BACTERIAL MICROCOMPARTMENT VIRUS-LIKE PARTICLES

FIELD OF THE INVENTION

The present invention relates to a bacterial microcompartment virus-like particle (VLP) carrying a cargo molecule, a method for producing the bacterial microcompartment VLP, an isolated plasmid or vector nucleic acid used to produce the VLP, a composition comprising at least one said VLP, a use of the said VLP, and a method of treatment using the said VLP.

BACKGROUND

Bacterial microcompartments (BMCs) are protein shells found in some species of bacteria and are thought to have evolved as a strategy to compartmentalize certain challenging biochemical reactions [Kerfeld, C. A., *et al.*, *Nature Reviews Microbiology* 16: 277-290 (2018)]. These protein complexes are comprised of hundreds to thousands of polypeptide subunits that self-assemble into polygonal structures, with diameters ranging from 40 nm to 400 nm. The carboxysome, found in cyanobacteria and some chemotrophic bacteria species, is the earliest known example of a BMC. This protein shell encapsulates ribulose-1,5-bisphosphate carboxylase (RuBisCO) and improves its catalytic efficiency by concentrating its substrates, CO_2 and ribulose-1,5-bisphosphate, in close proximity to RuBisCO. Carboxysomes are classified into two main groups depending on the class of RuBisCO that is encapsulated. Alpha-carboxysomes contain Form 1A RuBisCO that is found in α-cyanobacteria (generally saltwater cyanobacteria) and chemoautotrophs, while beta-carboxysomes house Form 1B RuBisCO that are observed in β-cyanobacteria (typically freshwater cyanobacteria) [Turmo, A., *et al.*, *FEMS Microbiol Lett* 364: (2017)].

Numerous atomic-scale structures of the subunits that comprise BMC shells, along with that of three intact shells – a reduced component beta-carboxysome from *Halothece* sp. PCC 7418, a synthetic glycyl-radical associated BMC group 2 (GRM2) from *Klebsiella pneumoniae* and a BMC of undetermined function from *Haliangium ochraceum* (HO-BMC) – have been reported in recent years [Kalnins, G., *et al.*, *Nature Communications* 11: 388 (2020); Sutter, M., *et al.*, *Science 356*(6344): 1293-1297 (2017); Sutter, M., *et al.*, *Plant Physiology* (2019)]. Despite the diversity in the appearances and functions of BMCs, the tertiary structures of the main building blocks

are conserved. The BMC-H domain protein (pfam00936) is the stoichiometrically major module and forms a homo-hexamer with C_6 geometry. The BMC-T protein is formed by a tandem repeat of two BMC-H domains and assembles as a trimer, or as a double stack of appressed trimers, with pseudo-hexagonal symmetry. The BMC-P domain unit (pfam03319) is a minor but important module in a BMC shell complex. The BMC-P protomers assemble into homo-pentamers with pyramidal geometry that occupy the vertices of the shell, capping the flat facets formed by the BMC-T and -H proteins. This gives rise to the polygonal appearance of BMCs. A detailed molecular understanding of the architecture of the components has contributed to the field of BMC shell engineering. Such endeavors include targeting heterologous protein cargo into the shell lumen by the use of encapsulation peptides (EP), which are short peptide sequences derived from cognate luminal proteins, or by protein engineering of the shell components [Lawrence, A. D., *et al.*, *ACS Synthetic Biology* 3: 454-465 (2014)]. These modifications are aimed at repurposing BMCs into intracellular nanoreactors or as scaffolds for delivery of biomolecules

The alpha-carboxysome from *Halothiobacillus neapolitanus* has been previously produced in *Escherichia coli* by transplantation of the entire alpha-carboxysome operon (*cso*) (Figure 1) into the recombinant host [Bonacci, W., *et al. PNAS* 109: 478-483 (2012)]. Genes found in the operon include three BMC-H paralogs (*cso1ABC*), one BMC-T protein (*csoS1D*) and two BMC-P paralogs (*csoS4AB*), along with the genes that encode the RuBisCO large and small units (*cbbLS*), carbonic anhydrase (*csoS3 / SCA*) and an intrinsically disordered protein (IDP), *csoS2*. This IDP is known to be critical for alpha-carboxysome assembly by facilitating interactions between the shell and luminal proteins [Cai, F. *et al. Life (Basel, Switzerland)* 5: 1141-1171 (2015)]. BMCs bereft of their native cargo are more suited for engineered utility, since heterologous cargo can be packaged more efficiently into the lumen. However, the *H. neapolitanus* alpha-carboxysome has never been recombinantly expressed in a structurally enclosed form with fewer than the aforementioned ten genes, despite decades of study into its structure and biochemical processes [Bonacci, W., *et al. PNAS* 109: 478-483 (2012)].

There is a need to provide bacterial microcompartment virus-like particles with improved efficiency of production in recombinant bacterial and yeast hosts, and alternative ways to encapsulate and/or present on the surface cargo molecules.

SUMMARY OF THE INVENTION

It has surprisingly been found that BMC VLPs can be formed using only two or three types of BMC protomers from *Halothiobacillus neapolitanus* and *Haliangium ochraceum*, respectively. The *H. neapolitanus* BMC VLPs are termed Cso-BMC and the *H. ochraceum* BMC VLPs are termed HO-BMC. In addition, cargo molecules may be encapsulated within the Cso-BMC using a novel short peptide derived from CsoS2 and termed S2CP or a variant thereof termed S2CP(30). Cso-BMCs which encapsulated a cargo molecule had a distinct shell conformation not observed in Cso-BMCs that did not encapsulate a cargo molecule. Of note, the peptide termini of the protomers of both shells face outwards, allowing genetic fusion of proteins of interest. Therefore, a cargo molecule may be displayed on the outer surface of the BMC VLPs of the invention by expressing it fused to the terminal end of a protomer or via the cargo molecule having a complementary binding partner to a biochemical tag attached to a protomer terminal end.

According to a first aspect, the present invention provides a method for producing a bacterial microcompartment virus-like particle (VLP) carrying a cargo molecule, said method comprising

A) introducing into a host cell or organism one or more heterologous polynucleotides comprising (i) a first sequence encoding bacterial microcompartment shell protomers; and (ii) a second sequence encoding a cargo molecule fused to an encapsulation peptide, wherein the encapsulation peptide comprises the amino acid sequence set forth in SEQ ID NO: 1 (SKITGSSGNDTQGSLITYSGGARG) or SEQ ID NO: 94 (KPEKPGSKITGSSGNDTQGSLITYSGGARG), or a functional variant thereof;

a) expressing the first and second sequences; and

b) forming a microcompartment that encapsulates the cargo molecule; or

B) introducing into a host cell or organism one or more polynucleotides comprising (i) a first sequence encoding bacterial microcompartment shell protomers; and (ii) a second sequence encoding at least one of said protomers fused with a cargo molecule or a biochemical tag;

a) expressing the first and second sequences; and

b) forming a microcompartment that expresses the cargo molecule on an exterior surface, or forming a microcompartment that expresses the biochemical tag on an exterior surface to which a cargo molecule comprising a complementary tag can bind.

In some embodiments, a functional variant of the encapsulation peptide set forth in SEQ ID NO: 1 comprises, at its amino terminus, 1, 2, 3, 4, or 5 of the additional amino acids at the amino terminus of SEQ ID NO: 94. For example, a variant of the encapsulation peptide of SEQ ID NO: 1 could comprise, at its amino terminus, 'G', 'PG', 'KPG' etc., and retain function. Such variants are intermediates between the sequences SEQ ID NO: 1 and SEQ ID NO: 94.

In some embodiments, said encapsulation peptide is encoded by a polynucleotide sequence having at least 80% identity, at least 85% identity, at least 90% identity, at least 95% identity, or 100% identity with the nucleic acid sequence set forth in SEQ ID NO: 7 or SEQ ID NO: 95 (S2CP(30)), respectively, due to redundancy of the genetic code.

In some embodiments, the bacterial microcompartment protomers are derived from *Halothiobacillus neapolitanus* or *Haliangium ochraceum*.

In some embodiments, the bacterial microcompartment protomers are CsoS1A (SEQ ID NO: 2) and CsoS4A (SEQ ID NO: 3) from *Halothiobacillus neapolitanus*; or HO-H (SEQ ID NO: 4), HO-P (SEQ ID NO: 5) and HO-T1 (SEQ ID NO: 6) from *Haliangium ochraceum*, and variants thereof.

In some embodiments, the cargo molecule is at least one peptide, such as an enzyme and/or fluorescent protein and/or immunogenic peptide.

In some embodiments, the biochemical tag may be selected from the group comprising Strep-Tag II (SII), SpyCatcher / SpyTag (SC / ST) pair and CC-Di-A / B (CCA / CCB) pair.

In some embodiments, the expression of CsoS1A is controlled by promoter PT7; CsoS4A is controlled by promoter P_{CON5}; HO-H is controlled by yeast promoter P_{TDH3}; HO-P is controlled by yeast promoter P_{PYK1} and HO-T1 is controlled by yeast promoter P_{YEF3}.

In some embodiments, the host organism is *E. coli* or *S. cerevisiae*.

According to a second aspect, the present invention provides an engineered bacterial microcompartment VLP carrying a cargo molecule, comprising: i) bacterial microcompartment shell protomers, and a cargo molecule fused to an encapsulation peptide, wherein the encapsulation peptide comprises the amino acid sequence set forth in SEQ ID NO: 1 (SKITGSSGNDTQGSLITYSGGARG) or SEQ ID NO: 94 (KPEKPGSKITGSSGNDTQGSLITYSGGARG), or a functional variant thereof; or ii) bacterial microcompartment shell protomers and a cargo molecule, wherein the cargo molecule is fused to an end of at least one of said protomers, or wherein at least one of said protomers is fused to a tag and a cargo molecule comprising a complementary tag is bound to it on the exterior surface of the VLP.

In some embodiments, the bacterial microcompartment protomers are derived from *Halothiobacillus neapolitanus* or *Haliangium ochraceum*.

In some embodiments, the bacterial microcompartment protomers are CsoS1A, comprising the amino acid sequence set forth in SEQ ID NO: 2, and CsoS4A, comprising the amino acid sequence set forth in SEQ ID NO: 3, from *Halothiobacillus neapolitanus*; or HO-H comprising the amino acid sequence set forth in SEQ ID NO: 4, HO-P comprising the amino acid sequence set forth in SEQ ID NO: 5 and HO-T1 comprising the amino acid sequence set forth in SEQ ID NO: 6 from *Haliangium ochraceum*, and variants thereof.

In some embodiments, the cargo molecule is at least one peptide, such as an enzyme and/or fluorescent protein and/or immunogenic peptide.

In some embodiments, the biochemical tag may be selected from the group comprising Strep-Tag II (SII), SpyCatcher / SpyTag (SC / ST) pair and CC-Di-A / B (CCA / CCB) pair.

According to a third aspect, the present invention provides an isolated plasmid or vector nucleic acid comprising:

a) a first DNA sequence that encodes bacterial microcompartment shell protomers, each of which is operably linked to a promoter, and

b) a second DNA sequence that encodes a cargo molecule fused to an encapsulation peptide, operably linked to a promoter, wherein the encapsulation peptide comprises the amino acid sequence set forth in SEQ ID NO: 1 (SKITGSSGNDTQGSLITYSGGARG) or SEQ ID NO: 94 (KPEKPGSKITGSSGNDTQGSLITYSGGARG), or a functional variant thereof; or

c) a first DNA sequence that encodes bacterial microcompartment shell protomers, each of which is operably linked to a promoter, and

d) a second DNA sequence encoding at least one of said protomers fused with a cargo molecule or a biochemical tag.

In some embodiments, the isolated plasmid or vector nucleic acid comprises bacterial microcompartment shell protomers, promoters, cargo molecules and tags as defined previously.

In some embodiments, the isolated plasmid or vector nucleic acid DNA sequences encoding said bacterial microcompartment shell protomers, cargo molecules and tags set forth in SEQ ID Nos: 1-6 and 94 have at least 70%, at least 80%, at least 90%, or 100% identity with the nucleic acid sequences set forth in SEQ ID Nos 7-12 and 95-S2CP(30), respectively, due to redundancy of the genetic code.

According to a fourth aspect, the present invention provides a composition or combination comprising at least one engineered VLP of any aspect of the invention for use in: a) the prophylaxis or treatment of disease in a subject; or b) a biochemical process.

In some embodiments, the at least one engineered VLP comprises an enzyme for conversion of a prodrug.

In some embodiments, the composition may comprise one or more additional therapeutic agents. The composition may be used as a vaccine.

According to a fifth aspect, the present invention provides a use of at least one engineered VLP of any aspect of the invention in the manufacture of a medicament for the prophylaxis or treatment of a disease in a subject.

According to a sixth aspect, the present invention provides a method of prophylaxis or treatment, comprising administering to a subject in need of such treatment an efficacious amount of an engineered VLP of any aspect of the invention.

It will be appreciated that the present invention is not limited to the specific embodiments described in detail below.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows a schematic of the alpha-carboxysome operon (cso) from *Halothiobacillus neapolitanus*. The dotted line indicates ten genes between csos1B and csoS1D that are unlikely to be associated with the carboxysome. Gene lengths and intervening distances are not drawn to scale.

Figure 2A-D shows schematics of developed plasmids. (**A**) TU acceptor plasmid, pES*X* contains a streptomycin selection marker ($Strep^R$) along with the *pUC* origin of replication. The RFP cassette is displaced by an incoming TU through digestion with the restriction enzyme (RE) BsmBI. (**B**) The pathway acceptor plasmid, pCKH, accepts TUs released from pES*X* plasmids by digestion with the RE BsaI. pCKH contains the kanamycin selection marker (Kan^R). (**C and D**) Modified HcKan_O plasmids that can attach either the SII or His_6 tag or one of four FPs (mT2, meGFP, mKOκ, mCh) at either N- or C-terminus of an ORF. After insertion of the ORF by BsaI, the ORF-tag fusion product (separated by a Gly-Ser-Ser linker), is released by BsmBI.

Figure 3 shows a schematic diagram of VLP pathways created. For Cso-BMC, all terminators used were T_{T7}. The greyscale intensities on the promoter arrows symbolize their relative strengths, with darker being stronger.

Figure 4A-B shows identification of a cargo targeting peptide sequence for the alpha-carboxysome system. (**A**) Multiple sequence alignment of CsoS2 orthologs (top 9 from different genera are shown) with that from *H. neapolitanus* reveals that the C-terminal region is highly conserved, as depicted by the sequence logo. (**B**) Pull-down assay of His6-meGFP-S2CP with the shell proteins CsoS1A-SII, SII-CsoS1D and CsoS4A-SII demonstrate that S2CP mediates interaction with CsoS1A-SII only.

Figure 5A-B shows the expression and purification of alpha-carboxysome shell components. (**A**) Schematics of the synthetic operons used to express the alpha-carboxysome components. The shell modules are also represented by their geometric

icons; Cso4A: pentagon; CsoS1D: trimeric hexagon; CsoS1A: hexameric hexagon. (**B**) Fluorescence micrographs of cells expressing the pathway Cso-P_{mCh}THC. Co-localization of meGFP-S2CP and CsoS4A-mCherry can be seen. DIC: differential interference contrast channel. Scale bars (white, bottom right) represent 2 µm. (**C-F**) TEM visualization of purified protein shells at the 0.4 M NaCl elution fraction following AIEX purification for (C) Cso-P_{mCh}THC, (D) Cso-P_{SII}THC, (E) Cso-P_{SII}TH and (F) Cso-P_{SII}H. Scale bars (black, bottom right) represent 50 nm.

Figure 6A-D shows how S2CP acts as an encapsulation peptide. (**A**) Schematic depicting how encapsulation of the $UmuD^{1-40}$ protease signal tagged GFP by S2CP may protect it from the endogenous ClpXP protease. (**B**) S2CP was able to target $UmuD^{1-40}$-meGFP into the lumen of the simplified carboxysome. Purified shells were subject to Western blot analysis by use of an anti-GFP antibody. $UmuD^{1-40}$meGFP was detected only from Cso-$P_{SII}THC_{U,S2CP}$ and not Cso-$P_{SII}THC_{U}$. Electron micrographs show that shells produced by (**C**) Cso-$P_{SII}THC_{U,S2CP}$ and (**D**) Cso-$P_{SII}THC_{U}$ were similar. Scale bars (black, bottom right) represent 50 nm.

Figure 7A-B shows atomic models of the simplified alpha carboxysome shells. (**A, B**) Surface representation of the shells, with the CsoS1A colored grey and CsoS4A in light grey. The right and down arrows over light grey indicate N- and C-termini of a CsoS4A monomer, respectively, and the upward and right arrows over grey those of a CsoS1A monomer, respectively.

Figure 8A-B shows fluorescence micrographs of the shell probes (**A**) CsoS4A-mCherry and (**B**) meGFP-S2CP used for investigating interactions between shell components in-vivo. Probes were generally homogenously distributed within the cytosol when expressed individually. Scale bars (bottom right) represent 2 µm.

Figure 9A-C shows anion-exchange (AIEX) chromatograms for the affinity-purified proteins from (**A**) Cso-P_{mCh}THC pathway construct, (**B**) Cso-P_{SII}THC and (**C**) CsoS4A-SII. Blue trace (left Y-axis) indicates absorbance (mAU) at 280 nm while green trace (right Y-axis) indicates percentage of AIEX buffer B (Tris 50 mM, NaCl 1.0 M, pH 7.9) used for the indicated elution volume. TEM micrographs on the right are viewings of elution fractions obtained at 0.3 M NaCl. It can be seen that CsoS4A-SII, expressed by itself, does not form protein shells. Scale bar (bottom right) represents 50 nm.

Figure 10A-B shows purification of Cso-P_{SII}TH. (**A**) AIEX chromatogram and (**B**) TEM micrograph for the elution fraction corresponding to 0.3 M NaCl. Scale bar (bottom right) represents 50 nm.

Figure 11A-E shows purification of Cso-P_{SII}H. (**A**) AIEX chromatogram for affinity purified proteins from Cso-P_{SII}H. (**B**) TEM micrograph for the elution fraction from Cso-P_{SII}H corresponding to 0.3 M NaCl. TEM micrographs for (**C**) CsoS1A-SII, (**D**) CsoS1A-SII co-expressed with CsoS1D and (**E**) CsoS4A-SII co-expressed with CsoS1D demonstrating that these combinations do not form protein shells. Scale bars (bottom right) represent 50 nm.

Figure 12A-D shows sodium dodecyl sulfate – polyacrylamide gel electrophoresis (SDS-PAGE) analyses of fractions collected from AIEX purification of (**A**) Cso-P_{mCh}THC, (**B**) P_{SII}THC, (**C**) P_{SII}TH and (**D**) P_{SII}H. Arrows indicate fractions used for TEM analysis, with the left and right arrows corresponding to [NaCl] = 0.3 M and 0.4 M respectively. Protein ladder lane is marked as L with the masses (kDa) indicated.

Figure 13 shows particle size distribution of protein shells as measured by dynamic light scattering. (A) Cso-P_{mCh}THC, (B) Cso-P_{SII}THC, (C) Cso-P_{SII}TH and (D) Cso-P_{SII}H.

Figure 14 shows a table summarizing the differences between the purified Cso-BMC and HO-BMC.

Figure 15 shows a close-up view of the Cso-BMC exterior demonstrating that the N- and C-termini of the hexameric subunits (grey) and pentameric subunits (light grey) point away from the shell lumen. The N-termini and C-termini of a selected hexameric chain and a pentameric chain are indicated by arrows. The topologies of the Cso-BMC hexameric and pentameric subunits shown are representative of the HO-BMC.

Figure 16 shows densitometric analysis of Cso-P_{SII}H shells co-expressed with either UmuD$^{1-40}$-GFP-S2CP or UmuD$^{1-40}$-GFP-S2CP(30). Approximately the same amount of shells (as judged by band peak area) were loaded per shell sample so that the relative amounts of UmuD$^{1-40}$-GFP-S2CP and UmuD$^{1-40}$-GFP-S2CP(30) could be directly compared. Arrows indicate UmuD$^{1-40}$-GFP-S2CP and UmuD$^{1-40}$-GFP-S2CP(30).

Figure 17A-D shows assessment of the stability of Cso-BMC shells against common denaturing factors. (**A-D**) DLS spectra of empty Cso-BMC tested against the conditions indicated. The baseline is vertically displaced by 0.2 for each subsequent spectrum so that all spectra can be seen in one graph.

Figure 18A-E shows loading of the APEX2 and LacZ enzymes into Cso-BMC shells. (**A-B**) SDS-PAGE and Western blot analyses (using anti-His_6 antibody) of Cso-BMC co-expressed with the enzymes. (**C-D**) TEM micrographs of Cso-BMC loaded with enzymes. Scale bars (black, bottom right) represent 50 nm. (**E**) DLS spectra of Cso-BMC co-expressed with enzymes, with empty Cso-BMC shells as a reference.

Figure 19 shows Michaelis-Menten kinetics of free and Cso-BMC encapsulated APEX2 and LacZ enzymes.

Figure 20A-D shows assessment of stabilizing effects of Cso-BMCs conferred on APEX2 and LacZ against denaturing conditions. Residual enzymes activities of free and encapsulated (+ shell) enzymes were obtained by normalizing the activity to that of the pristine sample, shown as (**A**) 23 °C, (**B**) 0% v/v methanol, (**C**) no freeze−thawing, and (**D**) pH 8. Error bars represent one standard deviation of the mean.

Figure 21A-E shows the purification of the HO-BMC shells: HO-HTP and HO-HT_{ST}P+GFP-SpyCatcher. (**A-B**) SDS-PAGE analysis of purified shells, (**C**) Western blot analysis (using anti-GFP antibody) indicating the presence of GFP-SpyCatcher in the HO-HT_{ST}P+GFP-SpyCatcher sample, (**D-E**) TEM micrographs of both HO-BMC constructs. Scale bars (black, bottom right) represent 50 nm.

DETAILED DESCRIPTION OF THE INVENTION

Bibliographic references mentioned in the present specification are for convenience listed in the form of a list of references and added at the end of the examples. The whole content of such bibliographic references is herein incorporated by reference.

Definitions

Certain terms employed in the specification, examples and appended claims are collected here for convenience.

As used herein, the term “comprising” or “including” is to be interpreted as specifying the presence of the stated features, integers, steps or components as referred to, but does not preclude the presence or addition of one or more features, integers, steps or components, or groups thereof. However, in context with the present disclosure, the term “comprising” or “including” also includes “consisting of”. The variations of the word “comprising”, such as “comprise” and “comprises”, and “including”, such as “include” and “includes”, have correspondingly varied meanings.

As used herein, the term Cso-P_{SII}H is used interchangeably with the term Cso-BMC.

The term "variant" as used herein, refers to an amino acid sequence that is altered by one or more amino acids, but retains the ability to function as an encapsulation peptide in the present invention. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties (e.g., replacement of leucine with isoleucine). More rarely, a variant may have "non-conservative" changes (e.g., replacement of glycine with tryptophan). Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, DNASTAR® software (DNASTAR, Inc. Madison, Wisconsin, USA). A type of variant is, for example, a peptide having the amino acid sequence set forth in SEQ ID NO: 94, which is longer than the sequence set forth in SEQ ID NO: 1, is also derived from CsoS2, and retains the encapsulation functionality of SEQ ID NO: 1. Other variants having amino acid sequences intermediate between SEQ ID NO: 1 and SEQ ID NO: 94, would be expected to retain functionality.

A composition or combination of the present invention will generally be administered as a pharmaceutical formulation in admixture with a pharmaceutically acceptable adjuvant, diluent or carrier, which may be selected with due regard to the intended route of administration and standard pharmaceutical practice. Such pharmaceutically acceptable carriers may be chemically inert to the active compounds and may have no detrimental side effects or toxicity under the conditions of use. Suitable pharmaceutical formulations may be found in, for example, Remington The Science and Practice of Pharmacy, 19th ed., Mack Printing Company, Easton, Pennsylvania (1995). For parenteral administration, a parenterally acceptable aqueous

solution may be employed, which is pyrogen free and has requisite pH, isotonicity, and stability. Suitable solutions will be well known to the skilled person, with numerous methods being described in the literature. A brief review of methods of drug delivery may also be found in e.g. Langer, (Science 249: 1527 (1990)).

Otherwise, the preparation of suitable formulations may be achieved routinely by the skilled person using routine techniques and/or in accordance with standard and/or accepted pharmaceutical practice.

The amount of a composition or combination in any pharmaceutical formulation used in accordance with the present invention will depend on various factors, such as the severity of the condition to be treated, the particular patient to be treated, as well as the compound(s) which is/are employed. In some embodiments the BMC-VLP displays an antigenic molecule on its surface and functions as a vaccine. In any event, the amount of a composition or combination in the formulation may be determined routinely by the skilled person.

For example, a solid oral composition such as a tablet or capsule may contain from 1 to 99% (w/w) active ingredient; from 0 to 99% (w/w) diluent or filler; from 0 to 20% (w/w) of a disintegrant; from 0 to 5% (w/w) of a lubricant; from 0 to 5% (w/w) of a flow aid; from 0 to 50% (w/w) of a granulating agent or binder; from 0 to 5% (w/w) of an antioxidant; and from 0 to 5% (w/w) of a pigment. A controlled release tablet may in addition contain from 0 to 90% (w/w) of a release-controlling polymer.

A parenteral formulation (such as a solution or suspension for injection or a solution for infusion) may contain from 1 to 50% (w/w) active ingredient; and from 50% (w/w) to 99% (w/w) of a liquid or semisolid carrier or vehicle (e.g. a solvent such as water); and 0-20% (w/w) of one or more other excipients such as buffering agents, antioxidants, suspension stabilisers, tonicity adjusting agents and preservatives.

Depending on the disorder, and the patient, to be treated, as well as the route of administration, compositions or combinations comprising BMC-VLPs of the invention may be administered at varying therapeutically effective doses to a patient in need thereof.

However, the dose administered to a mammal, particularly a human, in the context of the present invention should be sufficient to effect a therapeutic response in

the mammal over a reasonable timeframe. One skilled in the art will recognize that the selection of the exact dose and composition and the most appropriate delivery regimen will also be influenced by inter alia the pharmacological properties of the formulation, the nature and severity of the condition being treated, and the physical condition and mental acuity of the recipient, as well as the potency of the specific compound, the age, condition, body weight, sex and response of the patient to be treated.

Having now generally described the invention, the same will be more readily understood through reference to the following examples which are provided by way of illustration, and are not intended to be limiting of the present invention.

EXAMPLES

Standard molecular biology techniques known in the art and not specifically described were generally followed as described in Sambrook and Russel, Molecular Cloning: A Laboratory Manual, Cold Springs Harbor Laboratory, New York (2001).

EXAMPLE 1:

Materials and Methods

Bacterial strains and culture

E. coli Acella (DE3) (EdgeBio) cells were used for both molecular cloning and protein expression of Cso-BMC VLPs. Cells were grown in either lysogeny broth (LB) or Terrific broth (TB) supplemented with the appropriate antibiotics (kanamycin or streptomycin) at 50 µg/mL.

Saccharomyces cerevisiae (henceforth simply referred to as yeast) cells were used for both molecular cloning and protein expression of HO-BMC VLPs. Plasmid-based expression in yeast is based on nutritional selection, which requires a formulated growth medium. The medium lacks a key nutrient required by engineered yeast strains, which can be produced by a protein encoded by a gene on the plasmid. On pCKU, the gene product, Ura3p, produces uracil. As this chemically defined medium is expensive (about SGD 30/L), we sought to chromosomally integrate the pathways into the yeast genome so that the yeast strain can still express the pathway proteins on a non-defined culture medium (yeast-peptone-dextrose) that is less expensive (about SGD 5/L). We thus developed pGAU-*YMRWδ15* that installs homology sites flanking the pathway to be integrated into yeast. Using the endogenous homologous recombination machinery

in yeast, the desired pathway is inserted into the *YMRWδ15* chromosomal site in yeast and the proteins on the pathway can be expressed without the need for selection.

Golden Gate Assembly of plasmids

The Golden Gate one-pot plasmid assembly largely follows previously published protocols, with slight modifications [Guo, Y. *et al.*, *Nucleic Acids Res* 43: e88 (2015)]. For the insertion of one to three fragments, in a reaction pot was prepared 1 µL of T4 ligase buffer (NEB), 0.5 µL of 10x purified bovine serum albumin (BSA, NEB), 5 U of BsaI (NEB) or Esp3I (Thermo), 0.2 U of T4 ligase (Thermo), 15 ng of destination plasmid, 1 to 3 µL of insert(s) and topped up to 10 µL with water. The reaction pot was subject to a 37 °C to 18 °C thermocycling process with 5 min incubation at each step for 15 cycles, followed by a 55 °C step for 15 min to digest unassembled plasmids while inhibiting ligation so as to reduce the number of colonies harboring the original destination vector. For the assembly of more than three inserts, the amount of restriction enzyme and ligase was doubled, the amount of destination plasmid was increased to 75 ng, the number of thermocycles was increased to 70 and the inserts and destination plasmids were added in a 2:1 molar ratio rather than as fixed volumes. These were done with the aim of increasing the number of correctly assembled plasmids.

Codon-optimized BMC genes were synthesized (BioBasic) and cloned into HcKan_O. The promoters and terminator parts were amplified from various templates as PCR products and cloned into HcKan_P and HcKan_T respectively.

meGFP, which is the monomeric form of eGFP obtained through a A206K mutation, was created by site-directed mutagenesis (SDM) using NEBuilder® HiFi assembly. The plasmids pES1 – 7, pCKH and modified HcKan_O plasmids used for adding fluorescent protein, S2CP or purification tags to ORFs, were likewise created using HiFi assembly.

Primers for sequencing

The specific oligonucleotide primers used to sequence various plasmid constructs are shown in Table 1.

Table 1. Primers for sequencing

Primer name	Plasmid(s) used for	Sequence	SEQ ID NO
HcKan_chc_F'	HcKan	GATCCTTTGATTTTCTACCG	85
HcKan_chc_R'	HcKan	CTCGATAACTCAAAAAATACG	86
pES_Chc_F'	pES*X*	CGGAGCCTATGGAAAAACGC	87
pES_Chc_R'	pES*X*	CCGCAGTGTCTTGGGTCTCT	88
His_chc_F'	pCKH	TAGAGTGTACTAGAGGAGGCCAA	89
CEN_chc_R'	pCKH / pCKU	GGTGATGACGGTGAAAACCT	90
Ura_chc_F'	pCKU	TCTGTTCGGAGATTACCGAATCAA	91
pGau_chc_F'	pGAU-YMRWδ15 / pGAH-YPRCδ15	CCACCTCAGGCAGAGAACCT	92
pGau_chc_R'	pGAU-YMRWδ15 / pGAH-YPRCδ15	GGAAAAACGCCAGCAACGC	93

Sequence alignment

The CsoS2 sequences were aligned with Clustal Omega and output alignment file was prepared with JalView 2 [Waterhouse, A. M. *et al.*, *Bioinformatics* 25: 1189-1191 (2009); Sievers, F. and Higgins, D. G. *Methods in Molecular Biology (Clifton, N. J.)* 1079: 105-116 (2014)]. Accession numbers for the sequences used in sequence alignment are detailed in Table 2.

Table 2. GenBank accession numbers of the sequences shown and used in multiple sequence alignments.

Gene	Organism	GenBank accession number
csos2	*Halothiobacillus neapolitanus*	ACX95763.1
	Acidithiobacillus ferrivorans	OYV82041.1
	Burkholderiales bacterium	TNF63637.1
	Gallionellaceae bacterium	TAJ81120.1
	Hydrogenophilales bacterium	OZA28367.1
	Thiobacillus thioparus	WP_018507371.1
	Comamonadaceae bacterium	KJS73712.1
	Acidithiobacillus ferridurans	BBF66259.1
	Betaproteobacteria bacterium	TSA22668.1
	Ferrovum sp. Z-31	WP_062187313.1
csos1A	*Halothiobacillus neapolitanus*	WP_012823794.1

Hoch_5815 (BMC-H)	*Haliangium orchraceum*	WP_012830883.1
ccmK2	*Halothece sp. 7418*	WP_015227514.1
eutM	*Salmonella enterica*	VFS02811.1
pduA	*Citrobacter freundii*	WP_098065011.1
cmcC	*Klebsiella pneumoniae*	WP_004146125.1
csos4A	*Halothiobacillus neapolitanus*	WP_012823797.1
Hoch_5814 (BMC-P)	*Haliangium orchraceum*	WP_012830882.1
ccmL	*Halothece sp. 7418*	WP_015227516.1
eutN	*Salmonella enterica*	EBA6053551.1
pduN	*Citrobacter freundii*	WP_038641685.1
cmcD	*Klebsiella pneumoniae*	WP_009486245.1

Purification and cargo-loading analysis of VLPs

For Cso-BMC: Acella(DE3) cells were cultured in 500 mL terrific broth (TB, BioBasic) supplemented with kanamycin at 50 mg/L and shaken at 37 °C until the optical density (at λ = 600 nm) value of the culture reached approximately 0.6 to 1.0. The cultures were then cooled to 25 °C and isopropyl β-D-1-thiogalactopyranoside (IPTG, GoldBio) was added to 50 μM for protein induction. The cells were cultured at 25 °C for approximately 30 h before harvesting by centrifugation. Cells were lysed using a M-110P microfluidizer (Microfluidics) at 15,000 psi for three passes. To the cell lysate was added 0.1 mM of phenylmethylsulfonyl fluoride (PMSF) protease inhibitor. The lysate was spun twice at 20,000 xg for 20 min each time. The clarified lysate was loaded onto a StrepTrap™ HP 5 mL column (GE Life Sciences) at 1 mL/min linear flow rate. Purification was performed using an ÄKTA FPLC with 12 column volumes (CV) of washing with binding buffer (Tris·HCl 100 mM, NaCl 150 mM, pH 8.0) and 6 CVs of elution with the elution buffer (binding buffer supplemented with 2.5 mM desthiobiotin) at 3 mL/min linear flow rate.

In order to get high quality protein shells for structural studies, anion exchange (AIEX) chromatography was performed following StrepTrap™ affinity purification. AIEX Buffer A (Tris·HCl 50 mM, pH 8.0) was added to dilute the pooled StrepTrap™ elution fractions two-fold. The sample was loaded onto a 10 mL bed resin volume Q Sepharose (GE Life Sciences) column at 1 mL/min. A two-step gradient protocol consisting of 0 to 60 % AIEX Buffer B (Tris·HCl 50 mM, NaCl 1.0 M, pH 8.0) over 6 CV

and 60 to 100 % IEX Buffer B over 2 CV, both at 2 mL/min linear flow rate, was used for elution.

Proteins were analyzed using 13% stacking sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gels and stained using InstantBlue (Expedeon). Densitometric analysis was performed using the Bio-Rad image lab software in accordance with previous reports on BMC cargo quantification [Hagen, *et al.*, *Nature Communications* 9: 2881, doi:10.1038/s41467-018-05162-z (2018)]. Background subtraction was performed and peak areas corresponding to bands of interest were used for quantification. Absolute protein concentrations were measured using a DeNovix spectrophotometer using calculated molar attenuation coefficients at 280 nm (ε_{280}). The calculated ε_{280} of T = 3 shells, taken as the summation of the ε_{280} of its individual components, was 1 588 200 $M^{-1}\cdot cm^{-1}$. The ε_{280} of the T = 4 shell was calculated to be 1 677 600 $M^{-1}\cdot cm^{-1}$. The small difference in calculated ε_{280} values between both shell types is due to the low ε_{280} of CsoS1A (1490 $M^{-1}\cdot cm^{-1}$). Most of the ε_{280} contribution comes from CsoS4A (23 490 $M^{-1}\cdot cm^{-1}$), which has the same number of copies in both shell types. Determination of average number of GFP per shell using fluorescence was performed in accordance with a protocol described by Hagen *et. al.* [Hagen, *et al.*, *Nature Communications* 9: 2881, doi:10.1038/s41467-018-05162-z (2018)].

For HO-BMC: After growing in 8 L of YPD (yeast extract 1 %, peptone 2 %, glucose 2 %, BioBasic) for 48 h at 25 °C, yeast cells were pelleted and lysed using the M-110P microfluidizer at 20,000 psi for eight passes. The lysate was spun twice at 20,000 xg for 20 min each time and the clarified lysate was adjusted to pH 8 using 1 M Tris·HCl pH 12. The lysate was also incubated with 300 µL of biotin blocking buffer (IBA Lifesciences) with gentle stirring for 15 min. StrepTrap™ affinity purification was performed in the same way as described above.

Pull-down assay and immunoblotting

Purified His_6-meGFP-S2CP and His_6-meGFP were separately incubated with clarified *E. coli* lysate containing CsoS1A-SII, SII-CsoS1D or CsoS4A-SII with 1 h of gentle stirring at 25 °C. The lysate mixtures were purified using the abovementioned StrepTrap™ protocol. Immunoblotting detection of the eGFP epitope was done using GFP-HRP conjugated antibody (GF28R, Invitrogen) and for detection of the His_6

epitope, THE His Tag antibody conjugated with HRP (Genscript) was used. Detection was performed in accordance with the manufacturers' recommended protocols.

Fluorescence microscopy

E. coli cells were grown in accordance to the conditions described and 0.1 mL of culture was collected and pelleted. The pellet was resuspended in PBS with 1 % formaldehyde and allowed to stand at room temperature for 10 min. The cells were washed twice with PBS and resuspended in 0.5 mL PBS. A small amount (~ 3 µL) of cell suspension was mixed with an equivolume of ProLong™ Diamond Antifade Mountant (Thermo Scientific) before mounting onto poly-L-lysine microscope slides (Thermo Scientific). Samples were allowed to cure in the dark for at least 24 h before imaging. Slides were imaged using an Olympus FV1200 confocal microscope at 100x objective lens magnification. Co-localization analysis was performed using ImageJ software.

Transmission electron microscopy

Formvar / carbon coated copper grids were subject to glow discharge before 5 µL of purified protein samples (diluted to A_{280} ~ 0.05 or lower) were mounted for 60 s before the droplets were removed with filter paper. The grids were then negatively-stained by adding a 5 µL droplet of 2.5 % gadolinium(III) acetate, incubating for 90 s, and similarly blotted off. Grids were imaged using a JEOL JEM-1220 TEM.

Shell particle size and stability measurements

Particle size distribution was determined by dynamic light scattering (DLS) using the Uncle™ instrument (Unchained labs). Samples were diluted to 1 mg/mL in TBS-50/350 pH 8.0, (Tris·HCl 50 mM, NaCl 350 mM, pH 8.0), unless otherwise specified, and spun at 20,000 *g* for 5 min to remove aggregates prior to measurement. Care was taken to use the top-most supernatant for analysis. To the mini cuvette was added 9 µL of sample. All DLS measurements were done in triplicates and performed at 20 °C unless stated otherwise. Analysis of particle size distribution was done using the Uncle™ analysis software.

For shell stability measurements at various temperatures, shell samples were aliquoted into thin-walled PCR tubes and subject to temperatures ranging from 20 – 80

°C in 10 °C increments for 15 min in the Uncle™ instrument. At the end of the 15 min incubation, DLS spectra were taken.

For shell stability measurements in various buffer conditions, TBS-50/350 pH 8.0 buffers containing 10 % and 20 % (v/v) methanol were freshly prepared from stock solutions of Tris·HCl 1.0 M pH 8.0, NaCl 5.0 M and 99.8 % methanol (ACS reagent grade, Sigma) and used within the day of preparation. Due to the heat of mixing generated when methanol is mixed with water, methanol-containing buffers were allowed to equilibrate back to room temperature for at least 1 h after preparation. For making of buffers at various pH, the following components were used at 50 mM for the pH range indicated: glycine·HCl for pH 2 – 4; 4-morpholineethansulfonic acid (MES) sodium salt for pH 5 – 7; Tris·HCl for pH 8 – 9; *N*-cyclohexyl-3-aminopropanesulfonic acid (CAPS) sodium salt for pH 10 – 11; arginine·HCl for pH 12 – 13. All buffers contained 350 mM NaCl. Shells were incubated in the abovementioned buffers for 15 min to allow time for possible shell dissociation / protein denaturation before particle size measurement.

For freeze-thaw stability, shell samples in TBS-50/350 pH 8.0 were aliquoted into thin-walled PCR tubes and flash-frozen in liquid N_2. Samples were thawed at room temperature until no ice crystals were seen by standing for 15 min before re-freezing.

Enzyme steady-state kinetics assays.

For APEX2, all reagents were made to the appropriate working concentrations using TBS-50/350 pH 8.0. Working solutions of guaiacol and H_2O_2, both at 10 mM, were prepared on the day of assay. The guaiacol solution was shaken vigorously at 30 °C to ensure complete dissolution before equilibration back to room temperature. The assay concentration for APEX2 was 10 nM, H_2O_2 1 mM and that for guaiacol ranged from 0.20 to 2.0 mM. Total reaction volume was 200 µL. Reactions were monitored by the absorbance of tetraguaiacol formed at 470 nm using a BioTek Synergy™ HT microplate reader. Rate of tetraguaicol formation was found to be constant until 90 s. This time point was taken for initial rate, V_0, measurement. Kinetic constants were obtained using non-linear least square Michaelis-Menten fitting in GraphPad Prism software.

For steady-state kinetics of LacZ, all reagents were made to the appropriate working concentrations using TBS-50/350 pH 8.0 supplemented with 1 mM $MgCl_2$. Working solutions of ONPG at 10 mM were prepared from a 50 mM stock solution in

DMSO on the day of assay. The assay concentration for LacZ was 10 nM and that for ONPG (ortho-nitrophenyl-β-galactoside) ranged from 0.050 to 1.5 mM. Total reaction volume was 100 µL. Hydrolysis of ONPG was tracked by absorbance measurements at 405 nm. Rate of production formation was found to be constant until 60 s, and this time point was taken for V_0 measurement.

Enzyme / enzyme-shell activity and stability assays.

All enzyme activity measurements were performed at ambient temperature (23 °C) and working concentrations of enzymes were 10 nM. Measurements were performed in triplicates. Enzyme activity measurements were determined as initial rates of product formation using saturating substrate concentrations (i.e. near V_{max}). For APEX2, this was 1.4 mM guaiacol and 1 mM H_2O_2. For LacZ, this was 1.5 mM ONPG.

For heat shock assays, enzyme / enzyme-shell samples were aliquoted into thin-walled PCR tubes and subject to the elevated temperatures indicated (Fig. 6B) in a thermocycler for 15 min. Following incubation, samples were cooled to 20 °C and equilibrated back to ambient temperature for 15 min before assay.

For stability measurement in buffers containing methanol and under various pH conditions, enzyme / enzyme-shell samples were dialyzed to the various buffers as described in the particle size measurement section. Solutions were allowed to stand for at least 15 min before assay to allow time for possible protein denaturation.

For freeze-thaw stability, enzyme / enzyme- shell samples were aliquoted in thin-walled PCR tubes and flash-frozen in liquid N_2. Samples were thawed at room temperature until no ice crystals were seen by standing for 15 min before re-freezing or assay.

Cryo-electron microscopy and structural analysis

Protein solutions were diluted to a concentration of 0.5 mg/ml using ice-cold TBS-50/400 buffer (Tris·HCl 50 mM, NaCl 400 mM, pH 8.0). To glow discharged R 1.2/1.3 and R2/2 Molybdenum 200 grids with a holey carbon support film (Quantifoil) was applied 2.5 µL of the protein samples. Grids were transferred to a Leica EM GP plunge freezer, blotted over 2 s at 90% humidity and flash frozen in liquid ethane that was kept cool by liquid N_2. Grids were stored under liquid N_2 temperatures to prevent formation of crystalline ice.

The best cryoEM grid preparation conditions were screened for on a Talos™ Arctica Cryo-TEM (ThermoFisher Scientific) at the Institute for Protein Research, Osaka University, equipped with FEG operated at 200 kV at minimal dose system. Images were captured at an exposure time of 33.67 seconds giving a dose of approximately 20 $e^-/Å^2$ with 92,000 x magnification and defocus values of 1.6 to 2.5 µm. Images were recorded using a BM-Falcon3 camera in counting mode with exposure settings of 1.1 Å pixel size and fractions of 70 frames / individual images. For data collection, grids were prepared and imaged on a Titan™ Krios™ (FEI) (ThermoFisher Scientific), at the Research Center for Ultra-High Voltage Electron Microscopy (UHVEM), Osaka University, equipped with FEG operated at 300 kV and a minimal dose system. Imaging was done using the EPU software (FEI) attached to the Titan™ Krios™. Images were recorded at nominal magnification of 96,000, without using objective aperture, actual defocus range between 1.5 to 2.2 µm with a dose rate of 64.3 to 68.1 $e^-/Å^2$ and exposure time of 1 second with 8 image acquisitions per hole. Images were recorded using a Falcon II detector (FEI) at a pixel size of 0.86 Å/pixel and a frame rate of 17 frames / individual images.

About 2100 to 2500 raw movies were collected from different microscope sessions and processed in RELION 3.0 software [Zivanov, J. *et al.*, *eLife* 7: e42166 (2018)]. Drifts were motion corrected with MotionCor2 software and the CTF for each micrograph was estimated with CTFFind-4.1 software and Gctf software [Zhang, K. *J Struct Biol* 193: 1-12 (2016)]. Micrographs with good observed CTF estimations were selected for further processing. Shells were manually picked and extracted at a box size of 300 x 300 pixels in RELION 3.0. Particles from 2D classes displaying clear secondary structure elements were selected. Initial 3D reference models were prepared using RELION toolbox kit cylinder. 3D refinement was performed with a low pass filter of 20 Å with solvent flattening. CTF refinement was performed without particle polishing (no effect with particle polishing) and final 3D refinements performed. With post processing using solvent flattening and a soft mask, final resolutions were attained for the protein shells.

Model building and structure analysis

Biological assembly models for the pentamer (PDB ID: 2RCF) [Tanaka, S., *et al.*, *Science* 319: 1083-1086 (2008)] and hexamer (PDB ID: 2EWH) [Tsai, Y. *et al.*, *PLOS Biology* 5: e144 (2007)] were manually fit into the electron density maps using

UCSF Chimera [Pettersen, E. F., *et al.*, *J Comput Chem* 25: 1605-1612 (2004)]. Asymmetric units of the icosahedral reconstructions were extracted and rebuilt in COOT [Emsley, P., *et al.*, *Acta Crystallogr. D Biol. Crystallogr.* 66: 486-501 (2010)]. Whole-shell models were obtained by symmetry expansion and real space refinement in PHENIX [Liebschner, D., *et al.*, *Acta Crys D* 75: 861-877 (2019)] and CCP4 [Winn, M. D., *Acta Crystallogr. D Biol. Crystallogr.* 67: 235-242 (2011)].

EXAMPLE 2:

Genetic toolkit for modular construction of microcompartment parts

Our Golden-Gate cloning based genetic part assembly toolkit expands on the published YeastFab suite of plasmids, used for metabolic engineering in yeast [Guo, Y., *et al.*, *Nucleic Acids Res* 43: e88 (2015)]. Briefly, it is a hierarchical approach towards DNA assembly in which genetic part, i.e. promoters (Pro), open reading frames (ORF) and terminators (Ter) are modularized. These parts are termed as Level 0 plasmids. Level 1 plasmids link up Pro-ORF-Ter together to form a gene expression cassette and are termed as POT*X* plasmids, (*X* = 1 to 11). Level 2 plasmids chain up two or more expression cassettes to form a pathway combination. Level 3 plasmids are used for chromosomal integration of the pathways into the genome. For our yeast expression, we used the YeastFab Level 0 and 1 plasmid in the published YeastFab toolkit, but developed our own Level 2 and 3 plasmids to better suit our requirements. For *E. coli* expression we retained the use of the YeastFab Level 0 plasmids, but developed our own Level 1 and 2 plasmids. We did not develop Level 3 (genomic integration) plasmids for *E. coli*.

The Level 0 plasmids, termed HcKan_P, _O, and _T, are used for maintenance of Pro, ORF and Ter parts, respectively, for both *E. coli* and yeast as described in Table 3.

Table 3. List of genetic parts and corresponding maintenance plasmids (for storage and release of genetic parts) used for expression of our VLPs in *E. coli* and yeast

Genetic part / level	***SEQ ID NO* (for genetic part)**	***Name***	***SEQ ID NO* (for maintenance plasmids)**	***Function***
Pro / 0	84	P_{Con2}	83	Strong bacterial constitutive promoter from Anderson

				collection (Anderson, 2006), ID: BBa_J23100
	77	P_{Con3}	40	Strong bacterial constitutive promoter from Anderson collection, ID: BBa_J23108
	78	P_{Con4}	41	Moderate strength bacterial constitutive promoter from Anderson collection, ID: BBa_J23105
	24	P_{Con5}	42	Weak bacterial constitutive promoter from Anderson collection, ID: BBa_J23114
	23	P_{T7}	43	T7 bacteriophage promoter, regulated by the *lac* operon. Used for very high levels of protein production in *E. coli* strains with the λDE3 lysogen (Baneyx, 1999).
	25	P_{TDH3}	65	Strong yeast constitutive promoter
	27	P_{YEF3}	66	Moderate strength yeast constitutive promoter
	26	P_{PYK1}	67	Moderate strength yeast constitutive promoter
	104	P_{GPM1}	115	Moderate strength yeast constitutive promoter
ORF / 0	2, 8	CsoS1A	53	Hexameric BMC protomer
	3, 9	CsoS4A	54	Pentameric BMC protomer
	4 and 10	HO-H	68	Hexameric BMC protomer
	5 and 11	HO-P	69	Pentameric BMC protomer
	6 and 12	HO-T1	70	Trimeric BMC protomer
	97	HO-T1-SpyTag	116	Trimeric BMC protomer with internal SpyTag
	45, 46	meGFP	44	Monomeric enhanced green fluorescent protein.
	30 and 29	mCherry	28	Monomeric red fluorescent protein.
	48, 49	$UmuD^{1-40}$-meGFP-S2CP	47	$UmuD^{1-40}$ degradation tag fused to meGFP-S2CP
	51, 52	$UmuD^{1-40}$-meGFP	50	$UmuD^{1-40}$ degradation tag fused to meGFP
	99	GFP-SpyCatcher	121	GFP fused to SpyCatcher
	110	APEX2-	109	Engineered pea ascorbate

		S2CP(30)		peroxidase (Lam et al., 2015) fused to S2CP(30)
	103	LacZ-S2CP(30)	106	*E. coli* beta-galactosidase fused to S2CP(30)
Ter / 0	79	T_{T7}	55	HcKan_T-TT7 construct containing T7 transcriptional terminator (Banexyx, 1999).
	80	T_{RPL41B}	71	Yeast transcriptional terminator
	81	T_{HBT1}	72	Yeast transcriptional terminator
	82	T_{RPS20}	73	Yeast transcriptional terminator
	105	T_{YPT31}	119	Yeast transcriptional terminator
Gene expression / 1	N.A.	pES*N* (*N* = 1 – 7)	56 to 62	Receives Pro-ORF-Ter to form a gene expression cassette
	N.A.	POT*X* (*X* = 1 – 11)	N.A.	Receives Pro-ORF-Ter to form a gene expression cassette
Pathway / 2	N.A.	pCKH	63	Accepts pES*N* plasmids starting from *N* = 2, continuing with even numbered pES and ending with an odd numbered pES
	N.A.	pCKU	74	Accepts POT*X* plasmids starting from *N* = 1 or 2, continuing with even numbered POT and ending with an odd numbered POT
Integration / 3	N.A.	pGAU-*YMRWδ15*	75	Integrates pathways into *YMRWδ15* in the yeast genome
	N.A.	pGAU-*YPRCδ15*	76	Integrates pathways into *YPRCδ15* in the yeast genome

The Level 1 plasmids were modified to tailor to protein expression in *E. coli* by removing genetic elements from the POT*X* plasmids that can add unnecessary burden to the host cell. We term the *E. coli* Level 1 plasmids as pES*N* (*N* = 1 to 7), which contain the minimum genetic elements required for TU maintenance (Fig. 2A). For assembly of multiple Pro-ORF-Ter from POT*X* or pES*N* plasmids, we developed the Level 2 plasmids pCKU (SEQ ID NO: 74) and pCKH (SEQ ID NO: 63) (Fig. 2B), designated for yeast and *E. coli* respectively. Plasmid-based expression in yeast is based on nutritional selection, which requires a formulated growth medium. The medium lacks a key nutrient required by engineered yeast strains, which can be produced by a protein encoded by a gene on the plasmid. On pCKU, the gene product, Ura3p, produces uracil. As this chemically defined medium is expensive ($SGD 30/L),

RECTIFIED SHEET (RULE 91)

we sought to chromosomally integrate the pathways into the yeast genome so that the yeast strain can still express the pathway proteins on a non-defined culture medium (yeast-peptone-dextrose) that is less expensive ($SGD 5/L). We thus developed pGAU-*YMRWδ15* (SEQ ID NO: 75) that installs homology sites flanking the pathway to be integrated into yeast. Using the endogenous homologous recombination machinery in yeast, the desired pathway is inserted into the *YMRWδ15* chromosomal site in yeast and the proteins on the pathway can be expressed without the need for selection. For pathway expression in *E. coli*, we do not find pathway expression necessary at this point in time as plasmid selection in the bacterium is typically performed using the appropriate antibiotic (kanamycin in this case) in a non-defined culture medium (lysogeny broth or terrific broth).

We also modified the HcKan_O plasmids to install a fluorescent protein (FP), a biochemical/affinity tag, or an encapsulation peptide at either the amino or carboxy end of the ORF (Fig. 2C and D).

Four FPs were chosen – mTurquoise2 (mT2), monomeric-enhanced GFP (meGFP), monomeric Kusabira orange-kappa (mKOκ) and mCherry (mCh) – are known to exhibit monomeric behavior, which should reduce artefactual aggregation of the fusion product. An example of a modified HcKan_O plasmid is HcKan_O-CmCherry (SEQ ID NO: 28) which tags mCherry to the C-terminus of an ORF. Two affinity tags introduced were the hexahistidine (His_6) and Strep-tag II (SII) tags, permitting protein purification by immobilized metal affinity chromatography (IMAC) or by Strep-Tactin respectively. Examples of modified HcKan_O plasmids are HcKan_O-CHis6 (SEQ ID NO: 32) which tags His_6 to the C-terminus of an ORF, and HcKan_O-CSII (SEQ ID NO: 31) which tags Strep-Tag II to the C-terminus of an ORF.

Other tags include the SpyCatcher / SpyTag (ST / SC) pair (SEQ ID NOs: 13 and 16) and CC-Di-A / B (CCA / CCB) pair (SEQ ID Nos: 17-20). Examples of modified HcKan_O plasmids are HcKan_O-CSpyCatcher (SEQ ID NO: 37) which tags SpyCatcher to the C-terminus of an ORF; HcKan_O-CSpyTag (SEQ ID NO: 38) which tags SpyTag to the C-terminus of an ORF; HcKan_O-CCCDiA (SEQ ID NO: 35) which tags coiled-coil dimeric-A to the C-terminus of an ORF; and HcKan_O-CCCDiB (SEQ ID NO: 36) which tags coiled-coil dimeric-B to the C-terminus of an ORF. The SII tag (SEQ ID NOs: 21 and 22) has been widely used for purification of proteins and protein complexes while the ST / SC and CCA / CCB pairs have found use in the

functionalization of VLPs and other protein nanostructures [Fletcher, J. M. *et al.*, *Science* 340: 595-599 (2013); Keeble, A. H., & Howarth, M. *Methods in Enzymology, 617*, 443-461(2019)]. A protein tagged with SpyCatcher (SEQ ID NOs: 13 and 14) forms a covalent isoamide bond with another protein tagged with SpyTag,(SEQ ID Nos: 15 and 16) while a protein tagged with CC-Di-A (SEQ ID Nos: 17 and 18) forms strong intermolecular interactions (dissociation constant, K_d ~ 1 nM) with another tagged with CC-Di-B (SEQ ID NOs: 19 and 20) [Thomas, F., *et al.*, *Journal of the American Chemical Society* 135: 5161-5166, (2013)]. Installing one member of the SC / ST or CCA / CCB pairs on the surface of the VLPs allows guest proteins tagged with the other corresponding member in the pair to conjugate to the shell surface.

Controlling the intracellular stoichiometry of shell protomers is known to be important for successful assembly of BMC shells [Kerfeld, C. A., *et al.*, *Nature Reviews Microbiology* 16: 277 (2018)]. To tune the expression of each of the components, we incorporated five constitutively active promoters from the Anderson collection into HcKan_P (Table 4) [Anderson, J. C. *Anderson Promoter Library Registry of Standard Biological Parts* (2006)].

Table 4. List of constitutively active Anderson collection promoters (P_{CON2-5}) used in this study, along with the original identities and characterized relative strengths

Promoter	Anderson collection identity	Relative strength	SEQ ID NO.
P_{CON2}	BBa_J23100	1.00	84
P_{CON3}	BBa_J23108	0.50	77
P_{CON4}	BBa_J23105	0.24	78
P_{CON5}	BBa_J23114	0.10	24

We renamed these promoters P_{CON1} to P_{CON5} for brevity, with P_{CON2} being the strongest and P_{CON5} being the weakest. P_{CON2} to P_{CON5} sequences (SEQ ID Nos: 84, 77, 78 and 24, respectively) are shown in lowercase within SEQ ID Nos: 83 and 40 to 42, respectively. We also included the T7 promoter (P_{T7}; SEQ ID NO: 23), along with the *lacI* repressor and *lac* operator sequence (LacI+ P_{T7}), for inducible expression of genes

by addition of the inducer, isopropyl β-D-1-thiogalactopyranoside (IPTG). For transcription termination, we made use of the T7 terminator (T_{T7}) throughout all TUs. Using this multi-monocistronic system of DNA assembly, the expression levels of BMC components can be tailored in pES*N* plasmids (Table 5).

Table 5. Compilation of TUs (assembled in pES plasmids). The transcriptional units (TUs) are annotated with letters A – D. Abbreviations used are P_Y: P_{CONY} (Table 4); P_{T7}: P_{T7} with *lacI* and *lac* operator; meG: meGFP; mCh: mCherry. All TUs are terminated by T_{T7}.

pES	A	B	C	D
2	P_4-meG-S2CP	P_5-CsoS4A-SII	P_4-UmuD$^{1-40}$-meG-S2CP	P_4-UmuD$^{1-40}$-meG
3	P_{T7}-CsoS1A			
4	P_4-CsoS1D			
5	P_{T7}-CsoS1A			
6	P_5-CsoS4A-mCh	P_5-CsoS4A-SII		
7	P_{T7}-CsoS1A			

For encapsulation of cargo in Cso-BMC, we have identified an encapsulation peptide (EP) sequence (SKITGSSGNDTQGSLITYSGGARG; SEQ ID NO: 1), which we term S2CP, which mediates sequestration of protein cargo into the simplified carboxysomes. An example of a modified HcKan_O plasmid which tags S2CP to the C-terminus of an ORF is HcKan_O-S2CP (SEQ ID NO: 39). Details on identifying S2CP

as an EP are discussed later. For encapsulation of cargo in HO-BMC, although the reported EP for HO-BMC has been reported to function when *E. coli* was the recombinant host, we found that it did not work in yeast [Lassila, J. K., et al., *Journal of molecular biology* 426: 2217-2228 (2014)]. Schematics of the synthetic operons of the pathway used to make Cso-BMC in *E. coli* and the HO-ACB pathway used to express HO-BMC in yeast are shown in Figure 3.

The Golden Gate one-pot plasmid assembly largely follows previously published protocols, with slight modifications [Guo, Y., et al., *Nucleic Acids Research, 43*(13), e88 (2015)]. For the insertion of one to three fragments, in a reaction pot was prepared 1 µL of T4 ligase buffer, 0.5 µL of 10x purified bovine serum albumin (BSA), 5 U of BsaI (for Level 0 and 2 assembly) or Esp3I (for Level 1 assembly), 10 U of T4 ligase, 20 ng of destination plasmid, 1 to 3 µL of insert(s) and topped up to 10 µL with water. All enzymes and BSA used were from New England Biolabs (NEB). The reaction pot was subject to a 37 ^{o}C to 18 ^{o}C thermocycling process with 5 min incubation at each step for 70 cycles, followed by a 55 ^{o}C step for 15 min. Plasmids were transformed into *E. coli* Acella (DE3) strain (EdgeBio) and verified by Sanger sequencing.

Transformation of plasmids and chromosomal integration of genes into yeast was performed in accordance with the high-efficiency lithium acetate / single-stranded DNA / PEG-3350 protocol as described by Schiestl and co-workers [Gietz, R. D. and Schiestl, R. H. *Nature Protocols* 2: 31 (2007)].

EXAMPLE 3:

Identifying a targeting peptide for the Cso system

A key strategy for repurposing bacterial BMCs into intracellular nanoreactors is to encapsulate heterologous enzymes within the shells by installing EPs to the cargo. While EP sequences have been identified and characterized for some BMC systems, such a sequence has not been reported for the alpha-carboxysome [Kerfeld, C. A., *et al.*, *Nature Reviews: Microbiology, 16*, 277 (2018)]. An EP sequence has been suggested to reside on CsoS2 [Oltrogge, L. M., *et al.*, *Nature Structural & Molecular Biology* 27: 281-287 (2020)]. Studies performed on CsoS2 suggest that it initiates assembly of the carboxysome by recruiting luminal cargo through its N-terminus while its C-terminal region anchors to shell proteins [Oltrogge, L. M., *et al.*, *Nature Structural*

& Molecular Biology 27: 281-287 (2020)]. Multiple sequence alignment of 100 CsoS2 orthologs revealed that the C-terminal region is highly conserved, especially at the terminal residues (Fig. 4A). This suggests functional importance. We thus decided to interrogate the function of the terminal 24 residues of *H. neapolitanus* CsoS2 (SKITGSSGNDTQGSLITYSGGARG; SEQ ID NO: 1) and termed it "S2CP" as an abbreviation for Cso**S2** **C**-terminal **P**eptide. The nucleic acid sequence that encodes the S2CP peptide is set forth in SEQ ID NO: 7. We also considered the possibility that a slightly longer variant of S2CP might improve encapsulation efficacy without adding too much additional bulk on heterologous protein cargo. To this end, we chose the terminal 30 residues of *H. neapolitanus* CsoS2 (KPEKPG SKITGSSGNDTQGSLITYSGGARG SEQ ID NO: 94) as an encapsulation peptide variant and termed it "S2CP(30)". The nucleic acid sequence that encodes the S2CP(30) peptide is set forth in SEQ ID NO: 95.

We used a pull-down assay to investigate if non-native protein cargo tagged with S2CP can interact with CsoS1A, CsoS1D or CsoS4A – representing the BMC-H, BMC-T and BMC-P shell protein types, respectively. We created pES2-P_{con4}-His_6-meGFP-S2CP-T_{T7}, purified His_6-meGFP-S2CP and incubated the protein with *E. coli* lysates in which either CsoS1A-SII, SII-CsoS1D or CsoS4A-SII were expressed using P_{T7}. As negative controls, purified His_6-meGFP was similarly incubated with the same shell protein containing lysates. The mixtures were subject to purification via Strep-Tactin and purified fractions from the six mixtures were analyzed by Western blotting for the presence of GFP. It was found that His_6-meGFP-S2CP co-eluted with CsoS1A-SII, but not with SII-CsoS1D or CsoS4A-SII (Fig. 4B). His_6-meGFP was also not seen to co-elute with CsoS1A-SII, SII-CsoS1D or CsoS4A-SII. This demonstrates that S2CP was required for His_6-meGFP to interact with CsoS1A. While a previous report demonstrated that full-length CsoS2 interacts with CsoS1A (Cai et al., 2015), we have shown that just the terminal 24 residues of CsoS2 are sufficient for interaction. The association of S2CP with CsoS1A, a major shell module in the alpha-carboxysome, should allow this peptide sequence to target protein cargo to the shell complex. However, based on this result alone, it cannot yet be ascertained whether S2CP is able to mediate cargo encapsulation within the shell or merely targets it to the shell periphery.

EXAMPLE 4:

Recombinant formation of simplified alpha-carboxysome shells

We sought to investigate the interactions between Cso components with the aim of constructing a simplified microcompartment shell based on knowledge of the component structures. Our approach involved translational fusion of FPs to shell components and to S2CP to serve as probes of protein-protein interactions. Using the HcKan_O-FP plasmids, we fused the four FPs (mTurquoise2, meGFP, mKOκ and mCherry) to both the amino and carboxy termini of CsoS4A and expressed the hybrid proteins in *E. coli* using the P_{CON5} promoter from the pES6 plasmid. Only CsoS4A-mCherry was shown to be generally homogenously distributed within the cytosol (Fig. 8A). The rest of the fusion products demonstrated various degrees of aggregation (data not shown), making them less ideal for utilization as probes. Hence, CsoS4A-mChery was chosen as the shell component probe. We also expressed meGFP-S2CP using P_{CON4} in pES2 plasmid and found that the protein was generally diffuse within the cytosol (Fig. 8B).

With the shell (CsoS4A-mCherry) and targeting peptide (meGFP-S2CP) probes established, we next expressed CsoS1D and CsoS1A alongside these probes using the pathway plasmid pCKH-Cso-P_{mCh}THC (Fig. 5A, Table 6).

Table 6. List of pathways plasmids and the corresponding pES TU (Table 5) used in their assembly.

	Assembled TUs
Cso-P_{mCh}THC	2A-4A-6A-7A
Cso-P_{SII}THC	2A-4A-6B-7A
Cso-P_{SII}TH	2B-4A-5A
Cso-P_{SII}H	2B-3A
Cso-$P_{SII}THC_{U,S2CP}$	2C-4A-6B-7A
Cso-$P_{SII}THC_{U}$	2D-4A-6B-7A

In our pathway nomenclature, P_{mCh} represents the pentameric shell protein (CsoS4A) fused to mCherry, T represents trimeric (CsoS1D), H for hexameric (CsoS1A) and C for the cargo (meGFP-S2CP). In cells expressing these four components, it was seen that CsoS4A-mCherry and meGFP-S2CP co-localized upon addition of IPTG to 50 μM (Fig. 5B). We quantified the degree of co-localization using

the Mander's co-localization coefficients (MCC), [tM_1, tM_2], where tM_1 is the fraction of green signal found in areas where there is red signal, while tM_2 the fraction of red signal found in areas with green [Dunn, K. W., *et al.*, *American Journal of Physiology-Cell Physiology* 300: C723-C742 (2011)]. From the cells surveyed, the MCC values were found to be [0.688, 0.758], suggesting there was a significant proportion of co-localized probes.

We proceeded to determine if the fluorescent foci observed might be indicative of protein assemblies that can be purified. Two purification strategies were attempted. The first was to incubate *E. coli* lysate expressing Cso-P_{mCh}THC with pure CsoS4A-SII before purification by Strep-Tactin. The second was to replace CsoS4A-mCherry with CsoS4A-SII in the Cso-P_{mCh}THC pathway, thus creating a new pathway, Cso-P_{SII}THC. Proteins purified via Strep-Tactin were further purified by anion exchange ion chromatography (AIEX) using Q Sepharose. For both purification strategies, two elution peaks were seen at 0.3 M and 0.4 M NaCl in the AIEX chromatograms (Fig. 9A-B). Fractions from both peaks were viewed using transmission electron microscopy (TEM) and numerous capsid-like structures approximately 20 nm in diameter were seen in the 0.4 M NaCl elution fractions (Fig. 5C-D), with significantly fewer of such structures seen in the 0.3 M NaCl fractions (Fig. 9A-B). We reasoned that while the capsid-like structures eluted primarily at 0.4 M NaCl, some were observed at the 0.3 M NaCl fraction due to overlap of these two peaks. We also noted that CsoS4A-SII alone, when subject to the same AIEX procedure, eluted in a single peak at 0.3 M NaCl (Fig. 9C). Hence, the 0.3 M NaCl peak observed for Cso-P_{mCh}THC and Cso-P_{SII}THC likely corresponded to CsoS4A-SII not incorporated within the shells.

It has been proposed that CsoS2 is critical for the assembly of the alpha-carboxysome by recruiting shell proteins through its C-terminal [Oltrogge, L. M., *et al.*, *Nature Structural & Molecular Biology* 27: 281-287 (2020)]. In the Cso-P_{mCh}THC and Cso-P_{SII}THC constructs, the terminal 24 residues of CsoS2 (S2CP; SEQ ID NO: 1) could have assisted in shell assembly. We wished to investigate whether S2CP was essential for the formation of the shells derived from alpha-carboxysome components. We therefore constructed the pathway Cso-P_{SII}TH, in which S2CP was absent. In addition to similar AIEX chromatograms (Fig. 10A), capsid-like structures indistinguishable from those produced by Cso-P_{mCh}THC and Cso-P_{SII}THC were seen in the Cso-P_{SII}TH combination. These structures were again more abundant in the 0.4 M NaCl fraction than the 0.3 M NaCl one (Fig. 5E for 0.4 M NaCl, Figure 10B for 0.3 M

NaCl). These results demonstrate S2CP was unnecessary for the formation of the protein shells observed.

Next, we sought to determine the minimal components required for shell assembly. Given that CsoS1A and CsoS1D are constructed from the same protein domain, we considered the possibility that protein shells could be constructed from just CsoS1A and CsoS4A, each from a different protein domain. A new pathway combination, Cso-P_{SII}H, which expresses CsoS1A and CsoS4A-SII, was constructed (pCKH-Cso-BMC; SEQ ID NO: 64) and proteins were purified as before (Fig. 11A). Capsid-like structures that looked similar to those purified from previous pathway combinations were again seen (Fig. 11B for 0.3 M NaCl fraction, Fig. 5D for 0.4 M NaCl). No capsid-like structures were seen to assemble from CsoS1A-SII alone (Fig. 11C). In addition, constructs in which either CsoS1A and CsoS1D were co-expressed or CsoS1D and CsoS4A were co-expressed failed to produce protein shells (Fig. 11D-E). Taken together, these results indicate that CsoS1A and CsoS4A are necessary and sufficient for the assembly of the capsid-like shells.

EXAMPLE 5:

S2CP targets cargo into the lumen of the simplified carboxysome shells

To ascertain whether S2CP is able to target cargo into lumen of the simplified carboxysome shells, we fused the *E. coli* UmuD N-terminal degradation tag (residues 1 – 40) to the amino-end of meGFP-S2CP [Neher, S. B. *et al.*, *Proceedings of the National Academy of Sciences* 100: 13219-13224 (2003)]. We co-expressed $UmuD^{1-40}$-meGFP-S2CP with CsoS1A, CsoS1D and CsoS4A by constructing the plasmid pCKH-Cso-$P_{SII}THC_{U,S2CP}$. We hypothesized that if S2CP is able to target $UmuD^{1-40}$-meGFP into the carboxysome, $UmuD^{1-40}$-meGFP-S2CP (SEQ ID NO: 49) would be protected from proteolysis by the endogenous ClpXP protease that recognizes and degrades proteins tagged with the N-terminal region of UmuD (Fig. 6A). On the other hand, if S2CP only targets cargo to the shell exterior, $UmuD^{1-40}$-meGFP-S2CP would be exposed to ClpXP and be degraded. A similar construct, pCKH-Cso-$P_{SII}THC_U$, where the only difference was the absence of S2CP from $UmuD^{1-40}$-meGFP, served to account for stochastic encapsulation of $UmuD^{1-40}$-meGFP by the shells. Western blotting was used for detection of GFP (Fig. 6B). In the lane corresponding to purified proteins from Cso- $P_{SII}THC_{U,S2CP}$, $UmuD^{1-40}$-meGFP-

S2CP was detected. In the lane corresponding to the same quantity of purified proteins (determined by absorbance at 280 nm) from Cso-P_{SII}THC$_U$, UmuD$^{1-40}$-meGFP (SEQ ID NO: 52) could not be detected. As a further confirmatory analysis, similar protein shells could be seen in the elution fractions for both pathway combinations (Fig. 6C-D). This demonstrates that the protection of UmuD$^{1-40}$-meGFP from proteolysis was likely due to its encapsulation into the shells as mediated by S2CP.

EXAMPLE 6:

Atomic models of the simplified alpha-carboxysome shells

To better understand the molecular architecture of the simplified alpha-carboxysomes, cryo-electron microscopy (cryo-EM) was used to obtain near atomic-scale models of Cso-P_{SII}THC, Cso-P_{SII}TH and Cso-P_{SII}H. Two distinct shell sizes were seen for Cso-P_{SII}THC, corresponding to icosahedral capsid triangulation numbers T = 3 and T = 4. Shell models were obtained at resolutions of 3.24 and 2.90 Å respectively. For Cso-P_{SII}TH and Cso-P_{SII}H, only T = 3 shells were seen, and structures were obtained at resolutions of 3.35 and 3.14 Å respectively. The proportion of T = 3 shells observed in Cso-P_{SII}THC was 14.6 % while that for the T = 4 shell was 85.4 %. The reported X-ray crystal structures of *H. neapolitanus* CsoS1A (PDB: 2EWH) and CsoS4A (PDB: 2RCF) were used for model fitting [Tanaka, S. *et al.*, *Science* 319: 1083-1086 (2008); Tsai, Y. *et al.*, *PLOS Biology* 5: e144 (2007)]. *H. neapolitanus* CsoS1D is expected to assemble as double stacking layers of trimers, as deduced from the structure of CsoS1D from *Prochlorococcus marinus* MED4, with which it shares 60 % identical residues [Klein, M. G., *et al.*, *Journal of Molecular Biology* 392: 319-333 (2009)]. However, we were unable to discern double stacking layers in the electron density maps for Cso-P_{SII}THC and Cso-P_{SII}TH, suggesting CsoS1D was not incorporated within these shells. Electron density was also not detected for the meGFP-S2CP cargo in the luminal space of shells purified from Cso-P_{SII}THC. Nonetheless, in light of computational studies that suggest interactions between shell protomers and cargo affect shell size and shape, it is conceivable that formation of the T = 4 shell, which is seen only in Cso-P_{SII}THC, may be affected by cargo encapsulation while shells without cargo assemble as the smaller T = 3 form.

As there was no appreciable difference among the three pathway combinations used to obtain the *T* = 3 shell models, we focused on the shells produced by Cso-P_{SII}THC for model building and refinement (Table 7).

Table 7. Cryo-EM data collection, map and model refinement, and model validation

	Cso-P_{SII}THC		**Cso-P_{SII}TH**	**Cso-P_{SII}H**
	T = 3	*T* = 4	*T* = 3	*T* = 3
Accession codes				
Map (EMDB)	EMD-30384	EMD-30385	Not deposited, structure almost the same as Cso-P_{SII}THC *T* = 3	Not deposited, structure almost the same as Cso-P_{SII}THC *T* = 3
Coordinates (PDB)	7CKB	7CKC		
Data collection				
Microscope	Titan Krios (Research Center for Ultra-High Voltage Electron Microscopy, Osaka University, Osaka Prefecture, Japan			
Voltage (kV)	300 kV			
Detector	Falcon II			
Magnification	96 k			
Pixel size (Å)	0.86			
Defocus range (μm)	1.5~1.9	1.5~1.9	1.5~2.2	1.5~2.2
Electron exposure (e^{-}/Å^2)	64.3	64.3	68.1	68.1
Reconstruction				
Software				
Initial particle images (no.)	94129	94129	14401	15680
Final particle images (no.)	11468	67192	9349	7678
Box size (pixels)				
Symmetry imposed				
Accuracy rotations (°)				
Accuracy translations (pixels)				
Map resolution (Å)	3.24	2.9	3.35	3.14
FSC threshold	0.143	0.143	0.143	0.143
Map resolution range (Å)	∞ ~ 3.24	∞ ~ 2.90	∞ ~ 3.35	∞ ~ 3.14
Map sharpening *B* factor (Å^2)				
Model building & refinement				
Software	Chimera(Pettersen et al., 2004), Coot(Emsley & Cowtan, 2004),			

	Phenix(Adams et al., 2010)			
Initial model used (PDB code)	2RCF, 2EWH			
Model resolution (Å)	2.15, 1.40			
FSC threshold				
Model resolution range (Å)				
Model composition				
Non-hydrogen atoms	114840	150300		
Protein residues	15840	20820		
Ligands & water	0	0		
B factors overall (Å^2)	10.67	13.91		
R.m.s. deviations				
Bond lengths (Å)	0.0076	0.0076		
Bond angles (°)	0.87	1.23		
Validation				
MolProbity score	1.67	1.33		
Clashscore	6.97	5.49		
Poor rotamers (%)	1.05	0		
Cβ deviations (%)	0	0		
Ramachandran plot				
Favored (%)	96.01	97.84		
Allowed (%)	3.99	2.16		
Disallowed (%)	0	0		

The *T* = 3 shell contains 12 homo-pentamers of CsoS4A and 20 homo-hexamers of CsoS1A, with an external diameter of 217 Å, and a calculated molecular mass of 1.7 MDa (Fig. 7A). The *T* = 4 shell contains 12 homo-pentamers and 30 homo-hexamers, with an external diameter of 247 Å, and a mass of 2.3 MDa (Fig. 7B). Both shell types are largely similar in other aspects. The concave sides of CsoS1A and CsoS4A, wherein the N and C termini reside, face the outside of the shell (Fig. 7A).

EXAMPLE 7:

Determining encapsulation efficacies of S2CP and S2CP(30) for the Cso-BMC

Using the Cso-BMC shell structure, the average copy numbers of UmuD$^{1-40}$-GFP-S2CP and UmuD$^{1-40}$-GFP-S2CP(30) could be quantified via GFP fluorescence. Based on cryogenic electron microscopy observations, for all calculations involving shell molecular masses we presumed all shells co-expressed with cargo to be a mixture of T = 3 and 4 forms. As the proportions of shell forms may vary among

samples, two values, calculated by presuming all shells in a sample were either *T* = 3 or 4, are provided. It was determined that an average of 7.7 – 8.0 copies of UmuD$^{1-40}$-GFP-S2CP(30) were encapsulated per shell, compared to 1.6 – 1.7 copies of UmuD$^{1-40}$-GFP-S2CP. Densitometric analysis of Cso-P$_{SII}$H shells encapsulating either UmuD$^{1-40}$-GFP-S2CP or UmuD$^{1-40}$-GFP-S2CP(30) also indicated that there were approximately 4 times UmuD$^{1-40}$-GFP-S2CP(30) found within the shells compared to UmuD$^{1-40}$-GFP-S2CP (Fig. 16). Hence, S2CP(30) is a more efficacious encapsulation peptide than S2CP.

Table 8. Quantifying the average number of encapsulated UmuD$^{1-40}$-GFP cargo mediated by S2CP or S2CP(30). As the proportions of the shell forms are unknown, values calculated by presuming all shells are either *T* = 3 or 4 are given.

Encapsulation peptide	Number of residues	Average number of UmuD$^{1-40}$-GFP cargo per shell	
		T = 3	*T* = 4
S2CP	24	1.6	1.7
S2CP(30)	30	7.7	8.0

EXAMPLE 8:

Stabilization of enzymatic activities using Cso-BMCs

Protein shells have been garnering attention as platforms for conferring stability on enzymes against physical insults, such as heating or freezing, or chemical insults, such as the presence of organic co-solvents or non-physiological pH [Demchuk & Patel, *Biotechnology Advances*, 41: 107547 (2020); Silva, C., *et al*., *Critical Reviews in Biotechnology*, 38(3): 335-350 (2018)]. Enzyme confinement often reduces their conformational flexibility, which sometimes confers stability against structural changes that lead to denaturation [Das, Zhao, (2020) *Biochemistry*, 59(31): 2870-2881; Küchler, *et al*., *Nature Nanotechnology*, 11(5): 409-420 (2016)]. Currently, homomeric protein shells are more established for hosting enzymes, attributable to their relative ease of assembly and particle size homogeneity, which improves predictability and tractability during engineering [Patterson, D. P., Prevelige, P. E., & Douglas, T. (2012). *ACS Nano*, 6(6): 5000-5009; Patterson, D. P., Schwarz, B., El-Boubbou, K., van der Oost,

J., Prevelige, P. E., & Douglas, T. (2012). *Soft Matter*, 8(39): 10158-10166; Sánchez-Sánchez *et al.*, *Journal of Nanobiotechnology* 13(1): 66 (2015); Tan, Xue, & Yew, *Molecules* 26(5): 1389 (2021)]. Due to their heteromeric composition, minimal BMC-derived shells represent emerging scaffolds for hosting enzymes, as these shells can provide more avenues for purposeful modifications, while their generally homogenous particle size still confers predictability to facilitate engineering [Turmo, A., Gonzalez-Esquer, C. R., & Kerfeld, C. A. *FEMS Microbiology Letters*, 364(18): fnx176 (2017)]. However, minimal BMC-derived shells have yet to be explored for hosting heterologous enzymes [Cai, F., Bernstein, S. L., Wilson, S. C. & Kerfeld, C. A. *Plant Physiol* 170: 1868-1877 (2016); Hagen, A., *et al.*, *Nature Communications* 9: 2881, (2018)]. This encouraged us to investigate if the Cso-BMC could host and stabilize enzymes. Empty Cso-BMCs (Cso-P_{SII}H) were first tested for their stability against heat shock, freezing, presence of methanol co-solvent and environments with pH from 2 to 13. The DLS spectra of shells subject to these conditions were compared to that of shells in Tris·HCl-50/350 (Tris·HCl 50 mM pH 8.0, NaCl 350 mM). Significant changes in particle size distributions and/or appearance of multiple peaks indicate protein shell disassembly [Yu, Z., Reid, J. C., & Yang, Y.-P. *Journal of Pharmaceutical Sciences* 102(12): 4284-4290 (2013)]. Based on the conditions tested, Cso-BMCs were considered to be stable up to 70 ^{o}C for 15 min, 20 % v/v methanol, seven consecutive freeze-thawing and between pH 5 – 11 (Fig. 17).

To probe the capacity of the Cso-BMC for encasing enzymes of considerably different molecular sizes, an evolved pea cytosolic ascorbate peroxidase (APEX2), a 27.0 kDa monomer [Lam *et al.*, *Nature Methods* 12(1): 51-54 (2015)], and *E. coli* beta-galactosidase (LacZ), a 466.0 kDa homotetramer, were chosen for encapsulation [Golan, et al., *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1293(2): 238-242; Lam *et al.*, *Nature Methods* 12(1): 51-54 (2015)]. S2CP(30) was fused to the C-termini of the enzymes to mediate encapsulation as it was found to be more efficacious than S2CP for mediating encapsulation of recombinant proteins. Enzymes were also N-terminally tagged with the hexahistidine (His_6) tag to facilitate downstream removal of unencapsulated enzymes that might co-purify with shells (Nichols, Kennedy, & Tullman-Ercek, 2019). Cso-P_{SII}H shells co-expressed with enzymes were constructed and purified. SDS-PAGE analysis and Western blotting confirmed the presence of the target enzymes in shell samples (Fig. 18A-B) and the average copy numbers of enzymes per shell were estimated by Coomassie blue densitometry (Table 8) [Hagen, A., *et al.*, *Nature Communications* 9: 2881 (2018); Nichols *et al.*, *Methods in*

Enzymology, 617, 155-186 (2019)]. Encapsulation of these enzymes did not appear to significantly affect Cso-BMC size and morphology (Fig. 18C-E).

Encapsulation of an enzyme into a protein shell is known to alter the enzyme's catalytic properties in some cases. To probe how encapsulation by the Cso-BMC might affect the catalytic efficiencies of APEX2 and LacZ, we performed steady-state kinetics of both free and encapsulated enzymes and fitted the data to the Michaelis-Menten model to obtain the turnover number (k_{cat}), Michaelis-Menten constant (K_M), and catalytic efficiency (k_{cat}/K_M) (Table 8, Figure 19). For encapsulated APEX2, k_{cat}/K_M decreased to approximately 30 % of the free enzyme. For encapsulated LacZ, k_{cat}/K_M was not significantly different from the free enzyme. The kinetic constants, k_{cat} and $K_{M,}$, obtained for both free enzymes were in reasonable agreement with previous work, suggesting the presence of S2CP(30) did not affect the activities of the free enzymes [Juers, Hakda, Matthews, & Huber, *Biochemistry, 42*(46), 13505-13511 (2003); Lam *et al*., *Nature Methods* 12(1): 51-54 (2015)].

To determine possible stabilizing effects of the Cso-BMC on enzymes, free enzyme and shell-encapsulated enzyme samples were challenged with the abovementioned conditions under which empty shells were found to be stable. Enzyme activities were normalized to that of the pristine sample to determine the residual activity (Fig. 20). The Cso-BMC conferred a moderate level of thermostability on both enzymes. Encapsulated enzymes retained approximately 90 % of their activities following incubation at 40 ^{o}C for 15 min, in contrast to 40 % for free enzymes. At 50 ^{o}C, encapsulated APEX2 retained about half its activity while the free enzyme was essentially inactive. However, the activity of encapsulated LacZ was only marginally higher than free enzyme at 50 ^{o}C. At 60 ^{o}C and above, all enzyme samples were inactive. The Cso-BMC had a protective effect on APEX2 up to 20 % v/v methanol. On the other hand, for both free and encapsulated LacZ in methanol, increases in activities were seen. It has been reported that the presence of up to 40 % v/v methanol does not denature LacZ but rather enhances its activity [Shifrin & Hunn, *Archives of Biochemistry and Biophysics, 130*, 530-535 (1969)]. Hence, Cso-BMC was unlikely to have stabilized LacZ against methanol. For freeze-thaw stability, the Cso-BMC stabilized both enzyme species up to seven consecutive cycles.

Encapsulated enzymes displayed higher activities within the Cso-BMC at pH 10 – 11, but lower activities at pH 5 – 6. We reasoned the acidic microenvironment within the Cso-BMC likely shifted the pH-activity profile of encapsulated enzymes to more alkaline conditions compared to free enzyme. The influence of anionic scaffolds on the

pH-dependent activities of enzymes has been observed for synthetic maleic acid polymer scaffolds on trypsin and chymotrypsin, and more recently for the DNA polyphosphate backbone on the glucose oxidase-horseradish peroxidase (GOx-HRP) cascade [Goldstein, *Biochemistry* 11(22): 4072-4084 (1972); Goldstein, Levin, & Katchalski, *Biochemistry*, 3(12): 1913-1919 (1964); Zhang, Tsitkov, & Hess, *Nature Communications*, 7(1): 13982 (2016)].

To date, the Cso-BMC likely demonstrates the highest heterologous cargo loading via encapsulation peptides among minimal BMC-derived shells. The use of encapsulation peptides for such shells has been largely inefficacious and cargo often could not be detected via Coomassie blue staining, requiring more sensitive techniques such as immunoblotting or fluorescence [Cai, F., *et al.*, *Plant Physiol* 170: 1868-1877 (2016); Hagen, A., *et al.*, *Nature Communications* 9: 2881 (2018); Lassila, Bernstein, Kinney, Axen, & Kerfeld, (2014)]. In contrast, for the Cso-BMC and S2CP(30) system, all three heterologous protein cargo tested (GFP, APEX2, LacZ) could be clearly identified in Coomassie-blue stained gels (Figs. 16 and 18).

Table 8. Quantification of the average copy number of enzymes encapsulated per shell and kinetic constants of encapsulated and free enzymes. For average enzyme copy number per shell, values calculated by presuming all shells are either T = 3 or 4 are provided. Kinetic measurements were performed in triplicates and the mean values are shown with the standard error.

Enzyme sample	**Average enzyme copy number per shell (T = 3 / 4)**	**k_{cat} (s^{-1})**	**K_M (mM)**	**k_{cat}/K_M x 10^5 ($s^{-1} \cdot M^{-1}$)**
Free APEX2	N.A.	460 ± 20	0.58 ± 0.07	7.93 ± 1.02
APEX2 + shell	11.0 / 14.6	208 ± 14	0.79 ± 0.12	2.63 ± 0.44
LacZ	N.A.	258 ± 17	0 17 ± 0.02	15.2 ± 2.05
LacZ + shell	3.29 / 4.37	200 ± 14	0.14 ± 0.02	14.3 ± 2.27

EXAMPLE 9:

Production of HO-BMC VLPs in *S. cerevisiae*

Golden-Gate cloning system

Constructs to express HO-BMC VLPs comprised components described in Table 3, Figure 2 and Figure 3 and were assembled according to methods described in Example 2. Briefly, the yeast promoter P_{TDH3} was cloned into HcKan_P and designated HcKan_P- TDH3 (SEQ ID NO: 65). The yeast promoter P_{YEF3} was cloned into HcKan_P and designated HcKan_P-YEF3 (SEQ ID NO: 66). The yeast promoter P_{PYK1} was cloned into HcKan_P and designated HcKan_P- PYK1 (SEQ ID NO: 67). The yeast promoter $P_{GPM1}$was cloned into HcKan_P and designated HcKan_P-GPM1 (SEQ ID NO: 115)

The HO-H ORF was cloned into HcKan_O and designated HcKan_O-HO-H (SEQ ID NO: 68). The HO-P ORF was cloned into HcKan_O and designated HcKan_O-HO-P (SEQ ID NO: 69). The HO-T1 ORF was cloned into HcKan_O and designated HcKan_O-HO-T1 (SEQ ID NO: 70). The HO-T1-SpyTag ORF was cloned into HcKan_O and designated HcKan_O-HO-T1-SpyTag (SEQ ID NO: 116)

The yeast terminator T_{RPL41B} (SEQ ID NO: 80) was cloned into HcKan_T and designated HcKan_T-RPL41B (SEQ ID NO: 71). The yeast terminator T_{HBT1} (SEQ ID NO: 81) was cloned into HcKan_T and designated HcKan_T-HBT1 (SEQ ID NO: 72). The yeast terminator T_{RPS20} (SEQ ID NO: 82) was cloned into HcKan_T and designated HcKan_T-RPS20 (SEQ ID NO: 73). The yeast terminator T_{YPT31} (SEQ ID NO: 105) was cloned into HcKan_T and designated HcKan_T-YPT31 (SEQ ID NO: 119).

The promoter, ORF and terminator parts described above were assembled into the pathway assembly plasmid, pCKU (SEQ ID NO: 74). The assembled HO-BMC pathway was then sub-cloned into pGAU-YMRWδ15 (SEQ ID NO: 75), for chromosomal integration of pathways into the yeast YMRWδ15 site. The construct comprising HO-BMC for integration of the HO-BMC pathway into the yeast YMRWδ15 site was named pGAU-YMRWδ15-HO-BMC (SEQ ID NO: 76). The construct comprising P_{GPM1}-GFP-SpyCatcher-T_{RPS20} for integration of GFP-SpyCatcher into the yeast YPRCδ15 sites was named pGAH-YPRCδ15-GFP-SpyCatcher (SEQ ID NO: 121)

Transformation of plasmids and chromosomal integration of genes into yeast was performed in accordance with the high-efficiency lithium acetate / single-stranded DNA / PEG-3350 protocol as described by Schiestl and co-workers [Gietz, R. D. and Schiestl, R. H. *Nature Protocols* 2: 31 (2007)].

For encapsulation of cargo within HO-BMC, although the EP for HO-BMC has been reported to function when *E. coli* was the recombinant host, we found that it did not work in yeast [Lassila, J. K. *et al*., *Journal of Molecular Biology* 426: 2217-2228 (2014)]. Hence, an alternative methodology for cargo encapsulation, using the SpyCatcher / SpyTag protein conjugation system, into the HO-shell was adopted [Hagen, A., *et al*., *Nature Communications* 9: 2881 (2018)]. This method involved grafting the SpyTag sequence into a shell-facing peptide loop in HO-T1. This modified HO-T1 subunit is termed HO-T1-SpyTag. Cargo proteins with a fusion SpyCatcher domain can therefore form a covalent isopeptide bond with HO-T1-SpyTag and be encapsulated within the HO-shell. The transcriptional units that constitute the HO-shell in yeast are compiled in Table 9 and the yeast strains expressing the HO-shell pathways are compiled in Table 10. A schematic of the synthetic operons of the pathway used to make Cso-BMC in *E. coli* and the HO-ACB pathway used to express HO-BMC in yeast are shown in Figure 3.

Table 9. Compilation of transcriptional units (TUs) (assembled in POT plasmids). The (TUs) are annotated with letters A – D.

POT	A	B
2	P_{TDH3}-HO-H-T_{RPL41B}	
4	P_{YEF3}-HO-T1-T_{RPL41B}	P_{YEF3}-HO-T1-SpyTag-T_{RPL41B}
5	P_{PYK1}-HO-P-SII-T_{RPS20}	

Table 10. List of yeast strains expressing HO-shell pathways in relation to Table 9.

	Assembled TUs
HO-PTH	2A-4A-5A (integrated at *YMRWδ15*)
HO-PT$_{ST}$H+GFP-SpyCatcher	2A-4B-5A (integrated at *YMRWδ15*) + GFP-SpyCatcher (integrated at *YPRCδ15*)

Purification of VLPs

After growing in 8 L of YPD (yeast extract 1 %, peptone 2 %, glucose 2 %, BioBasic) for 48 h at 25 °C, yeast cells were pelleted and lysed using the M-110P microfluidizer at 20,000 psi for eight passes. The lysate was spun twice at 20,000 xg for 20 min each time and the clarified lysate was adjusted to pH 8 using 1 M Tris·HCl pH 12. The lysate was also incubated with 300 µL of biotin blocking buffer (IBA Lifesciences) with gentle stirring for 15 min. StrepTrap affinity purification was performed in the same way as described above.

Results:

Purification of Cso-BMC and HO-BMC

Using the synthetic operons constructed (Fig. 3), we purified Cso-BMC from *E. coli* and HO-BMC from yeast. To the inventor's knowledge, this is the first known instance of recombinant protein shell formation using only two components from the *H. neapolitanus cso* operon. While Silver and co-workers have reported the formation of *H. neapolitanus* carboxysomes in *E. coli*, it was done by transplanting the entire *cso* operon, which encodes ten genes, into *E. coli* [Bonacci, W. *et al.*, *Proceedings of the National Academy of Sciences* 109: 478-483 (2012)]. Our system has simplified the formation of protein shells to only two genes, *csoS1A*, and *csoS4A*. CsoS1A assembles into hexamers that form flat hexagonal tiles while CsoS4A assembles into pentamers that occupy the vertices of the shell, capping the flat tiles formed by CsoS1A and giving the shell its icosahedral geometry. While the resultant Cso-BMC shells are smaller (22 nm in diameter) than the native *H. neapolitanus* carboxysomes (90 to 110 nm in diameter), the synthetic shells are highly uniform in size, as evidenced by DLS measurements (Fig. 17).

Kerfeld and co-workers have reported recombinant expression of HO-BMC in *E. coli* obtained its atomic-scale structure using three shell protomers, HO-H, HO-P and

HO-T1 [Sutter, M. *et al.*, *Science* 356: 1293-1297 (2017)]. The structures and geometric functions of HO-H are similar to that of CsoS1A, while HO-P is analogous to CsoS4A. HO-T1, which resembles a tandem repeat of two HO-H, assembles into trimers that likewise form flat hexagonal tiles. We have managed to reconstruct the HO-BMC in yeast. Based on our current understanding of the literature, this is the first instance of recombinant expression of a BMC-derived protein shell in yeast. While recombinant protein titers in yeast are typically lower compared to *E. coli* (Fig. 14), expression of the HO-BMC VLP in yeast opens up avenues for tailoring by the eukaryotic post-translational modification machinery [Sudbery, P. E. *Curr Opin Biotechnol* 7 (1996)]. It is also noteworthy that many yeast-derived biomolecules, and the organism itself, are afforded the Generally Regarded as Safe (GRAS) status, placing HO-BMC in a good position for vaccine development [Sewalt, V. *et al.*, *Industrial Biotechnology* 12: 295-302 (2016)].

Viewed under the TEM, Cso-BMC appears as capsid-like structures, approximately 20 nm in diameter, with some having angled facets. This shape is reminiscent of the native *H. neapolitanus* carboxysomes, though as previously mentioned, the diameter of the synthetic Cso-BMC is about 20% of the native carboxysome. A plausible reason that accounts for the smaller size of the Cso-BMC is that its luminal space is empty. In the native carboxysome it is known that there are hundreds to thousands of proteins that are tightly packed within the shell [Bonacci, W. *et al. Proceedings of the National Academy of Sciences* 109: 478-483 (2012)]. For bioengineering purposes, it should be more desirable for VLPs to be bereft of their native luminal proteins so that recombinant protein cargo can be more efficiently encapsulated within these shells [Schwarz, B. *et al.*, *Advances in Virus Research* 97: 1-60 (2017)].

Our yeast expressed HO-BMC shells closely resemble those expressed from *E. coli* as reported by Kerfeld and co-workers in terms of size and shape [Sutter et al., *Science* 356: 1293-1297 (2017)]. Sodium dodecyl sulfate polyacrylamide gel (SDS-PAGE) analysis of the protein eluates from affinity purification of Cso-BMC and HO-BMC indicates the presence of the expected shell protomer proteins. As the hexameric (CsoS1A, HO-H) and pentameric (CsoS4A, HO-P) have similar molecular masses (10±1 kDa), they cannot be resolved well by SDS-PAGE. Nonetheless, given the presence of the protein shells, it can be deduced that both species are present in the ~10 kDa protein bands seen. Atomic-level structural details of both the Cso-BMC and

HO-shell indicate that these particles are largely uniform in size [Sutter, M. *et al.*, *Science* 356: 1293-1297 (2017); Tan, Ali, *et al.*, *Biomacromolecules* doi:10.1021/acs.biomac.1c00533 (2021)]. This uniformity in size is a useful feature in VLP engineering, as it translates to predictability when functionalizing VLPs as biomaterials [Schwarz, B. *et al.*, *Advances in Virus Research* 97: 1-60 (2017)].

Summary

BMCs are promising platforms for spatial programming of metabolic reactions in microbial cell factories, and can be repurposed into specialized biochemical delivery vehicles [Kerfeld C. A. *et al. Nature Reviews Microbiology* 16: 277 (2018)]. However, a major hurdle in utilizing these protein shells for such purposes is their often-intricate nature of assembly, which is not easily translatable into recombinant systems. The assembly of protein shells using two types of shell proteins is a significant reduction from the ten identified components from the *H. neapolitanus cso* operon that produced native-like alpha-carboxysomes [Bonacci, W. *et al. Proceedings of the National Academy of Sciences* 109: 478-483 (2012)]. Furthermore, we have identified a sequence, S2CP, which is able to target a heterologous protein cargo into the simplified carboxysome shell. An encapsulation peptide variant that contains 6 more residues, S2CP(30), is shown to be approximately 4 times more efficacious than S2CP at mediating encapsulation of a GFP cargo protein into the Cso-BMC. Hence, both S2CP and S2CP(30) are useful for controlling the quantity of heterologous protein cargo to be packaged within the Cso-BMC. The Cso-BMC is also able to stabilize two enzymes, APEX2 and LacZ, against common enzyme denaturing factors, such as thermal shock, the presence of methanol co-solvent, consecutive freeze-thaw cycles and highly alkaline environments. To the best of our knowledge, this is the first demonstration of utilizing a minimal component BMC-derived shell for hosting and stabilizing enzymes against such denaturing factors. The Cso-BMC expands the current scope of VLPs that can be used to encapsulate and stabilize enzymes [Demchuk & Patel, *Biotechnology Advances*, 41: 107547 (2020)].

We have also recombinantly expressed the HO-BMC in yeast and provide evidence that the shell is able to encapsulate recombinant protein cargo. To the best of our knowledge, this is the first demonstration of recombinant expression of BMC shells in yeast.

References

Any listing or discussion of an apparently prior-published document in this specification should not necessarily be taken as an acknowledgement that such document is part of the state of the art or is common general knowledge.

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica. Section D: Biological Crystallography, 66*(Pt 2), 213-221. doi:10.1107/s0907444909052925.

Anderson, J. C. (2006). Anderson Promoter Library Registry of Standard Biological Parts. Retrieved from parts.igem.org/Promoters/Catalog/Anderson.

Baneyx, F. (1999). Recombinant protein expression in *Escherichia coli*. *Current Opinion in Biotechnology, 10*(5), 411-421. doi:10.1016/S0958-1669(99)00003-8.

Bonacci, W., Teng, P. K., Afonso, B., Niederholtmeyer, H., Grob, P., Silver, P. A., & Savage, D. F. (2012). Modularity of a carbon-fixing protein organelle. *Proceedings of the National Academy of Sciences, 109*(2), 478-483. doi:10.1073/pnas.1108557109.

Cai, F., Bernstein, S. L., Wilson, S. C., & Kerfeld, C. A. (2016). Production and Characterization of Synthetic Carboxysome Shells with Incorporated Luminal Proteins. *Plant Physiology, 170*(3), 1868-1877. doi:10.1104/pp.15.01822.

Cai, F., Dou, Z., Bernstein, S. L., Leverenz, R., Williams, E. B., Heinhorst, S., Kerfeld, C. A. (2015). Advances in Understanding Carboxysome Assembly in *Prochlorococcus* and *Synechococcus* Implicate CsoS2 as a Critical Component. *Life (Basel), 5*(2), 1141-1171. doi:10.3390/life5021141.

Das, S., Zhao, L., Elofson, K., & Finn, M. G. (2020). Enzyme Stabilization by Virus-Like Particles. *Biochemistry, 59*(31), 2870-2881. doi:10.1021/acs.biochem.0c00435.

Demchuk, A. M., & Patel, T. R. (2020). The biomedical and bioengineering potential of protein nanocompartments. *Biotechnology Advances, 41*, 107547. doi:10.1016/j.biotechadv.2020.107547.

Dunn, K. W., Kamocka, M. M., & McDonald, J. H. (2011). A practical guide to evaluating colocalization in biological microscopy. *American Journal of Physiology-Cell Physiology, 300*(4), C723-C742. doi:10.1152/ajpcell.00462.2010.

Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica. Section D: Biological Crystallography, 60*(Pt 12 Pt 1), 2126-2132. doi:10.1107/s0907444904019158.

Fletcher, J. M., Harniman, R. L., Barnes, F. R. H., Boyle, A. L., Collins, A., Mantell, J., Woolfson, D. N. (2013). Self-Assembling Cages from Coiled-Coil Peptide Modules. *Science, 340*(6132), 595-599. doi:10.1126/science.1233936.

Gietz, R. D., & Schiestl, R. H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature Protocols, 2*, 31. doi:10.1038/nprot.2007.13.

Golan, R., Zehavi, U., Naim, M., Patchornik, A., & Smirnoff, P. (1996). Inhibition of *Escherichia coli* beta-galactosidase by 2-nitro-1-(4,5-dimethoxy-2-nitrophenyl) ethyl, a photoreversible thiol label. *Biochimica et Biophysica Acta (BBA) - Bioenergetics, 1293*(2), 238-242. doi:10.1016/0167-4838(95)00254-5.

Goldstein, L. (1972). Microenvironmental effects on enzyme catalysis. Kinetic study of polyanionic and polycationic derivatives of chymotrypsin. *Biochemistry, 11*(22), 4072-4084. doi:10.1021/bi00772a009.

Goldstein, L., Levin, Y., & Katchalski, E. (1964). A Water-insoluble Polyanionic Derivative of Trypsin. II. Effect of the Polyelectrolyte Carrier on the Kinetic Behavior of the Bound Trypsin*. *Biochemistry, 3*(12), 1913-1919. doi:10.1021/bi00900a022.

Guo, Y., Dong, J., Zhou, T., Auxillos, J., Li, T., Zhang, W., Dai, J. (2015). YeastFab: the design and construction of standard biological parts for metabolic engineering in Saccharomyces cerevisiae. *Nucleic Acids Research, 43*(13), e88. doi:10.1093/nar/gkv464.

Hagen, A., Sutter, M., Sloan, N., & Kerfeld, C. A. (2018). Programmed loading and rapid purification of engineered bacterial microcompartment shells. *Nature Communications, 9*(1), 2881. doi:10.1038/s41467-018-05162-z.

Juers, D. H., Hakda, S., Matthews, B. W., & Huber, R. E. (2003). Structural Basis for the Altered Activity of Gly794 Variants of *Escherichia coli* β-Galactosidase. *Biochemistry, 42*(46), 13505-13511. doi:10.1021/bi035506j.

Kalnins, G., Cesle, E.-E., Jansons, J., Liepins, J., Filimonenko, A., & Tars, K. (2020). Encapsulation mechanisms and structural studies of GRM2 bacterial microcompartment particles. *Nature Communications, 11*(1), 388. doi:10.1038/s41467-019-14205-y.

Keeble, A. H., & Howarth, M. (2019). Insider information on successful covalent protein coupling with help from SpyBank. *Methods in Enzymology, 617*, 443-461. doi:10.1016/bs.mie.2018.12.010.

Kerfeld, C. A., Aussignargues, C., Zarzycki, J., Cai, F., & Sutter, M. (2018). Bacterial microcompartments. *Nature Reviews: Microbiology, 16*, 277. doi:10.1038/nrmicro.2018.10.

Klein, M. G., Zwart, P., Bagby, S. C., Cai, F., Chisholm, S. W., Heinhorst, S., Kerfeld, C. A. (2009). Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport. *Journal of Molecular Biology, 392*(2), 319-333. doi:10.1016/j.jmb.2009.03.056.

Küchler, A., Yoshimoto, M., Luginbühl, S., Mavelli, F., & Walde, P. (2016). Enzymatic reactions in confined environments. *Nature Nanotechnology, 11*(5), 409-420. doi:10.1038/nnano.2016.54.

Lam, S. S., Martell, J. D., Kamer, K. J., Deerinck, T. J., Ellisman, M. H., Mootha, V. K., & Ting, A. Y. (2015). Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nature Methods, 12*(1), 51-54. doi:10.1038/nmeth.3179.

Lassila, J. K., Bernstein, S. L., Kinney, J. N., Axen, S. D., & Kerfeld, C. A. (2014). Assembly of robust bacterial microcompartment shells using building blocks from an organelle of unknown function. *Journal of Molecular Biology, 426*(11), 2217-2228. doi:10.1016/j.jmb.2014.02.025.

Lawrence, A. D., Frank, S., Newnham, S., Lee, M. J., Brown, I. R., Xue, W.-F., Warren, M. J. (2014). Solution Structure of a Bacterial Microcompartment Targeting Peptide and Its Application in the Construction of an Ethanol Bioreactor. *ACS Synthetic Biology, 3*(7), 454-465. doi:10.1021/sb4001118.

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I., Adams, P. D. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica Section D: Structural Biology, 75*(10), 861-877. doi:10.1107/S2059798319011471.

Neher, S. B., Sauer, R. T., & Baker, T. A. (2003). Distinct peptide signals in the UmuD and UmuD′ subunits of UmuD/D′ mediate tethering and substrate processing by the ClpXP protease. *Proceedings of the National Academy of Sciences, 100*(23), 13219-13224. doi:10.1073/pnas.2235804100.

Nichols, T. M., Kennedy, N. W., & Tullman-Ercek, D. (2019). Cargo encapsulation in bacterial microcompartments: Methods and analysis. *Methods in Enzymology, 617*, 155-186. doi:10.1016/bs.mie.2018.12.009.

Oltrogge, L. M., Chaijarasphong, T., Chen, A. W., Bolin, E. R., Marqusee, S., & Savage, D. F. (2020). Multivalent interactions between CsoS2 and Rubisco mediate α-carboxysome formation. *Nature Structural & Molecular Biology, 27*(3), 281-287. doi:10.1038/s41594-020-0387-7.

Patterson, D. P., Prevelige, P. E., & Douglas, T. (2012). Nanoreactors by Programmed Enzyme Encapsulation Inside the Capsid of the Bacteriophage P22. *ACS Nano, 6*(6), 5000-5009. doi:10.1021/nn300545z.

Patterson, D. P., Schwarz, B., El-Boubbou, K., van der Oost, J., Prevelige, P. E., & Douglas, T. (2012). Virus-like particle nanoreactors: programmed encapsulation of the thermostable CelB glycosidase inside the P22 capsid. *Soft Matter, 8*(39), 10158-10166. doi:10.1039/C2SM26485D.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera – a visualization system for

exploratory research and analysis. *Journal of Computational Chemistry, 25*(13), 1605-1612. doi:10.1002/jcc.20084.

Sánchez-Sánchez, L., Tapia-Moreno, A., Juarez-Moreno, K., Patterson, D. P., Cadena-Nava, R. D., Douglas, T., & Vazquez-Duhalt, R. (2015). Design of a VLP-nanovehicle for CYP450 enzymatic activity delivery. *Journal of Nanobiotechnology, 13*(1), 66. doi:10.1186/s12951-015-0127-z.

Schwarz, B., Uchida, M., & Douglas, T. (2017). Chapter One - Biomedical and Catalytic Opportunities of Virus-Like Particles in Nanotechnology. In M. Kielian, T. C. Mettenleiter, & M. J. Roossinck (Eds.), *Advances in Virus Research* (Vol. 97, pp. 1-60): Academic Press.

Sewalt, V., Shanahan, D., Gregg, L., La Marta, J., & Carrillo, R. (2016). The Generally Recognized as Safe (GRAS) Process for Industrial Microbial Enzymes. *Industrial Biotechnology, 12*(5), 295-302. doi:10.1089/ind.2016.0011.

Shifrin, S., & Hunn, G. (1969). Effect of alcohols on the enzymatic activity and subunit association of β-galactosidase. *Archives of Biochemistry and Biophysics, 130*, 530-535. doi:10.1016/0003-9861(69)90066-6.

Sievers, F., & Higgins, D. G. (2014). Clustal Omega, accurate alignment of very large numbers of sequences. *Methods in Molecular Biology, 1079*, 105-116. doi:10.1007/978-1-62703-646-7_6.

Silva, C., Martins, M., Jing, S., Fu, J., & Cavaco-Paulo, A. (2018). Practical insights on enzyme stabilization. *Critical Reviews in Biotechnology, 38*(3), 335-350. doi:10.1080/07388551.2017.1355294.

Sudbery, P. E. (1996). The expression of recombinant proteins in yeasts. *Current Opinion in Biotechnology, 7*. doi:10.1016/s0958-1669(96)80055-3.

Sutter, M., Greber, B., Aussignargues, C., & Kerfeld, C. A. (2017). Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell. *Science, 356*(6344), 1293-1297. doi:10.1126/science.aan3289.

Sutter, M., Laughlin, T. G., Sloan, N. B., Serwas, D., Davies, K. M., & Kerfeld, C. A. (2019). Structure of a synthetic beta-carboxysome shell. *Plant Physiology, 181*(3), 1050–1058. doi:10.1104/pp.19.00885.

Tan, Y. Q., Ali, S., Xue, B., Teo, W. Z., Ling, L. H., Go, M. K., . . . Yew, W. S. (2021). Structure of a Minimal α-Carboxysome-Derived Shell and Its Utility in Enzyme Stabilization. *Biomacromolecules*. doi:10.1021/acs.biomac.1c00533.

Tan, Y. Q., Xue, B., & Yew, W. S. (2021). Genetically Encodable Scaffolds for Optimizing Enzyme Function. *Molecules, 26*(5), 1389. Retrieved from wwwdotmdpidotcom/1420-3049/26/5/1389.

Tanaka, S., Kerfeld, C. A., Sawaya, M. R., Cai, F., Heinhorst, S., Cannon, G. C., & Yeates, T. O. (2008). Atomic-Level Models of the Bacterial Carboxysome Shell. *Science, 319*(5866), 1083-1086. doi:10.1126/science.1151458.

Thomas, F., Boyle, A. L., Burton, A. J., & Woolfson, D. N. (2013). A Set of de Novo Designed Parallel Heterodimeric Coiled Coils with Quantified Dissociation Constants in the Micromolar to Sub-nanomolar Regime. *Journal of the American Chemical Society, 135*(13), 5161-5166. doi:10.1021/ja312310g.

Tsai, Y., Sawaya, M. R., Cannon, G. C., Cai, F., Williams, E. B., Heinhorst, S., . . . Yeates, T. O. (2007). Structural Analysis of CsoS1A and the Protein Shell of the *Halothiobacillus neapolitanus* Carboxysome. *PLoS Biology, 5*(6), e144. doi:10.1371/journal.pbio.0050144.

Turmo, A., Gonzalez-Esquer, C. R., & Kerfeld, C. A. (2017). Carboxysomes: metabolic modules for CO_2 fixation. *FEMS Microbiology Letters, 364*(18), fnx176. doi:10.1093/femsle/fnx176.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics, 25*(9), 1189-1191. doi:10.1093/bioinformatics/btp033.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., . . . Wilson, K. S. (2011). Overview of the CCP4 suite and current developments.

Acta Crystallographica. Section D: Biological Crystallography, 67(Pt 4), 235-242. doi:10.1107/s0907444910045749.

Yu, Z., Reid, J. C., & Yang, Y.-P. (2013). Utilizing Dynamic Light Scattering as a Process Analytical Technology for Protein Folding and Aggregation Monitoring in Vaccine Manufacturing. *Journal of Pharmaceutical Sciences, 102*(12), 4284-4290. doi:10.1002/jps.23746.

Zhang, K. (2016). Gctf: Real-time CTF determination and correction. *Journal of Structural Biology, 193*(1), 1-12. doi:10.1016/j.jsb.2015.11.003

Zhang, Y., Tsitkov, S., & Hess, H. (2016). Proximity does not contribute to activity enhancement in the glucose oxidase–horseradish peroxidase cascade. *Nature Communications, 7*(1), 13982. doi:10.1038/ncomms13982

Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J. H., Lindahl, E., & Scheres, S. H. W. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife, 7*, e42166. doi:10.7554/eLife.42166.

What is claimed is:

1. A method for producing a bacterial microcompartment virus-like particle (VLP) carrying a cargo molecule, said method comprising

A) introducing into a host cell or organism one or more heterologous polynucleotides comprising

(i) a first sequence encoding bacterial microcompartment shell protomers; and

(ii) a second sequence encoding a cargo molecule fused to an encapsulation peptide, wherein the encapsulation peptide comprises the amino acid sequence set forth in SEQ ID NO: 1 (SKITGSSGNDTQGSLITYSGGARG) or SEQ ID NO: 94 (KPEKPGSKITGSSGNDTQGSLITYSGGARG), or a functional variant thereof;

a) expressing the first and second sequences; and

b) forming a microcompartment that encapsulates the cargo molecule; or

B) introducing into a host cell or organism one or more polynucleotides comprising

(i) a first sequence encoding bacterial microcompartment shell protomers; and

(ii) a second sequence encoding at least one of said protomers fused with a cargo molecule or a biochemical tag;

a) expressing the first and second sequences; and

b) forming a microcompartment that expresses the cargo molecule on an exterior surface, or

c) forming a microcompartment that expresses the biochemical tag on an exterior surface to which a cargo molecule comprising a complementary tag can bind.

2. The method of claim 1, wherein the functional variant of the encapsulation peptide set forth in SEQ ID NO: 1 comprises, at its amino terminus, 1, 2, 3, 4, or 5 of the additional amino acids at the amino terminus of SEQ ID NO: 94, wherein such variants are intermediates between the sequences SEQ ID NO: 1 and SEQ ID NO: 94.

3. The method of claim 1 or 2, wherein the bacterial microcompartment protomers are derived from *Halothiobacillus neapolitanus* or *Haliangium ochraceum.*

4. The method of claim 3, wherein the bacterial microcompartment protomers are CsoS1A (SEQ ID NO: 2) and CsoS4A (SEQ ID NO: 3) from *Halothiobacillus neapolitanus*; or HO-H (SEQ ID NO: 4), HO-P (SEQ ID NO: 5) and HO-T1 (SEQ ID NO: 6) from *Haliangium ochraceum*, and variants thereof.

5. The method of any one of claims 1 to 4, wherein the cargo molecule is at least one peptide, such as an enzyme and/or fluorescent protein and/or immunogenic peptide.

6. The method of any one of claims 1 to 5, wherein the biochemical tag may be selected from the group comprising Strep-Tag II (SII), SpyCatcher / SpyTag (SC / ST) pair and CC-Di-A / B (CCA / CCB) pair.

7. The method of any one of claims 4 to 6, wherein the expression of CsoS1A is controlled by promoter P_{T7}; CsoS4A is controlled by promoter P_{CON5}; HO-H is controlled by yeast promoter P_{TDH3}; HO-P is controlled by yeast promoter P_{PYK1} and HO-T1 is controlled by yeast promoter P_{YEF3}.

8. The method of any one of claims 1 to 7, wherein the host organism is *E. coli* or *S. cerevisiae.*

9. An engineered bacterial microcompartment VLP carrying a cargo molecule, comprising:
   i) bacterial microcompartment shell protomers, and a cargo molecule fused to an encapsulation peptide, wherein the encapsulation peptide comprises the amino acid sequence set forth in SEQ ID NO: 1 or SEQ ID NO: 94, or a functional variant thereof; or
   ii) bacterial microcompartment shell protomers and a cargo molecule, wherein the cargo molecule is fused to an end of at least one of said protomers, or wherein at least one of said protomers is fused to a tag and a cargo molecule comprising a complementary tag is bound to it on the exterior surface of the VLP.

10. The engineered VLP of claim 9, wherein the functional variant of the encapsulation peptide set forth in SEQ ID NO: 1 comprises, at its amino terminus, 1, 2, 3, 4, or 5 of the additional amino acids at the amino terminus of SEQ ID NO: 94, wherein such variants are intermediates between the sequences SEQ ID NO: 1 and SEQ ID NO: 94.

11. The engineered VLP of claim 9 or 10, wherein the bacterial microcompartment protomers are derived from *Halothiobacillus neapolitanus* or *Haliangium ochraceum.*

12. The engineered VLP of claim 11, wherein the bacterial microcompartment protomers are CsoS1A comprising the amino acid sequence set forth in SEQ ID NO: 2 and CsoS4A comprising the amino acid sequence set forth in SEQ ID NO: 3 from *Halothiobacillus neapolitanus*; or HO-H comprising the amino acid sequence set forth in SEQ ID NO: 4, HO-P comprising the amino acid sequence set forth in SEQ ID NO: 5 and HO-T1 comprising the

amino acid sequence set forth in SEQ ID NO: 6 from *Haliangium ochraceum*, and variants thereof.

13. The engineered VLP of any one of claims 9 to 12, wherein the cargo molecule is at least one peptide, such as an enzyme and/or fluorescent protein and/or immunogenic peptide.

14. The engineered VLP of any one of claims 9 to 13, wherein the biochemical tag is selected from the group comprising Strep-Tag II (SII), SpyCatcher / SpyTag (SC / ST) pair and CC-Di-A / B (CCA / CCB) pair.

15. An isolated plasmid or vector nucleic acid comprising:

a) a first DNA sequence that encodes bacterial microcompartment shell protomers, each of which is operably linked to a promoter, and

b) a second DNA sequence that encodes a cargo molecule fused to an encapsulation peptide, operably linked to a promoter, wherein the encapsulation peptide comprises the amino acid sequence set forth in SEQ ID NO: 1 or SEQ ID NO: 94, or a functional variant thereof; or

c) a first DNA sequence that encodes bacterial microcompartment shell protomers, each of which is operably linked to a promoter, and

d) a second DNA sequence encoding at least one of said protomers fused with a cargo molecule or a biochemical tag.

16. The isolated plasmid or vector of claim 13, wherein the functional variant of the encapsulation peptide set forth in SEQ ID NO: 1 comprises, at its amino terminus, 1, 2, 3, 4, or 5 of the additional amino acids at the amino terminus of SEQ ID NO: 94, wherein such variants are intermediates between the sequences SEQ ID NO: 1 and SEQ ID NO: 94.

17. The isolated plasmid or vector of claim 15 or 16, wherein the bacterial microcompartment shell protomers, promoters, cargo molecules and tags are as defined in any one of claims 1 to 7.

18. The isolated plasmid or vector of claim 17, wherein the DNA sequences encoding said bacterial microcompartment shell protomers, cargo molecules and tags have at least 70%, at least 80%, at least 90%, or 100% identity with SEQ ID Nos 7-12 and 95 due to redundancy of the genetic code.

19. A composition or combination comprising at least one engineered VLP of any one of claims 9 to 14 for use in:

a) the prophylaxis or treatment of disease in a subject; or

b) a biochemical process.

20. The composition or combination of claim 19, wherein the at least one engineered VLP comprises an enzyme for conversion of a prodrug.

21. The combination of claim 19 or 20, comprising one or more additional therapeutic agents.

22. The composition or combination of claim 19 or 21, which is a vaccine.

23. Use of at least one engineered VLP of any one of claims 9 to 14 in the manufacture of a medicament for the prophylaxis or treatment of a disease in a subject.

24. A method of prophylaxis or treatment, comprising administering to a subject in need of such treatment an efficacious amount of an engineered VLP of any one of claims 9 to 14.