Lawrence Berkeley National Laboratory

Title

Characterization of a widespread sugar phosphate-processing bacterial microcompartment.

Permalink

https://escholarship.org/uc/item/94q6414h

Journal

Communications Biology, 7(1)

Authors

Dwyer, Matthew

Sutter, Markus

Kerfeld, Cheryl

Publication Date

2024-11-24

DOI

10.1038/s42003-024-07287-y

Peer reviewed

Characterization of a widespread sugar phosphate-processing bacterial microcompartment

Check for updates

Matthew E. Dwyer [1,2], Markus Sutter [1,3,4] & Cheryl A. Kerfeld [1,2,3,4] ✉

Many prokaryotes form Bacterial Microcompartments (BMCs) that encapsulate segments of specialized metabolic pathways to enhance catalysis. The various functions of metabolosomes, catabolic BMCs, are dictated by the signature enzyme that processes initial substrates of the confined pathway. The components and native functions of several metabolosomes have been experimentally characterized; however one of the most prevalent across all bacteria has yet to be studied. Sugar Phosphate Utilizing (SPU) BMC loci encode enzymes predicted to be involved in sugar phosphate metabolism. The SPU genetic loci are found in organisms occupying habitats ranging from soils to hot springs, highlighting the ubiquity of the SPU BMC. We bioinformatically characterized seven SPU subtypes, all which contain an enzyme unique to SPU BMCs, a deoxyribose 5-phosphate aldolase (DERA). Here, we define the fundamental characteristics of SPU BMCs and have expressed, purified, and characterized a set of SPU core enzymes. These include a protein-protein complex formed between a SPU BMC DERA and a predicted ribose 5-phosphate isomerase. Further, we show that the SPU BMC DERA is catalytically active and propose that it acts as the universal signature enzyme for the SPU BMC, with implications for fundamental understanding and biotechnological applications of SPU BMCs.

Bacterial microcompartments (BMCs) are proteinaceous organelles utilized by diverse groups of bacteria to efficiently metabolize a substrate while confining or restricting access to cross-reactive intermediates; typically, they are niche-specific innovations enabling the organism to thrive in a given environment[1–5]. Common among all BMCs are the shell proteins forming the icosahedral, sometimes polyhedral bounding membrane. Shell facets are tiled with pfam00936 BMC-H hexamers[6] and BMC-T trimers[7]. BMC-T proteins are fusions of two pfam00936 domains, giving the same overall fold and appearance of a BMC-H hexagon. Two architectural types of BMC-T proteins can integrate into the BMC shell. BMC-T^s is a trimer that tiles with BMC-H proteins to form a single, planar shell membrane. BMC-T^{dp} proteins, however, protrude out of the shell membrane layer due to the dimerization of two BMC-T domains across their concave faces[8,9]. At the vertices are pfam03319 BMC-P pentamers, providing caps to the shell[10]. All BMCs share these protein shell architectural features, but the specific internal metabolic reactions differ depending on the encapsulated enzymes. Carboxysomes, the most well-known type, are the only type of anabolic BMC; they function at the heart of the carbon-concentrating mechanism in cyanobacteria and some chemoautotrophs[11,12]. Metabolosomes, discovered through shell protein sequence homology with carboxysomes[13], comprise the other class of bacterial microcompartments. The common core of these catabolic organelles consists of a signature enzyme(s) that generates an aldehyde that is further processed by the co-encapsulated aldehyde dehydrogenase[1,3,14,15]. Some metabolosome loci also encode the gene for a phosphotransacylase (PTAC) to regenerate Coenzyme A and create an acyl-phosphate[16], and an alcohol dehydrogenase for cofactor regeneration[15] within the organelle.

Encapsulation is critical for sequestering toxic intermediates generated by the enzymes inside the lumen of the BMC, and enhancing substrate-to-enzyme ratios[1,3,4,17–20]. Many enzymes encoded within BMC-containing loci contain sequence extensions, referred to as encapsulation peptides (EPs), that facilitate packaging inside the BMC[21–24]. These typically consist of an amphipathic alpha helix separated from the main protein domain by a flexible/poorly conserved linker[23]. These EPs tend to render the proteins insoluble when expressed in heterologous systems[16,25]. Functionally, EPs facilitate the aggregation of native BMC core proteins[24] in addition to binding to the grooves between tessellating shell protein tiles[26].

Recently, a comprehensive bioinformatic survey unveiled the widespread occurrence of a functionally uncharacterized BMC, the Sugar Phosphate Utilizing (SPU) BMC[14], spread across 26 bacterial phyla and clustered into seven subtypes[2]. Genomes of organisms from prominent phyla such as the Chloroflexi, Deltaproteobacteria, Elusimicrobia, Planctomycetes, Bacillus, Firmicutes, Clostridiales, and Gammaproteobacteria all contain SPU BMC loci[2,27]. Organisms encoding a SPU BMC locus are represented in a diverse range of habitats, including soils and hot springs, often in communities with cyanobacteria[27,28].

Central to the SPU BMC metabolism is a predicted DeoC-family class I aldolase enzyme (pfam01791) which converts deoxyribose 5-phosphate into glyceraldehyde 3-phosphate (GAP) and acetaldehyde, hence we refer to this enzyme as deoxyribose 5-phosphate aldolase (DERA). Among BMCs, DERA only occurs in SPU BMC loci. A second enzyme, also unique to SPU BMCs and often encoded adjacent to DERA, is a pfam02502 ribose 5-phosphate isomerase B family protein (hereafter RPI). This enzyme complement suggests that this BMC plays role in breaking down DNA degradation products, presumably in environments rich in detritus, and recycling the carbon back into the central metabolism through glyceraldehyde 3-phosphate[2]. DNA is widely available and in some cases quite stable[29] in soil environments, providing an abundant energy source for organisms containing SPU BMCs. Plants, soil fauna, and fungi, in addition to other bacteria contribute to the pool of environmental DNA within these habitats through various methods of secretion and cellular lysis[29,30]. Salvaging of these nucleotides from the environment is energetically favorable relative to the cost of de novo synthesis, and many organisms benefit from the recycling of DNA in external pools[30]. Upstream of 2-deoxy-alpha-D-ribose 5-phosphate, the presumed substrate of a SPU BMC, are precursors like 2-deoxy-alpha-D-ribose 1-phosphate that bacteria can convert into 2-deoxy-alpha-D-ribose 5-phosphate using a deoxyribose 1,5 phosphomutase (*deoB* gene)[31,32]. While some organisms will benefit from the symbiotic turnover of DNA precursors into more nucleic acid building blocks, bacteria-harboring SPU BMCs in the same environment could convert those breakdown products into cellular energy.

The SPU BMC has remained functionally uncharacterized despite being one of the most numerically abundant loci in available microbial genome sequence data. Here, we have performed a detailed analysis of the locus components of the seven SPU subtypes and propose the common core chemistry and modes of enzyme encapsulation. Further, we functionally characterized and validated the SPU BMC signature enzyme and accessory protein for solubility, oligomeric states, and enzymatic activity of the proposed substrates from a SPU6 representative informed from our expanded bioinformatic analysis. We propose a mechanism for SPU catabolic BMCs in the breakdown of deoxyribonucleosides into useful cellular building blocks. Additionally, the reversible aldolase reaction inside the BMC offers the potential for industrial use for the synthesis of organic compounds on immobilized enzyme scaffolds.

Results
SPU BMC locus analysis

We undertook a detailed analysis of the SPU loci compiled in our recent bioinformatic survey of all BMC types[2,27] to putatively assign their potential function in an organism. The previous study clustered the loci based on the amino acid sequences of all enzymes and shell proteins and listed locus fragments that clustered with SPU loci. Many of the loci originate from incomplete genomes and some contain no enzymes, only a few shell proteins. We therefore further curated the dataset, removing loci that were missing several genes typically observed, missing an aldehyde dehydrogenase, or were found to originate from obsolete datasets (see "Methods" for details). Our curated dataset includes 206 SPU BMC loci (from 280 in the original dataset) across seven subtypes (Supplemental Dataset S1). We then analyzed the composition of the loci to determine which enzymes were present across those seven types (Fig. 1A and Supplemental Table S1). RPI is less common than DERA in SPU4 (18 out of 29 loci) and completely absent in SPU5 (Supplemental Table S1). We also identified a RPI and DERA fusion protein in 20
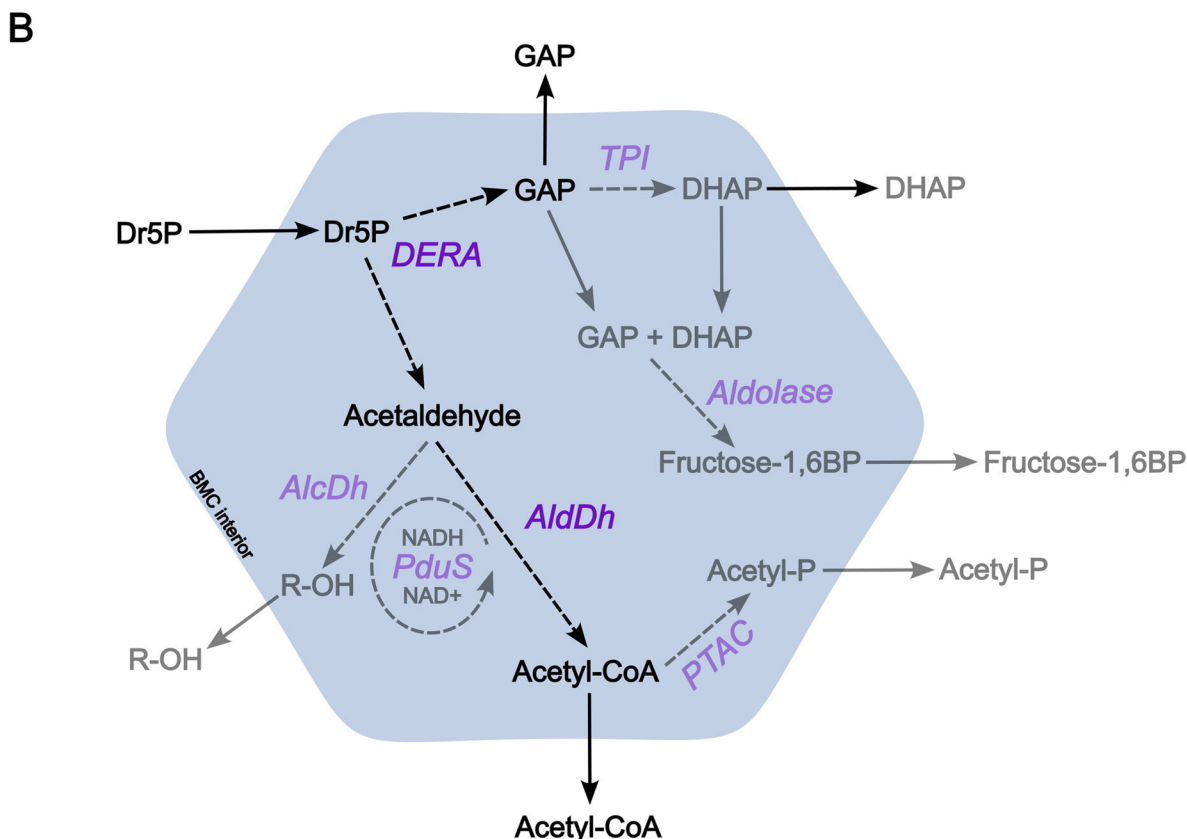
SPU3 loci (Supplemental Table S1). The fusion, as well as the observation that the genes for these two enzymes are often found adjacent in the locus, suggests a functional connection between them. All SPU BMC loci all contain the central metabolosome core enzyme aldehyde dehydrogenase as well as varying numbers of other typical BMC core enzymes such as a PTAC (pfam06130) that is found in almost all SPU1, SPU4, and SPU7 (Fig. 1A and Supplemental Table S1) and an alcohol dehydrogenase (pfam00465) in some SPU2, SPU4, and SPU6 (Supplemental Table S1). Other frequently associated enzymes are a PduS homolog in all but SPU6, a triosephosphate isomerase (pfam00121, TPI) that is prominently found in SPU1 and SPU4 as well as in some loci of the other types and a class II aldolase at a low occurrence in some types (Fig. 1A) (Supplemental Table S1).

The shells of the SPU subtypes have some distinctive differences that likely reflect their particular functions. In bioinformatic classifications, shell proteins have been found to cluster not necessarily by BMC type, so instead function agnostic colors were assigned to the clades of shell protein phylogenetic trees[27,33]. For BMC-H it was observed that a basal type of BMC-H found in the blue clade is typically the most abundant component of a given BMC type[2,33]. Across SPU subtypes, the most prevalent color clade for BMC-H proteins is the basal blue family[2,33]. SPU4 and SPU6 contain pink-type BMC-H proteins which are likewise a basal clade. Other members of the pink clade of BMC-H proteins are found in alpha-carboxysome loci of cyanobacteria and proteobacteria[33]. All SPU BMC subtypes encode at least one BMC-T[s] from various color clades, while genes for BMC-T[dp] are lacking in SPU1, SPU4, and SPU7 subtypes (Fig. 1A). The BMC-T[dp] proteins encoded in SPU BMC loci are diverse in color clade but phylogenetically similar to either the alpha-carboxysome BMC-T[dp] protein CsoS1D (SPU5) or its beta-carboxysomal homolog CcmP (SPU2/SPU3)[27,33]. In addition, all SPU BMC-encoding loci contain at least a triplet of BMC-P family proteins as described previously (Fig. 1A)[2], following the pattern of two BMC-P proteins from orange and gray color clades with the third represented by either purple, green, or blue color clades (Fig. 1A). Why an organism contains multiple paralogs of BMC-P-encoding genes remains unknown; their presence suggests they may have an additional structural or functional role because pentamers are generally a minor component of the shell (e.g., only 12 are needed to form a regular icosahedron).

To provide a framework for subsequent studies of the functions of the SPU subtypes, we bioinformatically developed schema for the core reactions and the metabolites involved. Based on the SPU locus enzyme complement, we predict that the DNA degradation product deoxyribose 5-phosphate enters the SPU BMC shell to be processed by DERA (Fig. 1A, B). The predicted entry metabolite, Dr5P, has the complementary charge to travel through positively charged pores of a SPU fuchsia BMC-H (Supplemental Fig. S1A) and primary SPU BMC-H (Supplemental Fig. S2) as well as an appropriate size for transit through the pore (Supplemental Fig. S1B), alternatively, or in addition Dr5P could enter through the larger pore of the BMC-T[dp] (Supplemental Fig. S1C)[27]. Hypothetically, after Dr5P enters the SPU BMC shell, the core SPU BMC metabolism initiated by DERA generates a toxic intermediate, acetaldehyde, that is further processed by the encapsulated aldehyde dehydrogenase to create Acetyl-CoA. Without an encapsulated PTAC and thus no further metabolism of Acetyl-CoA, this metabolite could exit the shell through the larger BMC-T[dp] pore (Fig. 1B). Interestingly, for SPU BMCs lacking putatively encapsulated PTACs, the operons typically encode for a BMC-T[dp] (Fig. 1A). Concurrently, DERA also creates the sugar-phosphate glyceraldehyde 3-phosphate that can leave the shell to participate in central carbon metabolism such as glycolysis (Fig. 1B). Due to the SPU BMC subtypes differing in their genetic composition beyond all subtypes containing DERA and AldDH, there are several hypothesized metabolites that exit a SPU BMC shell that require future experimental validation (Fig. 1A, B). The metabolic pathway initiated by DERA could proceed with TPI conversion of GAP to DHAP and subsequently a class II aldolase converting GAP and DHAP to fructose 1,6 bisphosphate to exit the shell for cytoplasmic metabolism. A fructose 1,6 bisphosphate molecule would possibly require passage through the larger, gated pore of the BMC-T[dp] for exiting the shell (Supplemental Fig S1D).

Fig. 1 | SPU BMCs contain enzymes relevant to nucleotide degradation and sugar processing. A List of enzymes, shell protein color clades, and predicted BMC-related entry and exit metabolites found across SPU BMC subtypes. Enzyme type is listed if it exists in over 50% of the SPU BMC subtype loci as found in Supplemental Table S1. Colors represent the broader color clades of the shell proteins found in each specific SPU BMC subtype and the numbers represent how many shell proteins are in the BMC subtype locus. **B** Schematic of a hypothetical SPU BMC with Dr5P substrate being processed by SPU DERA within the shell interior generating glyceraldehyde 3-phosphate and acetaldehyde. Acetaldehyde is then processed by core metabolosome enzyme, AldDh, to break down the toxic intermediate to Acetyl-CoA which could exit the shell. In gray are hypothetical metabolite fluxes for SPU BMCs containing additional enzymes within the shell-encoded loci. Dashed line arrows are enzymatic pathways and solid arrows are metabolite fluxes. DERA deoxyribose 5-phosphate aldolase, RPI ribose 5-phosphate isomerase, AldDH aldehyde dehydrogenase, PTAC phosphotransacetylase, PduS PduS/Complex 51 K/RnfC/SLBB domain-containing protein, TPI triosephosphate isomerase, Aldolase class II aldolase, GAP glyceraldehyde 3-phosphate, DHAP dihydroxyacetone phosphate, Fructose 1,6 BP fructose 1,6 bisphosphate.

Further, the Acetyl-CoA generated by AldDH could be converted to Acetyl-P by an encoded PTAC. If an AlcDH is present in the SPU BMC, the acetaldehyde can also be converted to an R-OH molecule along with NAD+ recycling. An alternate mode of NAD+ recycling can also be carried out by an encoded PduS, namely MNdh, as proposed recently[34]. The final exit metabolite would then depend on the SPU BMC subtype enzyme composition (Fig. 1B). In addition to our characterization of the universal SPU signature enzyme (below), substantial experimental effort can now be undertaken to validate all of these predictions for each SPU subtype.

Structural and bioinformatic analysis of SPU core enzymes

In order to better characterize the SPU enzymes, we predicted their structures. Using sequence alignment (Fig. 2A) and AlphaFold2 for structure prediction (Fig. 2B) we found that all SPU-associated DERA enzymes have an N-terminal extension (Fig. 2B and Supplemental Table S2) that consists of an alpha helix (~30 residues; pIDDT confident ~75–85) and two beta strands (pIDDT low confidence ~35–60). The predicted alpha helix is amphipathic, as characteristic of EPs[23], however at 25 amino acids it is slightly longer than typical EP helices[23,24]. This extension is not found in non-BMC-associated DERA, further implicating its function as EP, which are involved in cargo encapsulation, either by interacting with other enzymes to form the catalytic core, or interacting with the BMC shell, or both. While it initially seemed that this extension was not found on all BMC-associated DERA, when we examined the genomic sequences directly to manually assign start sites, we could consistently find this extension upstream of the computationally annotated start site. This highlights a
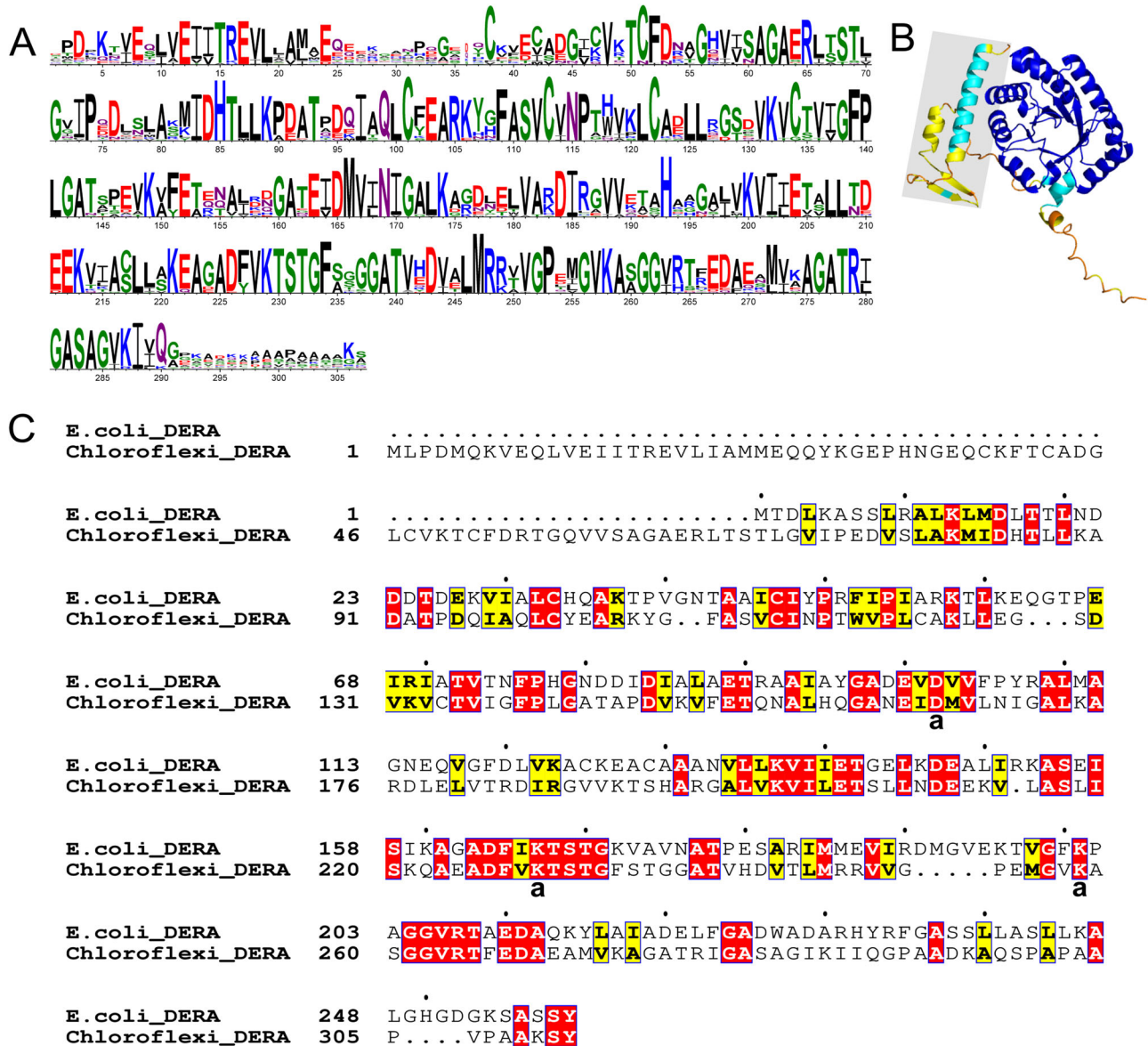
Fig. 2 | SPU6-associated DERA contains an extra N-terminal domain and has all active sites conserved with *E. coli* DERA. A Sequence conservation logo of the SPU6 DERA. The N-terminal extension spans 70 residues as seen in (**B**). B Alpha-Fold2 model of Chloroflexi DERA colored by AlphaFold2 secondary structure confidence (pIDDT) (Blue = very high, cyan = confident, yellow = low, orange = very low). The N-terminal extension is highlighted with a gray box. **C** Pairwise sequence alignment of *E. coli* and Chloroflexi DERA. All three *E. coli* DERA active sites (D102, K167, K201) are conserved in Chloroflexi DERA (D165, K229, K258). Dots represent every ten residues, yellow boxes represent similar residue substitutions, and red boxes represent conserved residues. Active sites are labeled as "a" in bold.

general issue with annotation of BMC-related enzymes that have extensions relative to non-BMC-associated homologs; automated assignments miss these extensions on BMC-encapsulated homologs because they rely on comparison with existing proteins or pfams that are likely based on proteins that lack those extensions. We therefore manually corrected sequence start sites for the other SPU BMC encoded enzymes and documented whether they contained putative encapsulation peptides or extensions (Supplemental Table S2), to further support the compartmentalization of key SPU BMC metabolic reactions.

The sequence alignment and structural prediction of the SPU-associated RPI reveals a typical N-terminal EP with an alpha helix connected to a less conserved, unstructured linker of around 75 residues before the typical RPI domain fold (Supplemental Fig. S3 and Supplemental Table S2). As with the DERA extension, automated annotations in databases frequently lack this extension. An EP can also be predicted for the SPU aldehyde dehydrogenases (C-terminal for SPU6, N-terminal for all

others). Furthermore, EPs can also be predicted for all SPU PTACs, SPU4 TPI (N-terminal), and SPU5 class II aldolase (C-terminal) (Supplemental Table S2).

We also found a chimera of the SPU signature enzyme DERA (pfam01791) and accessory protein RPI (pfam02502) that occurs in SPU3 loci[27]. AlphaFold2 modeling (Supplemental Fig. S4A) and sequence alignments (Supplemental Fig. S4B) of a representative SPU3 chimera from a *Verrucomicrobia bacterium* ADurb.Bin018 (UniProt ID BWX54) predicted an N-terminal RPI domain followed by a C-terminal DERA domain. What looks like a modeled "linker" region between RPI and DERA domains, is actually a structural element of the DERA domain (Supplemental Fig. S4A), as determined by structural superposition to Chloroflexi DERA (Supplemental Fig. S4B). This region might serve the function of an EP, given the evidence for putative N-terminal EPs on SPU BMC DERAs across all SPU subtypes and notably SPU5 where RPI is absent (Supplemental Table S2).
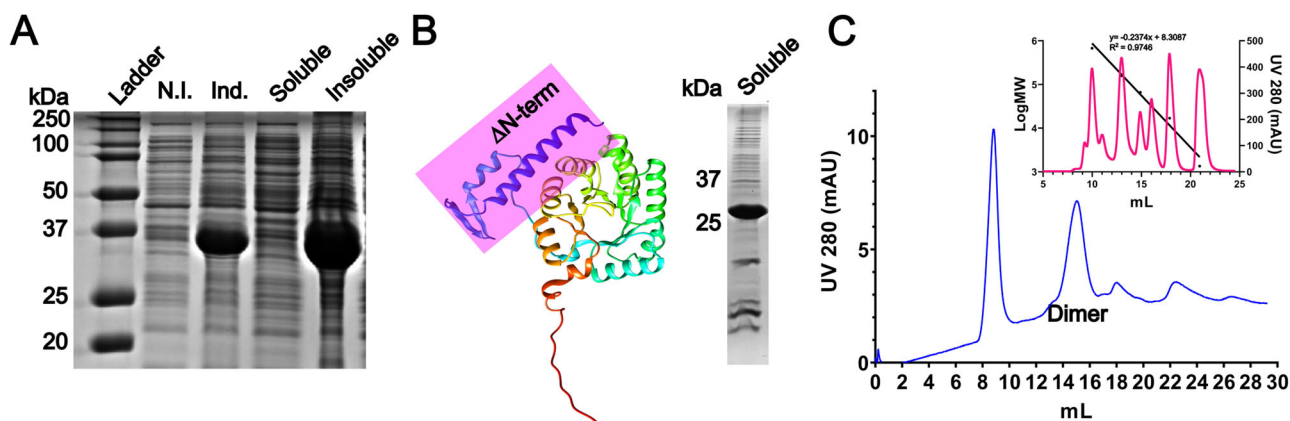
Fig. 3 | Purification of recombinantly expressed SPU6 Chloroflexi DERA.
A Coomassie-stained SDS-PAGE of HisTrap-affinity purification of HIS-DERA
(35.4 kDa). **B** After deletion of the N-terminal extension (pink box in AlphaFold2
model, N to C rainbow coloring), the protein purified in the soluble fraction as
observed on SDS-PAGE (MW of HIS-ΔN-term DERA: 26.9 kDa). **C** Soluble ΔN-
term DERA loaded onto a S200 Size Exclusion Column indicated a dimeric oligo-
meric state eluting at around 15 mL. Inset shows gel filtration standard analysis with
additional BSA protein to determine oligomeric state.

Expression, purification, and functional analysis of SPU6 DERA and RPI

To test our hypothesis for the initial SPU BMC reaction of utilizing
2-deoxyribose 5-phosphate, we selected an operon from the SPU6 subtype
out of the organism *Chloroflexi bacterium* GWB2_54_36 (UniProt: A2X24)
to characterize the DERA and RPI enzymes. Notably, this operon was
selected due to the interest in photosynthetic microbial communities and for
its relatively small operon size. However, other organisms containing dif-
ferent SPU BMC subtypes also warrant investigation. In initiating this study,
we did try expressing the set of SPU enzymes from a different organism
(Anaerolineae bacterium SM23_84, UniProt ID AMJ93) but the proteins
were insoluble. This *Chloroflexi bacterium* GWB2_54_36 BMC operon
contains a GntR-like family regulator protein followed by RPI, DERA, and
AldDH enzymes[27]. GntR-like family proteins consist of two major domains:
a N-terminal DNA-binding helix-turn-helix motif and C-terminal effector-
binding or oligomerization domain. The gene encoding for this GntR-like
protein in this Chloroflexi genome is annotated to have an N-terminal
pfam00392 GntR domain and C-terminal pfam07702 UTRA domain.
UTRA domains alter transcription upon interactions with small molecules
like sugar phosphates[35,36]. The shell proteins encoded in this locus include a
single fuchsia BMC-H, one bright blue BMC-T^{dp}, and three BMC-P genes
representing the squash, bordeaux, and silver color clades (Supplemental
Fig. S5). Both hypothetical proteins encoded in this SPU BMC operon do
not have homology to relevant BMC-related proteins or have pfam
assignments.

To functionally characterize the DERA and RPI enzymes from this
locus, we synthesized the coding regions with N-terminal HIS affinity tags
for purification. Heterologous expression of Chloroflexi HIS-DERA in *E.
coli* initially yielded insoluble protein (Fig. 3A). Because we predicted the
~70 residue N-terminal extension for SPU6 DERA enzymes to be a putative
EP (Fig. 2A–C), likely causing protein aggregation as previously observed
for a BMC-related PTAC[16], we deleted the extension (ΔN-term DERA). The
deletion yielded soluble DERA protein (Fig. 3B) that eluted in Size Exclusion
Chromatography (SEC) at about 50 kDa (other than the void volume of
HisTrap contaminants). This corresponds to a DERA dimer which is also
the oligomeric form of the *E. coli* ortholog[37] (Fig. 3C).

We were successful in purifying the full-length Chloroflexi RPI using
HIS-tag affinity purification (Supplemental Fig. S6A). Even though this
protein is also predicted to contain an EP (based on sequence and Alpha-
Fold2 models, Supplemental Fig. S3), it was soluble. In SEC the Chloroflexi
RPI eluted as a tetramer, similar to the *E. coli* ortholog[38] (Supplemental
Fig. S6B).

We next tested our bioinformatic predictions for DERA and RPI
enzymatic activities. The residues contributing to catalytic activity in *E. coli*

(D102, K167, K201) are conserved for the Chloroflexi DERA (D165, K229,
K258) (Fig. 2C)[37]. Using an NADH-coupled enzyme assay for aldolase
activity[39], we were able to measure activity for the substrate deoxyribose
5-phosphate proportional to the available substrate over a range of con-
centrations (Fig. 4). The enzyme V_{max} was 0.100 ± 0.020 at 5 nM con-
centration and K_m was 0.104 ± 0.010 for deoxyribose 5-phosphate substrate.
In contrast, the Chloroflexi RPI enzyme only has one of the conserved active
site residues responsible for the isomerization of ribose 5-phosphate to
ribulose 5-phosphate known for the *E. coli* ortholog[38]. The cysteine residue
(*E. coli* Cys66) important for catalytic isomerase activity is substituted by an
aspartic acid in the Chloroflexi RPI, and all SPU RPIs, while the other active
site residue (*E. coli* His99, Chloroflexi His152) is conserved (Supplemental
Fig. S3C). When isomerase activity was assayed by tracking the conversion
of ribose 5-phosphate to ribulose 5-phosphate at an absorbance of 290 nm,
negligible activity was seen for Chloroflexi RPI compared to the robust
activity measured for an *E. coli* RPI control (Supplemental Fig. S7).

Complex formation between SPU6 DERA and RPI

Because fusions of DERA and RPI exist in nature, we next tested whether the
two enzymes could form a complex. Similar to the insolubility of DERA due
to its N-terminal extension when expressed alone (Fig. 3A), co-expression of
full-length DERA and HIS-RPI yielded mostly insoluble protein. With
subsequent HIS-tag affinity purification the two enzymes co-eluted (Sup-
plemental Fig. S8). The protein complex became soluble when we co-
expressed Chloroflexi HIS-tagged RPI and Strep-tagged ΔN-term DERA in
E. coli and pulled down on DERA using a StrepII-tag affinity chromato-
graphy column. The elution fractions from the StrepII column were col-
lected and loaded onto a size exclusion column. SEC shows one major peak
around 131 kDa (Fig. 5A). SDS-PAGE analysis and Western blots of the
peak fractions confirmed the presence of both HIS-RPI and StrepII-ΔN-
term DERA at about equal ratios, additionally confirmed by mass spec-
trometry analysis (Supplemental Table S3). The elution volume corre-
sponding to about 131 kDa indicates that the complex likely contains at least
two subunits of each protein (Fig. 5A and Supplemental Table S3). We were
not able to find evidence of such a complex in literature nor public databases,
so this is likely the first experimentally characterized interaction between
DERA and RPI.

We then repeated the assays for aldolase and isomerase activity using
the size exclusion-separated DERA-RPI complex. The complex demon-
strated functional aldolase activity with a range of concentrations of deoxy-
ribose 5-phosphate substrate (Fig. 6). The enzyme complex V_{max} was
0.072 ± 0.018 at 5 nM concentration and K_m was 0.085 ± 0.029 for deox-
yribose 5-phosphate. Both the V_{max} and K_m values for the DERA alone
compared to the DERA-RPI complex are not statistically significant (V_{max}

Fig. 4 | SPU ΔN-term DERA is functional for the substrate deoxyribose 5-phosphate. NADH-coupled enzyme activity assay with varying concentrations of deoxyribose 5-phosphate (0.25 mM, 0.5 mM, 1 mM, 3 mM, 5 mM) were measured at 344 nm for oxidation of NADH upon addition of Chloroflexi ΔN-term DERA over 1 min to generate a Lineweaver–Burk plot ($R^2 = 0.89$, $y = 1.0453x + 10.032$) and Michaelis–Menton plot (inset, gray). $V_{max} = 0.100 \pm 0.020$ and $K_m = 0.104 \pm 0.010$. Each data point is representative of three independent biological purifications of the enzyme with two technical replicates for each substrate concentration ($n = 6$). All individual replicates represented in the Michaelis–Menten curve were averaged and plotted to generate a Lineweaver–Burk plot. Error bars represent the standard error of the mean (SEM). Means are represented by horizontal lines on the Michaelis–Menten plot and open circles on the Lineweaver–Burk plot.
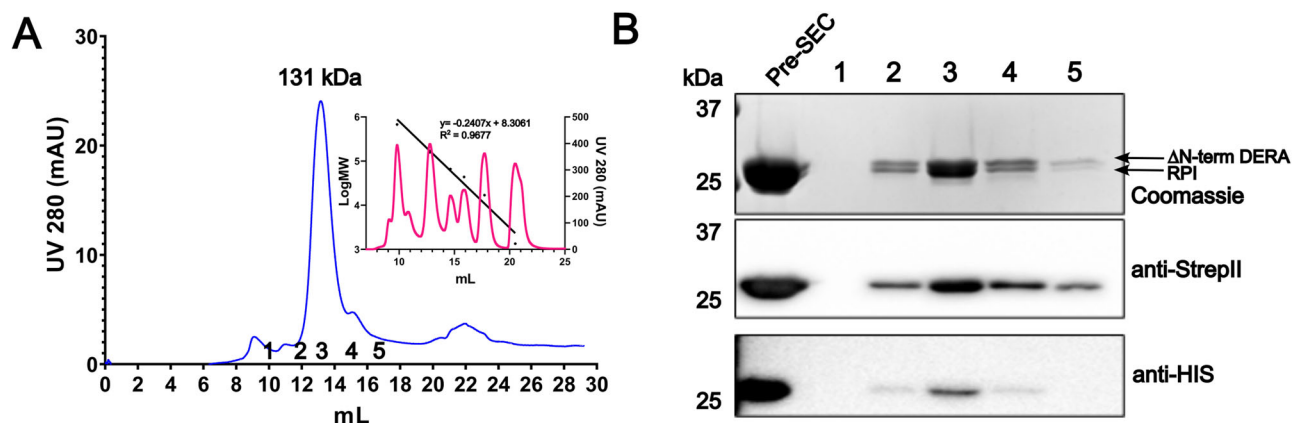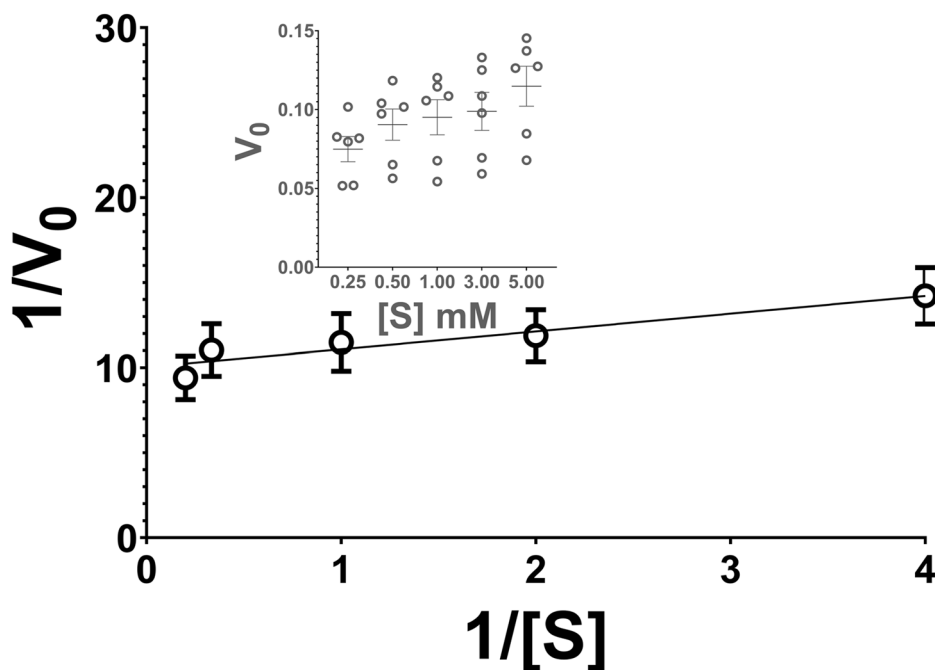
Fig. 5 | SPU signature enzyme DERA from Chloroflexi forms a complex with RPI when expressed in *E. coli*. A S200 size exclusion chromatography of StrepII-tag-purified ΔN-term DERA. The 131 kDa peak elution (fraction 3) contained both the ΔN-term-DERA (26.9 kDa) and HIS-RPI (23.8 kDa) enzymes when assayed by (**B**)

SDS-PAGE of fractions 1–5 followed by Coomassie stain and western blots for StrepII and HIS-tag antibodies. Inset contains gel filtration standards with additional BSA protein to determine oligomeric size.

$P = 0.17$; K_m $P = 0.51$; alpha set to 0.05). The SPU ΔN-term DERA-RPI complex had weak and inconsistent enzyme activity in the isomerization of ribose 5-phosphate (Supplemental Fig. S7).

Discussion

While the SPU BMC was bioinformatically identified previously[2,14], this metabolosome has received little attention despite its prevalence across bacterial phyla. Given the availability of DNA degradation products in diverse habitats ranging from soils to hot springs, we hypothesize the SPU BMC plays a major role in recycling organic detritus to derive metabolic energy. Interestingly, unlike many other BMCs, the SPU BMC tends to not co-occur with other types of BMCs, with the exception of the SPU4 subtype co-occurring with BUF3, EUT2C, EUT2E, EUT2x, and GRM1A[2]. In general, BMCs play a role in providing metabolic flexibility to organisms by encapsulating a given enzymatic reaction, whether it be through enhanced carbon fixation or targeted degradation of a substrate to yield fixed carbon or energy[1]. The SPU BMC exemplifies this role given its ability to utilize a sugar-phosphate substrate to derive energy in the form of glyceraldehyde

3-phosphate (GAP) (Fig. 1). GAP can then be further processed by metabolic pathways such as the Embden-Meyerhof-Parnas, Entner-Doudoroff, and pentose phosphate pathways to generate key energy sources like pyruvate (and thus acetyl-CoA) while forming ATP and NADH. Perhaps the prevalence of the SPU BMC and its tendency to not co-occur with other BMCs reflects its effectiveness in providing energy and carbon compounds from a ubiquitously available substrate[29,30].

Both the SPU BMC and the well-studied carboxysome are hypothesized to share a common requirement for sugar-phosphate transport across the shell. Structural modeling shows that the SPU BMC-H pore can accommodate the entry of the sugar-phosphate metabolite, Dr5P (Supplemental Fig. S1B). Given its smaller size, we presume that this same pore would allow exit of the small sugar-phosphate GAP. Conversely, the potential exit metabolite fructose 1,6 bisphosphate is unlikely to pass through the hexamer pore, as it is significantly larger than Dr5P (Supplemental Fig. S1B). As suggested recently by all-atom molecular simulations for synthetic beta-carboxysome shells showing that the permeation of ribulose 1,5-bisphosphate (RuBP) is minimal in a synthetic hexamer-

Fig. 6 | SPU DERA-RPI enzyme complex has functional activity for deoxyribose 5-phosphate substrate. NADH-coupled enzyme activity assay with varying concentrations of deoxyribose 5-phosphate (0.25 mM, 0.5 mM, 1 mM, 3 mM, 5 mM) were measured at 344 nm for oxidation of NADH upon addition of Chloroflexi ΔN-term DERA-RPI complex over 1 min to generate a Lineweaver–Burk plot (R^2 = 0.96, y = 1.185x + 13.84 and Michaelis–Menton curve (inset, gray). V_{max} = 0.072 ± 0.018 and K_m = 0.085 ± 0.029. Each data point is representative of three independent biological purifications of the enzyme with two technical replicates for each substrate concentration (n = 6). All individual replicates represented in the Michaelis–Menton curve were averaged and plotted to generate a Lineweaver–Burk plot. Error bars represent the standard error of the mean (SEM). Means are represented by horizontal lines on the Michaelis–Menton plot and open circles on Lineweaver–Burk plot.
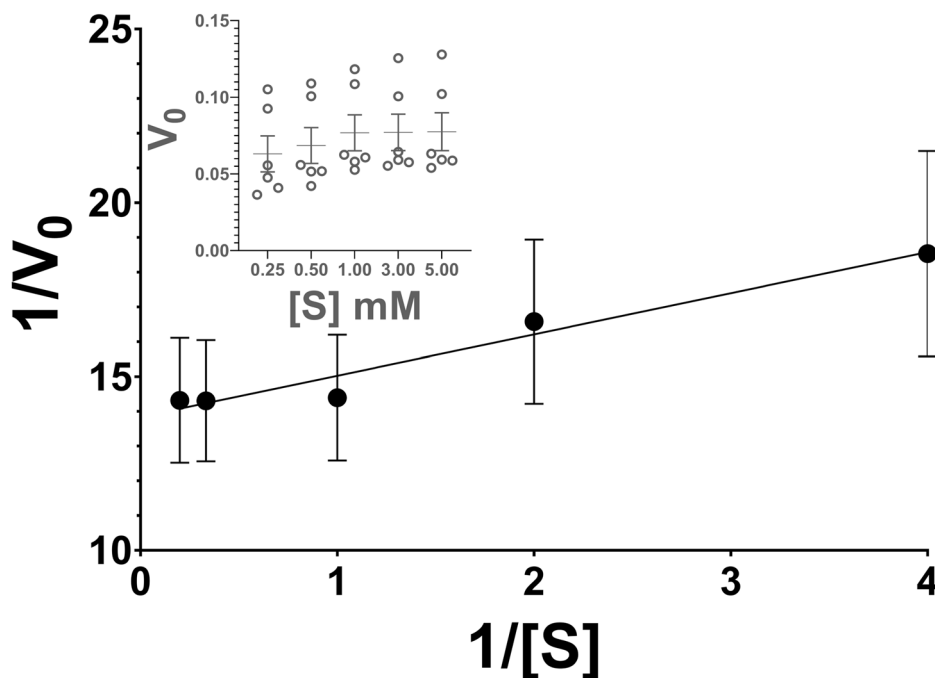
pentamer shell[40], the passage of larger phosphorylated sugars likely does not involve the pores of shell hexamers, but rather the gated BMC-T^{dp} proteins CcmP and CsoS1D. SPU BMC-T^{dp} proteins from the Chloroflexi bacteria are phylogenetically adjacent to CcmP and while Proteobacterial SPU T^{dp} clades are the closest relatives to CsoS1D[33], suggesting conserved modes of phosphorylated sugar transport (i.e., RuBP entering and 3-phosphoglyceric acid exiting the carboxysome shell). BMC-T^{dp} pseudohexamers dimerize across their concave surfaces creating a central nanocompartment that is hypothesized to be gated and function like an airlock that opens and closes based on environmental cues such as the presence of a substrate or product[7,9,41,42]. The pores of these trimers are ~14 Å in the open conformation, large enough for the passage of larger molecules like RuBP; the entry and exit of these substrates from the nanocompartment is likely regulated by the conformation of absolutely conserved residues at the pore which can also fully occlude the pore[7,41]. This feature shared between carboxysomes and SPU BMCs reinforces the hypothesis that sugar phosphates are involved in SPU metabolism, and is consistent with the functions of the encapsulated enzymes encoded across SPU BMC subtypes (Fig. 1A and Supplemental Table S1). SPU BMC subtypes that have the BMC-T^{dp} protein also contain a class II aldolase, which could hypothetically generate a larger sugar phosphate, fructose 1,6 bisphosphate, when combined with DERA and TPI activity (Fig. 1 and Supplemental Table S1); therefore, also requiring a larger pore for passage of this metabolite.

Across all SPU BMC subtypes, the common denominators are an aldehyde dehydrogenase and a deoxyribose 5-phosphate aldolase (DERA). Given the hallmark activity of a metabolosome signature enzyme as generating an aldehyde intermediate, here we define the core SPU BMC metabolism to involve these two enzymes. More specifically, SPU BMCs uniquely contain DERA, which is not found in any other bioinformatically-defined BMC enzymatic cores. Additionally, we observed multiple other enzymes in SPU BMC loci across other subtypes and were able to put them into the context of a generic SPU reaction scheme, but the diverse metabolic pathways of different SPU BMC subtypes will require future study to confirm their specific metabolic capabilities (Fig. 1 and Supplemental Table S1). We observe a triosephosphate isomerase (TPI) (pfam00121) across all SPU BMC loci subtypes (Fig. 1A and Supplemental Table S1); TPI can isomerize GAP and dihydroxyacetone phosphate (DHAP). Since GAP is the key hypothetical product of the SPU BMC, TPI could regulate which

molecule is needed for downstream metabolism. Furthermore, we found AlcDH, PTAC, a PduS homolog, and a class II aldolase in some SPU BMC subtypes, all of which fit linearly into the SPU BMC catabolic reaction sequence whether the final BMC exit product is GAP, or a further downstream sugar phosphate like fructose 1,6 bisphosphate (Fig. 1A, B and Supplemental Table S1).

Notably, most SPU BMC subtypes also have a ribose 5-phosphate isomerase (RPI) (Fig. 1A and Supplemental Table S1). RPI is a critical enzyme in the pentose phosphate pathway that interconverts D-ribose 5-phosphate (R5P) to D-ribulose 5-phosphate (Ru5P), reversibly[38,43,44]. Given its co-occurrence with DERA in many SPU loci (Fig. 1A and Supplemental Table S1) and the existence of a chimera protein (Supplemental Table S1 and Supplemental Fig. S4), we hypothesized that the two enzymes work together in the SPU BMC. Indeed, the two enzymes formed a complex in a likely one-to-one ratio when heterologously expressed in *E. coli* as determined by SDS-PAGE and semi-quantitative normalized values from mass spectrometry (Fig. 5 and Supplemental Table S3). However, the RPI in this complex had little to no isomerase activity for R5P (Supplemental Fig. S7), possibly due to the lack of conservation of the active site cysteine. There are other variants, such as in the organism *Mycobacterium tuberculosis* that have a substitution for the catalytic base in the RpiB gene product, but the residue (Glu, in this case) was still functional in isomerizing R5P[45]. While the LacA subunit of galactose 6-phosphate isomerase also has an aspartic acid substitution in the catalytic base region, it's the LacB subunit that provides the catalytic cysteine in the active site as a part of the heterodimer formation[46]. If the SPU RPI is active for R5P under different conditions, it still would not explain the lack of a direct enzymatic pathway connection from or to DERA within the BMC. To convert between R5P and GAP or vice versa, there would need to be a transketolase within the BMC[47]. Given that most RPIs have broad substrate specificities[44], it remains possible that SPU BMCs process other sugar phosphates that are rare and difficult to assay for, or perhaps there is a greater specificity for the non-phosphorylated sugar form of R5P, but there does not seem to be an obvious role for RPI enzymatic activity in SPU BMC chemistry.

Despite its enigmatic catalytic function in SPU BMC metabolism, the SPU RPI contains a typical encapsulation peptide indicating it is very likely part of the enzymatic core. Perhaps over time it has lost its native function but remained a vital component for properly encapsulating the core

proteins. Non-functional enzymes with EPs have been bioinformatically observed previously in some GRM BMCs[48] where one aldehyde dehydrogenase copy has an EP but lacks a critical active site residue[49]. Likewise in beta-carboxysomes, CcmM plays a critical role in core packaging and assembly[50] but has in many organisms lost its carbonic anhydrase function[51]. In the chimeric SPU3 DERA-RPI protein, we do not find the N-terminal EP on RPI but instead the intraenzyme helix derived from DERA may play the role of an EP. When co-expressed, SPU DERA and RPI form a stable complex (Fig. 5), and we are not aware of reports of this complex formation by non-BMC-associated DERA and RPI. Further, the complex formation did not enhance substrate specificity or enzyme efficiency in this system suggesting alternative roles for the complex beyond the BMC-encapsulated DERA. Altogether, the available evidence supports a role for the Chloroflexi RPI facilitating core enzyme packaging within the SPU BMC while other SPU BMC subtypes might rely on encapsulation by means of the DERA or its chimeric counterpart.

The prevalence of the SPU BMC in nature implies both its value as a metabolic module for deriving energy and its importance for microbial ecosystems. The capacity to take up ubiquitous DNA degradation products and recycle them into usable carbon provides growth advantages to the organism. Likewise, SPU BMCs also have implications for the bioengineering of industrially relevant organisms. The BMC shell system acts as a scaffold for encapsulated enzymatic reactions. To date, synthetic shell systems have been studied for their potential to aid in the role of enhanced carbon fixation[52], hydrogen production[53,54], pyruvate production[55], ethanol production[56], aromatic compound sequestration[57], phytonanotechnology[58,59], and rapid production of shells or shell proteins for multifunctional purposes and scaffolds[60–71]. In addition to synthetic shell systems, pathogenic bacteria harboring BMCs are being investigated to understand how to combat human diseases by targeting the BMCs of pathogens, or repurposing shells as therapeutic devices.[72–74] The SPU BMC can now be added to this growing list of industrially relevant BMCs, for example to produce GAP or simply for utilizing the well-studied aldolase mechanism that is encapsulated for enhanced activity on immobilized surfaces. In the SPU BMC, DERA presumably operates catabolically due to the nature of a metabolosome (Fig. 1B), however, DERA is a reversible enzyme capable of catalyzing the formation of organic compounds by means of many different aldehyde or ketone substrates[75,76]. Moreover, DERA is a demonstratedly versatile enzyme already used for many synthetic applications such as facilitating the creation of anti-tumor agent epothilone A[77], the role in statin production for cholesterol-lowering drugs[78], biosynthesis of (R)-1,3-butanediol[79], production of deoxy-ketoses, deoxy-sialic acid, and of course in the production of deoxysugars in general[75,80]. Characterization of DERAs from hyperthermophilic organisms also provides new options for biocatalysts under different temperature reaction conditions[81].

In summary, we described seven variants of the SPU BMC and characterized its signature enzyme DERA and the accessory protein RPI to shed light on the SPU core reaction mechanism. The catabolic function of a SPU BMC is processing deoxyribose 5-phosphate substrate via SPU DERA or the protein complex DERA-RPI, a novel mechanism across characterized BMC types. Future work is required to understand all of the accessory components of the SPU BMC and its role in nature across different bacterial species. For example, while the focus of this study was on the characterization of the SPU DERA and RPI core enzymes, there are additional proteins that could facilitate the packaging or organization of a SPU BMC given the predicted EPs on several other enzymes and the potential interaction partners that could be involved in SPU BMC biology. The SPU BMC subtypes vary in protein composition, eliciting the question of how these subtypes process the initial substrate and how the subtypes reflect adaptation to specific environments. The fundamental understanding of this widespread BMC provides insight into how a ubiquitous component of detritus is recycled as a carbon and energy source. Moreover, features of the SPU BMC poise it for application in industry with the goal of enhancing organic compound synthesis via BMC shell scaffolding.

Methods

Cloning
Both deoxyribose 5-phosphate aldolase (Integrated Microbial Genomes & Microbiomes (IMG) gene ID 2721888996) and ribose 5-phosphate isomerase B (IMG gene ID 2721888997) genes from *Chloroflexi bacterium GWB2_54_36* (UniProt ID A2X24_locus_1) were synthesized (not codon-optimized) and cloned into the *E. coli* expression vector pET-28a(+) by Twist Bioscience (www.twistbioscience.com). For co-expression analyses, genes were subcloned into a pCDFDUET vector for dual expression under the control of a T7 promoter. Constructs utilized for this study are shown in Supplemental Table S4.

Protein expression
E. coli BL21(DE3) strains harboring plasmids for either Chloroflexi HIS-DERA or HIS-RPI were grown in lysogeny broth to an OD600 of ~0.6 at 37 °C then induced with 250 μM of Isopropyl ß-D-1-thiogalactopyranoside and incubated at 18 °C for overnight growth. Cultures were then centrifuged at 8000×*g* for 10 min, and pellets were collected for subsequent protein purifications.

Protein purification
Bacterial pellets were resuspended in 1:1 lysis buffer per gram of pellet weight. Lysis buffer consisted of 50 mM Tris-HCl pH 7.5, 500 mM NaCl, 5% glycerol, DNase I (0.05 mg/mL; GoldBio; D-300-100), and 1× SigmaFast protease inhibitor cocktail (Sigma; S8820). After incubation in lysis buffer for 20 min, cells were passed twice through a French Press at 1100 psi. The lysate was clarified by centrifugation for 30 min at 45,000×*g* and the supernatant was filtered and imidazole was added to 10 mM final concentration before applying it to a HisTrap 5 mL column (Cytiva). Protein was eluted off the column with an imidazole concentration gradient from 50 mM to 500 mM. For purification using the StrepII tag, the soluble fraction was loaded onto a StrepTrap HP 5 mL column (Cytiva). StrepII-tag proteins were eluted using 2.5 mM D-Desthiobiotin (Sigma; D1411). Elution fractions were assayed by SDS-PAGE to determine which fractions contained the target protein. Fractions with the desired protein were pooled and concentrated using the appropriate MWCO Amicon depending on the size of the protein. RPI containing samples did not need to be concentrated due to their high concentration from the elution fractions. For assaying the oligomeric state, protein samples were filtered and loaded onto a Superdex 200 analytical size exclusion column (Cytiva) equilibrated in 50 mM Tris-HCl, 500 mM NaCl, 5% glycerol buffer followed by loading Bio-Rad gel filtration standards (Bio-Rad; Catalog #1511901) (including BSA) with the same buffer.

SDS-PAGE and western blot
Protein samples were boiled in sodium dodecyl sulfate (SDS)-loading buffer for 10 min before loading onto 18% polyacrylamide gels. Gels were run at 200 V for one hour in a Bio-Rad Mini-PROTEAN Tetra Vertical Electrophoresis Cell and then incubated with Brilliant Blue Coomassie stain for 30 min. Background stain was removed with destaining solution (10% acetic acid, 30% ethanol, 10% water) for several replicates until protein bands resolved for imaging with a Bio-Rad GelDoc Go System. For Western blot analysis, gels were transferred to nitrocellulose membrane using transfer buffer (1× Bio-Rad Tris/Glycine Native Gel Buffer, 10% ethanol, water) for one hour at 300 mA. The membrane was then stained with Ponceau-S for analysis of successful protein transfer. After staining, the membrane was blocked for one hour in 5% skim milk dissolved in Tris-Buffered Saline 0.1% Tween 20 (TBST). Primary anti-rabbit HIS antibody (OriGene; TA150087) was applied at 1:2000 dilution and incubated overnight at 4 °C with horizontal platform shaking (62 RPM) followed by secondary goat anti-rabbit-HRP antibody (Jackson ImmunoResearch; 111-035-003) incubation with three 5-min Tris-Buffered Saline 0.1% Tween 20 (TBST) washes in-between and after antibody applications. For StrepII detection, anti-rabbit-StrepII-HRP (Novagen; 71591-3) was applied at 1:4000 dilution. Blots were incubated with Amersham ECL Prime Western Blotting Detection Reagent

(RPN2232) for 5 min and imaged using a Bio-Rad ChemiDoc XRS+ Imaging System to detect chemiluminescence of the blotted proteins.

Enzyme assays

To assay aldolase activity, a NADH-coupled reaction was utilized, modeled after Valentin-Hansen et al.[39]. Solutions with deoxyribose 5-phosphate (Sigma; D3126; varied concentrations – 0.25 mM, 0.5 mM, 1 mM, 3 mM, 5 mM), 0.28 mM NADH (in excess) (GoldBio; N-035-1), 0.05 mg/mL yeast alcohol dehydrogenase (in excess) (Sigma; A3263), 0.05 M Tris pH 8.0, 0.1 mM EDTA pH 8.0, and water up to 440 μL total volume were measured at 344 nm for 3 min at 37 °C to determine a blank without enzyme. DERA enzyme at 5 nM (or DERA-RPI complex) was spiked in and the absorbance at 344 nm was tracked for 3 min at 37 °C. All measurements were conducted on a Cary 60 UV–Vis spectrophotometer (Agilent) with a 50 μL spectrophotometer cuvette (Starna Cells, 16.50-Q-10/Z15). To calculate enzyme kinetics and generate Michaelis–Menton and Lineweaver–Burk plots, the first minute of linear data from the spectrophotometer was used for three independent biological purification replicates of each enzyme/enzyme complex with two technical replicates per substrate concentration. The final plots are both the individual data points ($n = 6$) to represent the Michaelis–Menton plots which then are used for averaging and generation of the Lineweaver–Burk plots. The standard error of the mean (SEM) is used to display the variance between enzyme kinetic assays. Parameters were calculated using GraphPad Prism software (GraphPad, San Diego, CA) to fit non-linear curves for the Michaelis–Menton plots using the equation $v = V_{max}[S]/(K_m + [S])$ while Lineweaver–Burk plots were generated in Microsoft Excel plotting $1/V$ over $1/[S]$ to fit a linear curve and yield an equation for V_{max} and K_m calculations.

To assay for isomerase activity, the production of ribulose 5-phosphate was monitored at an absorbance of 290 nm with the starting substrate of ribose 5-phosphate (Sigma; 83875) modeled after Wood (1970)[82] and Zhang (2003)[38]. After measuring with no enzyme to establish a blank using 5 mM R5P substrate, 30 nM RPI (single or complex) was spiked in to measure any changes in absorbance for 15 min at 37 °C. All measurements were conducted on a Cary 60 UV–Vis spectrophotometer (Agilent), similar to the aldolase assay.

Statistics and reproducibility

A t test: paired two sample for means data analysis was performed in Microsoft Excel to determine statistical differences between V_{max} and K_m values between the single DERA enzyme and the DERA-RPI enzyme complex for $n = 3$ of each with an alpha value set to 0.05. As stated in the enzyme methods, three independent biological purification replicates of each enzyme/enzyme complex with two technical replicates per substrate concentration was used to calculate the final plot averages using standard error of the mean to display the variance between enzyme assays.

Mass spectrometry

Proteolytic digestion. Gel bands were digested in-gel according to Shevchenko, et. al. with modifications[83]. Briefly, gel bands were dehydrated using 100% acetonitrile and incubated with 10 mM dithiothreitol in 100 mM ammonium bicarbonate, pH ~8, at 56 °C for 45 min, dehydrated again and incubated in the dark with 50 mM chloroacetamide in 100 mM ammonium bicarbonate for 20 min. Gel bands were then washed with ammonium bicarbonate and dehydrated again. Sequencing grade modified typsin was prepared to 0.005 μg/μL in 50 mM ammonium bicarbonate and ~100 μL of this was added to each gel band so that the gel was completely submerged. Bands were then incubated at 37 °C overnight. Peptides were extracted from the gel by water bath sonication in a solution of 60% Acetonitrile (ACN) /1% Trifluoroacetic acid (TFA) and vacuum dried to ~2 μL. Samples were resuspended in 2% ACN/0.1% TFA to 20 μL and frozen.

LC/MS/MS analysis. An injection of 5 μL was automatically made using a Thermo (www.thermo.com) EASYnLC 1000 onto a Thermo Acclaim

PepMap RSLC 0.1 mm × 20 mm C18 trapping column and washed for ~5 min with buffer A. Bound peptides were then eluted over 35 min onto a Thermo Acclaim PepMap RSLC 0.075 mm × 150 mm resolving column with a linear gradient of 5%B to 19%B from 0 min to 19 min, 19%B to 40% B from 19 min to 24 min and 40%B to 90%B from 24 min to 25 min (Buffer A = 99.9% Water/0.1% Formic Acid, Buffer B = 80% Acetonitrile/0.1% Formic Acid/19.9% Water) at a constant flow rate of 300 nl/min. After the gradient, the column was washed with 90%B for the duration of the run. Eluted peptides were sprayed into a ThermoScientific Q-Exactive mass spectrometer (www.thermo.com) using a FlexSpray spray ion source. Survey scans were taken in the Orbi trap (35,000 resolution, determined at m/z 200) and the top 10 ions in each survey scan are then subjected to automatic higher energy collision-induced dissociation (HCD) with fragment spectra acquired at a resolution of 17,500.

Data analysis. The resulting MS/MS spectra are converted to peak lists Mascot Distiller, v2.8.5 (www.matrixscience.com) and searched against a database containing target protein sequences and all *E. coli* protein sequences available from Uniprot (www.uniprot.org, downloaded 2024-05-21) appended with common laboratory contaminants (downloaded from www.thegpm.org, cRAP project) using the Mascot searching algorithm, v2.8.3[84]. The Mascot output was then analyzed using Scaffold, v5.3.3 (www.proteomesoftware.com) to probabilistically validate protein identifications. Assignments validated using the Scaffold 1% FDR confidence filter are considered true.

Mascot parameters for all databases were as follows:
- allow up to 2 missed tryptic sites
- Fixed modification of Carbamidomethyl Cysteine,
- variable modification of Oxidation of Methionine,
- peptide tolerance of +/- 10ppm
- MS/MS tolerance of 0.02Da

FDR was calculated using randomized database search.

Bioinformatics

Locus data was obtained from Sutter et al.[2], as visualized in BMC Caller[27]. SPU loci were curated by selecting loci that contained at least one AldDh and one of either DERA or RPI. Loci that originated from datasets that have been retired in UniProt were also removed, as were loci that contained truncated protein sequences or loci that appeared fragmented. Sequence conservation logos were generated by aligning the sequences with clustalw 2.1[85], trimming with trimAl 1.2rev59[86] with parameters -gt 0.6 -cons 30 -w 3 and visualizing with WebLogo[87]. Pairwise sequence alignments to compare enzyme active sites were made with Clustal Omega through MEGA11 software[88] and visualized with ESPript 3.0[89]. For structural analyses, we used AlphaFold2[90] to generate protein structures and ChimeraX MatchMaker[91] and PyMol (The PyMOL Molecular Graphics System, Version 3.0 Schrödinger, LLC.) for protein structural visualization.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data supporting the findings of this study are available within the paper and its Supplementary Information. Methods detail which public databases were used for bioinformatic analyses. Our curated list of SPU BMC subtypes and their associated organisms/UniProt IDs are available in Supplemental Dataset S1. Uncropped gels or blots are shown in Supplemental Fig. S9 for Fig. 3A and Supplemental Fig. S10 for Fig. 5B. Mass spectrometry raw data is shared on FigShare: https://doi.org/10.6084/m9.figshare.27629535. Source values ($n = 6$) for the enzyme kinetic assays are available in Supplemental Dataset S2 as described in the Methods section and the figure legends for Figs. 4 and 6. All data are available from the corresponding author (or other sources, as applicable) on reasonable request.

References

1. Kerfeld, C. A., Aussignargues, C., Zarzycki, J., Cai, F. & Sutter, M. Bacterial microcompartments. *Nat. Rev. Microbiol.* **16**, 277–290 (2018).
2. Sutter, M., Melnicki, M. R., Schulz, F., Woyke, T. & Kerfeld, C. A. A catalog of the diversity and ubiquity of bacterial microcompartments. *Nat. Commun.* **12**, 3809 (2021).
3. Chowdhury, C., Sinha, S., Chun, S., Yeates, T. O. & Bobik, T. A. Diverse bacterial microcompartment organelles. *Microbiol. Mol. Biol. Rev.* **78**, 438–468 (2014).
4. Rae, B. D., Long, B. M., Badger, M. R. & Price, G. D. Functions, compositions, and evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate CO_2 fixation in cyanobacteria and some proteobacteria. *Microbiol. Mol. Biol. Rev.* **77**, 357–379 (2013).
5. Jakobson, C. M., Tullman-Ercek, D., Slininger, M. F. & Mangan, N. M. A systems-level model reveals that 1,2-Propanediol utilization microcompartments enhance pathway flux through intermediate sequestration. *PLoS Comput. Biol.* **13**, e1005525 (2017).
6. Kerfeld, C. A. et al. Protein structures forming the shell of primitive bacterial organelles. *Science* **309**, 936–938 (2005).
7. Klein, M. G. et al. Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport. *J. Mol. Biol.* **392**, 319–333 (2009).
8. Sutter, M., Greber, B., Aussignargues, C. & Kerfeld, C. A. Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell. *Science* **356**, 1293–1297 (2017).
9. Greber, B. J., Sutter, M. & Kerfeld, C. A. The plasticity of molecular interactions governs bacterial microcompartment shell assembly. *Structure* **27**, 749–763.e4 (2019).
10. Tanaka, S. et al. Atomic-level models of the bacterial carboxysome shell. *Science* **319**, 1083–1086 (2008).
11. Cannon, G. C. & Shively, J. M. Characterization of a homogenous preparation of carboxysomes from *Thiobacillus neapolitanus*. *Arch. Microbiol.* **134**, 52–59 (1983).
12. Shively, J. M., Ball, F., Brown, D. H. & Saunders, R. E. Functional organelles in prokaryotes: polyhedral inclusions (Carboxysomes) of *Thiobacillus neapolitanus*. *Science* **182**, 584–586 (1973).
13. Jorda, J., Lopez, D., Wheatley, N. M. & Yeates, T. O. Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. *Protein Sci.* **22**, 179–195 (2013).
14. Axen, S. D., Erbilgin, O. & Kerfeld, C. A. A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. *PLoS Comput. Biol.* **10**, e1003898 (2014).
15. Kerfeld, C. A. & Erbilgin, O. Bacterial microcompartments and the modular construction of microbial metabolism. *Trends Microbiol.* **23**, 22–34 (2015).
16. Erbilgin, O., Sutter, M. & Kerfeld, C. A. The structural basis of coenzyme a recycling in a bacterial organelle. *PLoS Biol.* **14**, e1002399 (2016).
17. Penrod, J. T. & Roth, J. R. Conserving a volatile metabolite: a role for carboxysome-like organelles in *Salmonella enterica*. *J. Bacteriol.* **188**, 2865–2874 (2006).
18. Havemann, G. D., Sampson, E. M. & Bobik, T. A. PduA is a shell protein of polyhedral organelles involved in coenzyme B_{12}-dependent degradation of 1,2-propanediol in *Salmonella enterica* Serovar Typhimurium LT2. *J. Bacteriol.* **184**, 1253–1261 (2002).
19. Rondon, M. R., Horswill, A. R. & Escalante-Semerena, J. C. DNA polymerase I function is required for the utilization of ethanolamine, 1,2-propanediol, and propionate by *Salmonella typhimurium* LT2. *J. Bacteriol.* **177**, 7119–7124 (1995).
20. Sampson, E. M. & Bobik, T. A. Microcompartments for B_{12}-dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate. *J. Bacteriol.* **190**, 2966–2971 (2008).
21. Fan, C. et al. Short N-terminal sequences package proteins into bacterial microcompartments. *Proc. Natl. Acad. Sci. USA* **107**, 7509–7514 (2010).
22. Fan, C. & Bobik, T. A. The N-terminal region of the medium subunit (PduD) packages adenosylcobalamin-dependent diol dehydratase (PduCDE) into the Pdu microcompartment. *J. Bacteriol.* **193**, 5623–5628 (2011).
23. Kinney, J. N., Salmeen, A., Cai, F. & Kerfeld, C. A. Elucidating essential role of conserved carboxysomal protein CcmN reveals common feature of bacterial microcompartment assembly. *J. Biol. Chem.* **287**, 17729–17736 (2012).
24. Aussignargues, C., Paasch, B. C., Gonzalez-Esquer, R., Erbilgin, O. & Kerfeld, C. A. Bacterial microcompartment assembly: the key role of encapsulation peptides. *Commun. Integr. Biol.* **8**, e1039755 (2015).
25. Liu, Y. et al. PduL is an evolutionarily distinct phosphotransacylase involved in B_{12}-dependent 1,2-propanediol degradation by *Salmonella enterica* Serovar Typhimurium LT2. *J. Bacteriol.* **189**, 1589–1596 (2007).
26. Gu, S., Bradley-Clarke, J., Rose, R.-S., Warren, M. J. & Pickersgill, R. W. Enzyme-cargo encapsulation peptides bind between tessellating tiles of the bacterial microcompartment shell. *J. Biol. Chem.* **300**, 107357 (2024).
27. Sutter, M. & Kerfeld, C. A. BMC Caller: a webtool to identify and analyze bacterial microcompartment types in sequence data. *Biol. Direct* **17**, 9 (2022).
28. Chen, I.-M. A. et al. The IMG/M data management and analysis system v.7: content updates and new features. *Nucleic Acids Res.* **51**, D723–D732 (2023).
29. Nielsen, K. M., Johnsen, P. J., Bensasson, D. & Daffonchio, D. Release and persistence of extracellular DNA in the environment. *Environ. Biosaf. Res.* **6**, 37–53 (2007).
30. Levy-Booth, D. J. et al. Cycling of extracellular DNA in the soil environment. *Soil Biol. Biochem.* **39**, 2977–2991 (2007).
31. Lomax, M. S. & Greenberg, G. R. Characteristics of the *deo* operon: role in thymine utilization and sensitivity to deoxyribonucleosides. *J. Bacteriol.* **96**, 501–514 (1968).
32. Jørgensen, P., Collins, J. & Valentin-Hansen, P. On the structure of the deo operon of *Escherichia coli*. *Mol. Gen. Genet.* **155**, 93–102 (1977).
33. Melnicki, M. R., Sutter, M. & Kerfeld, C. A. Evolutionary relationships among shell proteins of carboxysomes and metabolosomes. *Curr. Opin. Microbiol.* **63**, 1–9 (2021).
34. Sutter, M. et al. Electrochemical cofactor recycling of bacterial microcompartments. *Proc. Natl. Acad. Sci. USA* https://doi.org/10.1073/pnas.2414220121 (2024).
35. Fillenberg, S. B., Grau, F. C., Seidel, G. & Muller, Y. A. Structural insight into operator dre-sites recognition and effector binding in the GntR/HutC transcription regulator NagR. *Nucleic Acids Res.* **43**, 1283–1296 (2015).
36. Matilla, M. A., Velando, F., Martín-Mora, D., Monteagudo-Cascales, E. & Krell, T. A catalogue of signal molecules that interact with sensor kinases, chemoreceptors and transcriptional regulators. *FEMS Microbiol. Rev.* **46**, fuab043 (2022).
37. Heine, A., Luz, J. G., Wong, C.-H. & Wilson, I. A. Analysis of the class I aldolase binding site architecture based on the crystal structure of 2-deoxyribose-5-phosphate aldolase at 0.99 Å resolution. *J. Mol. Biol.* **343**, 1019–1034 (2004).
38. Zhang, R.-G. et al. The 2.2Å resolution structure of RpiB/AlsB from *Escherichia coli* illustrates a new approach to the ribose-5-phosphate isomerase reaction. *J. Mol. Biol.* **332**, 1083–1094 (2003).
39. Valentin-Hansen, P., Boëtius, F., Hammer-Jespersen, K. & Svendsen, I. The primary structure of *Escherichia coli* K12 2-deoxyribose 5-phosphate aldolase: nucleotide sequence of the deoC gene and the

amino acid sequence of the enzyme. *Eur. J. Biochem.* **125**, 561–566 (1982).

40. Sarkar, D. et al. Atomic view of photosynthetic metabolite permeability pathways and confinement in synthetic carboxysome shells. *Proc. Natl. Acad. Sci. USA* **121**, e2402277121 (2024).

41. Cai, F. et al. The structure of CcmP, a tandem bacterial microcompartment domain protein from the β-carboxysome, forms a subcompartment within a microcompartment. *J. Biol. Chem.* **288**, 16055–16063 (2013).

42. Larsson, A. M., Hasse, D., Valegård, K. & Andersson, I. Crystal structures of β-carboxysome shell protein CcmP: ligand binding correlates with the closed or open central pore. *J. Exp. Bot.* **68**, 3857–3867 (2017).

43. Zhang, R. et al. Structure of *Escherichia coli* ribose-5-phosphate isomerase. *Structure* **11**, 31–42 (2003).

44. Chen, J., Wu, H., Zhang, W. & Mu, W. Ribose-5-phosphate isomerases: characteristics, structural features, and applications. *Appl Microbiol. Biotechnol.* **104**, 6429–6441 (2020).

45. Roos, A. K. et al. Mycobacterium tuberculosis Ribose-5-phosphate Isomerase has a known fold, but a novel active site. *J. Mol. Biol.* **335**, 799–809 (2004).

46. Stern, A. L., Naworyta, A., Cazzulo, J. J. & Mowbray, S. L. Structures of type B ribose 5-phosphate isomerase from *Trypanosoma cruzi* shed light on the determinants of sugar specificity in the structural family. *FEBS J.* **278**, 793–808 (2011).

47. Sharkey, T. D. Pentose phosphate pathway reactions in photosynthesizing cells. *Cells* **10**, 1547 (2021).

48. Petit, E. et al. Involvement of a bacterial microcompartment in the metabolism of fucose and rhamnose by *Clostridium phytofermentans*. *PLoS ONE* **8**, e54337 (2013).

49. Zarzycki, J., Erbilgin, O. & Kerfeld, C. A. Bioinformatic characterization of glycyl radical enzyme-associated bacterial microcompartments. *Appl. Environ. Microbiol.* **81**, 8315–8329 (2015).

50. Cameron, J. C., Wilson, S. C., Bernstein, S. L. & Kerfeld, C. A. Biogenesis of a bacterial organelle: the carboxysome assembly pathway. *Cell* **155**, 1131–1140 (2013).

51. Cot, S. S.-W., So, A. K.-C. & Espie, G. S. A multiprotein bicarbonate dehydration complex essential to carboxysome function in cyanobacteria. *J. Bacteriol.* **190**, 936–945 (2008).

52. Borden, J. S. & Savage, D. F. New discoveries expand possibilities for carboxysome engineering. *Curr. Opin. Microbiol.* **61**, 58–66 (2021).

53. Li, T. et al. Reprogramming bacterial protein organelles as a nanoreactor for hydrogen production. *Nat. Commun.* **11**, 5448 (2020).

54. Jiang, Q. et al. Synthetic engineering of a new biocatalyst encapsulating [NiFe]-hydrogenases for enhanced hydrogen production. *J. Mater. Chem. B* **11**, 2684–2692 (2023).

55. Kirst, H. et al. Toward a glycyl radical enzyme containing synthetic bacterial microcompartment to produce pyruvate from formate and acetate. *Proc. Natl. Acad. Sci. USA* **119**, e2116871119 (2022).

56. Lawrence, A. D. et al. Solution structure of a bacterial microcompartment targeting peptide and its application in the construction of an ethanol bioreactor. *ACS Synth. Biol.* **3**, 454–465 (2014).

57. Doron, L., Sutter, M. & Kerfeld, C. A. Characterization of a novel aromatic substrate-processing microcompartment in Actinobacteria. *mBio* **14**, e01216-23 (2023).

58. Raba, D. A. & Kerfeld, C. A. The potential of bacterial microcompartment architectures for phytonanotechnology. *Environ. Microbiol. Rep.* **14**, 700–710 (2022).

59. Dwyer, M. E. et al. Towards chloroplastic nanofactories: formation of proteinaceous scaffolds for metabolic engineering. *Plant Biotechnol. J.* https://doi.org/10.1111/pbi.14462 (2024).

60. Aussignargues, C. et al. Structure and function of a bacterial microcompartment shell protein engineered to bind a [4Fe-4S] cluster. *J. Am. Chem. Soc.* **138**, 5262–5270 (2016).

61. Hagen, A., Sutter, M., Sloan, N. & Kerfeld, C. A. Programmed loading and rapid purification of engineered bacterial microcompartment shells. *Nat. Commun.* **9**, 2881 (2018).

62. Hagen, A. R. et al. In vitro assembly of diverse bacterial microcompartment shell architectures. *Nano Lett.* **18**, 7030–7037 (2018).

63. Plegaria, J. S. & Kerfeld, C. A. Engineering nanoreactors using bacterial microcompartment architectures. *Curr. Opin. Biotechnol.* **51**, 1–7 (2018).

64. Ferlez, B., Sutter, M. & Kerfeld, C. A. A designed bacterial microcompartment shell with tunable composition and precision cargo loading. *Metab. Eng.* **54**, 286–291 (2019).

65. Plegaria, J. S., Yates, M. D., Glaven, S. M. & Kerfeld, C. A. Redox characterization of electrode-immobilized bacterial microcompartment shell proteins engineered to bind metal centers. *ACS Appl. Bio Mater.* **3**, 685–692 (2020).

66. Doron, L. & Kerfeld, C. A. Bacterial microcompartments as a next-generation metabolic engineering tool: utilizing nature's solution for confining challenging catabolic pathways. *Biochem. Soc. Trans.* **52**, 997–1010 (2024).

67. Li, T. et al. Nanoengineering carboxysome shells for protein cages with programmable cargo targeting. *ACS Nano* **18**, 7473–7484 (2024).

68. Lee, M. J., Palmer, D. J. & Warren, M. J. Biotechnological advances in bacterial microcompartment technology. *Trends Biotechnol.* **37**, 325–336 (2019).

69. Stewart, A. M., Stewart, K. L., Yeates, T. O. & Bobik, T. A. Advances in the world of bacterial microcompartments. *Trends Biochem. Sci.* **46**, 406–416 (2021).

70. Frank, S., Lawrence, A. D., Prentice, M. B. & Warren, M. J. Bacterial microcompartments moving into a synthetic biological world. *J. Biotechnol.* **163**, 273–279 (2013).

71. Young, E. J. et al. Engineering the bacterial microcompartment domain for molecular scaffolding applications. *Front. Microbiol.* **8**, 1441 (2017).

72. Planamente, S. & Frank, S. Bio-engineering of bacterial microcompartments: a mini review. *Biochem. Soc. Trans.* **47**, 765–777 (2019).

73. Prentice, M. B. Bacterial microcompartments and their role in pathogenicity. *Curr. Opin. Microbiol.* **63**, 19–28 (2021).

74. Jakobson, C. M. & Tullman-Ercek, D. Dumpster diving in the gut: bacterial microcompartments as part of a host-associated lifestyle. *PLoS Pathog.* **12**, e1005558 (2016).

75. Haridas, M., Abdelraheem, E. M. M. & Hanefeld, U. 2-Deoxy-d-ribose-5-phosphate aldolase (DERA): applications and modifications. *Appl Microbiol. Biotechnol.* **102**, 9959–9971 (2018).

76. Voutilainen, S. et al. Substrate specificity of 2-deoxy-D-ribose 5-phosphate aldolase (DERA) assessed by different protein engineering and machine learning methods. *Appl. Microbiol. Biotechnol.* **104**, 10515–10529 (2020).

77. Liu, J. & Wong, C.-H. Aldolase-catalyzed asymmetric synthesis of novel pyranose synthons as a new entry to heterocycles and epothilones. *Angew. Chem. Int. Ed.* **41**, 1404–1407 (2002).

78. Gijsen, H. J. M. & Wong, C.-H. Unprecedented asymmetric aldol reactions with three aldehyde substrates catalyzed by 2-deoxyribose-5-phosphate aldolase. *J. Am. Chem. Soc.* **116**, 8422–8423 (1994).

79. Kim, T. et al. Rational engineering of 2-deoxyribose-5-phosphate aldolases for the biosynthesis of (R)-1,3-butanediol. *J. Biol. Chem.* **295**, 597–609 (2020).

80. Chambre, D. et al. 2-Deoxyribose-5-phosphate aldolase, a remarkably tolerant aldolase towards nucleophile substrates. *Chem. Commun.* **55**, 7498–7501 (2019).

81. Sakuraba, H. et al. Sequential aldol condensation catalyzed by hyperthermophilic 2-deoxy- D -ribose-5-phosphate aldolase. *Appl. Environ. Microbiol.* **73**, 7427–7434 (2007).

82. Wood, T. Spectrophotometric assay for d-ribose-5-phosphate ketol-isomerase and for d-ribulose-5-phosphate 3-epimerase. *Anal. Biochem.* **33**, 297–306 (1970).

83. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858 (1996).
84. Perkins, D. N., Pappin, D. J. C., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567 (1999).
85. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
86. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
87. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator: figure 1. *Genome Res.* **14**, 1188–1190 (2004).
88. Tamura, K., Stecher, G. & Kumar, S. MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **38**, 3022–3027 (2021).
89. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
90. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
91. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

Acknowledgements

Author contributions

M.E.D. and C.A.K. designed the experiments; M.E.D., M.S., and C.A.K. analyzed the data; M.E.D. wrote the manuscript; M.E.D., M.S., and C.A.K. edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-024-07287-y.

Correspondence and requests for materials should be addressed to Cheryl A. Kerfeld.

Peer review information *Communications Biology* thanks Denis Jallet and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Tobias Goris. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.