

RESEARCH ARTICLE

Exploring Bacterial Organelle Interactomes: A Model of the Protein-Protein Interaction Network in the Pdu Microcompartment

Julien Jorda¹, Yu Liu², Thomas A. Bobik², Todd O. Yeates^{1,3,4*}

1 UCLA-DOE Institute for Genomics and Proteomics, Los Angeles, California, United States of America, **2** Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, Iowa, United States of America, **3** Molecular Biology Institute, University of California, Los Angeles, Los Angeles, California, United States of America, **4** Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California, United States of America

* yeates@mbi.ucla.edu



Abstract

Bacterial microcompartments (MCPs) are protein-bound organelles that carry out diverse metabolic pathways in a wide range of bacteria. These supramolecular assemblies consist of a thin outer protein shell, reminiscent of a viral capsid, which encapsulates sequentially acting enzymes. The most complex MCP elucidated so far is the propanediol utilizing (Pdu) microcompartment. It contains the reactions for degrading 1,2-propanediol. While several experimental studies on the Pdu system have provided hints about its organization, a clear picture of how all the individual components interact has not emerged yet. Here we use co-evolution-based methods, involving pairwise comparisons of protein phylogenetic trees, to predict the protein-protein interaction (PPI) network governing the assembly of the Pdu MCP. We propose a model of the Pdu interactome, from which selected PPIs are further inspected via computational docking simulations. We find that shell protein PduA is able to serve as a “universal hub” for targeting an array of enzymes presenting special N-terminal extensions, namely PduC, D, E, L and P. The varied N-terminal peptides are predicted to bind in the same cleft on the presumptive luminal face of the PduA hexamer. We also propose that PduV, a protein of unknown function with remote homology to the Ras-like GTPase superfamily, is likely to localize outside the MCP, interacting with the protruding β -barrel of the hexameric PduU shell protein. Preliminary experiments involving a bacterial two-hybrid assay are presented that corroborate the existence of a PduU-PduV interaction. This first systematic computational study aimed at characterizing the interactome of a bacterial microcompartment provides fresh insight into the organization of the Pdu MCP.

OPEN ACCESS

Citation: Jorda J, Liu Y, Bobik TA, Yeates TO (2015) Exploring Bacterial Organelle Interactomes: A Model of the Protein-Protein Interaction Network in the Pdu Microcompartment. *PLoS Comput Biol* 11(2): e1004067. doi:10.1371/journal.pcbi.1004067

Editor: Sandor Vajda, Boston University, United States of America

Received: July 15, 2014

Accepted: December 1, 2014

Published: February 3, 2015

Copyright: Jorda et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was funded by grant AI081146 from the National Institutes of Health. TOY is supported by the BER program of the U.S. Department of Energy Office of Science. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors declare that no competing interests exist.

Author Summary

Many bacteria produce giant proteinaceous structures within their cells, which they use to carry out special metabolic reactions in their interior. Much has been learned recently about the individual components—shell proteins and encapsulated enzymes—that

assemble together, thousands of subunits in all, to make these bacterial microcompartments or MCPs. However, in order to carry out their biological functions, these systems must be highly organized through specific protein-protein interactions, and such a higher level understanding of organization in MCP systems is lacking. In this study, we use genomic data and phylogenetic analysis to predict the network of interactions between the approximately 20 different kinds of proteins and enzymes present in the Pdu MCP. Then, we use computational docking to examine a subset of those that are predicted to involve enzymes bound to the interior surface of the shell proteins, and show that the results are consistent with recent experimental data. We further provide new experimental evidence for one of the predicted protein-protein interactions. This study expands our understanding of a complex system of proteins serving as a metabolic organelle in bacterial cells, and provides a foundation for further experimental investigations.

Introduction

Cellular organization has long been considered to be much simpler in bacteria than in eukaryotic cells. However, accumulating evidence indicates a higher-order organization in terms of cellular compartmentalization [1–3] and cell structure [4,5]. In particular, electron microscopy and higher resolution structural studies have demonstrated that some bacteria can form polyhedral capsid-like bodies that are 80 to 150 nm in diameter [6,7]; reviewed in [8–11]. These polyhedral inclusions, known as bacterial microcompartments, are widely distributed across nearly 20% of known bacterial strains [9,12,13]. We refer here to bacterial microcompartments as MCPs; they are sometimes referred to as BMC's, but we reserve the latter name to refer to the family of shell proteins that comprise MCP shells. As opposed to membrane bound organelles characteristic of eukaryotic cells, MCPs are exclusively proteinaceous assemblies; they consist of a thin outer protein shell enclosing a metabolically active core of enzymes, earning them the status of bacterial organelles. MCPs fulfill diverse roles: enhancement of metabolic flux in their hosted enzymatic pathway [14], confinement of toxic or volatile intermediates [15–17] and shielding of interior enzymes from reactions with reactive or competing molecules [18].

The founding member of the MCP family, the carboxysome, was first isolated 40 years ago [19]. Carboxysomes are present in some chemotrophic bacteria and probably all cyanobacteria [18,20,21]. The carboxysome serves as an organelle for carbon fixation through the encapsulation of two enzymes: carbonic anhydrase and ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO). Several other kinds of MCPs are found dispersed across the bacterial kingdom, where they carry out metabolic pathways different from carbon fixation. These include the Pdu and the Eut microcompartment from *Salmonella* [22–24] and *E.coli* [25,26], which carry out the degradation of 1,2-propanediol and ethanolamine, respectively. These pathways rely on a similar mechanism: an initial substrate is first converted by a B₁₂-dependent enzyme to give an aldehyde intermediate, which is sequestered long enough to be converted to less toxic metabolites, e.g. an alcohol and/or carboxylic acid. However, these three relatively well-characterized MCPs (carboxysome, Pdu and Eut) constitute only a subset of the entire MCP family. Recent computational and experimental studies delineate at least seven kinds of MCPs, all with different metabolic purposes [13,27–30]. The accepted three-dimensional model of the Pdu MCP and its encapsulated metabolic pathway is summarized in Fig. 1B.

Though MCPs differ substantially according to their metabolic nature, they share a number of genomic and structural characteristics. In particular, most MCP proteins are encoded within

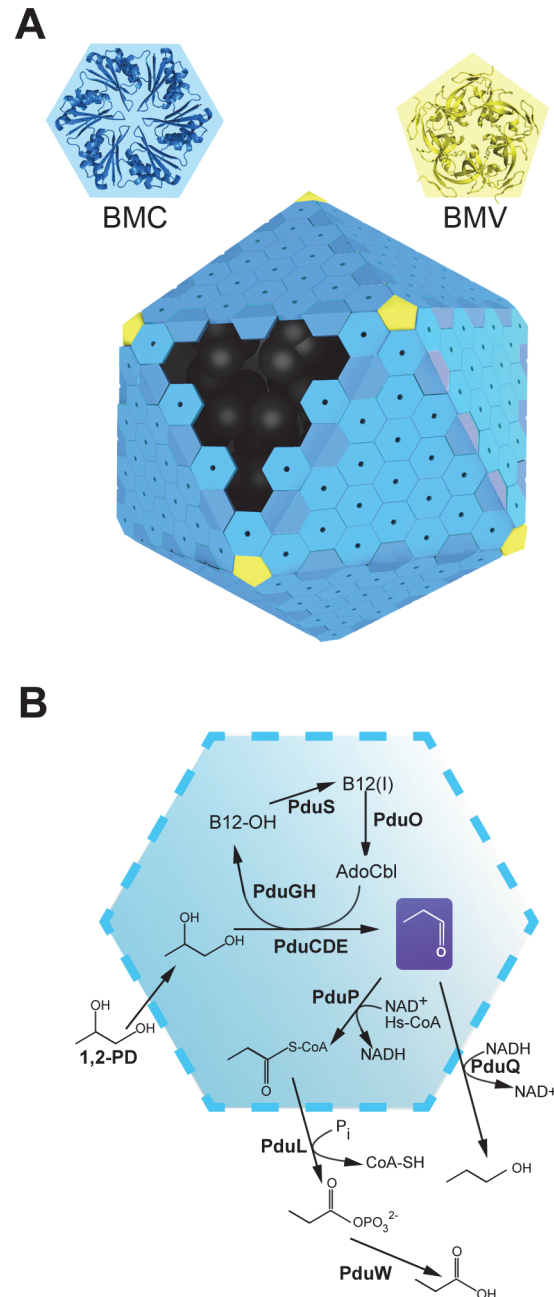


Figure 1. An idealized model of the Pdu MCP shell and its encapsulated pathway. The MCP shell is assembled from a few thousand copies of proteins belonging to the BMC (bacterial microcompartment) protein family. Several distinct paralogs from the BMC family are present within a single shell. BMC proteins self-assemble into cyclical hexamers (in blue). Also present in fewer copies are proteins from a distinct family, referred to as BMVs, which are pentameric proteins (in yellow) forming the vertices of the polyhedral structure. The polyhedron is shown here idealized as an icosahedron, while the Pdu MCP is typically less regular in shape. Sequentially acting enzymes (in black) carrying out the Pdu pathway are enclosed by the shell (A). The Pdu pathway degrades 1,2-propanediol to propionaldehyde via a B₁₂-dependent catalytic mechanism, the aldehyde being subsequently converted to 1-propanol or propionyl-phosphate (B).

doi:10.1371/journal.pcbi.1004067.g001

operons, which consist of multiple paralogous genes coding for the shell proteins alongside the genes for the associated enzymes. Consistent with this shared genomic signature, diverse MCPs share a similar organization and structure. Typically, each shell protein sequence is comprised by a bacterial microcompartment (BMC) domain, or sometimes two such domains duplicated in tandem. The first high resolution structures of BMC proteins shed light on the structural organization of the MCP shell [9–11,28,31–35]. A few thousand copies of these BMC proteins self-assemble into cyclic hexameric units packed side-by-side in a layer forming the essentially flat facets of the roughly icosahedral structure (Fig. 1A). The top and bottom sides of a BMC hexamer typically show distinctly different features: one face bears a central depression giving rise to a concave shape, whereas the other side is typically flatter and more polar in chemical character. Which of the two sides (convex or flat) faces inward to the MCP lumen is a question of key significance for MCP function [36–38]. Most often, the center of the hexamer is perforated by a narrow (4–7 Å) hydrophilic pore that is thought to act as a canal for molecular transport [32,38–40]. In addition to the main BMC shell proteins, other minor proteins have been found to be essential to the formation or closure of the shell. These proteins, which our group recently coined the bacterial microcompartment vertex (BMVs) proteins, assemble into pentamers suspected to close the vertices of the MCP [37,41,42] (Fig. 1A). Furthermore, a number of intriguing variations such as domain fusion, tandem duplication, circular permutation, or FeS cluster binding sites, have been revealed among the crystal structures of the paralogous BMC shell proteins [43–46]. Speculations on the roles of such variations support the idea that each type of BMC paralog has a defined role beyond simply assembling to form a physical barrier.

Interactions between the shell proteins and the encapsulated enzymes are vital for MCP function. Recent studies on the assembly of the α -type carboxysome suggest assembly of this type of MCP is initiated from the interior; the formation of enzymatic seeds precedes acquisition of the shell [47,48]. However, the processes governing the interactions between the encapsulated enzymes and the shell proteins are complex and apparently divergent between different types of MCPs. Specific interactions have been demonstrated in a few cases using pull-down assays and other experiments [36,49]. Fan *et al.* [50] first showed that short sequence extensions present at the N-terminus of numerous enzymes exist to bind enzymes to the MCP shell. A subsequent study showed that the C-terminal region of an α -carboxysomal protein (CcmN) interacted with the shell in that system [49]. Though enzyme targeting mechanisms are presumed to be widespread across the MCP systems, only a few enzyme-shell protein interactions have been specifically identified. Characterizing these interactions would open new perspectives on MCP biology and applications in synthetic biology [51,52]. Some progress has already been made along these lines. Fluorescent proteins and other proteins have been successfully directed to MCPs by appending terminal targeting peptides [29,50,53–55].

Despite knowing the identities of a few interactions between enzymes and shell proteins, atomic level detail is lacking. Attempts to isolate and determine the structures of cognate complexes have been unsuccessful. This has prompted us to undertake a computational study to develop interaction models for an MCP system. The ever-increasing genomic and structural data available for MCPs provides an unprecedented opportunity to apply computational methods to characterize the molecular networks ruling these extraordinary supramolecular machines. Over the last two decades, a handful of methods exploring genomic data have been developed for predicting functional linkages between different proteins in a cell. Popular methods such as protein phylogenetic profiles [56,57], gene fusion [58,59], gene neighborhood [60,61] or a combination of these [62–64], have been used extensively to make functional inferences about proteins. Indeed, one of our recent studies featured an adaptation of protein phylogenetic profile

methods for investigating co-occurrence patterns in MCP operons, and led to an articulated classification of existing MCP pathways [13].

Here, we aim to characterize the molecular network of physical protein-protein interactions (PPIs) in a single MCP type, the Pdu system. In this case, strategies relying on genomic context have limited application due to the high similarity of the genomic patterns found for different proteins across the Pdu operons; essentially all of the MCP shell proteins and enzymes typically found in the Pdu operon are functionally linked according to genomic context, but only a subset engage in direct physical PPIs. Other computational strategies are therefore required to develop models for direct physical PPIs. Detailed sequence variations within protein families can be analyzed via phylogenetic tree-based approaches, and indeed methods based on mining of phylogenetic features have proven useful for predicting PPIs in multiple cases, as recently reviewed [65,66]. A non-exhaustive list of such methods includes the so-called *mirror tree* [67], or its variant the *tol-mirror* [68], which compares trees—one for each protein of interest—by computing the pairwise correlation of their underlying evolutionary distance matrices. Others explore the topological similarity of the trees, coined congruence by Vienne *et al.* [69]. All follow the co-evolution hypothesis, where interacting protein families are expected to exhibit similar phylogenetic trees with similar patterns of amino acid sequence divergence.

In this work, we seek to identify new PPIs in the Pdu MCP with a coevolution-based machine learning algorithm. Specifically, we approach the PPI prediction problem within a binary classification framework: from the pairwise comparison of phylogenetic trees, coevolution features can be computed and subsequently mined by a decision tree classifier, a concept earlier described in Craig and Lio [70]. A group of PPIs that have been experimentally characterized recently in the Pdu system constitute a set of known positives for use as a “gold standard” for training the classifier. In the first part of this work, we design and train a Random Forests classifier to identify pairwise interactions of Pdu gene products, and then propose a model of the Pdu interactome. Following this genomic-based model, we further analyze selected predictions of PPIs and their probable binding modes via three-dimensional protein-protein docking calculation. We then provide new experimental data to support one of the predicted interactions.

Results

For each pair of Pdu gene products, we defined seven continuous-valued coevolution descriptors extracted from the pairwise comparison of their respective phylogenetic trees, and combined those seven values into a vector (Fig. 2). As an example, one of the seven descriptors is the linear correlation coefficient between two phylogenetic trees calculated by the *mirrortree* method. The other descriptors are variations on a similar theme (see Methods). Within this framework, and using experimental data on known interactions as a training set, we ran a binary classifier against these vectors of coevolution descriptors to identify positive PPIs.

Predicting PPIs of the Pdu interactome

We culled protein sequences from Pdu operons of 34 fully sequenced bacterial genomes, and collapsed them into 22 orthologous protein groups according to the canonical Pdu nomenclature [23]. For each of the 22 distinct protein families so identified, we inferred a phylogenetic tree from a multiple sequence alignment of its constitutive sequences. We refer to this as the ‘Pdu tree’ for that protein. Subsequently, for each pair of proteins seven co-evolution descriptors were computed from a comparison of their respective Pdu trees, following the general procedure depicted in Fig. 2. Pairwise combinations of the 22 orthologous protein groups resulted in 231 unique pairs that needed to be classified. For this purpose, we used a Random Forests

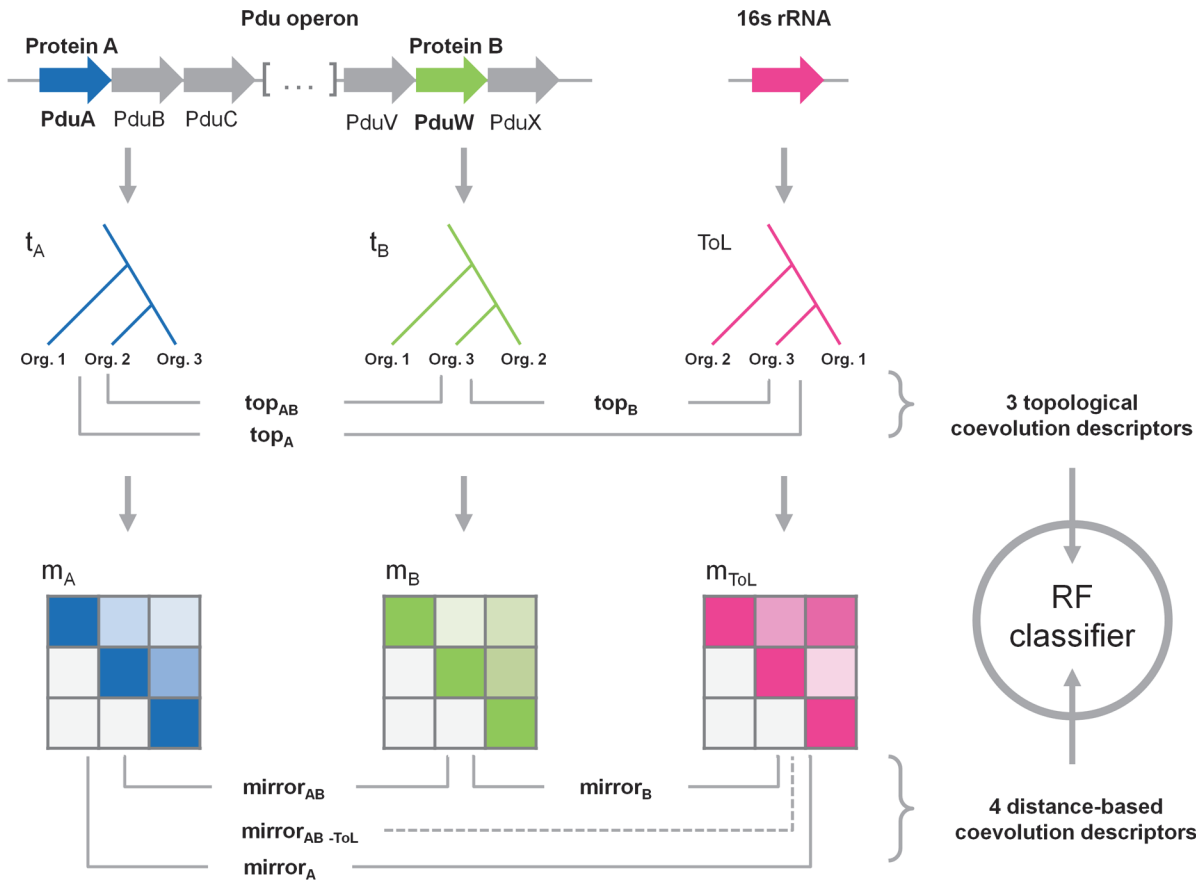


Figure 2. Description of the procedure for defining pairwise coevolution descriptors. Calculation of coevolution descriptors relies on the comparison of phylogenetic trees. For each given pair of Pdu gene products, three descriptors are extracted from a topological comparison of their respective phylogenetic trees (blue and green) and the Tree of Life (ToL, pink), while four other descriptors are calculated by comparing the distance matrices that underlie these trees. These seven descriptors are further combined into a vector for subsequent analysis by the RF classifier.

doi:10.1371/journal.pcbi.1004067.g002

classifier [71] exploring the seven descriptors, which after a training and cross-validation phase exhibited an area under the receiver-operator (ROC) curve of 0.75 (S1 Fig, suppl. Data), thereby demonstrating a reasonably good discriminative power. We also assessed whether similar classification performance could be obtained with fewer descriptors than the seven initially employed. We evaluated the discriminatory power of the descriptors individually by ranking their accuracies in the context of an unsupervised analysis (S2 Fig.). We found that the RF performs best when all seven of the descriptors are included in the classification analysis. Much of the signal can be recovered with just a few descriptors, but addition of subsequent descriptors does result in slight improvements in performance. When applied to the whole Pdu dataset, the classifier predicted a list of 109 positive PPIs along with their mean probabilities. To be conservative and increase the specificity of the classifier (even if at the expense of the sensitivity), we removed the putative PPIs with a probability less than 0.7, which reduced the final number of predicted PPIs to 51 (Suppl. data). From these results we modeled the Pdu interactome as a molecular network of 51 interactions and 22 nodes. The resulting network model is presented in Fig. 3.

An analysis of this model showed that 15 of the 16 experimentally characterized PPIs could still be retrieved under a high specificity criterion, and that they yielded the highest probabilities, confirming the robustness of the method. Furthermore, the missing positive interaction,



Figure 3. A model of the Pdu interactome. The Pdu PPI network, inferred from predictions made by the RF classifier in its analysis of coevolution descriptors. Individual Pdu gene products are represented as nodes. Enzymes are shown in light blue, while shell proteins are shown in gray; the shell proteins include several BMC type proteins and a single protein (PduN) from the BMV family presumed to be pentameric vertex proteins. Edges connecting two nodes correspond to predicted PPIs. The numerous PPIs emerging from the PduA node are highlighted in pink. It is not possible to fully convey the likely spatial relationships of all the proteins and enzymes (some of whose locations remain uncertain), but nodes for the shell proteins have been placed at the periphery of the layout to convey their outer locations.

doi:10.1371/journal.pcbi.1004067.g003

PduK-PduT, was initially predicted as positive by the classifier, but did not pass our 0.7 threshold. One striking feature of this model is the absence of a PPI connecting the PduX node to the network (Fig. 3). PduX is an enzyme involved in *de novo* synthesis of coenzyme B₁₂, an essential cofactor for enzymes of the Pdu pathway [72]. However, there is no evidence that PduX is directly associated with the MCP by any physical interaction [73]. Its tendency to occur within the Pdu operon (typically at the end) likely reflects an advantage of being under the influence of the Pdu promoter, rather than physical interaction with other MCP components. Two likely spurious findings involving interactions with PduF also appeared in our model, namely PduF-PduC and PduF-PduD. PduF is a propanediol/glycerol diffusion facilitator protein and is believed to be an integral membrane protein [74], making its physical presence in the MCP unlikely. Finally, after exclusion of the “gold standard” interactions and the suspected spurious predictions, the final dataset consisted of 36 predicted PPIs, with an average node connectivity of 4.8 partners, which can be loosely compared to results obtained with other interactome studies across whole cellular systems in yeast [75] or in cell junction complexes [76].

One intriguing observation is the hyperconnectivity of certain specific nodes, such as PduA (11 PPIs), PduC (9 PPIs) and PduG (9 PPIs). The central position of the propanediol dehydratase large subunit PduC in the Pdu pathway makes it an essential piece of the interactome (Fig. 1B). Likewise, PduG is the large subunit of the diol dehydratase-reactivating factor, which works in tight coordination with the propanediol dehydratase (PduCDE). In a complex with PduH, PduG is believed to reactivate the dehydratase by exchanging its B₁₂ cofactor, which becomes inactive during repeated catalytic cycling [73,77]. In our model, PduG was indeed predicted to interact with PduC and PduE but not with PduD. Although no structural information about this complex in *Salmonella* is available, crystal structures of highly similar homologs from *Klebsiella oxytoca* have been solved [22,78,79]. Studies with these same homologs demonstrated that the binding mechanism involved a subunit exchange between the dehydratase and

the reactivase, where one PduH subunit is released from the reactivase and replaced by one PduD subunit [80].

Particularly notable in our interactome model is the number of PPIs in which PduA [73,81], one of the most abundant shell proteins in the Pdu MCP, is predicted to be involved. Presently, PduP is the only enzyme in the Pdu MCP whose binding to individual shell proteins has been characterized. It was revealed that PduP interacts via its N-terminal region with PduA and PduJ, another major shell protein that shares high sequence identity (83%) with PduA [36]. Other Pdu enzymes besides PduP are suspected to carry such N-terminal extensions [50], but their shell protein partners have not been identified yet [54,82]. Sequence analysis as well as spectroscopic experiments on the PduP N-terminal segment show that it has a strong propensity to fold into an alpha-helical structure [36,55]. Here, we hypothesize that these structural features and associated binding mechanism are not specific to the PduP case, but that PduA (or PduJ) likely serves as a central binding hub for different enzymes carrying N-terminal extensions. To pursue this particular set of interactions further, we generated atomic models of predicted PduA-enzyme-tail complexes by molecular docking and analyzed their predicted modes of binding.

Additionally, we analyzed the PduU-PduV case, the only PPI in which PduV was predicted. PduU was the first BMC shell protein from a non-carboxysome MCP whose three-dimensional structure was determined [44]. Its topology involves a circularly permuted BMC domain, and the existence of a six-stranded β -barrel capping the central pore of the hexamer makes it unique in the BMC protein family. Previous speculations about the peculiar beta barrel include a possible role in gating an unusually wide pore, but further data are lacking. Additionally, PduV is a Ras-like GTPase that has been implicated in MCP dynamics within the cell by Parsons *et al.* [82]. In this case, PduV is believed to reside outside the shell, as opposed to the other Pdu enzymes that appear to be sequestered in the MCP interior. To clarify how these two might interact, as predicted by our interactome model, we modeled the PduU-PduV complex with docking simulations and compared the result to control calculations involving non-interacting protein pairs.

PduA: A universal hub for binding encapsulated enzymes

Of the 11 predicted interactions involving PduA, six include Pdu enzymes, namely PduC, PduD, PduE, PduG, PduP, and PduL (Fig. 3). As noted above, one of these interactions (PduA-PduP) has been demonstrated experimentally. Here we investigated whether enzymatic partners in addition to PduP are also able to bind PduA via terminal peptides, by attempting to model their presumptive binding modes computationally (see [Methods](#)).

As a first step, we searched for possible terminal peptidic extensions in the sequences of these six Pdu enzymes. Prediction of these extensions was done according to the method developed in Fan *et al.* [50]. The central idea is that enzymes that are targeted to the MCP exhibit extensions at their termini that are absent from homologous versions of the same enzyme that are not part of an MCP system. It is notable that among the six enzymes that are predicted by our model to bind to shell protein PduA (or its close homolog PduJ), our computational analysis indicates that five carry recognizable sequence extensions (PduC,D,E,P,L), (as reported in [50] and [54]). In contrast, none of the 15 enzymes that do not have predicted interactions with PduA (or PduJ) exhibit recognizable terminal sequence extensions. Sequence comparisons between the N-terminal peptide tails did not reveal strong similarities (less than 30% identity overall). However, *ab initio* predictions of their structures consistently modeled them as amphipathic α -helices. Experimental studies have already investigated the possible targeting of some of the Pdu enzymes; targeting by the N-terminal tail of PduP was noted above [50]. In

the case of PduD, experiments showed that its N-terminal peptide can be used as a targeting signal, but there was no evidence that it would fold identically to the PduP peptide or that its interaction would be with PduA [54]. In these same studies, PduE was implicated as having such terminal extensions, but fusion of GFP to its respective peptides did not provide clear evidence for targeting. In the case of PduC, Parsons, et al. showed that that enzyme could direct other proteins to the MCP when fused genetically, though the presence of a terminal tail on PduC was not indicated [81]. Despite the mixed findings on terminal targeting peptides on different enzymes in different experimental protocols, the presence on several of the Pdu enzymes in our bioinformatics analysis of extended termini with predicted alpha-helical propensities, and the prediction here of interactions between those enzymes and the PduA shell protein, supported the idea that some of these peptides likely recognize the interior surface of the shell using similar binding modes.

Since it was demonstrated that the targeting of PduP is mediated mostly by its terminal peptide segment [50], we sought to characterize the binding mode of the various implicated enzymes by docking their N-terminal peptides onto the hexameric structure of the PduA shell protein; 18-amino acid terminal segments were used in all cases. The benefits of using a model of the terminal peptide instead of a complete protein are twofold: (1) to avoid spurious modeling of full-length proteins in the absence of close structural homologs, and (2) to substantially reduce the size of the search space to be explored by the docking algorithm. In earlier work we proposed a model of the PduP N-terminal extension bound to the concave face of a PduA hexamer (proposed to be inward facing) [35]. However, this model was generated with a rigid-body approach, where the PduP peptide had only flexible side chains. Here we push further the flexibility limits of the docking simulation by additionally allowing conformational degrees of freedom for the peptide backbone. To do so, we employed a two-stage docking approach: a rough search by Autodock Vina [83] of the binding site in the PduA hexamer with a rigid helical model of the peptide, followed by a second docking phase with the FlexPepDock protocol from the Rosetta suite [84]. In this second step, the peptide is placed in its start position according to Vina's predictions; it is then simultaneously refolded and docked over the surface of the receptor. We applied this approach to the five identified PPIs and to a control case involving the N-terminal sequence from PduQ, an aldehyde dehydrogenase from the Pdu pathway that has no obvious targeting signal. In addition, the five peptides were alternatively docked on both faces of the PduA hexamer, with the expectation that meaningful results would have peptides docking to only one side of the PduA shell protein.

Results of the peptide docking simulations are overlaid in Fig. 4 along with their energy scores and their buried surface areas. Remarkably, when docked onto the concave (presumably luminal) face of PduA, all five peptides were predicted to bind the same binding cleft formed by the C-terminal segments of two adjacent PduA monomers in the hexamer (Fig. 4A). Moreover, with the exception of PduL, FlexPepDock automatically folded the peptides into well-defined α -helical structures. In the case of PduA-PduP, the model is similar to the one initially proposed in Yeates *et al.* [35], with a slight rotation and translation inside the cleft. Interestingly, the different peptides occupy the common binding cleft of PduA in different orientations: PduP and PduE have their N-termini pointing toward the pore, whereas PduC and PduD are docked in the opposite direction. The PduL peptide was also predicted to bind roughly the same region, but the flexible docking protocol did not automatically fold that peptide into a well-ordered alpha helix, leaving the veracity of the predicted binding mode in question in the case of PduL. In their computationally predicted bound configurations, most of the polar residues of the peptides are exposed to the solvent. A notable exception is an arginine recurrently found towards the center of each peptide, which is in all cases buried in the predicted interface and poised to form a salt-bridge with glutamate (E36) of either one of the two

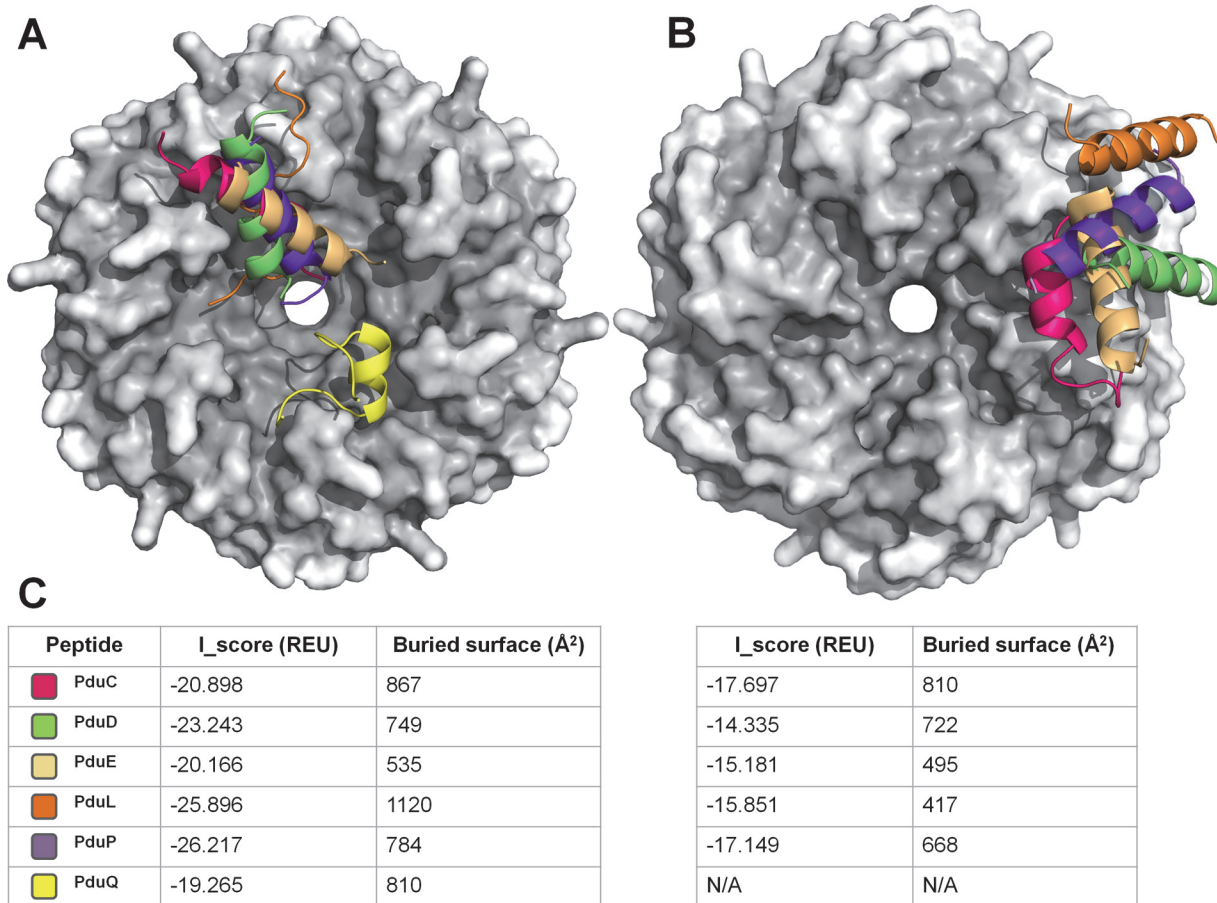


Figure 4. Models of N-terminal peptide extensions from different enzymes docked onto a PduA hexamer. All the models were aligned and overlaid using the PduA structure as guide. (A) Six N-terminal peptides are docked on the concave (presumptively luminal) face of the PduA hexamer. Four of the five identified earlier as probable targeting sequences (PduC, PduD, PduE, PduP) were folded into α -helices by the flexible docking procedure (see text and Methods) and were docked in the same cleft on the PduA surface. The tail of PduL adopted a less regular conformation during the simulation. The tail from PduQ, which was not predicted to act as a targeting sequence and thereby serves as a control, exhibits an apparently spurious binding mode. To convey depth, the surface of PduA is shaded according to diffusion accessibility [106]. (B) The five targeting peptides, when docked onto the other (flat) face of the PduA shell protein, were found scattered across the surface in arrangements exhibiting poorer interaction interfaces. (C) Binding statistics are reported for all the docking simulations. In all cases, both the predicted energy score (in Rosetta Energy Units) and the buried surface at the interface yielded better values when peptides were docked onto PduA's concave side. Because the shell protein hexamer is 6-fold symmetric, in all cases the solutions were rotated by multiples of 60° around the axis of symmetry to allow internal consistency.

doi:10.1371/journal.pcbi.1004067.g004

monomers constituting the binding cleft (Fig. 5A). The hydrophobic residues are oriented to interact with the C-terminal segment of PduA (Fig. 5B).

Various other docking calculations served as computational controls. In contrast to the results obtained for docking to the concave surface of the shell protein, docking of the peptides on the other (flat) side of PduA showed no consistent or compelling modes of binding. Those peptide models are instead scattered over the hexamer surface (Fig. 4B). Moreover, comparison of the energy scores and buried surface areas in both docking cases shows that the peptides have a significantly better fit to the concave surface (Fig. 4C). Another control consisted of docking the N-terminal 17 residues from PduQ (which was not predicted to have a targeting tail) following the same protocol. In the docking simulation the PduQ peptide partially folds into an α -helix, but does not seem to bind intimately in the canonical cleft (Fig. 4A). An additional calculation involved the docking of N-terminal peptides onto a layer of three PduA hexamers packed side-by-side, to verify that potential binding modes at the interfaces between

A

PduC	MRSKRFEALAKRPNVQDG
PduD	MEINEKLLRQIIEDVLRD
PduE	MNTDAIESMVRDVL SRMN
PduL	MDKELLQSTVRKVLDEMR
PduP	MNTSELETLIRITILSEQL

B

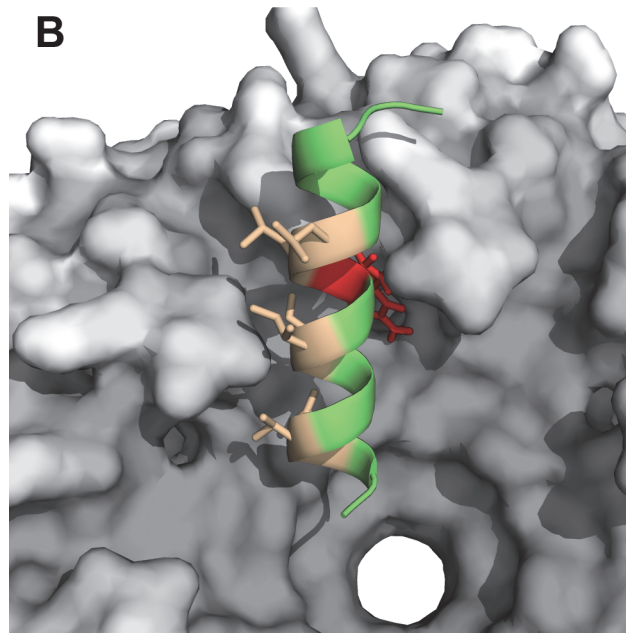


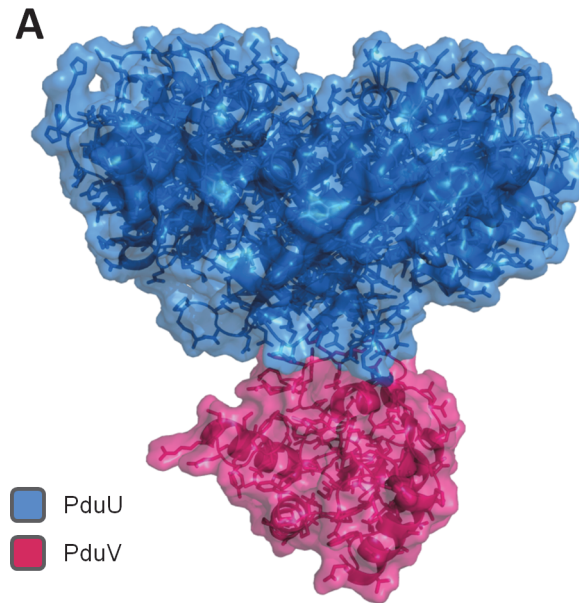
Figure 5. N-terminal extension sequences and atomic details of the PduD peptide docked onto a PduA hexamer. (A) Sequences of the five N-terminal extensions proposed to be acting as targeting peptides. An arginine is recurrently found near the center of the peptide (red). (B) The hydrophobic surface (in beige) of the PduD N-terminal tail peptide is predicted to interact with the C-terminal tail of PduA. A central arginine (in red), which is found in all of the N-terminal peptides predicted to dock in the cleft, is consistently oriented to make interactions with a glutamate in the BMC domain.

doi:10.1371/journal.pcbi.1004067.g005

hexamers were not overlooked. This simulation exhibited similar binding modes to those found with a single PduA hexamer. Overall, these computational predictions and control calculations support the hypothesis that the interior surface of PduA serves as a hub for binding multiple enzymes with terminal extensions. The findings are largely consistent with previous experimental data, while painting a more detailed picture of how interior enzymes in the Pdu MCP interact with PduA, as predicted by our coevolution analysis.

A predicted PduU-PduV complex

As an initial step in modeling a possible interaction between PduU and PduV, which was predicted by the coevolution analysis, a homology model had to be constructed for PduV. The PduV model was then docked into the crystal structure of the PduU hexamer using Rosetta-Dock [85] (see [Methods](#)). As a control, we ran two docking simulations under identical conditions on cases involving either PduU or PduV and non-interacting molecules: PduA-PduV, and PduU-ERA (the homologous GTPase used as the template for modeling of PduV). A



B

	I_score (REU)	Shape compl.	Buried surface (Å ²)
PduU-PduV	-5.378	0.59	714
PduUΔ17-PduV	-2.532	0.39	687
PduA-PduV	-4.15	0.46	614
PduU-ERA	-4.712	0.5	691

Figure 6. Model of PduV docked onto a PduU hexamer. (A) Docking calculations predict that the N-terminal region of PduV binds the PduU β -barrel that protrudes from the conserved BMC domain. Binding statistics for the PduU-PduV docking and three control simulations are reported in a separate table (B). Those latter, which included the docking of PduU to a non-cognate GTPase homolog of PduV (labeled PduU-ERA), a truncated version of PduU lacking the beta-barrel docking to PduV (labeled PduU Δ 17-PduV), and PduV docking to PduA instead of PduU (labeled PduA-PduV), all had substantially worse binding statistics than the PduU-PduV model.

doi:10.1371/journal.pcbi.1004067.g006

model of the PduU-PduV complex is proposed in Fig. 6A, along with statistics from the different docking simulations (Fig. 6B). Compared to the two controls, the predicted interface between PduU and PduV achieved a better Rosetta energy score. Likewise, the PduU-PduV complex featured a better shape complementarity and larger buried surface than the controls. In this model, PduV is sitting on the axis formed by the PduU beta-barrel; this PduU protuberance is exclusively contributing to the interface and precludes any interaction between PduV and the main BMC domain of PduU. Most of the interaction surface on PduV is formed by the N-terminal region spanning residue 13 to 35. This is consistent with preliminary results from Parsons *et al.*, where the first 42 amino acids from PduV were demonstrated to play a crucial role in PduV targeting to the MCP [82]. As a final control calculation, we investigated the binding mode of PduV after deleting the 17 N-terminal residues forming the β -barrel in the PduU hexamer. Here again, the model yielded worse interaction statistics than for the full-size PduU-PduV complex, supporting the model in which the β -barrel of PduU plays a crucial role in the interaction with PduV (Fig. 6B).

Experimental confirmation of a PduU-PduV interaction

Preliminary experimental assays were carried out on the PduU-PduV pair in parallel with our computational analysis. The BacterioMatch II two-hybrid system was used to test for interactions between these two proteins. In this system, a reporter strain is co-transformed with appropriate target and bait fusion genes. A protein-protein interaction between the target and bait activates the transcription of *HIS3*, an essential gene for histidine biosynthesis [86], thereby increasing the expression of the *HIS3* product to levels that are sufficient to allow growth on a selective medium lacking histidine and to overcome the effect of 3-amino-1,2,4-triazole (3-AT), a competitive inhibitor of the His3 enzyme. If a large number of colonies are obtained following co-transformation, an interaction between the target and bait proteins is indicated. When PduU and PduV were tested, the number of colonies obtained following co-transformation was comparable to that of a positive control with bait and prey proteins (LGF2 and Gal11^P) that are known to strongly interact (Table 1). Results showed that PduU and PduV also interacted in reciprocal tests where their roles as bait and prey were reversed (Table 1). Negative controls showed that PduU or PduV alone did not confer 3-AT resistance (Table 1). The positive result with the UV pair was confirmed by streptomycin resistance of co-transformed *E. coli* which requires expression of a second reporter gene, *aadA*. This experimental confirmation of a PduU-PduV interaction supports the Pdu MCP interactome model developed in the first (coevolution analysis) part of our work, while the docking calculations reveal a plausible mode of binding between those proteins.

Discussion

Proteins rarely carry out biological processes on their own. Instead, they typically participate with other proteins in the context of larger interaction networks. This is especially true for MCPs, where encapsulated pathways require coordination and spatial organization of their numerous components, from shell proteins to enzymes. Though structural studies of individual MCP components have paved the way to a better understanding of their assembly mechanism, a full comprehension of such metabolic systems requires investigation of their PPI networks. Unfortunately, experimental data for MCP protein complexes are still sparse, leading us to

Table 1. Two hybrid assay to test the interaction between PduU and PduV.

Pairs tested	NS ^a	NS 1:100 ^b	S ^c
pBT-LGF2, pTRG-Gal11 ^P	TNTC ^d	1000 ± 87	1000 ± 72
pBT-pduU, pTRG	TNTC	11 ± 3	0 ± 0
pBT-pduV, pTRG	TNTC	19 ± 10	0 ± 0
pBT, pTRG-pduU	TNTC	167 ± 21	0 ± 0
pBT, pTRG-pduV	TNTC	740 ± 75	7 ± 3
pBT-pduU, pTRG-pduV	TNTC	1000 ± 66	600 ± 12
pBT-pduV, pTRG-pduU	TNTC	840 ± 79	TNTC

Different combinations of bait(pBT) and prey (PTRG) genotypes are tested. The LGF2 and Gal11^P pair serves as a positive control. Tests with each pair were repeated three times and the numbers shown are the mean ± 1 standard deviation.

^acfu obtained on nonselective (NS) screening medium (no 3-AT) plates.

^bcfu obtained on nonselective screening medium plates with 1:100 diluted cells

^ccfu obtained on selective (S) screening medium (with 3-AT) plates.

^dtoo numerous to count

doi:10.1371/journal.pcbi.1004067.t001

turn to predictive methods. Here, we used coevolution calculations and a binary classifier to predict pairwise PPIs in the Pdu MCP, and proposed a model of its interactome. Approaches using binary classifiers for coevolution-based PPI predictions have been developed by others. Comparable approaches have been successfully applied to *E.coli* [87], and to the human genome [88]. Interpreting such networks is not a trivial task, considering that such methods are predictive in nature and can therefore include spurious predictions of PPIs or, *a contrario*, miss true interactions. Additionally, these methods cannot always distinguish direct (i.e. physical binding) and indirect (functional) correlations, a recurrent problem in coevolution studies that is illustrated here by the integration of PduF in our network. In order to mitigate the deficiencies of the computational methods we employed, a conservative approach was taken by considering only those predicted interactions that had the highest probability ($p \geq 0.7$). These cases were largely consistent with existing experimental data, where they were available. An example of a positive result is the agreement between our predictions and structural data relating to the reactivation mechanism of the diol dehydratase [80].

Extending on our predicted interactome model, we focused further analyses on PPIs emanating from the PduA shell protein node and involving Pdu enzymes (Fig. 3). Of these PPIs, five were identified as presenting an N-terminal extension, a characteristic of lumen-targeted enzymes. These N-terminal peptides, when docked onto a PduA hexamer, consistently bound the same cleft on the concave surface of the hexamer. Likewise, most of them folded into amphipathic α -helical structures, their hydrophobic faces oriented towards the C-terminal tail of the PduA shell protein, a region somewhat less conserved than the main BMC domain. These atomic details are depicted in Fig. 5, where for example the PduA-PduD case is more clearly pictured. These results are consistent with experimental studies by Fan *et al.*, which demonstrated the necessity of the PduA C-terminal helix in PduP binding and the role of hydrophobic residues in that interaction [36]. An exceptional case during these docking simulations was the PduL peptide, which did not fold into an amphipathic helix. With regard to our inability to obtain a robust docking result with a PduL peptide, it is notable that the interior vs exterior location of PduL remains unclear in current models of the Pdu MCP. If it is interior, its enzymatic reaction (depicted in Fig. 1B) could internally recycle the coenzyme A used by PduP for the conversion of propionaldehyde to propionyl-coA. Indeed, a similar mechanism is used for HS-CoA recycling by the Eut MCP [89] and has been demonstrated for NAD⁺ recycling by PduQ [90].

The results of our docking studies are of particular significance for the issue of sidedness of the MCP shell—i.e. which side of the shell proteins faces inward vs. outward. Previous arguments have suggested that the concave side of the shell protein faces into the MCP lumen [35,37,38]. Mutagenesis experiments by Fan *et al.* on the PduA C-terminal helix support that assignment [36]. In our present docking study, the consistent binding of the targeting peptides onto the concave side of the PduA hexamer, and the consistently better interface statistics compared to docking on the other side, strongly corroborate this idea.

PduA and PduJ, two highly similar paralogs of the BMC shell protein, are the two most abundant shell proteins after PduBB' in the Pdu system. As a consequence, they are suspected to play a critical structural role [73]. Indeed, while deletions of *pduK*, *pduT* or *pduU* do not affect the formation of the MCP, *pduA* mutations produce disrupted or enlarged shells [82,91]. Pull-down assays confirmed this architectural importance, where PduA was shown to interact with multiple other shell proteins [82]. Here we suggest that in addition to its transport and structural roles, PduA likely serves as a universal hub for a clique of cargo enzymes, attaching them to the shell via their N-terminal extensions. The highly similar shell protein PduJ is also predicted to interact with four of the same six enzymes associated with PduA. A possible interpretation is that the same clique of enzymes is able to bind both PduA and PduJ, some pairs

being more thermodynamically favored than others. Another explanation would be that these PPIs are in fact exclusive, but that our approach is not sensitive enough to discriminate PPIs involving close homologs. Note that the absence of an available structure for PduJ prevented a comparison by computational docking. Whether PduA and PduJ have similar or different affinities for various enzymatic partners will require further investigations, including experimental studies.

Attributing a special functional role to PduA (or PduJ) is consistent with the view that, though the multiple paralogous shell proteins in the MCP share a canonical BMC structure, each shell protein variant fulfills a specific task. For instance, tandem BMC proteins such as EutL are proposed to regulate the transport of metabolites via conformational changes and a gated pore [38,40,92,93]. The recent crystal structure of PduB, a EutL homolog, presents a view of a tandem domain shell protein from the Pdu system in a closed conformation [46]. Another apparently specialized shell protein is PduT, a tandem BMC domain shell protein that is suspected to bind an iron sulfur cluster in its central pore [39,45].

In this portrait of the Pdu family, the role of PduU remains to be elucidated. Here, we aimed to bring new clues by investigating the intriguing PduU-PduV case. Indeed, PduV is also poorly characterized compared to other Pdu components. Furthermore, from our predictions, PduV was the only enzyme exclusively interacting with a shell protein. The diverse docking simulations involving PduU and PduV all agreed with the existence of such a PPI, and predicted the N-terminal region of PduV binds directly to the PduU beta-barrel, consistent with recent experimental data on the importance of the N-terminus of PduV [82]. These predictions, coupled to our preliminary experimental data on a PduU-PduV interaction, fill a gap in understanding the role of the unique β -barrel in PduU.

To conclude, the present study brings further insights into the organization of the Pdu MCP, and constitutes the first systematic computational effort to describe an MCP interaction network. The basis of this work is predictive, but we have investigated one of the predicted interactions experimentally as part of this investigation, with a positive result. Further experimental studies will be required to more fully evaluate the interactome model developed here. Application of the same approach to other types of characterized MCPs might be of equal interest and could reveal similar insights.

Materials and Methods

Collection of Pdu operons and construction of phylogenetic trees

Protein orthologs were collected from 34 bacterial genomes in the KEGG database [94] and collapsed among the 22 types of MCP proteins known to be associated with the Pdu system: *pduABCDEFGHIJKLMNOQRSTUVWXYZ* (Suppl. Data). Incomplete or erroneous annotations of the Pdu gene products were corrected after sequence comparison with the Pdu operon from *Salmonella enterica* LT2, the best-characterized strain in terms of Pdu MCP.

For each ortholog group, its corresponding protein sequences were aligned with MUSCLE [95]. The multiple sequence alignments were subsequently input in PhyML [96] for the construction of phylogenetic trees using the Maximum Likelihood method. Since some of the co-evolution descriptors also involve the Tree of Life of the 34 genomes studied, sequences of their respective 16S ribosomal RNA were submitted to similar treatment. For amino acid and nucleotide-based tree construction in PhyML, we used the LG [97] and HKY85 [98] substitution matrices, respectively. Additionally, distance matrices were calculated for each tree, where the distance between two leaves corresponds to the sum of the branch lengths separating them.

Dataset construction

Seven coevolution descriptors measuring the pairwise tree similarities have been defined. Of these, four are based on pairwise comparison of the distance matrices, as defined in the *mirrortree* approach, while three others reflect topological similarities (Fig. 2). In the former class of descriptors, the metrics correspond to the linear correlation coefficient between the two matrices in consideration, while in the latter, it involves the congruence index I_{cong} as defined in Vienne *et al* [69]. Noteworthy is the fact that comparing two trees can be subject to artefacts and lead in some cases to spurious correlations if speciation events are not taken in account. For this reason, some of these descriptors involve comparisons of the individual proteins to the Tree of Life. Let A and B be the two MCP ortholog groups, m_A and m_B their respective matrices, t_A and t_B their trees, and ToL the Tree of Life of the 34 genomes. The parameter $mirror_{AB}$ is the correlation between m_A and m_B , $mirror_A$ is between m_A and ToL, and $mirror_B$ is between m_B and ToL. The fourth descriptor, $mirror_{AB-ToL}$, involves an adaptation of the mirror tree, also known as tol-mirror [68], which measures the correlation between m_A and m_B after removing the background similarity inherent to speciation events in the ToL. Since distances in the ToL are computed from a nucleotide-based substitution matrix, the distances in the ToL matrix have to be rescaled as proposed in [68] for proper comparison with the protein-based distance matrices.

Topological descriptors are derived from the I_{cong} index, defined as the probability that the Maximum Agreement Subtree (MAST) between two trees is arising by chance. Along the same idea, topological similarities were computed between tree A and ToL, tree B and ToL, and finally A and B (top_A, top_B, top_{AB}).

Binary classifier

We implemented a Random Forests (RF) classifier [71] from the Weka Library in Java [99]. Two classes were defined: *pos* for an interacting protein group pair and *neg* for those not interacting. Each of the ortholog group pairs sees its input vector of seven coevolution descriptors evaluated by the RF classifier. To classify a pair, its input vector is run through each decision tree of the forest and sees its mean probability attributed. The mean probability threshold for distinguishing the *pos* from the *neg* cases was set to 0.5, where a probability ≥ 0.5 will classify the pair as *pos*.

Gold standard and cross validation

The dataset used for training the RF classifier—the “gold standard”—was derived from experimental data found in the literature on the Pdu MCP. Manual mining of this data led to a total of 40 pairs of Pdu proteins whose physical interactions (or lack of interaction in many cases) could be verified experimentally via binding assays [36,82,90,100,101], complementation and expression studies [22] or crystallographic data [79]. An example of a verified non-interaction would be a direct binding experiment in which one protein component of a candidate pair failed to pull down the other. Among these, 16 are actual PPIs while the remaining 24 are non-interacting pairs. Of the 16 PPIs, 4, 6 and 6 pairs fall within the categories of: shell-enzyme (S-E) interactions, shell-shell (S-S) interactions and enzyme-enzyme (E-E) interactions, respectively. Likewise, the non-interacting pairs include 12 S-E, 12 S-S and zero E-E interactions. Each of these cases was assigned a class according to the rules defined earlier. The reported AUC value (0.75) for the classifier was calculated after a 10-fold cross validation. In parallel, we also carried out a 5 -fold cross validation that yielded a comparable AUC (0.73).

Interactome representation

The interactome was pictured as an undirected graph with the *igraph* library in R [102]. Nodes and edges were computed with a Fruchterman-Reingold layout [103].

Modeling of interacting partners with no structural information

While the structures of PduA and PduU are available in the PDB [104], structural information on the specific enzymes believed to interact with the shell proteins is limited. A recent NMR structure of the PduP tail showed an alpha helical structure, consistent with sequence-based predictions. Similar data are not available for the tails of the other enzymes of interest. We elected to assume as little as possible about the various tail structures and to model *ab initio* the first 18 residues of each enzyme with the PEP-FOLD server [105].

PduV was not presumed or predicted to bind by way of a terminal extension, so a model of that intact enzyme fold was required for docking analysis. The structure of PduV is presently unknown. Therefore, to enable computational docking, we built a homology model with I-TASSER [106] by threading the sequence of PduV onto two structural templates from the PDB (3IEV_A and 3R9W_A). The final model achieved a TM-score of 0.76, which is reasonable for further investigation by docking simulations [107].

Docking simulations

For protein-peptide docking, our approach relied mainly on the Rosetta-based protocol FlexPepDock [84]. Its ability to simultaneously fold and dock allows full flexibility of the peptide. However, FlexPepDock sees its accuracy decrease when the starting peptide conformation has an RMSD higher than 5.5 Å compared to the native structure. Mindful of this constraint, we designed a two-step method for docking the N-terminal enzymatic peptides onto the PduA hexamer. The first stage is a coarse-grained search of the approximate binding mode by AutoDock Vina [83]. This model is subsequently refined by an *ab initio* FlexPepDock run, where the Vina model is used as an input coordinates file. Vina has been designed for small molecule docking, which allows a ligand flexibility up to 32 rotatable bonds only, a limit not existing in FlexPepDock. However, it can still be used efficiently when medium-sized ligands like peptides are treated as semi-rigid for predicting an approximate binding region. File preparation for AutoDock Vina included a configuration file specifying an exhaustiveness of 10 and a 27000 Å³ grid box encompassing the surface of the hexamer and centered on the pore. Coordinate files in PDBQT format were generated from the PduA crystal structure and the PEP-FOLD models of each peptide. For the peptides, rotatable bonds were defined for the side chains while Kollman United Atom charges were assigned to both the hexamer and the peptides. The pose computed by Vina with the lowest energy score was subsequently considered as the starting point for FlexPepDock. In this second stage, we ran 10000 simulations where the peptide was completely refolded and docked into the PduA hexamer. After ranking the 10 000 poses by lowest Rosetta energy, the top 500 poses were collapsed into clusters for which the internal RMSD was less than 2.5 Å. Finally, we picked the definitive model as the one with the lowest energy among the two most populated clusters.

For the PduU-PduV case, we used a standard RosettaDock protocol where the input included coordinates of both partners in their unbound state, typically those from the PduU hexamer and the PduV homology model. The number of simulations, the ranking, clustering and selecting methods were identical to the FlexPepDock procedure, while the allowed flexibility in this case is limited to the side chains.

Two-hybrid assay on the PduU-PduV pair

To test for interactions between PduU and PduV, the BacterioMatch II two-hybrid system (Agilent technologies) was used according to the manufacturer's instructions with the following modification: co-transformation was carried out by using 30 ng each of the bait and prey vector. To construct the needed plasmids, *pduU* and *pduV* DNA sequences were amplified by PCR and then restricted and ligated into pBT for expression as fusions with the λ C1 protein, and into pTRG for expression as fusions with the RNAP α protein. Ligation reactions were used to transform *E. coli* XL1-Blue MRF'. Plasmid DNA was purified using a Qiagen mini prep kit, and all clones were verified by DNA sequencing. Self-activation by each recombinant bait and prey was tested before the two-hybrid interaction assays to determine if the bait or prey was capable of activating the reporter cassette on its own. Determination of protein-protein interaction was carried out by co-transforming BacterioMatch II validation reporter competent cells using recombinant bait and target.

Supporting Information

S1 Fig. Receiver Operating Characteristic(ROC) curves for the RF classifier with different combination of co-evolution descriptors. The quality of the RF classifier was assessed for three different combinations of coevolution descriptors: One combines the descriptors based on direct relationships between two proteins (A and B in Fig. 1) and exhibits an Area under the ROC Curve(AUC) of 0.47(green). The second scenario, which combines only the descriptors based on comparison between the Tree Of Life and one of the protein (A or B) obtains an AUC of 0.67(purple). A third case that uses all descriptors yields the best performance with an AUC of 0.75 (blue).

(TIF)

S2 Fig. Assessment of the classifier performance using incremental combinations of coevolution descriptors. AUC values were calculated after running the RF classifier with different incremental combinations of the descriptors, starting from the most accurate and adding the next best one at a time. Here again the classifier yields the best performance when combining all the descriptors.

(TIF)

S1 Dataset. List of probabilities of protein-protein interactions in the Pdu MCP. Protein-Protein interactions predicted by the RF classifier along with their respective mean probabilities (PPI in bold had $P > 0.7$ and were used for the construction of the Pdu interaction network pictured in Fig. 3).

(DOCX)

S2 Dataset. KEGG ID of the genes encoded within the Pdu operons analyzed in this study.

(DOCX)

Acknowledgments

The authors thank Dan E. McNamara, Michael C. Thompson and Eddy Kim for critical reading and helpful discussions.

Author Contributions

Conceived and designed the experiments: JJ TAB. Performed the experiments: JJ YL. Analyzed the data: JJ YL TAB TOY. Wrote the paper: JJ TOY YL TAB.

References

1. Fuerst JA, Webb RI (1991) Membrane-bounded nucleoid in the eubacterium *Gemmata obscuriglobus*. *Proc Natl Acad Sci U S A* 88: 8184–8188. PMID: [11607213](#)
2. Murat D, Quinlan A, Vali H, Komeili A (2010) Comprehensive genetic dissection of the magnetosome gene island reveals the step-wise assembly of a prokaryotic organelle. *Proc Natl Acad Sci U S A* 107: 5593–5598. doi: [10.1073/pnas.0914439107](#) PMID: [20212111](#)
3. Mullineaux CW (1999) The thylakoid membranes of cyanobacteria: structure, dynamics and function. *Funct Plant Biol* 26: 671–677.
4. Löwe J, Amos LA (1998) Crystal structure of the bacterial cell-division protein FtsZ. *Nature* 391: 203–206. PMID: [9428770](#)
5. Van den Ent F, Amos LA, Löwe J (2001) Prokaryotic origin of the actin cytoskeleton. *Nature* 413: 39–44. PMID: [11544518](#)
6. Gantt E, Conti SF (1969) Ultrastructure of blue-green algae. *J Bacteriol* 97: 1486–1493. PMID: [5776533](#)
7. Shively JM, Ball FL, Kline BW (1973) Electron Microscopy of the Carboxysomes (Polyhedral Bodies) of *Thiobacillus neapolitanus*. *J Bacteriol* 116: 1405–1411. PMID: [4127632](#)
8. Bobik TA (2006) Polyhedral organelles compartmenting bacterial metabolic processes. *Appl Microbiol Biotechnol* 70: 517–525. PMID: [16525780](#)
9. Kerfeld CA, Heinhorst S, Cannon GC (2010) Bacterial microcompartments. *Annu Rev Microbiol* 64: 391–408. doi: [10.1146/annurev.micro.112408.134211](#) PMID: [20825353](#)
10. Yeates TO, Kerfeld CA, Heinhorst S, Cannon GC, Shively JM (2008) Protein-based organelles in bacteria: carboxysomes and related microcompartments. *Nat Rev Microbiol* 6: 681–691. doi: [10.1038/nrmicro1913](#) PMID: [18679172](#)
11. Yeates TO, Crowley CS, Tanaka S (2010) Bacterial microcompartment organelles: protein shell structure and evolution. *Annu Rev Biophys* 39: 185–205. doi: [10.1146/annurev.biophys.093008.131418](#) PMID: [20192762](#)
12. Abdul-Rahman F, Petit E, Blanchard JL (2013) The Distribution of Polyhedral Bacterial Microcompartments Suggests Frequent Horizontal Transfer and Operon Reassembly Proteins. *J Phylogenetics Evol Biol*.
13. Jorda J, Lopez D, Wheatley NM, Yeates TO (2013) Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. *Protein Sci Publ Protein Soc* 22: 179–195. doi: [10.1002/pro.2196](#) PMID: [23188745](#)
14. Raushel FM, Thoden JB, Holden HM (2003) Enzymes with molecular tunnels. *Acc Chem Res* 36: 539–548. PMID: [12859215](#)
15. Penrod JT, Roth JR (2006) Conserving a Volatile Metabolite: a Role for Carboxysome-Like Organelles in *Salmonella enterica*. *J Bacteriol* 188: 2865–2874. doi: [10.1128/JB.188.8.2865-2874.2006](#) PMID: [16585748](#)
16. Rondon MR, Kazmierczak R, Escalante-Semerena JC (1995) Glutathione is required for maximal transcription of the cobalamin biosynthetic and 1,2-propanediol utilization (*cob/pdu*) regulon and for the catabolism of ethanolamine, 1,2-propanediol, and propionate in *Salmonella typhimurium* LT2. *J Bacteriol* 177: 5434–5439. PMID: [7559326](#)
17. Sampson EM, Bobik TA (2008) Microcompartments for B12-Dependent 1,2-Propanediol Degradation Provide Protection from DNA and Cellular Damage by a Reactive Metabolic Intermediate. *J Bacteriol* 190: 2966–2971. doi: [10.1128/JB.01925-07](#) PMID: [18296526](#)
18. Cannon GC, Bradburne CE, Aldrich HC, Baker SH, Heinhorst S, et al. (2001) Microcompartments in prokaryotes: carboxysomes and related polyhedra. *Appl Environ Microbiol* 67: 5351–5361. doi: [10.1128/AEM.67.12.5351-5361.2001](#) PMID: [11722879](#)
19. Shively JM, Ball F, Brown DH, Saunders RE (1973) Functional organelles in prokaryotes: polyhedral inclusions (carboxysomes) of *Thiobacillus neapolitanus*. *Science* 182: 584–586. PMID: [4355679](#)
20. Badger MR, Price GD (2003) CO₂ concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. *J Exp Bot* 54: 609–622. PMID: [12554704](#)
21. English RS, Lorbach SC, Qin X, Shively JM (1994) Isolation and characterization of a carboxysome shell gene from *Thiobacillus neapolitanus*. *Mol Microbiol* 12: 647–654. PMID: [7934888](#)
22. Bobik TA, Xu Y, Jeter RM, Otto KE, Roth JR (1997) Propanediol utilization genes (*pdu*) of *Salmonella typhimurium*: three genes for the propanediol dehydratase. *J Bacteriol* 179: 6633–6639. PMID: [9352910](#)
23. 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. PMID: [10498708](#)
24. Chen P, Ailion M, Bobik T, Stormo G, Roth J (1995) Five promoters integrate control of the cob/pdu regulon in *Salmonella typhimurium*. *J Bacteriol* 177: 5401–5410. PMID: [7559322](#)
 25. Kofoed E, Rappleye C, Stojiljkovic I, Roth J (1999) The 17-Gene Ethanolamine (eut) Operon of *Salmonella typhimurium* Encodes Five Homologues of Carboxysome Shell Proteins. *J Bacteriol* 181: 5317–5329. PMID: [10464203](#)
 26. Stojiljkovic I, Baumlner AJ, Heffron F (1995) Ethanolamine utilization in *Salmonella typhimurium*: nucleotide sequence, protein expression, and mutational analysis of the cchA cchB eutE eutJ eutG eutH gene cluster. *J Bacteriol* 177: 1357–1366. PMID: [7868611](#)
 27. Erbilgin O, McDonald KL, Kerfeld CA (2014) Characterization of a planctomycetal organelle: a novel bacterial microcompartment for the aerobic degradation of plant saccharides. *Appl Environ Microbiol* 80: 2193–2205. doi: [10.1128/AEM.03887-13](#) PMID: [24487526](#)
 28. Heldt D, Frank S, Seyedarabi A, Ladikis D, Parsons JB, et al. (2009) Structure of a trimeric bacterial microcompartment shell protein, EtuB, associated with ethanol utilization in *Clostridium kluyveri*. *Biochem J* 423: 199–207. doi: [10.1042/BJ20090780](#) PMID: [19635047](#)
 29. Lassila JK, Bernstein SL, Kinney JN, Axen SD, Kerfeld CA (2014) Assembly of Robust Bacterial Microcompartment Shells using Building Blocks from an Organelle of Unknown Function. *J Mol Biol*. Available: <http://www.sciencedirect.com/science/article/pii/S0022283614001260>. Accessed 20 March 2014.
 30. Petit E, LaTouf WG, Coppi MV, Warnick TA, Currie D, et al. (2013) Involvement of a Bacterial Microcompartment in the Metabolism of Fucose and Rhamnose by *Clostridium phytofermentans*. *PLoS ONE* 8: e54337. doi: [10.1371/journal.pone.0054337](#) PMID: [23382892](#)
 31. Cheng S, Liu Y, Crowley CS, Yeates TO, Bobik TA (2008) Bacterial microcompartments: their properties and paradoxes. *Bioessays* 30: 1084–1095. doi: [10.1002/bies.20830](#) PMID: [18937343](#)
 32. Kerfeld CA, Sawaya MR, Tanaka S, Nguyen CV, Phillips M, et al. (2005) Protein structures forming the shell of primitive bacterial organelles. *Science* 309: 936–938. PMID: [16081736](#)
 33. Tsai Y, Sawaya MR, Cannon GC, Cai F, Williams EB, et al. (2007) Structural analysis of CsoS1A and the protein shell of the *Halothiobacillus neapolitanus* carboxysome. *PLoS Biol* 5: e144. doi: [10.1371/journal.pbio.0050144](#) PMID: [17518518](#)
 34. Yeates TO, Thompson MC, Bobik TA (2011) The protein shells of bacterial microcompartment organelles. *Curr Opin Struct Biol* 21: 223–231. doi: [10.1016/j.sbi.2011.01.006](#) PMID: [21315581](#)
 35. Yeates TO, Jorda J, Bobik TA (2013) The shells of BMC-type microcompartment organelles in bacteria. *J Mol Microbiol Biotechnol* 23: 290–299. doi: [10.1159/000351347](#) PMID: [23920492](#)
 36. Fan C, Cheng S, Sinha S, Bobik TA (2012) Interactions between the termini of lumen enzymes and shell proteins mediate enzyme encapsulation into bacterial microcompartments. *Proc Natl Acad Sci* 109: 14995–15000. doi: [10.1073/pnas.1207516109](#) PMID: [22927404](#)
 37. Tanaka S, Kerfeld CA, Sawaya MR, Cai F, Heinhorst S, et al. (2008) Atomic-Level Models of the Bacterial Carboxysome Shell. *Science* 319: 1083–1086. doi: [10.1126/science.1151458](#) PMID: [18292340](#)
 38. Tanaka S, Sawaya MR, Phillips M, Yeates TO (2009) Insights from multiple structures of the shell proteins from the beta-carboxysome. *Protein Sci Publ Protein Soc* 18: 108–120. doi: [10.1002/pro.14](#) PMID: [19177356](#)
 39. Crowley CS, Cascio D, Sawaya MR, Kopstein JS, Bobik TA, et al. (2010) Structural Insight into the Mechanisms of Transport across the *Salmonella enterica* Pdu Microcompartment Shell. *J Biol Chem* 285: 37838–37846. doi: [10.1074/jbc.M110.160580](#) PMID: [20870711](#)
 40. Klein MG, Zwart P, Bagby SC, Cai F, Chisholm SW, et al. (2009) Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport. *J Mol Biol* 392: 319–333. doi: [10.1016/j.jmb.2009.03.056](#) PMID: [19328811](#)
 41. Sutter M, Wilson SC, Deutsch S, Kerfeld CA (2013) 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. doi: [10.1007/s1120-013-9909-z](#) PMID: [23949415](#)
 42. Wheatley NM, Gidaniyan SD, Liu Y, Cascio D, Yeates TO (2013) Bacterial microcompartment shells of diverse functional types possess pentameric vertex proteins. *Protein Sci* 22: 660–665. doi: [10.1002/pro.2246](#) PMID: [23456886](#)
 43. Cai F, Sutter M, Cameron JC, Stanley DN, Kinney JN, et al. (2013) The Structure of CcmP, a Tandem Bacterial Microcompartment Domain Protein from the β -Carboxysome, Forms a Subcompartment

- Within a Microcompartment. *J Biol Chem* 288: 16055–16063. doi: [10.1074/jbc.M113.456897](https://doi.org/10.1074/jbc.M113.456897) PMID: [23572529](https://pubmed.ncbi.nlm.nih.gov/23572529/)
44. Crowley CS, Sawaya MR, Bobik TA, Yeates TO (2008) Structure of the PduU shell protein from the Pdu microcompartment of Salmonella. *Struct Lond Engl* 16: 1324–1332. doi: [10.1016/j.str.2008.05.013](https://doi.org/10.1016/j.str.2008.05.013) PMID: [18786396](https://pubmed.ncbi.nlm.nih.gov/18786396/)
 45. Pang A, Warren MJ, Pickersgill RW (2011) Structure of PduT, a trimeric bacterial microcompartment protein with a 4Fe–4S cluster-binding site. *Acta Crystallogr D Biol Crystallogr* 67: 91–96. doi: [10.1107/S0907444910050201](https://doi.org/10.1107/S0907444910050201) PMID: [21245529](https://pubmed.ncbi.nlm.nih.gov/21245529/)
 46. Pang A, Liang M, Prentice MB, Pickersgill RW (2012) Substrate channels revealed in the trimeric *Lactobacillus reuteri* bacterial microcompartment shell protein PduB. *Acta Crystallogr D Biol Crystallogr* 68: 1642–1652. doi: [10.1107/S0907444912039315](https://doi.org/10.1107/S0907444912039315) PMID: [23151629](https://pubmed.ncbi.nlm.nih.gov/23151629/)
 47. Cameron JC, Wilson SC, Bernstein SL, Kerfeld CA (2013) Biogenesis of a Bacterial Organelle: The Carboxysome Assembly Pathway. *Cell* 155: 1131–1140. doi: [10.1016/j.cell.2013.10.044](https://doi.org/10.1016/j.cell.2013.10.044) PMID: [24267892](https://pubmed.ncbi.nlm.nih.gov/24267892/)
 48. Chen AH, Robinson-Mosher A, Savage DF, Silver PA, Polka JK (2013) The Bacterial Carbon-Fixing Organelle Is Formed by Shell Envelopment of Preassembled Cargo. *PLoS ONE* 8: e76127. doi: [10.1371/journal.pone.0076127](https://doi.org/10.1371/journal.pone.0076127) PMID: [24023971](https://pubmed.ncbi.nlm.nih.gov/24023971/)
 49. Kinney JN, Salmeen A, Cai F, Kerfeld CA (2012) Elucidating Essential Role of Conserved Carboxysomal Protein CcmN Reveals Common Feature of Bacterial Microcompartment Assembly. *J Biol Chem* 287: 17729–17736. doi: [10.1074/jbc.M112.355305](https://doi.org/10.1074/jbc.M112.355305) PMID: [22461622](https://pubmed.ncbi.nlm.nih.gov/22461622/)
 50. Fan C, Cheng S, Liu Y, Escobar CM, Crowley CS, et al. (2010) Short N-terminal sequences package proteins into bacterial microcompartments. *Proc Natl Acad Sci U S A* 107: 7509–7514. doi: [10.1073/pnas.0913199107](https://doi.org/10.1073/pnas.0913199107) PMID: [20308536](https://pubmed.ncbi.nlm.nih.gov/20308536/)
 51. Frank S, Lawrence AD, Prentice MB, Warren MJ (2013) Bacterial microcompartments moving into a synthetic biological world. *J Biotechnol* 163: 273–279. doi: [10.1016/j.jbiotec.2012.09.002](https://doi.org/10.1016/j.jbiotec.2012.09.002) PMID: [22982517](https://pubmed.ncbi.nlm.nih.gov/22982517/)
 52. Sargent F, Davidson FA, Kelly CL, Binny R, Christodoulides N, et al. (2013) A synthetic system for expression of components of a bacterial microcompartment. *Microbiology* 159: doi: [10.1099/mic.0.069922-0](https://doi.org/10.1099/mic.0.069922-0) PMID: [24014666](https://pubmed.ncbi.nlm.nih.gov/24014666/)
 53. Choudhary S, Quin MB, Sanders MA, Johnson ET, Schmidt-Dannert C (2012) Engineered Protein Nano-Compartments for Targeted Enzyme Localization. *PLoS ONE* 7. Available: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3299773/>. Accessed 20 March 2014. doi: [10.1371/journal.pone.0033342](https://doi.org/10.1371/journal.pone.0033342) PMID: [22428024](https://pubmed.ncbi.nlm.nih.gov/22428024/)
 54. Fan C, Bobik TA (2011) The N-Terminal Region of the Medium Subunit (PduD) Packages Adenosylcobalamin-Dependent Diol Dehydratase (PduCDE) into the Pdu Microcompartment ▽. *J Bacteriol* 193: 5623–5628. doi: [10.1128/JB.05661-11](https://doi.org/10.1128/JB.05661-11) PMID: [21821773](https://pubmed.ncbi.nlm.nih.gov/21821773/)
 55. Lawrence AD, Frank S, Newnham S, Lee MJ, Brown IR, et al. (2014) Solution Structure of a Bacterial Microcompartment Targeting Peptide and Its Application in the Construction of an Ethanol Bioreactor. *ACS Synth Biol*. Available: <http://dx.doi.org/10.1021/sb4001118>. Accessed 20 March 2014.
 56. Huynen MA, Bork P (1998) Measuring genome evolution. *Proc Natl Acad Sci U S A* 95: 5849–5856. PMID: [9600883](https://pubmed.ncbi.nlm.nih.gov/9600883/)
 57. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci U S A* 96: 4285–4288. PMID: [10200254](https://pubmed.ncbi.nlm.nih.gov/10200254/)
 58. Enright AJ, Iliopoulos I, Kyrpides NC, Ouzounis CA (1999) Protein interaction maps for complete genomes based on gene fusion events. *Nature* 402: 86–90. PMID: [10573422](https://pubmed.ncbi.nlm.nih.gov/10573422/)
 59. Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, et al. (1999) Detecting protein function and protein-protein interactions from genome sequences. *Science* 285: 751–753. PMID: [10427000](https://pubmed.ncbi.nlm.nih.gov/10427000/)
 60. Dandekar T, Snel B, Huynen M, Bork P (1998) Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem Sci* 23: 324–328. PMID: [9787636](https://pubmed.ncbi.nlm.nih.gov/9787636/)
 61. Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N (1999) The use of gene clusters to infer functional coupling. *Proc Natl Acad Sci* 96: 2896–2901. doi: [10.1073/pnas.96.6.2896](https://doi.org/10.1073/pnas.96.6.2896) PMID: [10077608](https://pubmed.ncbi.nlm.nih.gov/10077608/)
 62. Bowers PM, Pellegrini M, Thompson MJ, Fierro J, Yeates TO, et al. (2004) Prolinks: a database of protein functional linkages derived from coevolution. *Genome Biol* 5: R35. doi: [10.1186/gb-2004-5-5-r35](https://doi.org/10.1186/gb-2004-5-5-r35) PMID: [15128449](https://pubmed.ncbi.nlm.nih.gov/15128449/)
 63. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, et al. (2009) STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* 37: D412–416. doi: [10.1093/nar/gkn760](https://doi.org/10.1093/nar/gkn760) PMID: [18940858](https://pubmed.ncbi.nlm.nih.gov/18940858/)

64. Von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, et al. (2003) STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res* 31: 258–261. PMID: [12519996](#)
65. De Juan D, Pazos F, Valencia A (2013) Emerging methods in protein co-evolution. *Nat Rev Genet* 14: 249–261. doi: [10.1038/nrg3414](#) PMID: [23458856](#)
66. Pazos F, Valencia A (2008) Protein co-evolution, co-adaptation and interactions. *EMBO J* 27: 2648–2655. doi: [10.1038/emboj.2008.189](#) PMID: [18818697](#)
67. Pazos F, Valencia A (2001) Similarity of phylogenetic trees as indicator of protein-protein interaction. *Protein Eng* 14: 609–614. PMID: [11707606](#)
68. Pazos F, Ranea JAG, Juan D, Sternberg MJE (2005) Assessing protein co-evolution in the context of the tree of life assists in the prediction of the interactome. *J Mol Biol* 352: 1002–1015. PMID: [16139301](#)
69. de VienneDM, Giraud T, Martin OC (2007) A congruence index for testing topological similarity between trees. *Bioinformatics* 23: 3119–3124. PMID: [17933852](#)
70. Craig RA, Liao L (2007) Phylogenetic tree information aids supervised learning for predicting protein-protein interaction based on distance matrices. *BMC Bioinformatics* 8: 6. doi: [10.1186/1471-2105-8-6](#) PMID: [17212819](#)
71. Breiman L (2001) Random Forests. *Mach Learn* 45: 5–32.
72. Fan C, Fromm HJ, Bobik TA (2009) Kinetic and Functional Analysis of l-Threonine Kinase, the PduX Enzyme of *Salmonella enterica*. *J Biol Chem* 284: 20240–20248. doi: [10.1074/jbc.M109.027425](#) PMID: [19509296](#)
73. 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. doi: [10.1128/JB.185.17.5086-5095.2003](#) PMID: [12923081](#)
74. Chen P, Andersson DI, Roth JR (1994) The control region of the pdu/cob regulon in *Salmonella typhimurium*. *J Bacteriol* 176: 5474–5482. PMID: [8071226](#)
75. Grigoriev A (2003) On the number of protein-protein interactions in the yeast proteome. *Nucleic Acids Res* 31: 4157–4161. PMID: [12853633](#)
76. Paris L, Bazzoni G (2008) The Protein Interaction Network of the Epithelial Junctional Complex: A System-Level Analysis. *Mol Biol Cell* 19: 5409–5421. doi: [10.1091/mbc.E08-05-0477](#) PMID: [18923145](#)
77. Honda S, Toraya T, Fukui S (1980) In situ reactivation of glycerol-inactivated coenzyme B12-dependent enzymes, glycerol dehydratase and diol dehydratase. *J Bacteriol* 143: 1458–1465. PMID: [6997273](#)
78. Shibata N, Masuda J, Tobimatsu T, Toraya T, Suto K, et al. (1999) A new mode of B12 binding and the direct participation of a potassium ion in enzyme catalysis: X-ray structure of diol dehydratase. *Struct Lond Engl* 1993 7: 997–1008. PMID: [10467140](#)
79. Shibata N, Mori K, Hieda N, Higuchi Y, Yamanishi M, et al. (2005) Release of a damaged cofactor from a coenzyme B12-dependent enzyme: X-ray structures of diol dehydratase-reactivating factor. *Struct Lond Engl* 1993 13: 1745–1754. PMID: [16338403](#)
80. Mori K, Hosokawa Y, Yoshinaga T, Toraya T (2010) Diol dehydratase-reactivating factor is a reactivase—evidence for multiple turnovers and subunit swapping with diol dehydratase. *FEBS J* 277: 4931–4943. doi: [10.1111/j.1742-4658.2010.07898.x](#) PMID: [21040475](#)
81. Havemann GD, Sampson EM, Bobik TA (2002) PduA is a shell protein of polyhedral organelles involved in coenzyme B(12)-dependent degradation of 1,2-propanediol in *Salmonella enterica* serovar typhimurium LT2. *J Bacteriol* 184: 1253–1261. doi: [10.1128/JB.184.5.1253-1261.2002](#) PMID: [11844753](#)
82. Parsons JB, Frank S, Bhella D, Liang M, Prentice MB, et al. (2010) Synthesis of empty bacterial microcompartments, directed organelle protein incorporation, and evidence of filament-associated organelle movement. *Mol Cell* 38: 305–315. doi: [10.1016/j.molcel.2010.04.008](#) PMID: [20417607](#)
83. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31: 455–461. doi: [10.1002/jcc.21334](#) PMID: [19499576](#)
84. Raveh B, London N, Zimmerman L, Schueler-Furman O (2011) Rosetta FlexPepDock ab-initio: simultaneous folding, docking and refinement of peptides onto their receptors. *PLoS One* 6: e18934. doi: [10.1371/journal.pone.0018934](#) PMID: [21572516](#)
85. Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, et al. (2003) Protein-Protein Docking with Simultaneous Optimization of Rigid-body Displacement and Side-chain Conformations. *J Mol Biol* 331: 281–299. PMID: [12875852](#)

86. Joung JK, Ramm EI, Pabo CO (2000) A bacterial two-hybrid selection system for studying protein–DNA and protein-protein interactions. *Proc Natl Acad Sci* 97: 7382–7387. doi: [10.1073/pnas.110149297](https://doi.org/10.1073/pnas.110149297) PMID: [10852947](https://pubmed.ncbi.nlm.nih.gov/10852947/)
87. De Vienne DM, Azé J (2012) Efficient Prediction of Co-Complexed Proteins Based on Coevolution. *PLoS ONE* 7: e48728. doi: [10.1371/journal.pone.0048728](https://doi.org/10.1371/journal.pone.0048728) PMID: [23152796](https://pubmed.ncbi.nlm.nih.gov/23152796/)
88. Liu CH, Li K-C, Yuan S (2013) Human protein–protein interaction prediction by a novel sequence-based co-evolution method: co-evolutionary divergence. *Bioinformatics* 29: 92–98. doi: [10.1093/bioinformatics/bts620](https://doi.org/10.1093/bioinformatics/bts620) PMID: [23080115](https://pubmed.ncbi.nlm.nih.gov/23080115/)
89. Huseby DL, Roth JR (2013) Evidence that a metabolic microcompartment contains and recycles private cofactor pools. *J Bacteriol* 195: 2864–2879. doi: [10.1128/JB.02179-12](https://doi.org/10.1128/JB.02179-12) PMID: [23585538](https://pubmed.ncbi.nlm.nih.gov/23585538/)
90. Cheng S, Fan C, Sinha S, Bobik TA (2012) The PduQ Enzyme Is an Alcohol Dehydrogenase Used to Recycle NAD⁺ Internally within the Pdu Microcompartment of *Salmonella enterica*. *PLoS ONE* 7: e47144. doi: [10.1371/journal.pone.0047144](https://doi.org/10.1371/journal.pone.0047144) PMID: [23077559](https://pubmed.ncbi.nlm.nih.gov/23077559/)
91. Cheng S, Sinha S, Fan C, Liu Y, Bobik TA (2011) Genetic Analysis of the Protein Shell of the Microcompartments Involved in Coenzyme B₁₂-Dependent 1,2-Propanediol Degradation by *Salmonella*. *J Bacteriol* 193: 1385–1392. doi: [10.1128/JB.01473-10](https://doi.org/10.1128/JB.01473-10) PMID: [21239588](https://pubmed.ncbi.nlm.nih.gov/21239588/)
92. Sagermann M, Ohtaki A, Nikolakakis K (2009) Crystal structure of the EutL shell protein of the ethanolamine ammonia lyase microcompartment. *Proc Natl Acad Sci* 106: 8883–8887. doi: [10.1073/pnas.0902324106](https://doi.org/10.1073/pnas.0902324106) PMID: [19451619](https://pubmed.ncbi.nlm.nih.gov/19451619/)
93. Takenoya M, Nikolakakis K, Sagermann M (2010) Crystallographic Insights into the Pore Structures and Mechanisms of the EutL and EutM Shell Proteins of the Ethanolamine-Utilizing Microcompartment of *Escherichia coli*. *J Bacteriol* 192: 6056–6063. doi: [10.1128/JB.00652-10](https://doi.org/10.1128/JB.00652-10) PMID: [20851901](https://pubmed.ncbi.nlm.nih.gov/20851901/)
94. Kanehisa M (2002) The KEGG database. *Novartis Found Symp* 247: 91–101; discussion 101–103, 119–128, 244–252. PMID: [12539951](https://pubmed.ncbi.nlm.nih.gov/12539951/)
95. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792–1797. doi: [10.1093/nar/gkh340](https://doi.org/10.1093/nar/gkh340) PMID: [15034147](https://pubmed.ncbi.nlm.nih.gov/15034147/)
96. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59: 307–321. doi: [10.1093/sysbio/syq010](https://doi.org/10.1093/sysbio/syq010) PMID: [20525638](https://pubmed.ncbi.nlm.nih.gov/20525638/)
97. Le SQ, Gascuel O (2008) An Improved General Amino Acid Replacement Matrix. *Mol Biol Evol* 25: 1307–1320. doi: [10.1093/molbev/msn067](https://doi.org/10.1093/molbev/msn067) PMID: [18367465](https://pubmed.ncbi.nlm.nih.gov/18367465/)
98. Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 22: 160–174. PMID: [3934395](https://pubmed.ncbi.nlm.nih.gov/3934395/)
99. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, et al. (2009) The WEKA Data Mining Software: An Update. *SIGKDD Explor Newsl* 11: 10–18.
100. Parsons JB, Lawrence AD, McLean KJ, Munro AW, Rigby SEJ, et al. (2010) Characterisation of PduS, the pdu metabolosome corrin reductase, and evidence of substructural organisation within the bacterial microcompartment. *PloS One* 5: e14009. doi: [10.1371/journal.pone.0014009](https://doi.org/10.1371/journal.pone.0014009) PMID: [21103360](https://pubmed.ncbi.nlm.nih.gov/21103360/)
101. Cheng S, Bobik TA (2010) Characterization of the PduS cobalamin reductase of *Salmonella enterica* and its role in the Pdu microcompartment. *J Bacteriol* 192: 5071–5080. doi: [10.1128/JB.00575-10](https://doi.org/10.1128/JB.00575-10) PMID: [20656910](https://pubmed.ncbi.nlm.nih.gov/20656910/)
102. Csardi G, Nepusz T (2006) The igraph software package for complex network research. *InterJournal Complex Syst* 1695: 1695.
103. Fruchterman TMJ, Reingold EM (1991) Graph drawing by force-directed placement. *Softw Pract Exp* 21: 1129–1164. PMID: [16805262](https://pubmed.ncbi.nlm.nih.gov/16805262/)
104. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, et al. (2002) The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr* 58: 899–907. PMID: [12037327](https://pubmed.ncbi.nlm.nih.gov/12037327/)
105. Maupetit J, Derreumaux P, Tuffery P (2009) PEP-FOLD: an online resource for de novo peptide structure prediction. *Nucleic Acids Res* 37: W498–503. doi: [10.1093/nar/gkp323](https://doi.org/10.1093/nar/gkp323) PMID: [19433514](https://pubmed.ncbi.nlm.nih.gov/19433514/)
106. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5: 725–738. doi: [10.1038/nprot.2010.5](https://doi.org/10.1038/nprot.2010.5) PMID: [20360767](https://pubmed.ncbi.nlm.nih.gov/20360767/)
107. Zhang Y, Skolnick J (2004) Scoring function for automated assessment of protein structure template quality. *Proteins* 57: 702–710. PMID: [15476259](https://pubmed.ncbi.nlm.nih.gov/15476259/)