta Crystallogr. Sect.*
662 *D Biol. Crystallogr.* 68, 1642–1652
- 663 49 Samborska, B. and Kimber, M.S. (2012) A dodecameric CcmK2 structure
664 suggests β -carboxysomal shell facets have a double-layered organization.
665 *Structure* 20, 1353–62
- 666 50 Lassila, J.K. *et al.* (2014) Assembly of Robust Bacterial Microcompartment
667 Shells using Building Blocks from an Organelle of Unknown Function. *J. Mol.*
668 *Biol.* 426, 2217-2228
- 669 51 Parsons, J.B. *et al.* (2008) Biochemical and structural insights into bacterial
670 organelle form and biogenesis. *J. Biol. Chem.* 283, 14366–75
- 671 52 Parsons, J.B. *et al.* (2010) Characterisation of PduS, the *pdu* metabolosome
672 corrin reductase, and evidence of substructural organisation within the
673 bacterial microcompartment. *PLoS One* 5, e14009
- 674 53 Cheng, S. and Bobik, T. a (2010) Characterization of the PduS cobalamin
675 reductase of *Salmonella enterica* and its role in the Pdu microcompartment. *J.*
676 *Bacteriol.* 192, 5071–80

677 54 Cai, F. *et al.* (2009) The pentameric vertex proteins are necessary for the
678 icosahedral carboxysome shell to function as a CO₂ leakage barrier. *PLoS One*
679 4, e7521

680 55 Sampson, E.M. and Bobik, T.A. (2008) Microcompartments for B₁₂-Dependent
681 1,2-Propanediol Degradation Provide Protection from DNA and Cellular
682 Damage by a Reactive Metabolic Intermediate. *J. Bacteriol.* 190, 2966–71

683 56 Penrod, J.T. and Roth, J.R. (2006) Conserving a Volatile Metabolite: a Role for
684 Carboxysome-Like Organelles in *Salmonella enterica*. *J. Bacteriol.* 188, 2865 –
685 2874

686 57 Cheng, S. *et al.* (2012) The PduQ enzyme is an alcohol dehydrogenase used to
687 recycle NAD⁺ internally within the Pdu microcompartment of *Salmonella*
688 *enterica*. *PLoS One* 7, e47144

689 58 Huseby, D.L. and Roth, J.R. (2013) Evidence that a metabolic
690 microcompartment contains and recycles private cofactor pools. *J. Bacteriol.*
691 195, 2864–79

692 59 Chen, A.H. *et al.* (2013) The bacterial carbon-fixing organelle is formed by
693 shell envelopment of preassembled cargo. *PLoS One* 8, e76127

694 60 Peña, K.L. *et al.* (2010) Structural basis of the oxidative activation of the
695 carboxysomal gamma-carbonic anhydrase, CcmM. *Proc. Natl. Acad. Sci. U. S. A.*
696 107, 2455–60

697 61 Zarzycki, J. *et al.* (2013) Cyanobacterial-based approaches to improving
698 photosynthesis in plants. *J. Exp. Bot.* 64, 787–798

699 62 Roberts, E.W. *et al.* (2012) Isolation and characterization of the
700 *Prochlorococcus* carboxysome reveal the presence of the novel shell protein
701 CsoS1D. *J. Bacteriol.* 194, 787–95

702 63 Johnson, C.L. V *et al.* (2001) Functional Genomic , Biochemical, and Genetic
703 Characterization of the *Salmonella pduO* Gene, an ATP:Cob(I)alamin
704 Adenosyltransferase Gene. *J. Bacteriol.* 183, 1577–1584

705 64 Buan, N.R. *et al.* (2004) The *eutT* gene of *Salmonella enterica* encodes an
706 oxygen-labile, metal-containing ATP:corrinoid adenosyltransferase enzyme. *J.*
707 *Bacteriol.* 186, 5708–14

708 65 McGrath, J.M. and Long, S.P. (2014) Can the cyanobacterial carbon-
709 concentrating mechanism increase photosynthesis in crop species? A
710 theoretical analysis. *Plant Physiol.* DOI: 10.1104/pp.113.232611

711 66 Menon, B.B. *et al.* (2008) *Halothiobacillus neapolitanus* carboxysomes
712 sequester heterologous and chimeric RubisCO species. *PLoS One* 3, e3570

713 67 Baker, S.H. *et al.* (1998) Insertion Mutation of the Form I *cbbL* Gene Encoding
714 Ribulose Biphosphate Carboxylase / Oxygenase (RuBisCO) in *Thiobacillus*
715 *neapolitanus* Results in Expression of Form II RuBisCO , Loss of Carboxysomes
716 , and an Increased CO₂ Requirement fo. *J. Bacteriol.* 180, 4133–4139

717 68 Choudhary, S. *et al.* (2012) Engineered protein nano-compartments for
718 targeted enzyme localization. *PLoS One* 7, e33342

719 69 Li, Y. *et al.* (2005) Control of virus assembly: HK97 “Whiffleball” mutant
720 capsids without pentons. *J. Mol. Biol.* 348, 167–82

721 70 Newcomb, W. and Brown, J. (1991) Structure of the herpes simplex virus
722 capsid : effects of extraction with guanidine hydrochloride and partial
723 reconstitution of extracted capsids. *J. Virol.* 65, 613–620

724 71 Fan, C. *et al.* (2010) Short N-terminal sequences package proteins into
725 bacterial microcompartments. *Proc. Natl. Acad. Sci. U. S. A.* 107, 7509–14

726 72 Fan, C. and Bobik, T.A. (2011) The N-Terminal Region of the Medium Subunit
727 (PduD) Packages Adenosylcobalamin-Dependent Diol Dehydratase (PduCDE)
728 into the Pdu Microcompartment. *J. Bacteriol.* 193, 5623–5628

729 73 Fan, C. *et al.* (2012) Interactions between the termini of lumen enzymes and
730 shell proteins mediate enzyme encapsulation into bacterial
731 microcompartments. *Proc. Natl. Acad. Sci. U. S. A.* 109, 14995–5000

732 74 Sargent, F. *et al.* (2013) A synthetic system for expression of components of a
733 bacterial microcompartment. *Microbiology* 159, 2427–36

734 75 Zhao, X. *et al.* (1995) In vitro Assembly of Cowpea Chlorotic Mottle Virus from
735 Coat Protein Expressed in *Escherichia coli* and *in Vitro*-transcribed viral cDNA.
736 *Virology* 207, 486–494

737 76 Cameron, J.C. *et al.* (2013) Biogenesis of a bacterial organelle : the
738 carboxysome assembly pathway. *Cell* 155, 1131–1140

739 77 Cot, S.S.-W. *et al.* (2008) A multiprotein bicarbonate dehydration complex
740 essential to carboxysome function in cyanobacteria. *J. Bacteriol.* 190, 936–45

741 78 Kaneko, Y. *et al.* (2006) Intact Carboxysomes in a Cyanobacterial Cell
742 Visualized by Hilbert Differential Contrast Transmission Electron Microscopy
743 Intact Carboxysomes in a Cyanobacterial Cell Visualized by Hilbert

744 Differential Contrast Transmission Electron Microscopy. DOI:
745 10.1128/JB.188.2.805

746 79 Whitty, A. (2008) Cooperativity and biological complexity. *Nature* 4, 435–439

747 80 Marianayagam, N.J. *et al.* (2004) The power of two: protein dimerization in
748 biology. *Trends Biochem. Sci.* 29, 618–25

749 81 Marsh, J. a *et al.* (2013) Protein complexes are under evolutionary selection to
750 assemble via ordered pathways. *Cell* 153, 461–70

751 82 Tobimatsu, T. *et al.* (2005) The N-Terminal Regions of β and γ Subunits Lower
752 the Solubility of Adenosylcobalamin-Dependent Diol Dehydratase. *Biosci.*
753 *Biotechnol. Biochem.* 69, 455–462

754 83 Shibata, N. *et al.* (2010) Crystal structures of ethanolamine ammonia-lyase
755 complexed with coenzyme B12 analogs and substrates. *J. Biol. Chem.* 285,
756 26484–93

757 84 Sauvageot, N. *et al.* (2002) Purification, characterization and subunits
758 identification of the diol dehydratase of *Lactobacillus collinoides*. *Eur. J.*
759 *Biochem.* 269, 5731–5737

760 85 Leal, N.A. *et al.* (2003) PduP is a coenzyme-a-acylating propionaldehyde
761 dehydrogenase associated with the polyhedral bodies involved in B12-
762 dependent 1,2-propanediol degradation by *Salmonella enterica* serovar
763 Typhimurium LT2. *Arch. Microbiol.* 180, 353–61

764 86 Enright, A.J. *et al.* (1999) Protein interaction maps for complete genomes
765 based on gene fusion events. *Nature* 402, 86–90

766 87 Marcotte, E.M. (1999) Detecting protein function and protein-protein
767 interactions from genome sequences. *Science* 285, 751–753

768 88 Iancu, C. V *et al.* (2010) Organization, structure, and assembly of alpha-
769 carboxysomes determined by electron cryotomography of intact cells. *J. Mol.*
770 *Biol.* 396, 105–17

771 89 Lawrence, J.G. and Roth, J.R. (1996) Selfish operons: horizontal transfer may
772 drive the evolution of gene clusters. *Genetics* 143, 1843–60

773 90 Porwollik, S. *et al.* (2002) Evolutionary genomics of *Salmonella*: gene
774 acquisitions revealed by microarray analysis. *Proc. Natl. Acad. Sci. U. S. A.* 99,
775 8956–61

- 776 91 De Carvalho, L.P.S. *et al.* (2010) Metabolomics of *Mycobacterium tuberculosis*
777 reveals compartmentalized co-catabolism of carbon substrates. *Chem. Biol.* 17,
778 1122–31
- 779 92 Bonacci, W. *et al.* (2012) Modularity of a carbon-fixing protein organelle. *Proc.*
780 *Natl. Acad. Sci. U. S. A.* 109, 478–83
- 781 93 Lin, M.T. *et al.* (2014) β -carboxysomal proteins assemble into highly
782 organized structures in *Nicotiana* chloroplasts. *Plant J.* DOI:
783 10.1111/tpj.12536
- 784 94 Long, B.M. *et al.* (2007) Analysis of carboxysomes from *Synechococcus*
785 PCC7942 reveals multiple Rubisco complexes with carboxysomal proteins
786 CcmM and CcaA. *J. Biol. Chem.* 282, 29323–29335
- 787 95 Long, B.M. *et al.* (2010) Functional Cyanobacterial β -Carboxysomes Have an
788 Absolute Requirement for Both Long and Short Forms of the CcmM Protein.
789 *Plant Physiol.* 153, 285–293
- 790 96 Lin, M.T. *et al.* (2014) A faster Rubisco with potential to increase
791 photosynthesis in crops. *Nature* 513, 547–550
- 792 97 Luo, L.H. *et al.* (2011) Identification and characterization of the propanediol
793 utilization protein PduP of *Lactobacillus reuteri* for 3-hydroxypropionic acid
794 production from glycerol. *Appl. Microbiol. Biotechnol.* 89, 697–703
- 795 98 Anantharaman, V. *et al.* (2003) Emergence of diverse biochemical activities in
796 evolutionarily conserved structural scaffolds of proteins. *Curr. Opin. Chem.*
797 *Biol.* 7, 12–20
- 798 99 Yeates, T.O. *et al.* (2013) The shells of BMC-type microcompartment
799 organelles in bacteria. *J. Mol. Microbiol. Biotechnol.* 23, 290–9
- 800 100 Urano, N. *et al.* (2011) Genetic analysis around aminoalcohol dehydrogenase
801 gene of *Rhodococcus erythropolis* MAK154: a putative GntR transcription
802 factor in transcriptional regulation. *Appl. Microbiol. Biotechnol.* 89, 739–46
- 803 101 Kataoka, M. *et al.* (2006) A novel NADP⁺-dependent L-1-amino-2-propanol
804 dehydrogenase from *Rhodococcus erythropolis* MAK154: a promising enzyme
805 for the production of double chiral aminoalcohols. *Lett. Appl. Microbiol.* 43,
806 430–5

807

808

809

810 **Figure Legends**

811 **Figure 1. Bacterial microcompartments are metabolic modules embedded**
812 **within members of bacterial communities. (A)** Representation of a microbial
813 consortium, each member with a distinct complement of metabolic pathways and
814 sharing metabolites (filled in shapes). The metabolism of some members involves
815 BMCs (open hexagons) that require specific transporters (open triangles), and are
816 linked to central metabolism. **(B)** Cartoon representation of an icosahedral BMC ,
817 and how the three types of shell proteins fit together to form the shell. **(C) and (D)**
818 Characterized metabolic models for **(C)** carboxysome and **(D)** metabolosome
819 function, depicting the flux of specific metabolites across the shell, and integration
820 with cellular metabolism (e.g., Calvin-Benson cycle). In the carboxysome, ribulose
821 biphosphate is abbreviated RuBP, and 3-phosphoglyceraldehyde is abbreviated 3-
822 PGA. **(E)** Cartoon representation of the genetic and functional modularity of the
823 components of a BMC locus. Each symbol may represent one or more genes and/or
824 proteins (e.g., the core module for metabolosomes consists of the AldDH, AlcDH, and
825 PTAC enzymes). **(F)** Cartoon representation of how each genetic module is
826 constituted of one or more protein domain(s) (shown as puzzle pieces), with each
827 domain interacting with others either physically or through biochemical
828 intermediates. **(G)** Potential utility of the genetic, functional, and domain-based
829 modularity of BMC loci: a domain within the native signature enzyme could be
830 replaced by a modified structurally similar domain with novel function.

831

832 **Figure 2. Known building blocks of BMC shells.** For all shell proteins, individual
833 polypeptide chains are colored differently, with the lysine of the D/N-X-X-X-K motif
834 colored red for BMC-H and BMC-T proteins. **(A)** Views of PduA (PDB ID: 3NGK), a
835 representative BMC-H protein. **(B)** Views of CsoS1D (PDB ID: 3F56), a
836 representative BMC-T protein of the clade that forms stacked trimers. The views
837 show the potential gating mechanism and the compartment formed between the
838 stacked trimers. **(C)** Views of CcmL (PDB ID: 2QW7), a representative BMC-P
839 protein.

840

841 **Figure 3. Proposed BMC assembly pathways.** **(A)** Shell first: initially an empty
842 shell is formed (presumably by self-assembly), after which encapsulated proteins
843 and enzymes are targeted to the BMC shell via the encapsulation peptide. **(B)** Core
844 first: a protein and enzyme core is formed (e.g., pro-carboxysome), around which
845 the BMC shell assembles, precipitated by the encapsulation peptide. **(C)**
846 Concomitant shell and core assembly: core proteins and enzymes associate with
847 shell subunits in simultaneous self-assembly. **(D)** Example micrographs of synthetic
848 empty shells, **(E)** partial BMCs (black arrow), and **(F)** whole β -carboxysomes in
849 cyanobacteria.

850 **Table 1. Biochemical and phylogenetic diversity of bacterial microcompartments**

Name	Function	Signature enzyme	Metabolites	Cofactors	Present in phyla	Evidence	Refs
α- and β-Carboxysome	Carbon fixation	RuBisCO, carbonic anhydrase	Ribulose biphosphate, CO ₂ ^a , 3-phosphoglycerdehyde	ATP ^b	Actinobacteria, Cyanobacteria, Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria	Genetic ^c , EM ^d	[4,61]
PDU	Propanediol	Vitamin B12-dependent propanediol dehydratase	Propanediol, propionaldehyde ^a , propionyl-CoA ^a , propionyl phosphate, propionate	Vitamin B ₁₂ , NAD, CoA, flavin	Actinobacteria, Firmicutes, Fusobacteria, Gammaproteobacteria, Deltaproteobacteria, Spirochaetes, Synergistetes	Genetic ^c , EM ^d	[3,14,16,52,53,99]
PDU-like					Firmicutes	Genomic ^e	[14]
EUT1-3	Ethanolamine	Vitamin B12-dependent ethanolamine ammonia lyase	Ethanolamine, acetaldehyde ^a , acetyl-CoA ^a , acetyl phosphate, acetate	Vitamin B ₁₂ , NAD, CoA,	Actinobacteria, Chloroflexi, Firmicutes, Fusobacteria, Gammaproteobacteria, Synergistetes	Genetic ^c , EM ^d	[3,14,16,17]
PDU/EUT fusion	Ethanolamine and propanediol	Same as PDU and EUT	Same as PDU and EUT	Same as PDU and EUT	Actinobacteria, Firmicutes, Fusobacteria	Genomic ^e , transcriptional	[14]
ETU	Ethanol	N/A	Ethanol, acetaldehyde ^a , acetate	NAD ^b , CoA ^b	Firmicutes	EM ^d , genomic ^e	[3]
GRP^f/GRM5	Fucose/rhamnose – anaerobic	SAM-dependent propanediol dehydratase (GRE) ^g	Fucose/rhamnose, DHAP ^h , lactaldehyde, propanediol, propionaldehyde ^a , propionyl-CoA ^a , propionyl-phosphate, propionate	SAM ^b , NAD ^b , CoA ^b , flavin ^b	Firmicutes	Transcriptional, EM ^e	[7,14,16,17]
CUT/GRM1	Choline	SAM-dependent choline lyase	Choline, trimethylamine, acetaldehyde ^a , acetyl-CoA ^a , acetyl-	SAM ^b , NAD ^b , CoA ^b , flavin ^b	Actinobacteria, Gammaproteobacteria, Firmicutes, Deltaproteobacteria	Genetic ^c , genomic ^e	[5,6,14]

		(GRE) ^g	phosphate, acetate				
GRM2	Unknown	GRE ^{b, g}	Aldehyde ^b , alcohol ^b , acyl phosphate ^b	SAM ^b , NAD ^b , CoA ^b	Gammaproteobacteria	Genomic ^e	[14]
GRM3	Unknown	GRE ^{b, g}	Aldehyde ^b , alcohol ^b , acyl phosphate ^b	SAM ^b , NAD ^b , CoA ^b , flavin ^b	Firmicutes, Alphaproteobacteria, Gammaproteobacteria	Genomic ^e	[14]
GRM4	Unknown	GRE ^{b, g}	Aldehyde ^b , alcohol ^b , acyl phosphate ^b	SAM ^b , NAD ^b , CoA ^b	Gammaproteobacteria	Genomic ^e	[14]
PVM	Fucose/rhamnose – aerobic	Aldolase	Fucose/rhamnose, DHAP ^h or Pyruvate, lactaldehyde ^a , lactyl-CoA ^a , lactyl phosphate, lactate	NAD ^b , CoA ^b	Planctomycetes, Verrucomicrobia,	Genetic ^c , EM ^d	[9]
PVM-like					Acidobacteria, Firmicutes, Ignavibacteria	Genomic ^e	[14]
SPU	Sugar phosphate ^b	Aldolase ^b	Sugar phosphate(s) ^b , aldehyde ^b , acyl phosphate ^b	NAD ^b , CoA ^b	Atribacteria, Gemmatimonadetes, Marinimicrobia	Genomic ^e	[14]
SPU-like					Latescibacteria,	Genomic ^e	[14]
RMM1, 2	Amino-2-propanol ^b	Aminotransferase ^b	Amino-2-propanol ^b , aminoacetone ^b , methylglyoxal ^b , pyruvyl CoA ^b	NAD ^b , CoA ^b , flavin ^b (RMM2)	Actinobacteria	Genetic ^c , genomic ^e	[100,101]
MIC	Unknown	No prediction	Aldehyde ^b	NAD ^b , CoA ^b	Actinobacteria, Firmicutes, Caldithrix	Genomic ^e	[14]
MUF	Unknown	No prediction	Aldehyde ^b , alcohol ^b , acyl phosphate ^b	NAD ^b , CoA ^b , flavin ^b	Firmicutes	Genomic ^e	[14]
BUF	Unknown	No prediction	No prediction	No prediction	Firmicutes	Genomic ^e	[14]
Satellite loci	N/A	N/A	N/A	N/A	Cyanobacteria, Acidobacteria, Firmicutes, Fusobacteria, Ignavibacteria, Planctomycetes, Proteobacteria, Verrucomicrobia, BRC1, Poribacteria, Caldithrix	Genomic ^e	[14]

851 ^a Sequestered metabolite.

852 ^b Prediction based on bioinformatics and/or related studies (that did not directly examine BMC function); see cited references.

853 ^c Genetic studies.

854 ^d Electron microscopy.

855 ^e Genomic evidence refers to bioinformatic evidence based on genomes only.

856 ^f GRP, Glycyl radical-based propanediol [17]

857 ^g GRE, glycyl radical enzyme.

858 ^h DHAP, dihydroxyacetone phosphate.

859

Box 1. Milestones in BMC research

While the bacterial microcompartment field is almost 60 years old, the field has seen exponential growth only in the past decade (Figure I). Select milestones in the BMC field are highlighted below:

Figure I. Number of publications related to BMCs over time. Values were generated by searching www.scopus.com for publications with titles, abstracts, or keywords matching the query: metabolosome, carboxysome, bacterial microcompartment, (polyhedral AND Rubisco, Cyanobacteria, propanediol, or ethanolamine), but not virus; commas represent the Boolean operator OR. The publications predating 1973 were added in manually. Solid orange represents number of publications as of June 2014. The open orange symbol represents the extrapolated value for publications in 2014, generated by dividing the current number of publications in 2014 by 6 months, and multiplying by 12 months.

1956:

- Polyhedral bodies described in Cyanobacteria.

1973:

- Polyhedral bodies contain RuBisCO, and are named carboxysomes.

1998:

- PDU and EUT operons sequenced, found to encode homologs to carboxysome shell proteins.

2005:

- First BMC shell protein structure solved and bioinformatics survey reveal prevalence of BMCs.

2008:

- BMC-P structures of the carboxysome shell solved, icosahedral model for BMC shells based on geometrical constraints proposed.
- Chimeric RuBisCO encapsulated into carboxysomes.
- First empty BMC shells formed.
- Higher order structures observed when BMC shells are overexpressed.

- 883 • First heterologous expression of an entire BMC operon.
- 884 2009
- 885 • First BMC-T structure determined indicating that shell proteins can be gated.
- 886
- 887 2010:
- 888 • Carboxysomes and PDU metabolosomes labeled with GFP.
- 889 • BMC encapsulation peptide identified.
- 890 2012:
- 891 • BMC with a glycyl-radical enzyme implicated in choline degradation (CUT).
- 892 • Foreign proteins are targeted to BMC shells using the encapsulation
- 893 peptide.
- 894 • Encapsulation peptides features identified and are proposed to be shared by many
- 895 different BMC enzymes.
- 896 2013:
- 897 • β -Carboxysome assembly pathway elucidated.
- 898 • BMC with a glycyl-radical enzyme implicated in anaerobic fucose/rhamnose to
- 899 propanediol degradation (GRP).
- 900 2014:
- 901 • BMC implicated in aerobic fucose/rhamnose degradation (PVM).
- 902 • Robust synthetic shells from obscure BMC locus produced and purified.
- 903 • Chimeric shells produced.
- 904 • Carboxysome proteins expressed and assembled in chloroplasts.
- 905 • A comprehensive taxonomy of BMC loci is described.
- 906
- 907

908 **Box 2. Outstanding questions**

- 909 • What function(s) do(es) the BMC play in host–microbe interactions?
- 910 • What are the physical characteristics of the shell (e.g., mechanical properties or
- 911 tensile strength)?
- 912 • Is the shell a single layer of protein, double layer, or a mixture?
- 913 • Which proteins are in the lumen versus associated with the shell; are some
- 914 associated with the external surface?
- 915 • What is the sidedness of the shell proteins: concave- vs. convex-in, or a mixture of
- 916 both orientations?
- 917 • What is the purpose of the multiple BMC-P paralogs observed in some loci?
- 918 • What other functional paradigms for BMCs (other than carbon-fixing or aldehyde-
- 919 degrading) exist?
- 920 • Is metabolosome assembly similar to β -carboxysome assembly (core first)?
- 921 • How dynamic is the shell; does the composition change with environmental
- 922 conditions?
- 923 • What are the flux properties of shells?
- 924 • Are all BMCs descendants of a common ancestor?
- 925 • What is the structure of intact organelle – what are the specific pairwise interactions
- 926 among the shell proteins that noncovalently hold the shell together? What
- 927 constitutes the edges between facets?

928

Figure 1

Figure 2

Figure 3