Supporting Information

Fan et al. 10.1073/pnas.0913199107

SI Materials and Methods

General Molecular Methods. Plasmid DNA was purified using QIAGEN products according to the manufacturer's instructions. Following PCR amplification, DNA was purified by Promega Wizard PCR Preps. Restriction digests were carried out using standard protocols. For ligation of DNA fragments, T4 DNA ligase was used according to the manufacturer's instructions (New England Biolabs).

Western Blots. Ten micrograms of purified MCPs was loaded in each lane on a 10–20% gradient SDS/PAGE gel (Bio-Rad). Proteins were transferred to a nitrocellulose membrane and detected by a primary antibody from rabbit and goat anti-rabbit conjugated to alkaline phosphatase or HRP as a secondary antibody (Bio-Rad). Chromogenic developing agents were used following the manufacturer's instructions (Bio-Rad).

MCP Purification. Strains were grown in 400 mL of no-carbon-E medium supplemented with 1 mM $MgSO_4$, 0.5% succinate, and 0.6% 1,2-PD. The inoculum was a 1-mL LB culture, and incubation was at 37 °C with shaking at 275 rpm. After cultures reached optical density at 600 nm between 1 and 1.2, cells were harvested by centrifugation. MCP purification was described previously (1). Purified MCPs were stored at 4 °C before analysis.

PduP Enzyme Assays. Assays were carried out as previously described (2). Reaction mixtures contained 50 mM N-cyclohexyl-2-aminoethanesulfonic acid (pH 9.5), 10 mM propionaldehyde, 1 mM DTT, 2 mM NAD^+, 500 μM coenzyme A, and an appropriate amount of protein. Assays were incubated at 37 °C. Enzyme activity was measured by following the conversion of NAD^+ to NADH by monitoring the absorbance of reaction mixtures at 340 nm, and the amount of NADH formed was determined using $\Delta\ \varepsilon_{340} = 6.22\ mM^{-1}{\cdot}cm^{-1}$.

EM. For EM, cells were cultured as described above for MCP purification. For immunogold localization of the GFP and GST, cells were sectioned, fixed, observed, and photographed as described previously (3). The primary antibodies used were rabbit polyclonal anti-GFP, anti-GST, or anti-diol dehydratase diluted 1:100 in PBS. The secondary antibody used was goat anti-rabbit IgG conjugated to 6-nm and 10-nm colloidal gold (Jackson ImmunoResearch Laboratories, Inc.) diluted 1:30 in PBS.

Bioinformatic Analysis of Targeting to MCPs. Among the known bacterial MCPs, some, including the Pdu MCP analyzed here, participate in complex metabolic pathways. Numerous enzymes are involved, some acting in the cytosol and some in the lumen of the MCP. However, it has been possible in only a minority of cases to determine enzyme content by purifying intact MCPs. The encapsulated enzymes define which metabolic intermediates are formed and consumed within the MCP interior and which small molecules must be transported across the protein shell. Interest in accurately modeling the metabolic functions of MCPs therefore motivates efforts to identify which specific enzymatic steps occur inside. The finding reported here (in the text) that N-terminal sequence extensions targeting PduP to the Pdu MCP suggests a general bioinformatic approach to predict which specific enzymes are likely to be encapsulated in yet uncharacterized MCPs.

The bioinformatic analysis here examines the N-termini of a set of homologous enzymes and compares homologues occurring in MCP operons with those occurring outside the context of an MCP operon. This is possible because enzymes that are associated with MCPs in some bacteria typically also occur in other bacteria that lack MCPs. Our hypothesis is that N-terminal extensions are indicative of MCP encapsulation.

To measure relative N-terminal sequence lengths, a multiple sequence alignment was performed on a set of orthologues predicted to be MCP-associated or not MCP-associated (based on gene proximity) as a single group. The relative lengths of extensions at the N terminus were determined by counting unaligned N-terminal residue positions. Individual proteins are classified (probabilistically) as MCP-associated if their encoding genes are nearby genes that encode members of a conserved family of proteins that form the shells of diverse MCPs (BMC proteins). This approach is used because known MCP-associated enzymes are nearly always encoded in operons that include BMC shell genes. In this study, we treated proteins encoded within 10 genes upstream or downstream of a BMC shell protein as MCP-associated.

To analyze targeting to the Pdu MCP, the 15 non-BMC genes in the *pdu* operon were examined. Homologues for each protein were retrieved by a BLAST query of the National Center for Biotechnology Information (NCBI) database. Individual sequences were classified as belonging to an MCP operon according to their chromosomal proximity to BMC shell genes using annotated fully sequenced bacterial genomes in the NCBI database. The full set of sequences was then aligned using the program MUSCLE (4) for examination of N-terminal sequences.

The sequence analysis results show that for three Pdu enzymes, PduP (in the text), PduD, and PduE, homologues presumed to be MCP-associated contain extended N-terminal sequences compared with homologues lacking an MCP association. These N-terminal extensions range from ≈20 residues (PduP) to ≈30 residues (PduD and PduE). The terminal sequence extensions for PduD, PduE, and PduP appear to be distinct. For the cases of PduD and PduE, the terminal extensions from different homologues show some sequence conservation. In contrast, the terminal extensions from different PduP homologues are highly divergent. The significance of these sequence conservation trends is unknown. Whether specific elements of the MCP (e.g., on the interior surface of the protein shell) serve as recognition elements for these tails has yet to be determined. In contrast to PduP, PduD, and PduE, analysis of PduG or PduH gave no indication of an extended tail for targeting. Likewise, PduO does not encode any apparent N-terminal targeting sequence but appears to encode a roughly 100 aa-long domain of unknown function in its C terminus.

This bioinformatic approach was also applied to proteins of the Eut MCP from *Escherichia coli* K-12, whose contents have not been established experimentally. The Eut MCP shares some similar enzymes with the Pdu MCP, but it metabolizes ethanolamine instead of 1,2-PD. From the analysis of Eut proteins, N-terminal sequence extensions were detected in the EutC and EutG proteins. EutC encodes ≈45 additional N-terminal amino acids compared with homologues not occurring in MCP operons, whereas EutG encodes ≈15 additional N-terminal amino acids. The identification of putative targeting sequences in these two proteins is consistent with current models for Eut MCP function. EutC together with EutB forms the ethanolamine ammonia lyase, which converts ethanolamine to acetaldehyde. EutG, an alcohol dehydrogenase, reduces acetaldehyde to ethanol. Because acetaldehyde is understood to be sequestered inside the MCP, the EutBC complex and EutG are both presumed to operate in the MCP lumen. Definitive results were not obtained for other Eut proteins.

To investigate the existence of terminal sequence extensions further, one additional putative MCP, which has not been experimentally characterized, was similarly analyzed. A locus encoding a pyruvate formate lyase enzyme (*pfl2*) among several BMC shell proteins was described in the sequenced genome of *Vibrio furnissi*i (5). Our bioinformatic analysis suggests that the *pfl2* gene encodes a roughly 35-residue N-terminal extension compared with homologues not encoded in MCP operons. Based on this observation, its compartmentalization within a MCP is predicted here.

Several caveats should be noted. The predictions of targeted enzymes presented here is not likely to be inclusive. Other enzymes could be targeted to the MCP by distinct mechanisms. For example, a targeting sequence was not detected in PduC. However, PduC is a subunit of the diol dehydratase enzyme together with PduD and PduE; hence, it could be encapsulated via the N-terminal extensions of PduD and PduE. Similar targeting could also be the case for the EutB and EutC subunits of ethanolamine ammonia lyase. Furthermore, the presence of a targeting sequence does not prove interior location. It is possible that some sequence motifs could target enzymes or other protein components to the exterior of the MCP. The structures of MCP shells are not understood completely enough to know whether this is a possibility. Finally, the bioinformatic analysis is probabilistic. This is emphasized by the results for PduP (Fig. 2). Four of the five groups of PduP homologues identified as being MCP-associated show prominent terminal extensions, but the group represented by the *Clostridium botulinum* sequence reveals only a short or negligible extension. The significance of this observation is unknown. Likewise, the sequences in the non-BMC-proximal groups are generally consistent with the conclusion that they operate in reactions not associated with compartmentalization; in addition to the absence of nearby BMC shell genes, they are generally not found in operons encoding enzymes similar to known Pdu or Eut enzymes. However, the *Shewanella frigidimarina* protein and its close homologues occur in operons that do not reveal BMC shell genes but do encode enzymes similar to those that have been implicated in compartmentalized pyruvate metabolism.

In Dataset S1, full sequence alignments are provided for the genes noted above as having clear indications for (or against) targeting. MCP-associated sequences appear in the top rows of the alignment and are distinguished by a "++" in the sequence identification following "gi." The NBCI gene accession identification and annotation follow these labels. BMC shell proteins, other presumably nonmetabolic proteins (e.g., transcriptional regulators), or uncharacterized proteins were not included in the analysis.

1. Havemann GD, Bobik TA (2003) Protein content of polyhedral organelles involved in coenzyme B_{12}-dependent degradation of 1,2-propanediol in *Salmonella enterica serovar* Typhimurium LT2. *J Bacteriol* 185:5086–5095.
2. Walter D, Ailion M, Roth J (1997) Genetic characterization of the *pdu* operon: Use of 1,2-propanediol in *Salmonella typhimurium*. *J Bacteriol* 179:1013–1022.
3. Bobik TA, Havemann GD, Busch RJ, Williams DS, Aldrich HC (1999) The propanediol utilization (*pdu*) operon of *Salmonella enterica* serovar Typhimurium LT2 includes genes necessary for formation of polyhedral organelles involved in coenzyme $B_{(12)}$-dependent 1,2-propanediol degradation. *J Bacteriol* 181:5967–5975.
4. Edgar, RC (2004) MUSCLE: multiple sequence alignment with high accuracy and throughput. *Nucl Acids Res* 32:1792–1797.
5. Wackett LP, Frias JA, Seffernick JL, Sukovich DJ, Cameron SM (2007) Genomic and biochemical studies demonstrating the absence of an alkane-producing phenotype in *Vibrio furnissii* M1. *Appl Environ Microbiol* 73:7192–7198.

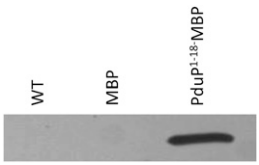

Fig. S1. Association of PduP$^{1-18}$-MBP with purified MCPs. Western blots were used to test for the presence of MBP. MCPs were purified from (left to right) WT *S. enterica*, Δ*pduP*/pLac22-MBP, and Δ*pduP*/pLac22-PduP$^{1-18}$-MBP. MBP produced from pLac22 was not associated with purified MCPs. In contrast, PduP$^{1-18}$-MBP was associated with purified MCPs. PduP$^{1-18}$-MBP, a fusion protein consisting of 18 N-terminal amino acids from PduP fused to MBP.

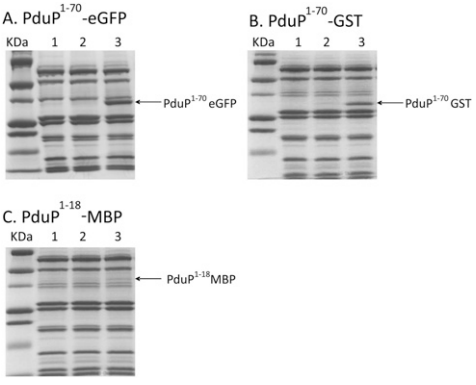

Fig. S2. SDS/PAGE of purified MCPs from strains producing selected fusion proteins. (*A*) WT *S. enterica*/plac22-PduP$^{1-70}$-eGFP induced with 0 (lane 1), 0.01 mM (lane 2), and 0.1 mM (lane 3) IPTG. (*B*) WT *S. enterica*/plac22-PduP$^{1-70}$-GST induced with 0 (lane 1), 0.01 mM (lane 2), and 0.1 mM (lane 3) IPTG. (*C*) WT *S. enterica*/plac22-PduP$^{1-18}$-MBP induced with 0 (lane 1), 0.01 mM (lane 2), and 0.1 mM (lane 3) IPTG. Other fusion proteins used in this study comigrated with MCP proteins, making detection by SDS/PAGE problematical and necessitating detection by Western blot (in the text). Native eGFP, GST, and MBP produced from pLac22 did not associate with purified MCPs (in the text). PduP$^{1-70}$-eGFP, amino acids 1–70 of PduP fused to eGFP and similarly for the other fusion proteins.

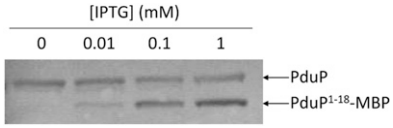

Fig. S3. Competition between PduP$^{1-18}$-MBP and PduP for MCP binding. Increasing IPTG concentrations (0, 0.01, 0.1, and 1 mM) were used to increase production of PduP$^{1-18}$-MBP by *S. enterica*/pLac22-PduP$^{1-18}$-MBP. Western blots were used to detect MBP in purified MCPs. PduP$^{1-18}$-MBP, a fusion protein consisting of amino acids 1–18 from PduP fused to MBP.

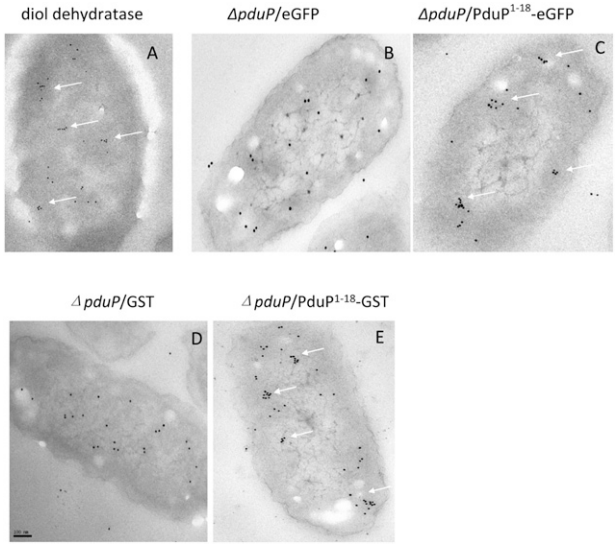

Fig. S4. Immunogold localization of PduP$^{1-18}$-eGFP and PduP$^{1-18}$-GST indicates MCP localization. (*A*) Positive control that shows immunogold localization of B_{12}-dependent diol dehydratase, which was previously shown to be MCP-associated [Havemann GD, Bobik TA (2003) Protein content of polyhedral organelles involved in coenzyme B12-dependent degradation of 1,2-propanediol in Salmonella enterica serovar Typhimurium LT2. J Bacteriol 185:5086–5095]. (*B* and *C*) Strains producing eGFP, PduP$^{1-18}$-eGFP. eGFP was randomly distributed in cells (*B*), but PduP$^{1-18}$-eGFP localized to discrete areas (*C*) similar to diol dehydratase, indicating that PduP$^{1-18}$-eGFP was MCP-associated but eGFP was not. Similar results were obtained using GST (*D*) or PduP$^{1-18}$-GST (*E*), further supporting a role for the N terminus of PduP in MCP targeting. The MCPs are not visible in the immuno-EM procedure because the required fixation protocol imparts poor contrast.

Table S1. Bacterial strains used in this study

S. enterica serovar Typhimurium LT2	Genotype	Source
BE269	Δ*pduP659*/plac22(no insert)	T.A. Bobik laboratory collection
BE270	Δ*pduP659*/plac22-*pduP*	T.A. Bobik laboratory collection
BE760	Δ*pduP659*/plac22-*pduP*	T.A. Bobik laboratory collection
BE1086	Δ*pduP659*/plac22-*pduP*$^{11–464}$	This study
BE1087	Δ*pduP659*/plac22-*pduP*$^{15–464}$	This study
BE1171	Δ*pduP659*/plac22-*eGFP*	This study
BE1175	Δ*pduP659*/plac22-*pduP*$^{1–10}$-*eGFP*	This study
BE1179	Δ*pduP659*/plac22-*pduP*$^{1–14}$-*eGFP*	This study
BE1172	Δ*pduP659*/plac22-*pduP*$^{1–18}$-*eGFP*	This study
BE1173	Δ*pduP659*/plac22-*pduP*$^{1–70}$-*eGFP*	This study
BE1187	LT2/plac22-*pduP*$^{1–18}$-*eGFP*	This study
BE1224	Δ*pduP659*/plac22-*GST*	This study
BE1125	Δ*pduP659*/plac22-*pduP*$^{1–10}$-*GST*	This study
BE1126	Δ*pdu659P*/plac22-*pduP*$^{1–14}$-*GST*	This study
BE1127	Δ*pduP659*/plac22-*pduP*$^{1–18}$-*GST*	This study
BE1128	Δ*pduP659*/plac22-*pduP*$^{1–70}$-*GST*	This study
BE1129	Δ*pduP659*/plac22-*MBP*	This study
BE1130	Δ*pduP659*/plac22-*pduP*$^{1–18}$-*MBP*	This study
BE1375	Δ*pduP659*/plac22-*pduP*$^{19–464}$	This study

Other Supporting Information Files

Dataset S1 (PDF)