

# Biogenesis of a bacterial metabolosome for propanediol utilization

Mengru Yang<sup>1</sup>, Nicolas Wenner<sup>2</sup>, Gregory Dykes<sup>1</sup>, Yan Li<sup>2</sup>, Xiaojun Zhu<sup>2</sup>, Yaqi Sun<sup>1</sup>, Fang Huang<sup>1</sup>, Jay C. D. Hinton<sup>2</sup>, Lu-Ning Liu<sup>1,3,\*</sup>

<sup>1</sup> Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Crown Street, Liverpool L69 7ZB, United Kingdom

<sup>2</sup> Institute of Infection, Veterinary & Ecological Sciences, University of Liverpool, Crown Street, Liverpool L69 7ZB, United Kingdom

<sup>3</sup> College of Marine Life Sciences, and Frontiers Science Center for Deep Ocean Multispheres and Earth System, Ocean University of China, 266003 Qingdao, China

\* Correspondence: [luning.liu@liverpool.ac.uk](mailto:luning.liu@liverpool.ac.uk)

## Abstract

Bacterial metabolosomes are a family of protein organelles in bacteria. Elucidating how thousands of proteins self-assemble to form functional metabolosomes is essential for understanding their significance in cellular metabolism and pathogenesis. Here we investigate the *de novo* biogenesis of propanediol-utilization (Pdu) metabolosomes and characterize the roles of the key constituents in generation and intracellular positioning of functional metabolosomes. Our results demonstrate that the Pdu metabolosome undertakes both “Shell first” and “Cargo first” assembly pathways, unlike the  $\beta$ -carboxysome structural analog which only involves the “Cargo first” strategy. Shell and cargo assemblies occur independently at the cell poles. The internal cargo core is formed through the ordered assembly of multiple enzyme complexes, and exhibits liquid-like properties within the metabolosome architecture. Our findings provide mechanistic insight into the molecular principles driving bacterial metabolosome assembly and expand our understanding of liquid-like organelle biogenesis.

## Keywords

Bacterial microcompartment, metabolosome, propanediol utilization, Salmonella, organelle biogenesis

## Introduction

Compartmentalization of metabolic pathways within eukaryotic and prokaryotic cells both enhances and regulates energy and metabolism<sup>1</sup>. Unlike eukaryotes that employ lipid-bound organelles to perform specific metabolic functions, many bacteria have evolved proteinaceous organelles, termed bacterial microcompartments (BMCs), that allow key metabolic processes to function in the cytoplasm of bacterial cells<sup>2-4</sup>. BMCs consist of a single-layer proteinaceous shell that encapsulates both enzymes and metabolites. The virus-like polyhedral shell serves as a physical barrier to control the passage of metabolites and facilitates catabolism in a sequestered micro-environment<sup>5-8</sup>. The highly evolved structural and assembly features of BMCs enable the special bacterial organelles to play pivotal roles in CO<sub>2</sub> fixation, infection, and microbial ecology across diverse bacterial species<sup>2,9,10</sup>.

The best-characterized BMCs are the carboxysomes (anabolic BMCs, including  $\alpha$ - and  $\beta$ -carboxysomes based on the types of Rubisco encapsulated), which enhance CO<sub>2</sub> fixation in all cyanobacteria and many chemoautotrophs<sup>11-13</sup>. In contrast, the 1,2-propanediol (1,2-PD) utilization microcompartment (Pdu BMC) is a model for catabolic BMCs or metabolosomes. A critical stage of *Salmonella* pathogenesis is the colonization of the mammalian gastrointestinal tract, which requires the catabolism of a number of carbon sources, including 1,2-PD<sup>14</sup>. The Pdu BMC plays a vital role in the breakdown of 1,2-PD during gastrointestinal pathogenesis, and promotes bacterial fitness in specific anoxic environmental niches<sup>15-22</sup>.

The architecture of Pdu BMCs is constructed by more than ten thousand protein subunits of 18 to 20 different types via intrinsic protein self-assembly<sup>23-26</sup>. Based on current knowledge, the Pdu BMC shell in the pathogenic bacterium *Salmonella enterica* serovar Typhimurium (*S. Typhimurium*) contains nine types of protein paralogs: PduA, PduB, PduB', PduM, PduN, PduJ, PduK, PduT, and PduU<sup>27,28</sup>. The hexameric protein PduJ is the most abundant shell protein within the Pdu BMC in *S. Typhimurium* LT2, followed by PduA and PduBB'<sup>24</sup>. PduB and PduB' are encoded by overlapping genes, and PduB has an N-terminal 37 amino acid extension that is absent from PduB'<sup>29</sup>. The pentameric protein PduN occupies the vertices of the polyhedral

shell<sup>30</sup>. PduBB', PduA or PduJ, and PduN are the shell components that are essential for Pdu BMC formation and structure<sup>27,31</sup>. PduM is a structural protein, which is much less abundant than PduN<sup>24</sup> and has an unknown function within the Pdu BMC<sup>28</sup>. PduK is another minor shell protein within the Pdu BMC<sup>24</sup>.

The enzymes involved in 1,2-PD degradation within the Pdu BMC comprise PduCDE (diol dehydratase)<sup>29</sup>, PduL<sup>32</sup>, PduP<sup>33</sup>, PduQ<sup>34</sup>, and PduW<sup>35</sup>. Additional enzymes required for reactivation of diol dehydratase and vitamin B<sub>12</sub> recycling are PduS<sup>36</sup>, PduO<sup>37</sup>, PduGH<sup>15</sup>, and PduX (L-threonine kinase)<sup>38</sup>. PduV was proposed to connect with the filament-associated movement of BMCs within the cell<sup>39</sup>. Previous studies illustrated that the short N-terminal sequences of PduD, PduP, and PduL function as encapsulation peptides and are important for directing enzymes into the Pdu BMC<sup>40-42</sup>. The PduB N-terminus has been reported to bind the shell to cargo enzymes<sup>43</sup>.

For decades, it has remained unclear how thousands of protein components self-recognize and assemble in the highly dynamic cellular cytoplasm to generate intact, functional BMCs. Our knowledge of BMCs is largely based on studies on  $\beta$ -carboxysome biogenesis in cyanobacteria. Assembly of the anabolic  $\beta$ -carboxysomes adopts an “Inside out” or “Cargo first” pathway with the assistance of chaperones: Rubisco enzymes first assemble, by interacting with the linker protein CcmM, to form a liquid-like matrix, and then the Rubisco condensate is encapsulated by shell proteins to form a defined carboxysome structure<sup>44-48</sup>. The architecture and protein organization of functional  $\beta$ -carboxysomes exhibit variability in response to changing environments<sup>49</sup>. Unlike the assembly pathway of  $\beta$ -carboxysomes,  $\alpha$ -carboxysome assembly appears to follow either the “Shell first” or “Concomitant shell–core assembly” routes: the shell assembles alone without Rubisco packing<sup>50</sup> or concomitantly with nucleation of Rubisco<sup>51,52</sup>. Consistently, empty  $\alpha$ -carboxysome shells have been reconstituted in heterologous hosts in the absence of cargo enzymes<sup>53</sup>. In contrast, little is known about the assembly principles that mediate the formation of bacterial metabolosomes, which are required for the catabolism of various substrates and are associated with numerous human diseases.

Here we designed a system to systematically investigate the structural roles of individual Pdu protein components and characterized the spatiotemporal biogenesis of Pdu metabolosomes in *S. Typhimurium* LT2. Genetic analysis, live-cell fluorescence imaging, electron microscopy (EM), and growth assays revealed that the Pdu components undergo independent self-assembly to form the cargo and shell aggregates. The biogenesis of Pdu metabolosomes adopts a combination of the “Shell first” and “Cargo first” pathways and occurs at the cell poles. We showed how the PduB N-terminus and PduM mediate the physical binding between the shell and internal enzyme assemblies, and how PduK impacts the subcellular positioning of Pdu metabolosomes. We also found that the ordered assembly of cargo enzymes generates an internal enzymatic core that possesses a liquid-like status, a physical property that is fundamental for the phase transitions of multicomponent systems. Our findings provide mechanistic insight into the assembly principles of Pdu BMCs, which can be extrapolated to a range of bacterial metabolosomes.

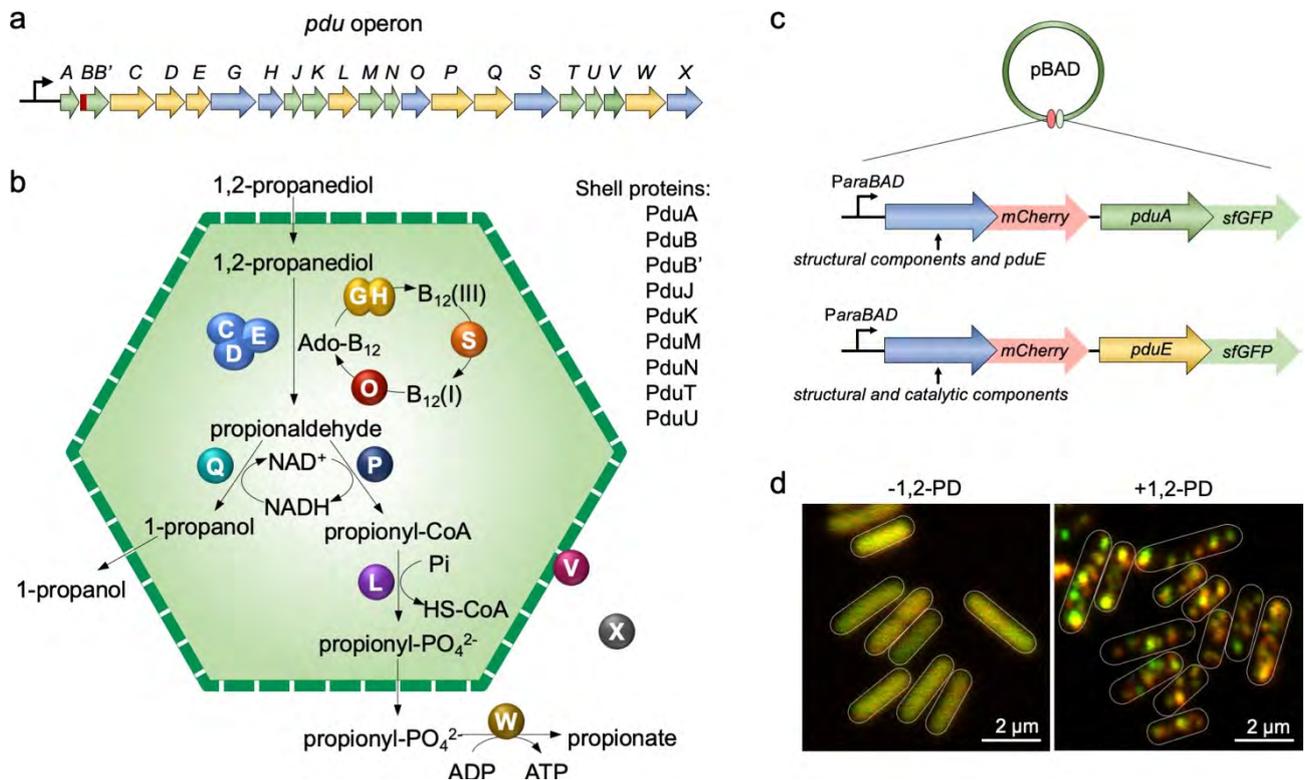
## Results

### A system to investigate Pdu BMC biogenesis in *S. Typhimurium* LT2

The genes encoding the proteins of the Pdu BMCs that mediate 1,2-PD degradation are located in the *pdu* operon (*pduA-X*) in the *S. Typhimurium* chromosome (Fig. 1a and 1b). Transcription of the *pdu* operon occurs in the presence of 1,2-PD during anaerobic growth, and is activated by Crp and Arc<sup>54,55</sup>. To investigate the biogenesis pathway of Pdu BMCs in *S. Typhimurium* LT2, we designed a system based on the pBAD/*Myc*-His vector with the arabinose-inducible *ParaBAD* promoter<sup>56,57</sup> and labeled individual Pdu proteins with mCherry and super-folder GFP (sfGFP) (Fig. 1c). We used the sfGFP tag to visualize the assembly of the PduA shell protein and the PduE cargo protein. In the PduA-sfGFP background, we fused mCherry to other structural components (PduB/B'/J/K/M/N/T/U) and PduE individually. In the PduE-sfGFP background, we fused mCherry to the structural proteins (PduB/B'/J/K/M/N/T/U) and other cargo components (PduG/L/O/P/Q/S/V) individually (Table S1 and S2).

To investigate the effect of increasing levels of protein expression, we detected the fluorescence of PduA-sfGFP at various L-arabinose concentrations in microcompartment-inducing media (MIM) (Fig. S1). In the absence or a low concentration (0.00002%, g·mL<sup>-1</sup>) of L-arabinose, PduA-sfGFP appeared as discrete patches in cells, the typical subcellular distribution of Pdu BMCs as reported previously<sup>24,58</sup>. This expression of PduA-sfGFP probably resulted from the low-level transcription of the *ParaBAD* promoter in the absence of arabinose<sup>59</sup>. Higher concentrations of L-arabinose caused the formation of large puncta, showing that high levels of

expression of Pdu proteins could cause aggregation or interference with the assembly of Pdu BMCs. Accordingly, the fluorescently tagged Pdu proteins were expressed in *S. Typhimurium* LT2 without addition of L-arabinose in this study.



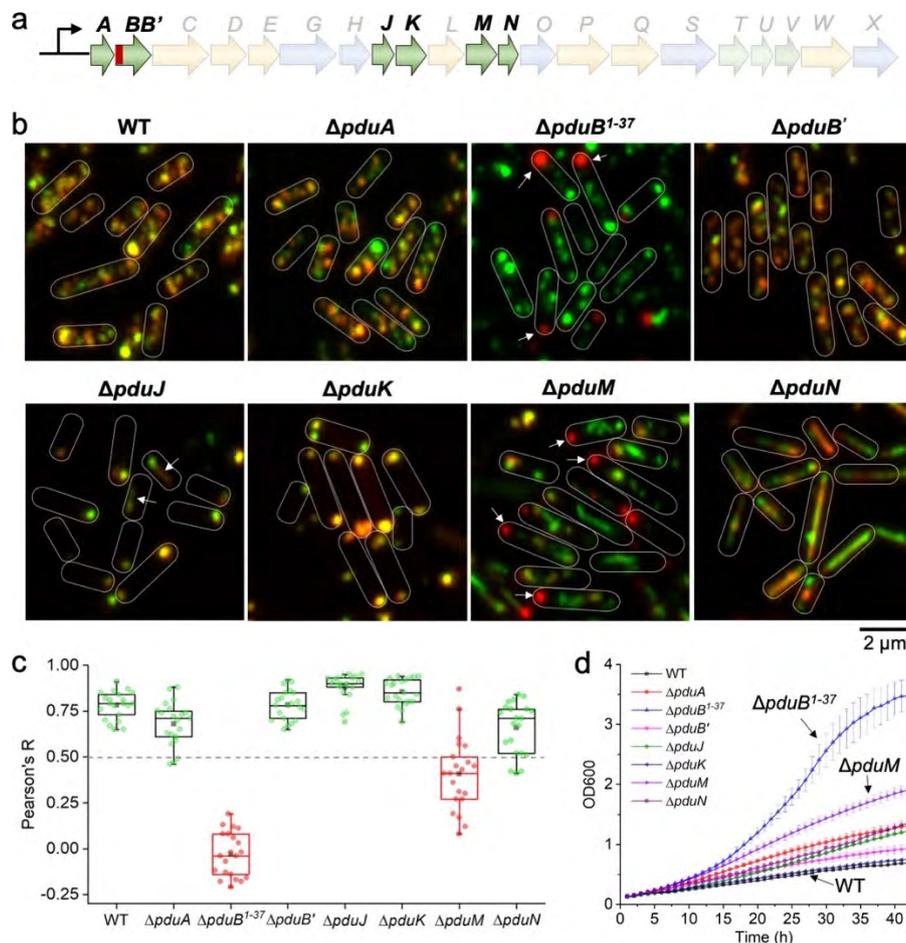
**Fig. 1. A system to investigate Pdu BMC biogenesis.** (a) The chromosomal *pdu* operon of *S. Typhimurium* LT2 encodes structural genes (*pduABB'JKMNTU*), catalytic genes (*pduCDELQPW* for 1,2-propanediol degradation, and the *pduGHOSX* B<sub>12</sub> recycling genes), and *pduV* (proposed to connect with filament-associated BMC movement<sup>39</sup>). PduM is listed as a structural protein according to previous studies<sup>28</sup>. (b) The Pdu BMC shell encapsulates key enzymes that are required for 1,2-propanediol metabolism. (c) All Pdu proteins are visible by dual-labeling with fluorescence proteins (mCherry and sfGFP). Two series of vectors were constructed. First, PduA (a major shell protein) was labeled with sfGFP, and other structural components or PduE (a major enzymatic component) were labeled with mCherry. Second, PduE was labeled with sfGFP, and the structural components or other catalytic components were labeled with mCherry. (d) The presence of 1,2-propanediol (1,2-PD) is required to visualize Pdu BMCs. The fluorescence images show WT *S. Typhimurium* LT2 carrying pBAD-PduE-mCherry/PduA-sfGFP grown in microcompartment-inducing media (MIM), in the absence (-1,2-PD) or presence (+1,2-PD) of 1,2-propanediol.

Growth of *S. Typhimurium* LT2 in the absence of 1,2-PD does not stimulate Pdu BMC formation<sup>15,24</sup>. In our experimental model, growth in minimal media that lacked 1,2-PD (MIM-1,2-PD) allowed the visualization of expressed Pdu proteins diffused throughout the cytosol, confirming no formation of Pdu BMCs (Fig. 1d, Fig. S2). Growth in the presence of 1,2-PD induced the expression of endogenous Pdu proteins and the formation of Pdu BMCs with a typical clustered distribution, as indicated by *in vivo* colocalization of PduE-mCherry and PduA-sfGFP fluorescence (Fig. 1d, Fig. S3). This system and the generated *S. Typhimurium* LT2 mutant strains provided an ideal platform to explore the assembly and roles of individual Pdu proteins during Pdu BMC biogenesis.

### Roles of shell proteins during Pdu BMC biogenesis

To characterize the roles of shell proteins in Pdu BMC assembly, we systematically generated a series of mutants that lacked individual shell proteins ( $\Delta pduA$ ,  $\Delta pduB^{1-37}$ ,  $\Delta pduB'$ ,  $\Delta pduJ$ ,  $\Delta pduK$ ,  $\Delta pduM$ , and  $\Delta pduN$ , Fig. 2a, Fig. S4). pBAD-PduE-mCherry/PduA-sfGFP was then transformed into the deletion mutants, and Pdu BMCs were induced by growth in the presence of 1,2-PD, to investigate the assembly of cargos and shell proteins during Pdu BMC formation. For the  $\Delta pduA$  strain, we specifically used PduE-mCherry/PduJ-sfGFP. The correlation between the assembly of shell proteins and cargos was evaluated by live-cell confocal imaging and colocalization analysis of sfGFP and mCherry fluorescence (Fig. 2b and 2c, Fig. S5).

As a control, we investigated PduE-mCherry and PduA-sfGFP in the WT background and found that the individual proteins were colocalized into discrete patches (Fig. 2b and 2c). Correlation between PduE-mCherry and PduA-sfGFP (PduJ-sfGFP for  $\Delta pduA$ ) were also observed in the  $\Delta pduA$ ,  $\Delta pduB'$  (the start codon of PduB' was replaced by GCT),  $\Delta pduJ$ ,  $\Delta pduK$  and  $\Delta pduN$  strains (Fig. 2b and 2c). The Pdu BMC assemblies in the  $\Delta pduA$  and  $\Delta pduB'$  cells possessed a similar distribution to that seen in the WT background. The absence of PduJ, the most abundant shell protein of the Pdu BMC <sup>24</sup>, resulted in formation of polar aggregates and occasionally elongated Pdu BMC structures (Fig. 2b, arrows), confirming that PduJ is required for both shell encapsulation and Pdu BMC formation <sup>27</sup>. In the absence of PduK, Pdu BMC assemblies were mostly located at the cell poles, implying a role for PduK in the spatial localization of Pdu BMCs (see detailed analysis below). In  $\Delta pduN$ , the Pdu BMC assemblies possessed elongated structures predominately, revealing that the pentameric protein PduN shaped the Pdu BMC polyhedral architecture in the same way as CcmL in the  $\beta$ -carboxysome <sup>60</sup>. In the  $\Delta pduB^{1-37}$  strain, in contrast, PduE assembled to form a large aggregate at one of the cell poles, whereas PduA formed several assemblies in the cytosol (Fig. 2b and 2c), explicitly revealing that the assembly of shell and cargo proteins were spatially separated. In the  $\Delta pduM$  cells, PduE also formed a polar aggregate but features a relatively higher level of colocalization with PduA compared with that in  $\Delta pduB^{1-37}$  (Fig. 2b and 2c). These results indicate that PduB<sup>1-37</sup> and PduM play important roles in correlating assembly between the shell and cargo proteins of Pdu BMCs.



**Fig. 2. Role of *pdu* shell gene products in Pdu BMC biogenesis.** (a) Relevant shell-encoding genes of relevance to (b, c, d) are highlighted in the context of the *pdu* operon of *S. Typhimurium* LT2. (b) PduA-sfGFP (green, PduJ-sfGFP for  $\Delta pduA$ ) and PduE-mCherry (red) were visualized in WT and deletion mutants that lacked individual shell genes, following growth in MIM+1,2-PD. (c) Colocalization analysis of sfGFP and mCherry fluorescence in (b). A Pearson's R value close to 1 reflects reliable colocalization, while values near zero indicate uncorrelated fluorescence distributions. Shell (PduA) and cargo (PduE) were colocalized in all strains except the  $\Delta pduB^{1-37}$  and  $\Delta pduM$  mutants. The Pearson's R values for all strains were:  $0.78 \pm 0.07$  (WT);  $0.68 \pm 0.12$  ( $\Delta pduA$ );  $-0.04 \pm 0.12$  ( $\Delta pduB$ );  $0.78 \pm 0.08$  ( $\Delta pduB'$ );  $0.88 \pm 0.07$  ( $\Delta pduJ$ );  $0.86 \pm 0.07$  ( $\Delta pduK$ );  $0.42 \pm 0.19$  ( $\Delta pduM$ );  $0.66 \pm 0.14$  ( $\Delta pduN$ ). Data are represented as mean  $\pm$  SD. Boxplot center lines correspond to the median value; upper and lower whiskers extend from the box to the largest or smallest value correspondingly, but no more than 1.5 times the interquartile range; mean values show as the square symbol.  $n = 20$ ,  $n$  represents the number of cells. (d) Growth curves of WT *S. Typhimurium* LT2 and Pdu shell gene-deletion mutants on 1,2-PD with limiting B<sub>12</sub> (20 nM) in NCE medium (containing 0.6% 1,2-PD; 0.3 mM each of leucine, isoleucine, threonine, and valine; 50  $\mu$ M ferric citrate).

The Pdu BMC shell acts as a selectively permeable barrier that controls the influx/efflux of substrates and cofactors<sup>6,18,61</sup>. To evaluate the effects of shell proteins on shell integrity and enzyme activity, we performed cell growth assays on the WT and gene-deletion mutants in the presence of 1,2-PD at a limiting level of vitamin B<sub>12</sub> (Fig. 2d, see Methods). The activity of PduCDE is dependent on the vitamin B<sub>12</sub> level, the absence or defect of the shell structure caused by deletion of shell proteins may elevate the availability of the large molecule vitamin B<sub>12</sub> and thus increase the PduCDE activity<sup>10,27</sup>; the level of PduCDE activity enhancement is limited by the vitamin B<sub>12</sub> level, and the production of the toxic propionaldehyde is likely retained at a moderate level to ensure the mutants survive. These could explain why some Pdu shell gene-deletion mutants grow more rapidly than the WT at a limiting level of vitamin B<sub>12</sub> as reported previously<sup>27</sup>.

At a limiting level of vitamin B<sub>12</sub>, the  $\Delta pduB^{1-37}$  and  $\Delta pduM$  strains exhibited significantly enhanced growth compared with other deletion mutants and the WT strain, confirming that the assembly of shell and cargo proteins in the  $\Delta pduB^{1-37}$  and  $\Delta pduM$  strains had been separated, as we visualized by confocal microscopy (Fig. 2b and 2c). The  $\Delta pduA$ ,  $\Delta pduB'$ ,  $\Delta pduJ$  and  $\Delta pduN$  cells grew more rapidly than the WT, suggesting that the resulting Pdu BMC shells had altered structures to control the access and availability of vitamin B<sub>12</sub>. The growth rate of the  $\Delta pduK$  mutant is comparable to that of the WT, confirming that PduK is not essential for controlling shell integrity<sup>27,62</sup>.

However, during growth in the media containing a saturating level of vitamin B<sub>12</sub>, the Pdu BMC shell no longer acts as a physical barrier that separates vitamin B<sub>12</sub> from the PduCDE protein complexes, and so the PduCDE enzyme activity is maximized, which leads to enhanced propionaldehyde production. Cell growth assays at a saturating level of vitamin B<sub>12</sub> revealed that the intact Pdu BMC shells of WT *Salmonella* effectively sequestered propionaldehyde, as reflected by a higher growth rate; whereas the  $\Delta pduB^{1-37}$  and  $\Delta pduM$  mutants, in which cargo enzymes and shell proteins assemble separately, exhibited significant growth arrest between 18 and 32 h (Fig. S6).

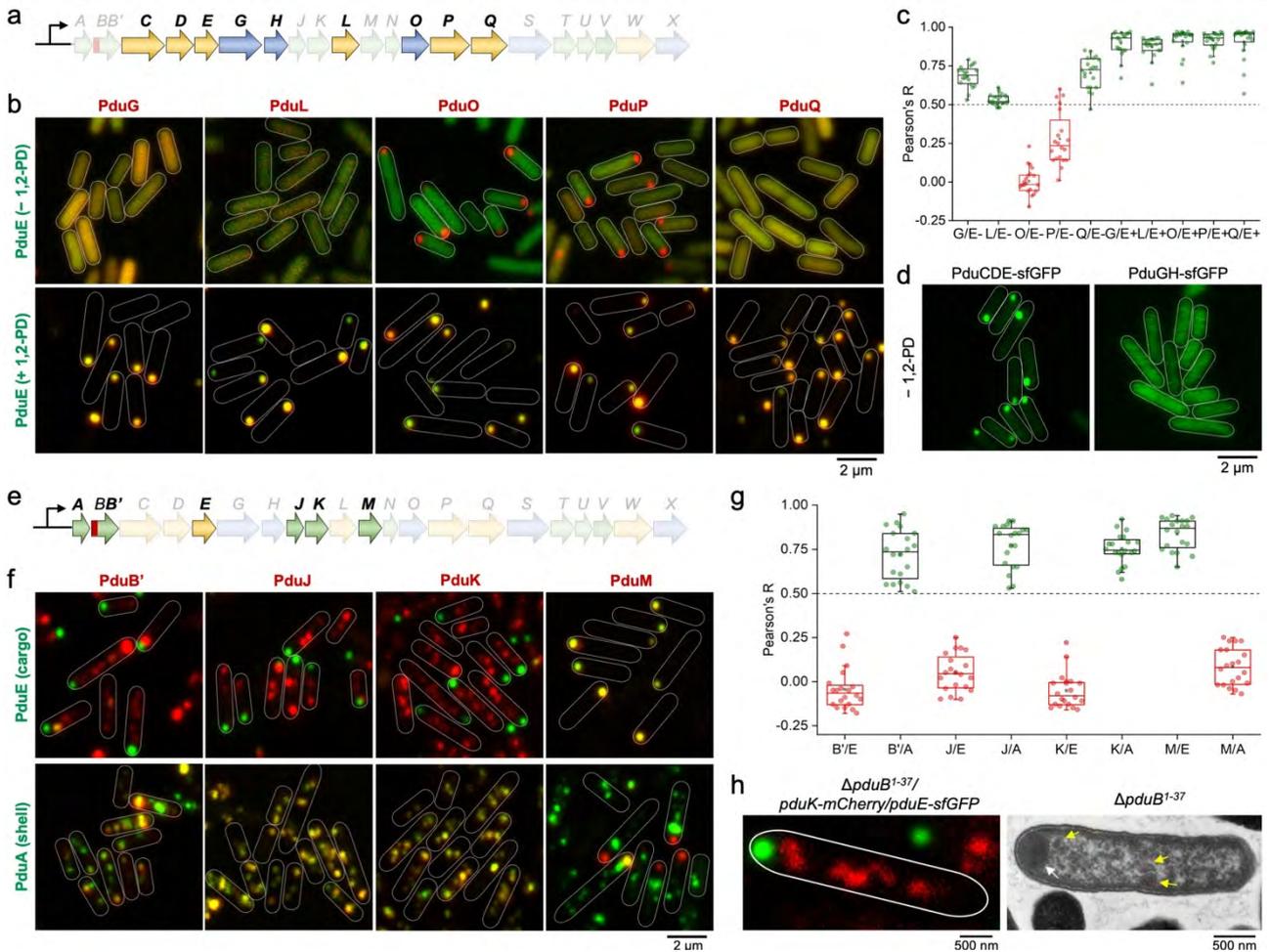
### **Independent assembly of shell and cargo proteins and PduB<sup>1-37</sup>-mediated binding between the shell and the enzymatic core**

Our results revealed that in the absence of the PduB N-terminus (PduB<sup>1-37</sup>), the shell protein PduA and the cargo PduE were spatially disassociated and each of them self-assembled to form one or several aggregates (Fig. 2). To understand the composition of these protein assemblies, we studied the localization of individual Pdu proteins in the  $\Delta pduB^{1-37}$  strain. Catalytic cargo components were tagged individually with mCherry in the PduE-sfGFP/ $\Delta pduB^{1-37}$  background (Fig. 3a). When Pdu BMC expression was not stimulated (in the absence of 1,2-PD), the cargo proteins PduE, PduG, PduL, and PduQ were diffusely distributed throughout the cytosol (Fig. 3b and 3c). In contrast, both PduO and PduP formed aggregates at a single pole of the cell, accompanied by a low level of cytosolic distribution (Fig. 3b and 3c). In the presence of Pdu protein assemblies (following 1,2-PD induction), PduG, PduL, PduO, PduP, and PduQ colocalized with PduE at one pole of the  $\Delta pduB^{1-37}$  cell (Fig. 3b and 3c, Fig. S5), suggesting that the self-association of these Pdu cargo enzymes is required to form the enzymatic core located at the cell pole. Moreover, we determined the subcellular localization of multiple enzyme complexes, PduCDE and PduGH, in the absence of Pdu BMC structures formed in cells (Fig. 3d). PduCDE formed an aggregate close to the cell pole, whereas PduGH had a dispersed distribution in the cytosol, similar to PduG alone (Fig. 3b).

The aggregation of PduCDE-sfGFP and PduP-mCherry was likely mediated by the native encapsulation peptides located at the N-termini of the cargo proteins<sup>89</sup>. Further, we speculate that the aggregation of PduO reflects either the self-aggregation or insolubility of expressed PduO, or is a consequence of the fusion to the GFP protein. The difference between the fluorescence phenotypes of PduCDE-sfGFP (PduE tagged with sfGFP, Fig. 3d) and PduE-sfGFP (Fig. 3b, -1,2-PD) could reflect the fact that PduC and PduD were expressed from the pBAD vector instead of being endogenously expressed in the absence of 1,2-PD.

Collectively, our results show that intrinsic protein self-assembly is pivotal to the formation of the multi-subunit Pdu cargo complexes. We found that the fluorescence signal of the minor components PduS and PduV in the  $\Delta pduB^{1-37}$  strain was weak (Fig. S7), similar to the fluorescence observed in the WT background (Fig. S3), thereby reflecting the low expression levels and partial distribution of these proteins in the cytosol. Furthermore, PduS and PduV did not colocalize with PduE (Fig. S7), consistent with direct interactions between the PduS and PduV proteins and the shell<sup>39,63</sup>.

$\Delta pduB^{1-37}$



**Fig. 3. The PduB N-terminus links between the shell and the enzymatic core.** (a) Relevant cargo-encoding genes of relevance to Fig. 3b-3d are highlighted in the context of the *pdu* operon of *S. Typhimurium* LT2. (b) PduE-sfGFP (green) and mCherry (red, labeling different catalytic components) were visualized in  $\Delta pduB^{1-37}$  following growth in MIM in the absence and presence of 1,2-PD. (c) Colocalization analysis of sfGFP and mCherry fluorescence in (b). Note that in panels c and g, the first capital letter of the name refers to the Pdu protein tagged with mCherry and the second capital letter of the name refers to the Pdu protein tagged with sfGFP. ‘+’ and ‘-’ represents the presence and absence of 1,2-PD, respectively. For example: ‘G/E+’ stands for PduG-mCherry/PduE-sfGFP in the presence of 1,2-PD. Pearson’s R values for all the strains are:  $0.68 \pm 0.07$  (G/E-);  $0.53 \pm 0.04$  (L/E-);  $0.01 \pm 0.09$  (O/E-);  $0.28 \pm 0.17$  (P/E-);  $0.70 \pm 0.10$  (Q/E-);  $0.90 \pm 0.08$  (G/E+);  $0.87 \pm 0.07$  (L/E+);  $0.91 \pm 0.08$  (O/E+);  $0.91 \pm 0.05$  (P/E+);  $0.91 \pm 0.10$  (Q/E+). Data are represented as mean  $\pm$  SD.  $n = 20$ ,  $n$  represents the number of cells. (d) The localization of Pdu proteins with multiple subunits, PduCDE and PduGH, were characterized following growth in MIM-1,2-PD media. The PduCDE and PduGH complexes were expressed from the pBAD plasmid, with PduE and PduH subunits tagged with sfGFP, respectively. (e) Relevant genes encoding shell and catalytic proteins in Fig. 3f and 3g, highlighted in the context of the *pdu* operon. (f) PduE-sfGFP (green) or PduA-sfGFP (green) and mCherry (red, labeling different structural components) were visualized in  $\Delta pduB^{1-37}$ , following growth in MIM+1,2-PD media. (g) Colocalization analysis of sfGFP and mCherry fluorescence in (f). Structure components (PduB’, J, K) colocalized with PduA (shell), but not colocalized with PduE (cargo). While PduM, a structural component colocalized with PduE. The Pearson’s R values for all the strains are:  $-0.05 \pm 0.12$  (B’/E);  $0.72 \pm 0.14$  (B’/A);  $0.05 \pm 0.10$  (J/E);  $0.77 \pm 0.12$  (J/A);  $-0.05 \pm 0.10$  (K/E);  $0.75 \pm 0.08$  (K/A);  $0.84 \pm 0.09$  (M/E);  $0.08 \pm 0.10$  (M/A). Data are represented as mean  $\pm$  SD.  $n = 20$ ,  $n$  represents the number of cells. (h) Microscopic characterization of the  $\Delta pduB^{1-37}$  strain following growth in MIM+1,2-PD media, showing a polar enzymatic core and several empty shell structures (arrows). Left, confocal image of the pBAD-PduK-mCherry/PduE-sfGFP transformed into  $\Delta pduB^{1-37}$  strain. Right, EM image of the  $\Delta pduB^{1-37}$  strain. Note: Boxplot center lines correspond to the median value; and upper and lower whiskers extend from the box to the largest or smallest value correspondingly, but no more than 1.5 times the interquartile range; mean values show as the square symbol.

To probe the shell assembling structures, individual shell components were labeled with mCherry in both the PduE-sfGFP/ $\Delta pduB^{1-37}$  and PduA-sfGFP/ $\Delta pduB^{1-37}$  background (Fig. 3e and 3f). Following the formation of Pdu BMCs, PduB’, PduJ, and PduK showed a patchy distribution; they colocalized with the shell protein PduA,

but had no spatial correlation with the enzymatic core (indicated by PduE-sfGFP fluorescence, Fig. 3f and 3g). The minor shell proteins PduN, PduT, and PduU<sup>24</sup> had a stronger propensity to colocalize with PduA than PduE (Fig. S7). These results demonstrate that the shell proteins can self-associate to form multi-complex structures in the  $\Delta pduB^{1-37}$  strain, which resemble the empty shell-like structures observed previously<sup>43</sup>. We corroborated the fluorescence data with thin-section EM, which identified a polar aggregate (likely the core structure) and several shell-like particles in the  $\Delta pduB^{1-37}$  cell (Fig. 3h).

We detected the spatial separation of the assembly of shell proteins and cargos in the  $\Delta pduB^{1-37}$  cells (Fig. 3); on the contrary, the absence of PduB' had no major effects on the shell-cargo association (Fig. S8). These findings pinpointed the importance of the PduB N-terminus itself in the physical binding between shell and cargo proteins. Complementation experiments by expressing PduB in the  $\Delta pduB^{1-37}$  strain showed the recovery of Pdu BMC formation, as evidenced by confocal imaging, thin-section EM, and growth assay (Fig. S9). The formation of an enzymatic core without encapsulation of the shell explained the increased growth rate of the  $\Delta pduB^{1-37}$  strain in a B<sub>12</sub>-limiting environment, compared to that of the WT strain (Fig. 2d).

Taken together, our data demonstrate that the Pdu shell and cargo proteins have the intrinsic tendency to self-assemble resulting in the formation of independent shell-like structures and the enzymatic core. Moreover, PduB<sup>1-37</sup> serves as a linker peptide that acts as a bridge between the enzymatic core and shell assemblies, playing an essential role in the biogenesis of Pdu BMCs. Comparative genomic analysis revealed that PduB<sup>1-37</sup> is conserved throughout the *Klebsiella*, *Escherichia*, *Citrobacter*, and *Salmonella* genera that carry the *pdu* operon (Fig. S10). In these four genera, the PduB<sup>1-37</sup> linker protein was conserved as highly as the other structural or essential Pdu proteins (PduA/B/D/J/N/U). In contrast, the PduX protein, which has an unknown function, was not highly conserved<sup>38</sup>. Because the predicted structures of PduB<sup>1-37</sup> from different genera were very similar (Fig. S10), our findings suggest that PduB<sup>1-37</sup> may play a ubiquitous role in the assembly of the Pdu BMC in different bacterial species.

### **PduM plays a role in the binding between shell and enzyme core**

Our results showed that when the Pdu BMC shell-cargo association is impeded by the absence of PduB<sup>1-37</sup>, PduM formed large puncta at the cell pole and colocalized with PduE but not with PduA (Fig. 3f and 3g). This suggests a strong interaction between PduM and cargo enzymes within the enzymatic core, and no significant association between PduM and shell assemblies. To investigate the role of PduM in more detail, we studied the assembly of structural and catalytic Pdu proteins in the  $\Delta pduM$  background (Fig. 4a and 4b). We found that the catalytic enzymes PduG, PduO, PduP, and PduQ colocalized with PduE, whereas the shell proteins PduB, PduB', PduJ and PduK assembled with PduA (Fig. 4b and 4c, Fig. S11), recapitulating the results observed in the  $\Delta pduB^{1-37}$  strain (Fig. 3). Moreover, confocal images of the  $\Delta pduM$  strain expressing PduE-mCherry and PduA-sfGFP revealed a relatively low level of physical correlation between shell assemblies and cargos (Fig. 2b and 2c).

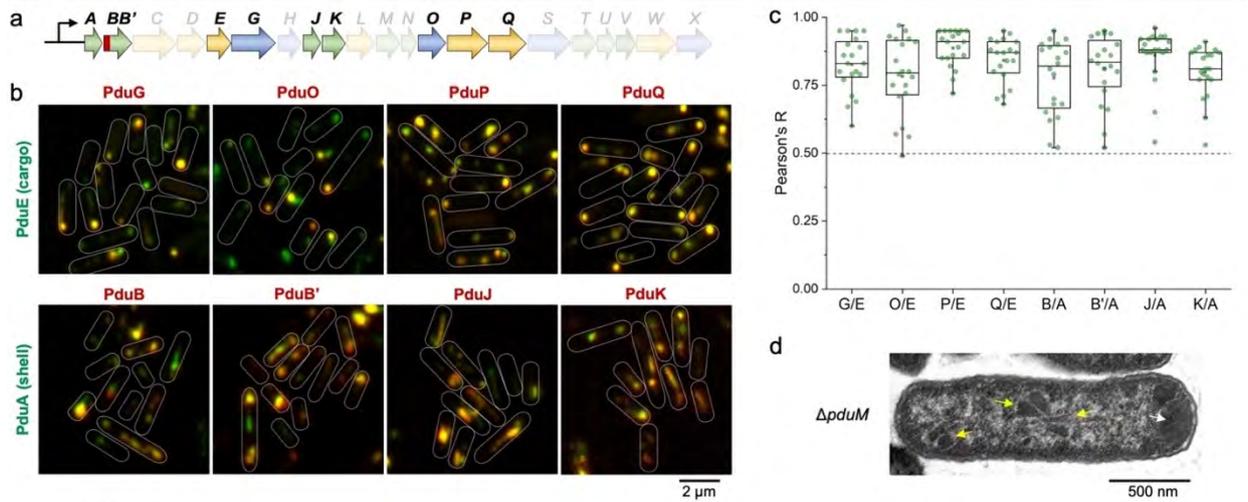
These results indicate that PduB<sup>1-37</sup> and PduM play similar roles in mediating the shell-cargo assembly. In addition, in the absence of PduM, both PduB and PduB' colocalized with PduA in the resulting shell assemblies, suggesting that PduM sits between PduB<sup>1-37</sup> and cargo enzymes to promote the shell-cargo binding. However, the function of PduM in the shell-cargo assembly appears less significant, as Pearson's R was higher in the absence of PduM ( $\Delta pduM$ ) than that in the absence of PduB N-terminus ( $\Delta pduB^{1-37}$ ) (Fig. 2c). This would be consistent with other associations occurring between PduB<sup>1-37</sup> and cargos, such as the binding of PduB<sup>1-37</sup> to cargo proteins via specific encapsulation peptides at the N-termini of enzymes localized within the Pdu BMC

40-42

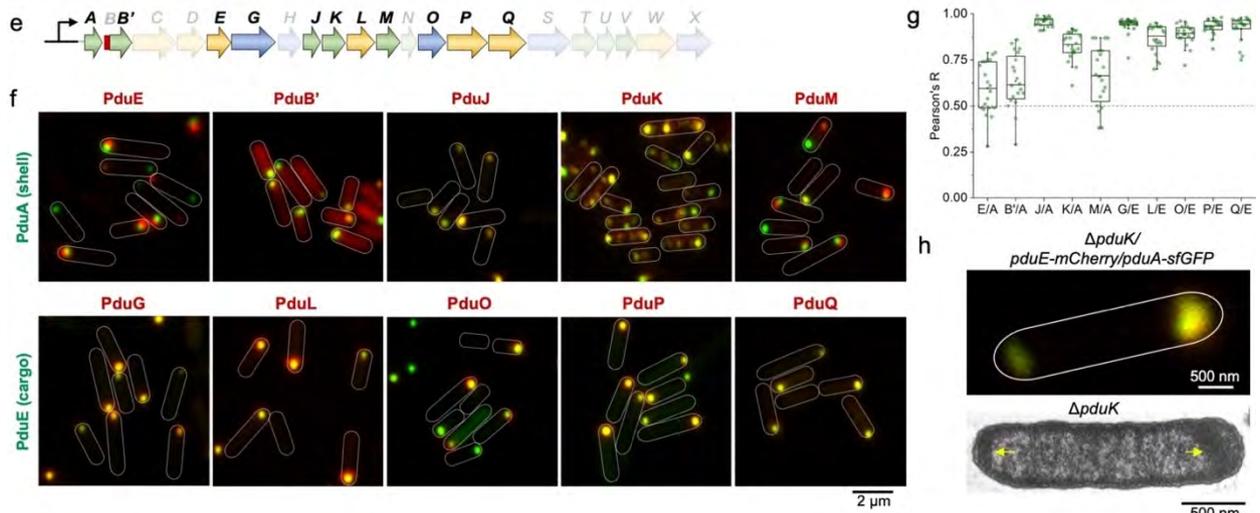
In addition to the shell-like structures formed by PduA, PduB, PduB', PduJ and PduK in the  $\Delta pduM$  mutant, some elongated or enlarged shell assemblies were also seen (Fig. 4b and 4d). This observation suggests that PduM is required for the defined polyhedral structure of the Pdu BMC, which fits with previous findings<sup>28</sup>. Expressing PduM in the  $\Delta pduM$  strain partially rescued the assembly of Pdu BMCs, as evidenced by a higher colocalization and a slower growth rate compared with those of  $\Delta pduM$  (Fig. S9). Interestingly, the aberrant assembly structures were still observed in the complementation experiment (Fig. S9), perhaps reflecting the high level of PduM expression. It has previously been shown that only a low amount of PduM is required to define the assembly of native Pdu BMCs<sup>24,28</sup>. We searched for orthologous PduM proteins amongst 61 bacterial genomes that contain the *pdu* operon, and found that PduM was conserved (> 50% identity at the

amino acid level) in all but one strain (Fig. S10). The high similarity of the predicted PduM structures from these genera (Fig. S10) suggests that PduM may play a critical role in Pdu BMC assembly.

#### *ΔpduM*



#### *ΔpduB<sup>1-37</sup>-pduK*



**Fig. 4. Roles of PduM and PduK in the biogenesis of Pdu BMCs.** (a) Relevant genes encoding shell and catalytic proteins of relevance to (b) and (c), highlighted in the context of the *pdu* operon of *S. Typhimurium* LT2. (b) PduE-sfGFP (green) and mCherry-fused catalytic components (red), and PduA-sfGFP (green) and mCherry-fused structural components (red) were visualized in  $\Delta pduM$  grown in MIM+1,2-PD media. (c) Colocalization analysis of sfGFP and mCherry fluorescence in (b). The graph is labeled as described in the legend to Fig. 3. Structure components (PduB, B', J, K) colocalized with PduA (shell), and catalytic components (PduG, O, P, Q) colocalized with PduE (cargo). The Pearson's R values for all the strains are:  $0.83 \pm 0.10$  (G/E);  $0.78 \pm 0.14$  (O/E);  $0.89 \pm 0.07$  (P/E);  $0.84 \pm 0.08$  (Q/E);  $0.78 \pm 0.13$  (B/A);  $0.81 \pm 0.12$  (B'/A);  $0.86 \pm 0.10$  (J/A);  $0.80 \pm 0.09$  (K/A). Data are represented as mean  $\pm$  SD.  $n = 20$ ,  $n$  represents the number of cells. (d) EM of the  $\Delta pduM$  strain following growth in MIM+1,2-PD media, showing a polar enzymatic core (white arrow) and various shapes of the shell structure (yellow arrows). (e) Related shell and catalytic genes in Fig. 4f and 4g were highlighted in the *pdu* operon. (f) PduA-sfGFP (green) and mCherry (red, labeling PduE and different structural components), and PduE-sfGFP (green) and mCherry (red, labeling different catalytic components), were visualized in  $\Delta pduB^{1-37}\Delta pduK$  strain growth in MIM+1,2-PD media. Note: PduB' in  $\Delta pduB^{1-37}\Delta pduK$  mainly formed a polar aggregation, which is explicitly colocalized with PduA; the cytosolic distribution of PduB' was also observed in the WT background, indicating that a portion of PduB' could not be integrated into Pdu BMCs. (g) Colocalization analysis of sfGFP and mCherry fluorescence in (f). Structural components (PduJ, K) colocalized with PduA (shell), and catalytic components (PduG, PduL, PduO, PduP, PduQ) colocalized with PduE (cargo). PduE and PduM were partially not colocalized with PduA. The Pearson's R values for all the strains are:  $0.60 \pm 0.14$  (E/A);  $0.63 \pm 0.15$  (B'/A);  $0.96 \pm 0.02$  (J/A);  $0.83 \pm 0.08$  (K/A);  $0.66 \pm 0.16$  (M/A);  $0.94 \pm 0.05$  (G/E);  $0.86 \pm 0.07$  (L/E);  $0.89 \pm 0.05$  (O/E);  $0.93 \pm 0.04$  (P/E);  $0.92 \pm 0.07$  (Q/E). Data are represented as mean  $\pm$  SD.  $n = 20$ ,  $n$  represents the number of cells. (h) The polar aggregates of Pdu proteins visualized in confocal microscopy (top) and EM (bottom, yellow arrows) of the  $\Delta pduK$  strain grown in MIM+1,2-PD media. Top, confocal image of the pBAD-PduE-mCherry/PduA-sfGFP transformed into  $\Delta pduK$  strain. Bottom, EM image of the  $\Delta pduK$  strain without fluorescence tagging. Note: Boxplot center lines correspond to the median value; and upper and lower whiskers extend from the box to the largest or smallest value correspondingly, but no more than 1.5 times the interquartile range; mean values show as the square symbol.

### **PduK is vital for the subcellular distribution of Pdu BMCs**

PduK is another minor shell protein within the Pdu BMC<sup>24</sup>. The N-terminal region of PduK has a high sequence similarity to the hexameric shell protein PduA, and the C-terminal extension of PduK has an unknown function (Fig. S10). In the absence of PduK, Pdu BMCs were restricted to the cell poles (Fig. 2b)<sup>27</sup>, and SDS-PAGE and EM results revealed that the Pdu BMC structures are well-formed in the  $\Delta pduK$  cells (Fig. S12), suggesting that PduK is essential for the spatial localization of Pdu BMCs.

To explore the role of PduK in the biogenesis of Pdu BMCs, we determined the distribution of individual Pdu proteins in the  $\Delta pduB^{1-37}/\Delta pduK$  mutant (Fig. 4e). The major shell proteins PduB' and PduJ colocalized with PduA, forming assemblies at one cell pole (Fig. 4f and 4g, Fig. S11). Apart from those colocalized with PduA, some PduB' proteins were diffusely distributed throughout the cytosol of the  $\Delta pduB^{1-37}/\Delta pduK$  cells, which was also seen in the WT background (Fig. S3). These findings are consistent with the comparison of Pdu protein abundance in cell extracts and isolated Pdu BMCs that indicated that a portion of PduBB' could not be assembled into Pdu BMC structures<sup>24</sup>.

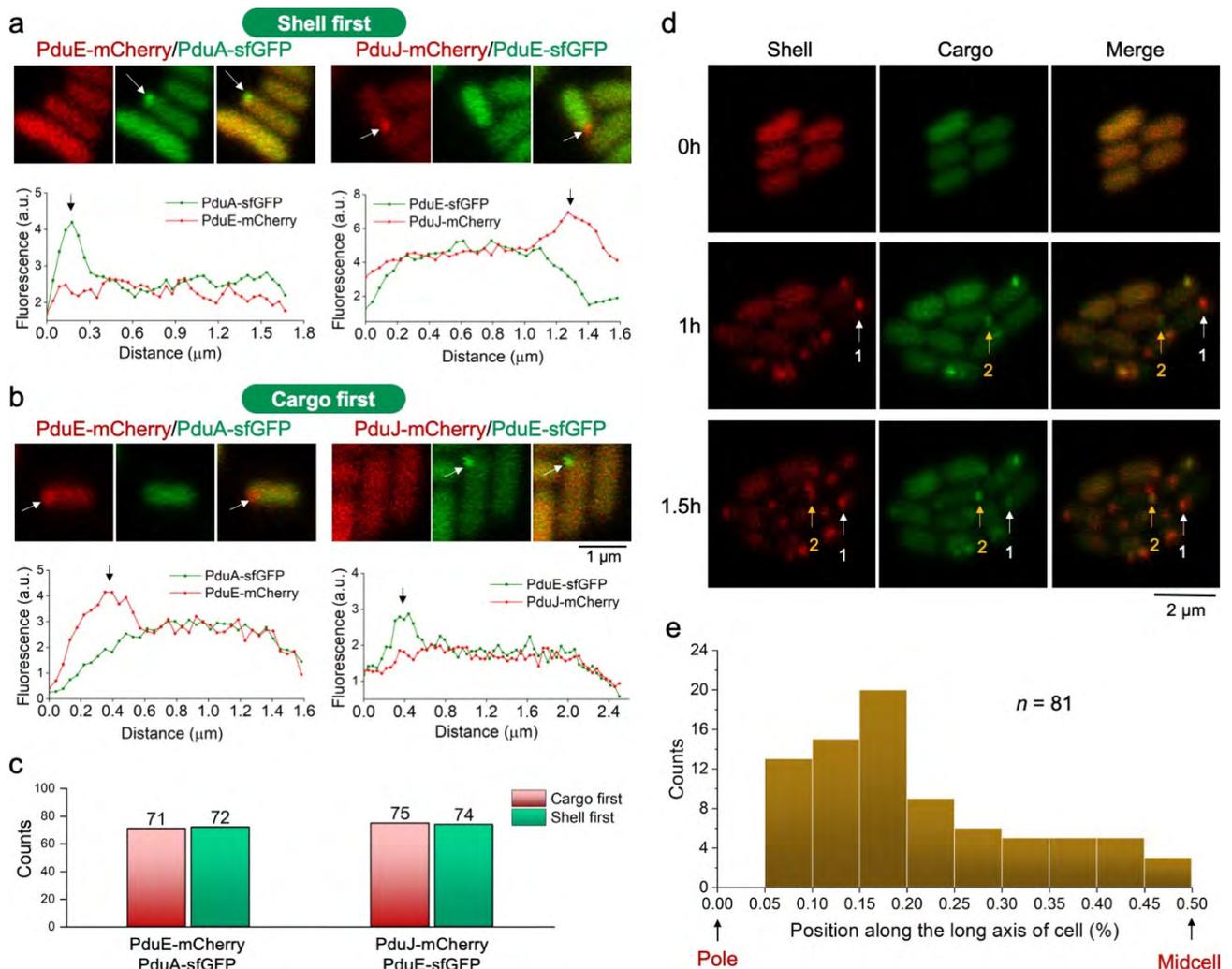
The localization of cargo enzymes in  $\Delta pduB^{1-37}/\Delta pduK$  was comparable to that seen in  $\Delta pduB^{1-37}$ : PduG, PduL, PduO, PduP, and PduQ colocalized with PduE within a large aggregate located at one cell pole (Fig. 4f and 4g), indicating that PduK is not essential for the formation of the enzymatic core. However, unlike the  $\Delta pduB^{1-37}$  cells that showed several patches of shell assemblies distributed in the cytosol, in  $\Delta pduK$  shell proteins were prone to assemble into large aggregates located at the cell poles (Fig. 4f and 4g). As depicted in PduA-sfGFP/PduE-mCherry/ $\Delta pduK$  and PduA-sfGFP/PduM-mCherry/ $\Delta pduK$ , the shell and cargo aggregates were either located at the opposite poles of the cell, or next to each other at the same pole but not exactly colocalized, suggesting that PduK plays a role in facilitating the relocation of *in situ* synthesized shell proteins for the generation of additional shell structures in the cytosol.

Consistent with the fluorescence results, EM showed polar aggregations of Pdu proteins in the  $\Delta pduK$  strain (Fig. 4h), in agreement with previous observations<sup>27</sup>. Complementation experiments showed that expressing PduK in the  $\Delta pduK$  strain rescued the distribution of Pdu BMCs, and the growth rate remained similar to WT (Fig. S9). Expression of PduK-mCherry in  $\Delta pduB^{1-37}/\Delta pduK$  also recovered the aberrant distribution of Pdu BMC shells (Fig. 4f).

Taken together, our results confirmed that the assembly of Pdu BMC shell proteins and cargos occur independently *in vivo*, and revealed that PduK is important for the assembly and subcellular distribution of both shell assemblies and entire Pdu BMCs. PduK orthologs were highly conserved and are predicted to have significant structural similarity between the bacterial genera that carry the *pdu* operon (Fig. S10), suggesting a universal role of PduK in Pdu BMC assembly.

### **A combination of two biogenesis pathways of the Pdu BMC**

To understand the precise nature of the Pdu BMC biogenesis, we studied the temporal order of the initial assembly of Pdu shell proteins and cargos, by applying time-lapse confocal imaging on cells that express either PduE-mCherry/PduA-sfGFP or PduJ-mCherry/PduE-sfGFP for cross-validation. After 1h 1,2-PD induction to stimulate the expression of Pdu BMCs, fluorescence foci of shell proteins (indicated by PduA-sfGFP or PduJ-mCherry) and cargo proteins (indicated by PduE-mCherry or PduE-sfGFP) appeared at the cell poles from the cytosolic-distribution fluorescence background. In some cells, the fluorescent spot of shell proteins appeared earlier than that of cargo proteins, termed the “Shell first” event (Fig. 5a), whereas in some cells the fluorescent spot of enzymes formed first, termed the “Cargo first” event (Fig. 5a and 5b). Image analysis revealed that the “Shell first” events occurred as frequently as the “Cargo first” events (Fig. 5c,  $n = 72$  vs  $71$  and  $n = 74$  vs  $75$ , respectively). Time-lapse fluorescence microscopy confirmed that the “Shell first” and “Cargo first” events co-occurred in different cells after 1h 1,2-PD induction. The colocalization of shell and cargo fluorescence was then visualized after 1.5h induction with 1,2-PD, indicating the subsequent import of cargo enzymes and encapsulation of shell structures, respectively, prior to the formation of the Pdu BMCs (Fig. 5d). We also observed the colocalization of shell and core assemblies occurring at 1h, and speculate that a concomitant assembly or a fusion event of shell/cargo-independent assembly has occurred (Fig. 5d, Table S4). Approximately 70.4% of the initial assembly events occurred between the pole and the quarter positions along the longitudinal axis of the cell (Fig. 5e), reminiscent of the “birth” events of  $\beta$ -carboxysomes in cyanobacteria



**Fig. 5. Pdu BMCs undergo a combination of the “Shell first” and “Cargo first” assembly pathways.** (a, b) Aggregation of the shell (PduA-sfGFP or PduJ-mCherry) and cargo (PduE-sfGFP or PduE-mCherry) in the WT strain following induction with 1,2-PD (addition of 0.6% 1,2-PD to MIM media at  $t = 0$ ). (c) The distribution of the number of “Shell first” and “Cargo first” events in PduE-mCherry/ PduA-sfGFP and PduJ-mCherry/ PduE-sfGFP. (d) Progression of shell and cargo formation in WT strain with pBAD-PduJ-mCherry/PduE-sfGFP. “Shell first” (white arrow, number 1) and “Cargo first” (yellow arrow, number 2) events co-occurred 1h after addition of 1,2-PD to MIM media. Shell and cargo fluorescence colocalized at 1.5h after addition of 1,2-PD (white and yellow arrows). (e) Histogram of the spatial distribution of the first Pdu BMC along the long axis of the cell.  $n = 81$ ,  $n$  represents the number of cells.

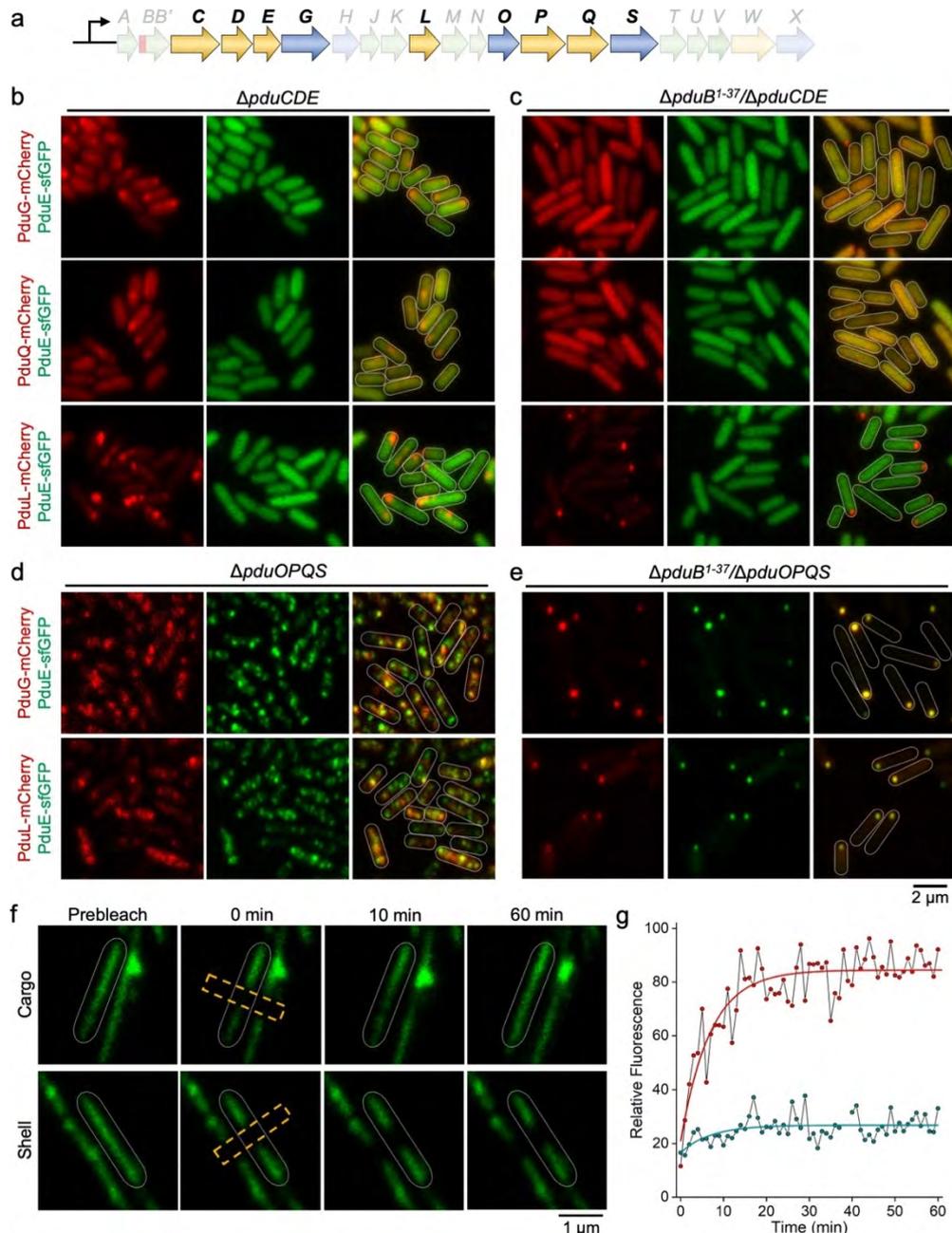
Overall, our data prove that the Pdu BMC undertakes both the Shell-first and Cargo-first assembly pathways, which is preceded by initial shell and cargo assembly that occurs independently, and focuses at the cell poles.

### PduCDE triggers hierarchical assembly of the enzymatic core

Our data showed that the cargo enzymes (PduCDE/PduGH/L/Q) have inherent affinity to assemble to form the multi-complex enzymatic core (Fig. 3b and 3d), providing a means for enhancing catalytic efficiency of 1,2-PD degradation pathways of native Pdu BMCs. To investigate the mechanisms underlying the self-assembly of cargo enzymes, we characterized the subcellular locations of cargo proteins after removing the major enzyme PduCDE or minor enzymes PduO/PduP/PduQ/PduS in both the WT and  $\Delta pduB^{1-37}$  background (Fig. 6a). Without PduCDE in the WT background with addition of 1,2-PD, only PduE distributed throughout the cytosol (Fig. 6b), reminiscent of previous findings suggesting that the PduD N-terminus is important for directing PduCDE into the Pdu BMC<sup>41</sup>. Both PduG and PduQ formed large puncta in addition to substantial cytosolic distribution, differing from the typical distribution of Pdu BMCs in WT. PduL still assembled to construct several aggregates in most cells, given that the N-terminus of PduL serves as an encapsulation peptide<sup>42</sup>. Removal of both PduB<sup>1-37</sup> and PduCDE resulted in the cytosolic distribution of PduG and PduQ, differing from the polar aggregate in the  $\Delta pduB^{1-37}$  strain (Fig. 3b). In contrast, the localization of PduL was not greatly affected (Fig. 6c).

On the contrary, the deletion of PduO/PduP/PduQ/PduS in the WT background did not affect the assembly of PduCDE, PduGH, and PduL within Pdu BMC structures (Fig. 6d). After removal of PduB<sup>1-37</sup> and PduO/PduP/PduQ/PduS, PduG and PduL still formed a polar aggregate within the enzyme core (Fig. 6e), indicating that PduO/PduP/PduQ/PduS are not involved in the assembly of PduGH and PduL enzymes.

Together, our data shed light on the hierarchy of cargo assembly during the formation of the higher-ordered enzyme core, and demonstrate that self-assembly of the major cargo enzyme complex PduCDE triggers the physical association of other cargo enzymes.



**Fig. 6. Interactions and liquid-like association of internal enzymes.** (a) Relevant genes encoding catalytic proteins of relevance to Fig. 6b-6e, highlighted in the context of the *pdu* operon of *S. Typhimurium* LT2. (b, c) PduE-sfGFP (green) and mCherry (red, labeling different internal enzymes) were visualized in  $\Delta pduCDE$  (b) and  $\Delta pduB^{1-37}/\Delta pduCDE$  (c) grown in MIM+1,2-PD media. (d, e) PduE-sfGFP (green) and mCherry (red, labeling different internal enzymes) were visualized in  $\Delta pduOPQS$  (d) and  $\Delta pduB^{1-37}/\Delta pduOPQS$  (e) grown in MIM+1,2-PD media. (f) Representative fluorescence images before bleaching, immediately after bleaching of PduE-sfGFP (cargo) and PduA-sfGFP (shell), at various time lapses. The yellow rectangular boxes indicated the bleaching area. (g) Representative time course of fluorescence recovery of bleached regions of PduE-sfGFP (red) and PduA-sfGFP (cyan). The y-axis indicates fluorescence values relative to the fluorescence intensity of the selected region prior to bleaching. The recovery of sfGFP fluorescence is shown as circles and fitted to an exponential function.

## Liquid-like association of internal enzymes within the Pdu BMC

What is the organizational status of the enzymatic core made of multiple cargo complexes within the Pdu BMC? To tackle this question, we determined the *in vivo* diffusion dynamics of internal enzymes and shell proteins of Pdu BMCs, using fluorescence recovery after photobleaching (FRAP). We chose the elongated Pdu BMC structures in the  $\Delta pduN$  strain (Fig. S13), which are suitable for photobleaching and quantitative analysis<sup>64</sup>. Growth assays suggest that the resulting Pdu BMCs in  $\Delta pduN$  are metabolically functional (Fig. 2d). Fig. 6f and 6g show the representative FRAP image sequences of PduE-sfGFP (cargo) and PduA-sfGFP (shell) in the  $\Delta pduN$  cells. The cargo enzymes exhibited a distinctly higher mobile fraction ( $83\% \pm 6\%$ ,  $n = 20$ ), compared with shell proteins that showed lower mobile fraction ( $6\% \pm 3\%$ ,  $n = 20$ ) (Fig. 6f and 6g, Table S4). The average diffusion coefficient of cargos within the Pdu BMCs of  $\Delta pduN$  is  $4.02 \pm 1.89 \times 10^{-4} \mu\text{m}^2\cdot\text{s}^{-1}$  ( $n = 20$ ), more than 14-fold faster than that of shell proteins ( $0.28 \pm 0.15 \times 10^{-4} \mu\text{m}^2\cdot\text{s}^{-1}$ ,  $n = 20$ ).

In comparison, FRAP of the PduGH-sfGFP proteins that are dispersedly distributed in the cytoplasm (Fig. 3d) in the absence of Pdu BMC formation revealed that the PduGH-sfGFP fluorescence recovered rapidly after only  $\sim 3$  seconds (Fig. S14, Table S4), indicating that cargo enzymes exhibited highly dynamic mobility in the cytosolic solution than inside the Pdu BMC. While protein interactions and the internal arrangement of bacterial metabolosomes may differ between the canonical Pdu BMCs in the WT and the elongated structures of the Pdu BMCs in the  $\Delta pduN$  strain, our results reveal the organizational dynamics of internal enzymes relative to the stable shell structure, suggesting that the cargo condensate within the Pdu BMC has a liquid-like property. A similar liquid-like state has been proposed to be the foundation of the assembly of  $\alpha$ - and  $\beta$ -carboxysomes<sup>47,65</sup>.

## Discussion

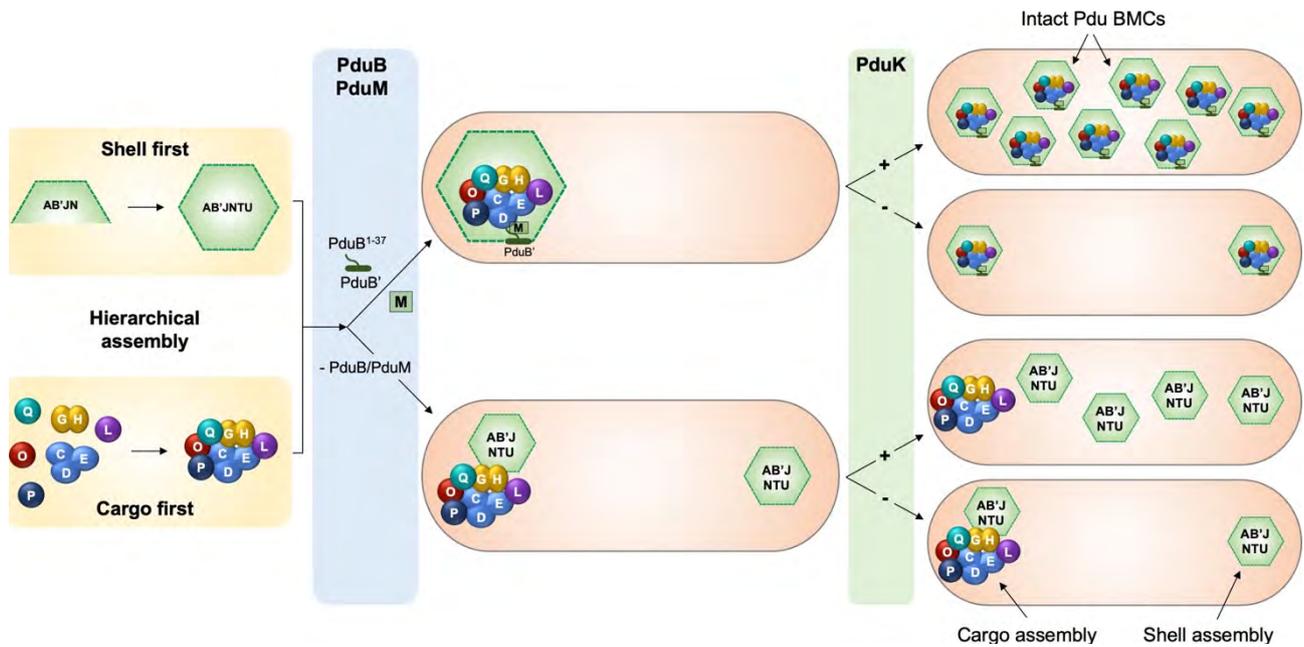
Pdu BMCs are self-assembling protein-based organelles ( $\sim 202$  MDa) composed of more than ten thousand protein polypeptides that sequester toxic intermediates to provide a key metabolic capability and reduce cytotoxicity<sup>18,24,25</sup>. The ability to utilize 1,2-propanediol is required for the gastrointestinal pathogenesis of *S. Typhimurium* and other pathogens<sup>19,66</sup>. However, it has remained unclear how the large set of protein components of the Pdu BMC self-recognize and assemble so rapidly and efficiently to form such defined, functional architecture.

Previous attempts using genetic, biochemical, and microscopic analysis, and synthetic biology provided fragmented knowledge about the protein components essential for the assembly and structures of Pdu BMCs<sup>27,39-41,43,67-70</sup>. Here, we developed experimental and analytical tools to systematically study the roles and hierarchical assembly of Pdu BMC proteins in the native host. This study provides the first insight into the defined, stepwise biogenesis pathway of bacterial metabolosomes. First, the Pdu cargo and shell components assembled independently, and then the Pdu metabolosomes formed by a combination of the “Shell first” and “Cargo first” assembly pathways. The ordered consolidation of internal enzymes drives the formation of a liquid-like cargo aggregate to provide the framework for organelle assembly.

## A model for the hierarchical Pdu BMC biogenesis

The ordered assembly of protein complexes has a general significance in many intracellular biological systems, including  $\beta$ -carboxysomes<sup>44,71</sup> and likely the glycol-radical enzyme-associated microcompartments<sup>72</sup>. We propose a new model of Pdu BMC biogenesis based on our findings that successive protein self-assembly and hierarchical association drive the formation of higher-order structural intermediates and entire Pdu BMCs (Fig. 7). The Pdu internal enzymes and shell proteins first self-associate independently to form cargo and shell aggregates. We found that shell proteins formed several assembly intermediates in the cytosol of *S. Typhimurium* LT2 cells in the absence of PduB<sup>1-37</sup>, including empty shell structures (Fig. 3). It has been suggested that PduJ, PduN, PduA, and PduB could first assemble to form scaffolding structures to trigger association of other shell proteins, as PduJ and PduN are essential for formation of the shell and Pdu BMC structures<sup>27</sup>, and PduA, PduB, and PduJ can form nanotube shell structures<sup>67,73,74</sup>.

The enzymatic core of the Pdu BMC is clustered into six distinct types of enzyme complexes. Some internal enzymes, including PduP, PduD, and PduL, feature N-terminal extensions that serve as encapsulation peptides to ensure recognition and loading of internal enzymes<sup>40-42</sup>. The initial assembly of the main PduCDE enzyme complex may create high-affinity binding sites for cargo enzymes through the N-termini of PduL and PduP, to promote the ultimate formation of the enzymatic core near the cell poles (Fig. 6).



**Fig. 7. Model of the Pdu BMC assembly pathway of *S. Typhimurium*.** The Pdu internal enzymes and shell proteins independently self-associate to form cargo and shell aggregates, respectively, at the cell pole. PduB<sup>1-37</sup> binds the shell and cargo aggregates, mainly via PduM that strongly associates with cargo enzymes. PduK is involved in the relocation of the assembled structural intermediates at the cell pole to other cellular positions, to permit the generation of additional Pdu BMCs.

The N-terminal short peptide of the shell protein PduB (PduB<sup>1-37</sup>) plays an essential role in binding the enzymatic core to the shell (Fig. 2 and 3), as suggested earlier<sup>43</sup>. PduB<sup>1-37</sup> mainly interacts with the internal core through PduM; PduM in turn strongly binds with cargo enzymes, likely the N-terminus of PduD that serves as an encapsulation peptide. Both PduB<sup>1-37</sup> and PduM play multiple roles, being important for shell-interior interaction, shell encapsulation and full assembly of intact Pdu BMC structures, reminiscent of the  $\beta$ -carboxysome CcmM and CcmN that serve as the essential linker proteins to bind the shell and Rubisco matrix during  $\beta$ -carboxysome assembly<sup>44</sup>. The fact that orthologs of PduB<sup>1-37</sup>, PduK and PduM were identified in *Klebsiella*, *Escherichia*, *Citrobacter*, and *Salmonella* suggests that the general assembly principle of Pdu metabolosomes that we have identified may be pervasive among these genera (Fig. S10), and might extend to other metabolosomes that have multiple interior enzymes possessing structurally-similar encapsulation peptides.

Although diverse BMCs share a common shell architecture, the assembly pathways of the distinct types of BMCs differ<sup>75</sup>. The biogenesis of  $\beta$ -carboxysomes adopts a “Cargo first” pathway, with the assistance of chaperones<sup>44-46,48</sup>. The assembled Rubisco enzymes first nucleate by interactions with CcmM to form a preorganized cargo matrix, which is then encapsulated by shell proteins to form the functional carboxysome. In contrast, assembly of  $\alpha$ -carboxysomes may follow the “Shell first” or “concomitant shell–core assembly” pathway, as partially-formed  $\alpha$ -carboxysomes have been visualized *in vivo*<sup>50-52,76</sup>, and empty  $\alpha$ -carboxysome shells have been reconstituted in heterologous hosts<sup>53</sup>. In the present study, both pathways were visualized equally during live-cell confocal imaging (Fig. 5). These results provided the first evidence that both the “Shell-first” and “Cargo-first” assembly pathways are required in Pdu BMC biosynthesis. These findings are supported by the successful construction of empty Pdu BMC shells, and the fact that encapsulation peptides are involved in Pdu BMC assembly<sup>39,41</sup>. During the shell-first and cargo-first events that underpin the formation of mature Pdu BMCs, we hypothesize that shell or cargo proteins begin by forming partially assembled structures, which then dynamically associate with cargo enzymes or shell proteins, respectively. Such a mechanism would be supported by the partially formed  $\alpha$ -carboxysomes that have been observed previously<sup>51</sup>.

#### **De novo biogenesis of Pdu metabolosomes occurs at the cell poles**

We found that generation of the Pdu BMC enzyme and shell assembly intermediates generally occurred at a cellular pole. Consistently, the procarboxysome structures formed during  $\beta$ -carboxysome biogenesis and degradation of inactive  $\beta$ -carboxysomes also occur at the polar regions of cyanobacterial cells<sup>44,77</sup>. Likewise,

$\alpha$ -carboxysomes were observed exclusively at the cell poles of chemoautotrophs when lacking the capacity of functional positioning in the absence of the McdAB system<sup>78</sup>. These observations suggest that the cell pole represents a universal, spatially-confined region for BMC biogenesis in bacterial cells.

The intracellular positioning of BMCs is of physiological importance for cellular metabolism and can ensure equal segregation of BMCs during cell division<sup>79-81</sup>. It has been proposed that specific interactions between carboxysomes and the cytoskeletal protein ParA and McdAB mediate partitioning of both  $\alpha$ - and  $\beta$ -carboxysomes in cyanobacteria and proteobacteria<sup>78,79,82,83</sup>. Our results suggest that PduK is involved in the relocation of the assembled structural intermediates from the cell pole to other cellular positions to generate additional Pdu BMCs. The PduK component of the Pdu BMC could interact with other intracellular elements via its C-terminal flexible extension at the concave side facing outside of the shell, to control the spatial positioning and mobility of Pdu BMCs in bacterial cells.

### **The internal enzyme aggregate within Pdu metabolosomes possesses a liquid-like nature**

Liquid-liquid phase separation (LLPS) has been increasingly recognized as a general mechanism that mediates the formation and organization of protein assemblies and membrane-free organelles<sup>84-86</sup>. LLPS has also been proposed to play important roles in determining the cytoplasmic behavior of bacterial cells, with broad impacts on the metabolism of Enterobacteriaceae<sup>87</sup>. The intriguing features of LLPS involve the local cytosolic accumulation of cargo molecules and the presence of intrinsically-disordered protein peptides. Cargo nucleation and condensation are mediated by specific multivalent weak interactions between the intrinsically-disordered protein peptides and structural proteins<sup>86,88</sup>.

The Pdu metabolosome contains several short N-terminal protein extensions that function as encapsulation peptides for the recruitment of multiple cargo enzyme complexes and represent a model system for the assessment of the LLPS-mediated ordered assembly of protein organelles. We note that the secondary structures of these encapsulation peptides (an amphipathic  $\alpha$ -helix) do not perfectly meet the definition of conventional proteins that undergo LLPS, which often contain large disordered regions of low-complexity, composed of repetitive and biased amino acids<sup>70,89-91</sup>, and might represent a novel class of LLPS systems.

Our data demonstrate the hierarchical assembly of Pdu enzyme complexes, and provide *in vivo* experimental evidence suggesting that the cargo enzymes form a dynamic protein condensate within the Pdu BMC. The organizational dynamic rate of Pdu cargo proteins is lower than that of free GFP in the *Escherichia coli* (*E. coli*) cytosol<sup>92</sup> and that of membrane complexes in thylakoids<sup>93</sup>. We speculate that Pdu metabolosomes leverage the flexible and dynamic internal organization to gain a balance between stable association and fluidity of cargo enzymes. Such a liquid-like nature would enhance enzyme activities of the catabolic pathways, and would permit reorganization of internal proteins and metabolites between the multiple sequestered pathways to modulate metabolism in the fluctuating environments experienced by *Salmonella* bacteria during infection<sup>94</sup>.

A similar liquid-like cargo condensate mediated by linker proteins has been proposed for CO<sub>2</sub> fixation in anabolic BMCs and algal pyrenoids. In  $\alpha$ - and  $\beta$ -carboxysomes, Rubisco enzymes form dynamic interactions with the intrinsically-disordered CsoS2 N-terminus and CcmM35 respectively, to generate the liquid-like condensate matrix of Rubisco within the carboxysome<sup>47,65</sup>. Similarly, within the CO<sub>2</sub>-fixing pyrenoid organelles of eukaryotic algae, the liquid-like state of Rubisco clusters is facilitated by the intrinsically disordered protein EPYC1<sup>95,96</sup>. LLPS-facilitated cargo sorting mechanisms have also been identified recently in plant chloroplasts<sup>97</sup>.

This contextualization of our latest findings with other prokaryotic and eukaryotic organelles leads us to propose that LLPS plays a universal role in the formation of metabolic organelles that span diverse architectures, biogenesis pathways and functions.

## **Methods**

### **Bacterial strains and growth conditions**

The bacterial strains used in this study derived from *S. enterica* subsp *enterica* serovar Typhimurium LT2<sup>98,99</sup>. The rich medium was LB-Lennox medium (10 g·L<sup>-1</sup> tryptone (Appleton Woods), 5 g·L<sup>-1</sup> yeast extract

(Appleton Woods), 5 g·L<sup>-1</sup> sodium chloride, pH 7.0), and the minimal medium was no-carbon-E (NCE) medium<sup>100</sup>. The microcompartment-inducing media (MIM) was NCE medium supplemented with 1 mM MgSO<sub>4</sub>, 0.5% succinate, 50 μM Fe(III) citrate, and 0.6% 1,2-PD (if applicable)<sup>28,43</sup>. The MIM medium does not contain vitamin B<sub>12</sub>, as succinate serves as the carbon source instead of 1,2-PD and vitamin B<sub>12</sub> is not required for *S. Typhimurium* LT2 growth in MIM. All medium components were from Sigma-Aldrich, except where specified.

Intracellular visualization of fluorescently-tagged Pdu BMC proteins was done following growth of the relevant plasmid-carrying strains in MIM containing 0.6% 1,2-PD (MIM+1,2-PD) under aerobic conditions. An overnight LB culture was inoculated 1:100 in 100 μL of MIM in a 2-mL Eppendorf tube in the absence of 1,2-PD (MIM-1,2-PD) shaken horizontally at 220 rpm overnight (Innova 2300 Platform shaker). Unless otherwise specified, 1 μL of this culture was sub-inoculated to 100 μL of MIM in a 2-mL Eppendorf tube, both in the absence and in the presence of 1,2-PD, shaken aerobically at 37°C for 10 h until OD<sub>600</sub> reaching 1.0 to 1.2 (DS-11 Spectrophotometer/Fluorometer Series from DeNovix); for birth event detection, the sub-inoculated culture was shaken aerobically at 37°C for 1 h.

Antibiotics were added to liquid media as required at the following final concentrations: ampicillin at 100 μg·mL<sup>-1</sup>, kanamycin at 50 μg·mL<sup>-1</sup>, and gentamicin at 20 μg·mL<sup>-1</sup> in ddH<sub>2</sub>O, chloramphenicol at 25 μg·mL<sup>-1</sup> in ethanol, and tetracycline at 25 μg·mL<sup>-1</sup> in methanol.

### Construction of chromosomal mutations and fluorescence tagging vectors

LT2- $\Delta pduB^{1-37}$  and LT2-*pduB'* were constructed by genome editing technique developed previously<sup>101</sup>. The PduB<sup>1-37</sup> truncation (LT2- $\Delta pduB^{1-37}$ ), resulting in expression of the N-terminus truncated PduB (namely PduB'), was achieved by deleting the 5' terminal region of *pduB*, without modification of the *pduB'* gene promoter, RBS and starting codon as confirmed by DNA sequencing. LT2-*pduB'* was achieved by replacing the start codon for PduB' by GCT (alanine). The pEMG and pSW-2 plasmids were used. First, DNA fragments (700-800 bp) flanking the chromosome regions of interest were PCR amplified and inserted into the pEMG suicide plasmid by Gibson assembly (NEBuilder HiFi DNA Assembly kit). The pEMG-derivative suicide plasmids were mobilized from *E. coli* S17-1 *λpir* to *S. Typhimurium* by conjugation. *S. Typhimurium* transconjugants clones that have integrated the suicide plasmid were selected on solid M9 minimal medium supplemented with 0.2% of glucose and 50 μg/mL of kanamycin. pSW-2, purified from LT2-WT strain, was then introduced into transconjugants by electroporation. Transformants were selected on LB agar medium supplemented with 20 μg/mL gentamicin and 1 mM of *m*-toluate. Colonies were screened for kanamycin resistance and sensitive clones were tested by specific PCR. pSW-2 was cured from the resulting strains by two passages in LB in the absence of gentamicin.

Other *Salmonella* single mutants were constructed by the gene disruption method established previously<sup>102</sup>. pKD4 and pKD3 plasmid were used as the template for PCR. DNA fragments with antibiotic resistance cassette were PCR amplified and transformed into *S. Typhimurium* harboring pSIM5-*tet*<sup>103</sup> for *lambda* red recombination. P22 transduction was performed to move the individual mutation into a clean genetic background<sup>104</sup>. Then the antibiotic resistance cassette was eliminated using the FLP recombinase expressing plasmid pCP20. Thermo-sensitive pCP20 was cured from the resulting strains by two passages in LB in the absence of ampicillin at 42 °C. Double mutants were constructed following the same protocol but using  $\Delta pduB^{1-37}$  as the background strain instead of WT.

pBAD/*Myc*-His was used as a backbone for construction of visible vectors. First, this plasmid was digested by NcoI and HindIII. Then PCR-amplified DNA fragments of individual *pdu* genes, mCherry, and sfGFP were cloned into the linear vector by Gibson assembly<sup>105</sup>. pXG10-SF containing a constitutive P<sub>LetO-1</sub> promoter was employed as a backbone of expression vectors for complementary experiments<sup>106</sup>. The linear pXG10-SF was obtained by PCR cloning. The coding sequences of PduB (M38A), PduBB', PduM and PduK were cloned into linear pXG10-SF via PCR and Gibson assembly.

Strains and plasmids used in this study are listed in Table S1. A complete list of primers used is in Table S2. All mutants and vectors were verified by PCR and DNA sequencing of PCR-amplified genomic DNA or plasmid sequencing (Fig. S4).

### Bacterial growth assays

Overnight LB cultures were inoculated 1:1000 in 10 mL of LB supplemented with 0.6% 1,2-PD and shaken aerobically in 50 mL Falcon tubes for 6 h at 220 rpm. The cells were washed three times with a mixture of 0.6% 1,2-PD and 1 mM MgSO<sub>4</sub> in the NCE medium and then resuspended in NCE medium (containing 0.6% 1,2-PD; 0.3 mM each of leucine, isoleucine, threonine, and valine; 50 μM ferric citrate; and 20 nM or 150 nM CN-B<sub>12</sub>) to an OD<sub>600</sub> of 0.15. At 20 nM of CN-B<sub>12</sub> (a limiting level of vitamin B<sub>12</sub>), 200 μL of the culture at OD<sub>600</sub> = 0.15 was grown aerobically at 37°C with intermittent shaking in a microplate reader (FLUOstar Omega, BMG LABTECH).

At 150 nM CN-B<sub>12</sub> (saturating level of vitamin B<sub>12</sub>), 250 μL of the culture at OD<sub>600</sub> = 0.15 was grown aerobically at 37°C with intermittent rapid shaking, using the System Duetz technology platform (Growth Profiler 960, EnzyScreen). The OD<sub>600</sub> readings were taken hourly. At least three biological replicates of each growth curve were obtained.

### **Pdu BMC purification**

The Pdu BMCs from both *S. Typhimurium* LT2-WT and LT2- $\Delta$ pduK were isolated by detergent treatment and differential centrifugation<sup>28</sup>. Briefly, 400 mL cells grown in MIM (OD<sub>600</sub> = 1.0-1.2) were harvested and washed with buffer A (50 mM Tris-HCl pH 8.0, 500 mM KCl, 12.5 mM MgCl<sub>2</sub>, and 1.5% 1,2-PD), and lysed with the bacteria-specific reagent (BPER-II). Subsequently, Pdu BMCs were separated from cell debris by sequential centrifuge steps (12,000 g to pellet cell debris, 20,000 g to pellet Pdu BMCs). The isolated Pdu BMCs were resuspended in buffer B (50 mM Tris-HCl pH 8.0, 50 mM KCl, 5 mM MgCl<sub>2</sub>, 1% 1,2-PD) containing protease inhibitor cocktail (Sigma-Aldrich).

### **SDS-PAGE analysis**

Standard procedures for sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) were employed. Proteins were separated by 15% polyacrylamide gels and stained with Coomassie Brilliant Blue G-250.

### **Transmission electron microscopy**

The *S. Typhimurium* LT2 WT and mutant cells were characterized using thin-section EM<sup>80</sup>. *Salmonella* cells were pelleted by centrifugation (6,000 g, 10 mins) and processed for thin section using a Pelco BioWave Pro laboratory microwave system. The cells are first fixed with 0.1 M sodium cacodylate buffer (pH 7.2) supplemented with 2.5% glutaraldehyde using two steps of 100W for 1 minute each (P1). Samples were then embedded in 4% agarose, followed by staining with 2% osmium tetroxide and 3% Potassium Ferrocyanide using three steps of 100W for 20 seconds each (P2). The reduced osmium stain was then set using a solution of 1% Thiocarbohydrazide for 10 minutes. The second osmium stain was applied using P2 with 2% osmium tetroxide. The sample was made electron dense with 2% Uranyl Acetate incubated at 4°C overnight. Dehydration was operated with a series of increasing alcohol concentrations (30 to 100%) before cells embedding in medium resin. Thin sections of 70 nm were cut with a diamond knife. The structures of purified Pdu MCPs were characterized using negative staining EM as described previously<sup>45,107-109</sup>. Images were recorded on an FEI 120Kv Tecnai G2 Spirit BioTWIN transmission electron microscope equipped with a Gatan Rio 16 camera.

### **Confocal microscopy**

The cells were imaged using a Zeiss LSM780 confocal microscopy with a 63 × oil-immersion objective excitation at 488 nm and/or 561 nm. GFP and mCherry fluorescence were detected at 500-520 nm and 660-700 nm, respectively. Images were captured as 512 x 512 pixels at 16 bits by Zeiss Zen 2010 software. The pinhole was set to ~1 μm to make sure that all fluorescence signals of the *Salmonella* cells were recorded. The temperature was controlled at 37°C during the whole imaging process. All images were captured from at least three different cultures. Colocalization analysis was performed by using the Coloc2 plugin in ImageJ to generate Pearson's correlation coefficient R and scatterplot<sup>110</sup>. The values of Pearson's R close to 1 reflect reliable colocalization, whereas the values near zero indicate distributions of fluorescence uncorrelated with one other<sup>111</sup>. Scatterplots are generated by plotting the intensity value of each pixel of mCherry along the x-axis and the intensity value of the same pixel location of sfGFP on the y-axis using Coloc2 plugins in ImageJ. The scatterplots describe the relationship between the fluorescent signals. If the dots on the diagram appearing as a cloud clustered on a line indicates a strong colocalization, and Pearson's R is close to 1 in this case. If the scattered distributions of the pixels close to both axes indicate a mutual exclusion, and Pearson's R is near zero. Fluorescence intensity profiles were generated using ImageJ software.

To monitor the progression of shell and cargo formation of Pdu BMCs, overnight MIM (- 1,2-PD) culture was diluted 1:100 in MIM (+ 1,2-PD) pre-incubated at 37°C and 10 µL of this culture was dropped onto MIM agar plate (+ 1,2-PD) and left to dry at 37°C. The MIM agar with cell patches was cut out and attached to a 0.17 mm glass coverslip, followed by sandwiched within a 35mm glass-bottom dish. The temperature of the microscope system was set to 37°C for cell growth. The images were taken at 0 h when the cells started to grow in the presence of 1,2-PD, at 1 h when the shell and cargo started to assemble, and at 1.5 h for further tracking.

### Fluorescence recovery after photobleaching

$\Delta pduN/pBAD-pduE-sfGFP$  and  $\Delta pduN/pBAD-pduA-sfGFP$  cells in stationary phase were diluted to 0.2 to 0.4 OD<sub>600 nm</sub>, and 2 µL was spotted onto a 2% low-melting agarose ampicillin plate that was preincubated at 37°C. After cells were dried, the pad was converted onto a glass-bottom dish (35 mm), which was covered by a cover glass. Then cells were imaged with a Zeiss LSM780 confocal microscopy. 100% laser power confocal laser was used to bleach a line across the center of the cell. Images were taken every 1 min for 60 min to ensure the stationary of fluorescence intensity. FRAP data were analyzed following the previous procedure<sup>93,112</sup>. The fluorescent profiles within a cell were obtained by ImageJ software and were normalized to the same total fluorescence to compare fluorescence distributions before and after bleaching. The mobile proportion (M) was given by

$$M = \frac{(\text{Final fluorescence}) - (\text{Postbleach fluorescence})}{(\text{Scaled prebleach fluorescence}) - (\text{Postbleach fluorescence})}$$

Fluorescence profiles along the long axis of the cell were extracted and normalized to the same total fluorescence against that of the prebleached cell. This normalization aimed to minimize the fluctuations caused by the microscopy and the decrease of fluorescence intensity caused by laser scanning. Then we subtracted the scaled pre-bleach profile from the post-bleach profile to give a different profile, which allowed us to know the final fluorescence intensity. The different profile was plotted versus cell length to find the minimum value (usually the center of the bleached area in the x-position). The fluorescence profile values at the center of the bleached area were plotted against time to give the experimental fluorescence recovery curve. The single exponential function is given by

$$f(t) = A(1 - e^{-\tau t})$$

Half time is given by

$$\tau_{1/2} = \frac{\ln 0.5}{-\tau}$$

Based on the post-bleaching and final fluorescence profiles, it is possible to simulate the evolution of the fluorescence profile. This was done by running the experimental different profile in a home-made and iterative computer routine on SigmaPlot14 software, assuming an arbitrary diffusion coefficient for random diffusion<sup>113,114</sup>. The routine can predict how the fluorescence profile will evolve in small time increments, based on the basic diffusion equation:

$$\frac{dC}{dt} = D \frac{\delta^2 C}{\delta X^2}$$

The routine estimates the value of  $\delta^2(C)/\delta(X^2)$  at each point by calculating the difference between points in the profile and its two neighbors. This value was then multiplied by a factor related to the diffusion coefficient, and was used to predict the increment in C at that point. Where the mobile fraction is less than 1, the routine calculates the "mobile profile" and predicts how it will evolve. Then a set of predicted fluorescence profiles were generated to estimate the diffusion coefficient. Once the simulated fluorescence curve was overlapped with the experimentally observed fluorescence recovery curve, the diffusion coefficient of the simulation was recorded.

### Bioinformatic analysis and structural prediction

To identify the bacterial genomes that harbor the Pdu operon, the protein sequence of PduA was queried against the complete genomes ( $n = 23,521$ ) of the Refseq bacterial database (<https://ftp.ncbi.nlm.nih.gov/genomes/refseq/bacteria/>, accessed on 24/06/2021) using tBLASTn v2.5.0+<sup>115</sup>. Four bacterial genera were found to carry PduA gene: *Citrobacter*, *Escherichia*, *Salmonella* and *Klebsiella*. From the four genera, 61 strains were selected to represent different species or sub-species (see Source Data). The selected genome sequences were downloaded from Refseq database. The protein sequences of the

genomes were produced using Prokka v1.14.6<sup>116</sup>. The Pdu proteins of *S. Typhimurium* strain LT2 were used as references, including PduA, PduB, PduD, PduJ, PduK, PduM, PduN, PduU, and PduX. The Pdu protein sequences were queried against the proteomes using BLASTp v2.5.0+<sup>115</sup> with a threshold of 90% amino acid identity. The similarity of the proteins was summarized from the BLAST results.

To build the phylogenetic tree of the 61 bacterial genomes, the universal single-copy orthologs of Gammaproteobacteria were obtained from each genome using BUSCO v5.2.2<sup>117</sup>. *Vibrio cholerae* MS6 was used as an outgroup to root the phylogenetic tree. From the BUSCO result, 327 genes were found to be present in all the genomes. Each gene set was aligned with Mafft v7.475<sup>118</sup>. The alignments were concatenated with SeqKit v0.15.0<sup>119</sup>, then trimmed with Trimal v1.4<sup>120</sup> with the automatic method. A phylogenetic tree was constructed from the alignment with Fasttree v2.1.10<sup>121</sup> using the JTT+Gamma model.

The protein sequences of PduB1-37, PduM, and PduK were extracted from *S. Typhimurium* LT2, *Klebsiella pneumoniae* HS11286, *Escherichia coli* O104:H21 strain ATCC\_BAA-178, and *Citrobacter freundii* ATCC\_8090 as representatives of each genus. The structure of each protein was predicted by AlphaFold2<sup>122</sup>, accessed via ColabFold<sup>123</sup>. The structures were visualized in ChimeraX<sup>124</sup>. The protein structures were colored by the conservation value of each residue in alignment with the proteins from the 61 isolates. The conservation values were automatically calculated by ChimeraX using the entropy-based measure from AL2CO<sup>125</sup>.

## Acknowledgments

We thank the Liverpool Centre for Cell Imaging and Biomedical Electron Microscopy Unit for providing technical assistance and provision for microscopic imaging. This work was supported by the Royal Society (URF\R\180030, UF120411, RGF\EA\181061 and RGF\EA\180233 to LNL), the Biotechnology and Biological Sciences Research Council (BBSRC) (BB/V009729/1, BB/M024202/1 and BB/R003890/1 to LNL), the China Scholarship Council and, in part, by a Wellcome Trust Senior Investigator award [grant number 106914/Z/15/Z] to JH. For the purpose of open access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

## Author contributions

M.Y., N.W., G.D., Y.L., X.Z., F.H., and Y.S. performed research; M.Y. and L.-N.L. analyzed data; M.Y., J.C.D.H, and L.-N.L. designed research; M.Y., J.C.D.H, and L.-N.L. wrote the article with contributions from other authors.

## Competing interests

The authors declare no competing interests.

## References

1. Chen AH, Silver PA. Designing biological compartmentalization. *Trends in cell biology* **22**, 662-670 (2012).
2. Yeates TO, Jorda J, Bobik TA. The shells of BMC-type microcompartment organelles in bacteria. *Journal of molecular microbiology and biotechnology* **23**, 290-299 (2013).
3. Kerfeld CA, Aussignargues C, Zarzycki J, Cai F, Sutter M. Bacterial microcompartments. *Nat Rev Microbiol* **16**, 277-290 (2018).
4. Liu LN, Yang M, Sun Y, Yang J. Protein stoichiometry, structural plasticity and regulation of bacterial microcompartments. *Curr Opin Microbiol* **63**, 133-141 (2021).
5. Kerfeld CA, *et al.* Protein structures forming the shell of primitive bacterial organelles. *Science* **309**, 936-938 (2005).
6. Chowdhury C, *et al.* Selective molecular transport through the protein shell of a bacterial microcompartment organelle. *Proceedings of the National Academy of Sciences* **112**, 2990-2995 (2015).
7. Crowley CS, Sawaya MR, Bobik TA, Yeates TO. Structure of the PduU shell protein from the Pdu microcompartment of Salmonella. *Structure* **16**, 1324-1332 (2008).
8. Faulkner M, *et al.* Molecular simulations unravel the molecular principles that mediate selective permeability of carboxysome shell protein. *bioRxiv*, (2020).

9. Axen SD, Erbilgin O, Kerfeld CA. A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. *PLoS Comput Biol* **10**, e1003898 (2014).
10. Chowdhury C, Sinha S, Chun S, Yeates TO, Bobik TA. Diverse bacterial microcompartment organelles. *Microbiology and Molecular Biology Reviews* **78**, 438-468 (2014).
11. Hennacy JH, Jonikas MC. Prospects for Engineering Biophysical CO<sub>2</sub> Concentrating Mechanisms into Land Plants to Enhance Yields. *Annu Rev Plant Biol* **71**, 461-485 (2020).
12. Rae BD, *et al.* Progress and challenges of engineering a biophysical CO<sub>2</sub>-concentrating mechanism into higher plants. *J Exp Bot* **68**, 3717-3737 (2017).
13. Liu LN. Advances in the bacterial organelles for CO<sub>2</sub> fixation. *Trends in Microbiology* <https://doi.org/10.1016/j.tim.2021.10.004>, (2021).
14. Taylor SJ, Winter SE. *Salmonella* finds a way: Metabolic versatility of *Salmonella enterica* serovar Typhimurium in diverse host environments. *PLoS Pathog* **16**, e1008540 (2020).
15. Bobik TA, Havemann GD, Busch RJ, Williams DS, Aldrich HC. The Propanediol Utilization (pdu) Operon of *Salmonella enterica* Serovar Typhimurium LT2 Includes Genes Necessary for Formation of Polyhedral Organelles Involved in Coenzyme B12-Dependent 1, 2-Propanediol Degradation. *Journal of bacteriology* **181**, 5967-5975 (1999).
16. Stewart AM, Stewart KL, Yeates TO, Bobik TA. Advances in the World of Bacterial Microcompartments. *Trends Biochem Sci*, (2021).
17. Lee MJ, Palmer DJ, Warren MJ. Biotechnological Advances in Bacterial Microcompartment Technology. *Trends Biotechnol* **37**, 325-336 (2019).
18. Liu LN. Bacterial metabolosomes: new insights into their structure and bioengineering. *Microbial Biotechnology* **14**, 88-93 (2021).
19. Faber F, *et al.* Respiration of Microbiota-Derived 1,2-propanediol Drives *Salmonella* Expansion during Colitis. *PLoS Pathog* **13**, e1006129 (2017).
20. Faber F, *et al.* Respiration of microbiota-derived 1, 2-propanediol drives *Salmonella* expansion during colitis. *PLoS pathogens* **13**, e1006129 (2017).
21. Price-Carter M, Tingey J, Bobik TA, Roth JR. The Alternative Electron Acceptor Tetrathionate Supports B12-Dependent Anaerobic Growth of *Salmonella enterica* Serovar Typhimurium on Ethanolamine or 1, 2-Propanediol. *Journal of bacteriology* **183**, 2463-2475 (2001).
22. Jeter RM. Cobalamin-dependent 1, 2-propanediol utilization by *Salmonella typhimurium*. *Microbiology* **136**, 887-896 (1990).
23. Cheng S, Liu Y, Crowley CS, Yeates TO, Bobik TA. Bacterial microcompartments: their properties and paradoxes. *Bioessays* **30**, 1084-1095 (2008).
24. Yang M, *et al.* Decoding the stoichiometric composition and organisation of bacterial metabolosomes. *Nature Communications* **11**, 1-11 (2020).
25. Stewart KL, Stewart AM, Bobik TA. Prokaryotic Organelles: Bacterial Microcompartments in *E. coli* and *Salmonella*. *EcoSal Plus* **9**, (2020).
26. Sutter M, *et al.* Visualization of bacterial microcompartment facet assembly using high-speed atomic force microscopy. *Nano Letters* **16**, 1590-1595 (2016).
27. Cheng S, Sinha S, Fan C, Liu Y, Bobik TA. Genetic analysis of the protein shell of the microcompartments involved in coenzyme B12-dependent 1, 2-propanediol degradation by *Salmonella*. *Journal of bacteriology* **193**, 1385-1392 (2011).
28. Sinha S, Cheng S, Fan C, Bobik TA. The PduM protein is a structural component of the microcompartments involved in coenzyme B12-dependent 1, 2-propanediol degradation by *Salmonella enterica*. *Journal of bacteriology* **194**, 1912-1918 (2012).
29. Havemann GD, Bobik TA. Protein content of polyhedral organelles involved in coenzyme B12-dependent degradation of 1, 2-propanediol in *Salmonella enterica* serovar Typhimurium LT2. *Journal of bacteriology* **185**, 5086-5095 (2003).
30. Wheatley NM, Gidaniyan SD, Liu Y, Cascio D, Yeates TO. Bacterial microcompartment shells of diverse functional types possess pentameric vertex proteins. *Protein Science* **22**, 660-665 (2013).
31. Kennedy NW, Ikonomova SP, Lee MS, Raeder HW, Tullman-Ercek D. Self-assembling shell proteins PduA and PduJ have essential and redundant roles in bacterial microcompartment assembly. *Journal of Molecular Biology* **433**, 166721 (2021).
32. Liu Y, Leal NA, Sampson EM, Johnson CL, Havemann GD, Bobik TA. PduL is an evolutionarily distinct phosphotransacylase involved in B12-dependent 1, 2-propanediol degradation by *Salmonella enterica* serovar Typhimurium LT2. *Journal of bacteriology* **189**, 1589-1596 (2007).
33. Leal NA, Havemann GD, Bobik TA. PduP is a coenzyme-a-acylating propionaldehyde dehydrogenase associated with the polyhedral bodies involved in B12-dependent 1, 2-propanediol degradation by *Salmonella enterica* serovar Typhimurium LT2. *Archives of microbiology* **180**, 353-361 (2003).
34. Cheng S, Fan C, Sinha S, Bobik TA. The PduQ enzyme is an alcohol dehydrogenase used to recycle NAD<sup>+</sup> internally within the Pdu microcompartment of *Salmonella enterica*. *PLoS One* **7**, e47144 (2012).

35. Palacios S, Starai VJ, Escalante-Semerena JC. Propionyl coenzyme A is a common intermediate in the 1, 2-propanediol and propionate catabolic pathways needed for expression of the prpBCDE operon during growth of *Salmonella enterica* on 1, 2-propanediol. *Journal of bacteriology* **185**, 2802-2810 (2003).
36. Sampson EM, Johnson CL, Bobik TA. Biochemical evidence that the pduS gene encodes a bifunctional cobalamin reductase. *Microbiology* **151**, 1169-1177 (2005).
37. Maurice MS, Mera PE, Taranto MP, Sesma F, Escalante-Semerena JC, Rayment I. Structural characterization of the active site of the PduO-type ATP: Co (I) rrinoid adenosyltransferase from *Lactobacillus reuteri*. *Journal of Biological Chemistry* **282**, 2596-2605 (2007).
38. Fan C, Bobik TA. The PduX enzyme of *Salmonella enterica* is an L-threonine kinase used for coenzyme B12 synthesis. *Journal of Biological Chemistry* **283**, 11322-11329 (2008).
39. Parsons JB, *et al.* Synthesis of empty bacterial microcompartments, directed organelle protein incorporation, and evidence of filament-associated organelle movement. *Molecular cell* **38**, 305-315 (2010).
40. Fan C, *et al.* Short N-terminal sequences package proteins into bacterial microcompartments. *Proceedings of the National Academy of Sciences* **107**, 7509-7514 (2010).
41. Fan C, Bobik TA. The N-terminal region of the medium subunit (PduD) packages adenosylcobalamin-dependent diol dehydratase (PduCDE) into the Pdu microcompartment. *Journal of bacteriology* **193**, 5623-5628 (2011).
42. Liu Y, Jorda J, Yeates TO, Bobik TA. The PduL phosphotransacylase is used to recycle coenzyme A within the Pdu microcompartment. *Journal of bacteriology* **197**, 2392-2399 (2015).
43. Lehman BP, Chowdhury C, Bobik TA. The N terminus of the PduB protein binds the protein shell of the Pdu microcompartment to its enzymatic core. *Journal of bacteriology* **199**, e00785-00716 (2017).
44. Cameron JC, Wilson SC, Bernstein SL, Kerfeld CA. Biogenesis of a bacterial organelle: the carboxysome assembly pathway. *Cell* **155**, 1131-1140 (2013).
45. Huang F, *et al.* Rubisco accumulation factor 1 (Raf1) plays essential roles in mediating Rubisco assembly and carboxysome biogenesis. *Proc Natl Acad Sci USA* **117**, 17418-17428 (2020).
46. Huang F, *et al.* Roles of RbcX in carboxysome biosynthesis in the cyanobacterium *Synechococcus elongatus* PCC7942. *Plant Physiol* **179**, 184-194 (2019).
47. Wang H, *et al.* Rubisco condensate formation by CcmM in beta-carboxysome biogenesis. *Nature* **566**, 131-135 (2019).
48. Chen AH, Robinson-Mosher A, Savage DF, Silver PA, Polka JK. The bacterial carbon-fixing organelle is formed by shell envelopment of preassembled cargo. *PloS one* **8**, (2013).
49. Sun Y, Wollman AJM, Huang F, Leake MC, Liu LN. Single-organelle quantification reveals the stoichiometric and structural variability of carboxysomes dependent on the environment. *Plant Cell* **31**, 1648-1664 (2019).
50. Menon BB, Dou Z, Heinhorst S, Shively JM, Cannon GC. Halothiobacillus neapolitanus carboxysomes sequester heterologous and chimeric RubisCO species. *PLoS One* **3**, e3570 (2008).
51. Iancu CV, Morris DM, Dou Z, Heinhorst S, Cannon GC, Jensen GJ. Organization, structure, and assembly of  $\alpha$ -carboxysomes determined by electron cryotomography of intact cells. *Journal of molecular biology* **396**, 105-117 (2010).
52. Dai W, *et al.* Visualizing Individual RuBisCO and Its Assembly into Carboxysomes in Marine Cyanobacteria by Cryo-Electron Tomography. *J Mol Biol* **430**, 4156-4167 (2018).
53. Li T, *et al.* Reprogramming bacterial protein organelles as a nanoreactor for hydrogen production. *Nature Communications* **11**, 5448 (2020).
54. Ailion M, Bobik TA, Roth JR. Two global regulatory systems (Crp and Arc) control the cobalamin/propanediol regulon of *Salmonella typhimurium*. *J Bacteriol* **175**, 7200-7208 (1993).
55. Chen P, Ailion M, Bobik T, Stormo G, Roth J. Five promoters integrate control of the cob/pdu regulon in *Salmonella typhimurium*. *J Bacteriol* **177**, 5401-5410 (1995).
56. Guzman L-M, Belin D, Carson MJ, Beckwith J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *Journal of Bacteriology* **177**, 4121-4130 (1995).
57. Lee N, Francklyn C, Hamilton EP. Arabinose-induced binding of AraC protein to araI2 activates the araBAD operon promoter. *Proceedings of the National Academy of Sciences* **84**, 8814-8818 (1987).
58. Kim EY, Jakobson CM, Tullman-Ercek D. Engineering transcriptional regulation to control Pdu microcompartment formation. *PloS one* **9**, e113814 (2014).
59. Meisner J, Goldberg JB. The *Escherichia coli* rhaSR-PrhaBAD Inducible Promoter System Allows Tightly Controlled Gene Expression over a Wide Range in *Pseudomonas aeruginosa*. *Appl Environ Microbiol* **82**, 6715-6727 (2016).
60. Sutter M, Wilson SC, Deutsch S, Kerfeld CA. Two new high-resolution crystal structures of carboxysome pentamer proteins reveal high structural conservation of CcmL orthologs among distantly related cyanobacterial species. *Photosynth Res* **118**, 9-16 (2013).
61. Jakobson CM, Tullman-Ercek D, Slininger MF, Mangan NM. A systems-level model reveals that 1,2-Propanediol utilization microcompartments enhance pathway flux through intermediate sequestration. *PLoS Comput Biol* **13**, e1005525 (2017).
62. Havemann GD, Sampson EM, Bobik TA. PduA is a shell protein of polyhedral organelles involved in coenzyme B12-dependent degradation of 1, 2-propanediol in *Salmonella enterica* serovar Typhimurium LT2. *Journal of Bacteriology* **184**, 1253-1261 (2002).

63. Parsons JB, Lawrence AD, McLean KJ, Munro AW, Rigby SE, Warren MJ. Characterisation of PduS, the pdu metabolosome corrin reductase, and evidence of substructural organisation within the bacterial microcompartment. *PLoS One* **5**, e14009 (2010).
64. Mullineaux CW, Sarcina M. Probing the dynamics of photosynthetic membranes with fluorescence recovery after photobleaching. *Trends in plant science* **7**, 237-240 (2002).
65. Oltrogge LM, Chaijarasphong T, Chen AW, Bolin ER, Marqusee S, Savage DF. Multivalent interactions between CsoS2 and Rubisco mediate alpha-carboxysome formation. *Nat Struct Mol Biol* **27**, 281-287 (2020).
66. Connolly JPR, *et al.* Host-associated niche metabolism controls enteric infection through fine-tuning the regulation of type 3 secretion. *Nat Commun* **9**, 4187 (2018).
67. Kennedy NW, Ikonomova SP, Slininger Lee M, Raeder HW, Tullman-Ercek D. Self-assembling Shell Proteins PduA and PduJ have Essential and Redundant Roles in Bacterial Microcompartment Assembly. *J Mol Biol* **433**, 166721 (2021).
68. Wade Y, Daniel RA, Leak DJ. Heterologous Microcompartment Assembly in Bacillaceae: Establishing the Components Necessary for Scaffold Formation. *ACS Synth Biol* **8**, 1642-1654 (2019).
69. Kennedy NW, Ikonomova SP, Lee MS, Raeder HW, Tullman-Ercek D. Self-assembling shell proteins PduA and PduJ have essential and redundant roles in bacterial microcompartment assembly. *bioRxiv*, (2020).
70. Juodeikis R, *et al.* Effect of metabolosome encapsulation peptides on enzyme activity, coaggregation, incorporation, and bacterial microcompartment formation. *MicrobiologyOpen* **9**, e1010 (2020).
71. Marsh JA, *et al.* Protein complexes are under evolutionary selection to assemble via ordered pathways. *Cell* **153**, 461-470 (2013).
72. Kalnins G, Cesle EE, Jansons J, Liepins J, Filimonenko A, Tars K. Encapsulation mechanisms and structural studies of GRM2 bacterial microcompartment particles. *Nat Commun* **11**, 388 (2020).
73. Uddin I, Frank S, Warren MJ, Pickersgill RW. A generic self-assembly process in microcompartments and synthetic protein nanotubes. *Small* **14**, e1704020 (2018).
74. Pang A, Frank S, Brown I, Warren MJ, Pickersgill RW. Structural insights into higher order assembly and function of the bacterial microcompartment protein PduA. *J Biol Chem* **289**, 22377-22384 (2014).
75. Kerfeld CA, Erbilgin O. Bacterial microcompartments and the modular construction of microbial metabolism. *Trends in microbiology* **23**, 22-34 (2015).
76. Cai F, *et al.* Advances in understanding carboxysome assembly in Prochlorococcus and Synechococcus implicate CsoS2 as a critical component. *Life* **5**, 1141-1171 (2015).
77. Hill NC, Tay JW, Altus S, Bortz DM, Cameron JC. Life cycle of a cyanobacterial carboxysome. *Sci Adv* **6**, eaba1269 (2020).
78. MacCready JS, Tran L, Basalla JL, Hakim P, Vecchiarelli AG. The McdAB system positions  $\alpha$ -carboxysomes in proteobacteria. *Mol Microbiol* **16**, 277-297 (2021).
79. Savage DF, Afonso B, Chen AH, Silver PA. Spatially ordered dynamics of the bacterial carbon fixation machinery. *Science* **327**, 1258-1261 (2010).
80. Sun Y, *et al.* Light modulates the biosynthesis and organization of cyanobacterial carbon fixation machinery through photosynthetic electron flow. *Plant Physiol* **171**, 530-541 (2016).
81. Rillema R, MacCready JS, Vecchiarelli AG. Cyanobacterial growth and morphology are influenced by carboxysome positioning and temperature. *bioRxiv*, 2020.2006.2001.127845 (2020).
82. MacCready JS, *et al.* Protein gradients on the nucleoid position the carbon-fixing organelles of cyanobacteria. *Elife* **7**, e39723 (2018).
83. MacCready JS, Basalla JL, Vecchiarelli AG. Origin and Evolution of Carboxysome Positioning Systems in Cyanobacteria. *Mol Biol Evol* **37**, 1434-1451 (2020).
84. Shin Y, Brangwynne CP. Liquid phase condensation in cell physiology and disease. *Science* **357**, (2017).
85. Greening C, Lithgow T. Formation and function of bacterial organelles. *Nature Reviews Microbiology* **18**, 677-689 (2020).
86. Alberti S, Gladfelter A, Mittag T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419-434 (2019).
87. Parry BR, Surovtsev IV, Cabeen MT, O'Hern CS, Dufresne ER, Jacobs-Wagner C. The bacterial cytoplasm has glass-like properties and is fluidized by metabolic activity. *Cell* **156**, 183-194 (2014).
88. Li P, *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336-340 (2012).
89. Lee MJ, Brown IR, Juodeikis R, Frank S, Warren MJ. Employing bacterial microcompartment technology to engineer a shell-free enzyme-aggregate for enhanced 1, 2-propanediol production in Escherichia coli. *Metabolic engineering* **36**, 48-56 (2016).
90. Erbilgin O, Sutter M, Kerfeld CA. The structural basis of coenzyme A recycling in a bacterial organelle. *PLoS biology* **14**, e1002399 (2016).
91. Jakobson CM, Slininger Lee MF, Tullman-Ercek D. De novo design of signal sequences to localize cargo to the 1, 2-propanediol utilization microcompartment. *Protein Science* **26**, 1086-1092 (2017).
92. Elowitz MB, Surette MG, Wolf P-E, Stock JB, Leibler S. Protein Mobility in the Cytoplasm of Escherichia coli. *Journal of bacteriology* **181**, 197-203 (1999).
93. Casella S, *et al.* Dissecting the native architecture and dynamics of cyanobacterial photosynthetic machinery. *Molecular plant* **10**, 1434-1448 (2017).

94. Kroger C, *et al.* An infection-relevant transcriptomic compendium for *Salmonella enterica* Serovar Typhimurium. *Cell Host Microbe* **14**, 683-695 (2013).
95. Freeman Rosenzweig ES, *et al.* The Eukaryotic CO<sub>2</sub>-Concentrating Organelle Is Liquid-like and Exhibits Dynamic Reorganization. *Cell* **171**, 148-162 e119 (2017).
96. He S, *et al.* The structural basis of Rubisco phase separation in the pyrenoid. *Nat Plants* **6**, 1480-1490 (2020).
97. Ouyang M, *et al.* Liquid-Liquid Phase Transition Drives Intra-chloroplast Cargo Sorting. *Cell* **180**, 1144-1159 e1120 (2020).
98. McClelland M, *et al.* Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* **413**, 852-856 (2001).
99. Zinder ND, Lederberg J. Genetic exchange in *Salmonella*. *J Bacteriol* **64**, 679-699 (1952).
100. Vogel HJ, Bonner DM. Acetylornithinase of *Escherichia coli*: partial purification and some properties. *Journal of Biological Chemistry* **218**, 97 (1956).
101. Martínez-García E, de Lorenzo V. Engineering multiple genomic deletions in Gram-negative bacteria: analysis of the multi-resistant antibiotic profile of *Pseudomonas putida* KT2440. *Environmental microbiology* **13**, 2702-2716 (2011).
102. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 6640 (2000).
103. Koskiniemi S, Pránting M, Gullberg E, Näsval J, Andersson DI. Activation of cryptic aminoglycoside resistance in *Salmonella enterica*. *Molecular Microbiology* **80**, 1464-1478 (2011).
104. Schmieger H. A method for detection of phage mutants with altered transducing ability. *Molecular and General Genetics MGG* **110**, 378-381 (1971).
105. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* **6**, 343-345 (2009).
106. Corcoran CP, Podkaminski D, Papenfort K, Urban JH, Hinton JC, Vogel J. Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. *Molecular microbiology* **84**, 428-445 (2012).
107. Faulkner M, *et al.* Direct characterization of the native structure and mechanics of cyanobacterial carboxysomes. *Nanoscale* **9**, 10662-10673 (2017).
108. Fang Y, *et al.* Engineering and modulating functional cyanobacterial CO<sub>2</sub>-fixing organelles. *Front Plant Sci* **9**, 739 (2018).
109. Chen T, *et al.* Incorporation of Functional Rubisco Activases into Engineered Carboxysomes to Enhance Carbon Fixation. *ACS Synth Biol*, <https://doi.org/10.1021/acssynbio.1021c00311> (2021).
110. Bolte S, Cordelières FP. A guided tour into subcellular colocalization analysis in light microscopy. *Journal of microscopy* **224**, 213-232 (2006).
111. Dunn KW, Kamocka MM, McDonald JH. A practical guide to evaluating colocalization in biological microscopy. *American Journal of Physiology-Cell Physiology*, (2011).
112. Mullineaux CW, Tobin MJ, Jones GR. Mobility of photosynthetic complexes in thylakoid membranes. *Nature* **390**, 421-424 (1997).
113. Kirchhoff H, Haferkamp S, Allen JF, Epstein DB, Mullineaux CW. Protein diffusion and macromolecular crowding in thylakoid membranes. *Plant physiology* **146**, 1571-1578 (2008).
114. Casella S, *et al.* Dissecting the native architecture and dynamics of cyanobacterial photosynthetic machinery. *Mol Plant* **10**, 1434-1448 (2017).
115. Camacho C, *et al.* BLAST+: architecture and applications. **10**, 421-421 (2009).
116. Seemann T. Prokka: rapid prokaryotic genome annotation. **30**, 2068-2069 (2014).
117. Manni M, Berkeley MR, Seppely M, Simão FA, Zdobnov EM. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. **38**, 4647-4654 (2021).
118. Katoh K, Misawa K, Kuma Ki, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. **30**, 3059-3066 (2002).
119. Shen W, Le S, Li Y, Hu F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. **11**, 1-10 (2016).
120. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. **25**, 1972-1973 (2009).
121. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. **5**, e9490-e9490 (2010).
122. Jumper J, *et al.* Highly accurate protein structure prediction with AlphaFold. **596**, 583-589 (2021).
123. Mirdita M, Ovchinnikov S, Steinegger M. ColabFold-Making protein folding accessible to all. (2021).
124. Pettersen EF, *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science* **30**, 70-82 (2021).
125. Pei J, Grishin NV. AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics* **17**, 700-712 (2001).