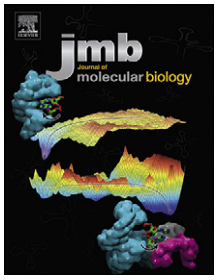


Identification of a Unique Fe-S Cluster Binding Site in a Glycyl-Radical Type Microcompartment Shell Protein

Michael C. Thompson [1], Nicole M. Wheatley [2], Julien Jorda [3], Michael R. Sawaya [3], Soheil D. Gidaniyan [1], Hoda Ahmed [1], Zhongyu Yang [4], Krystal N. McCarty [1], Julian P. Whitelegge [5] and Todd O. Yeates [1,3,6]

*1 - **Department of Chemistry and Biochemistry,** University of California, Los Angeles, CA 90095, USA*
*2 - **Molecular Biology Interdepartmental Ph.D. Program,** University of California, Los Angeles, CA 90095, USA*
*3 - **University of California,** Los Angeles–Department of Energy Institute for Genomics and Proteomics, University of California, Los Angeles, CA 90095, USA*
*4 - **Jules Stein Eye Institute,** University of California, Los Angeles, CA 90095, USA*
*5 - **Pasarow Mass Spectrometry Laboratory,** Neuropsychiatric Institute–Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, CA 90095, USA*
*6 - **Molecular Biology Institute,** University of California, Los Angeles, CA 90095, USA*

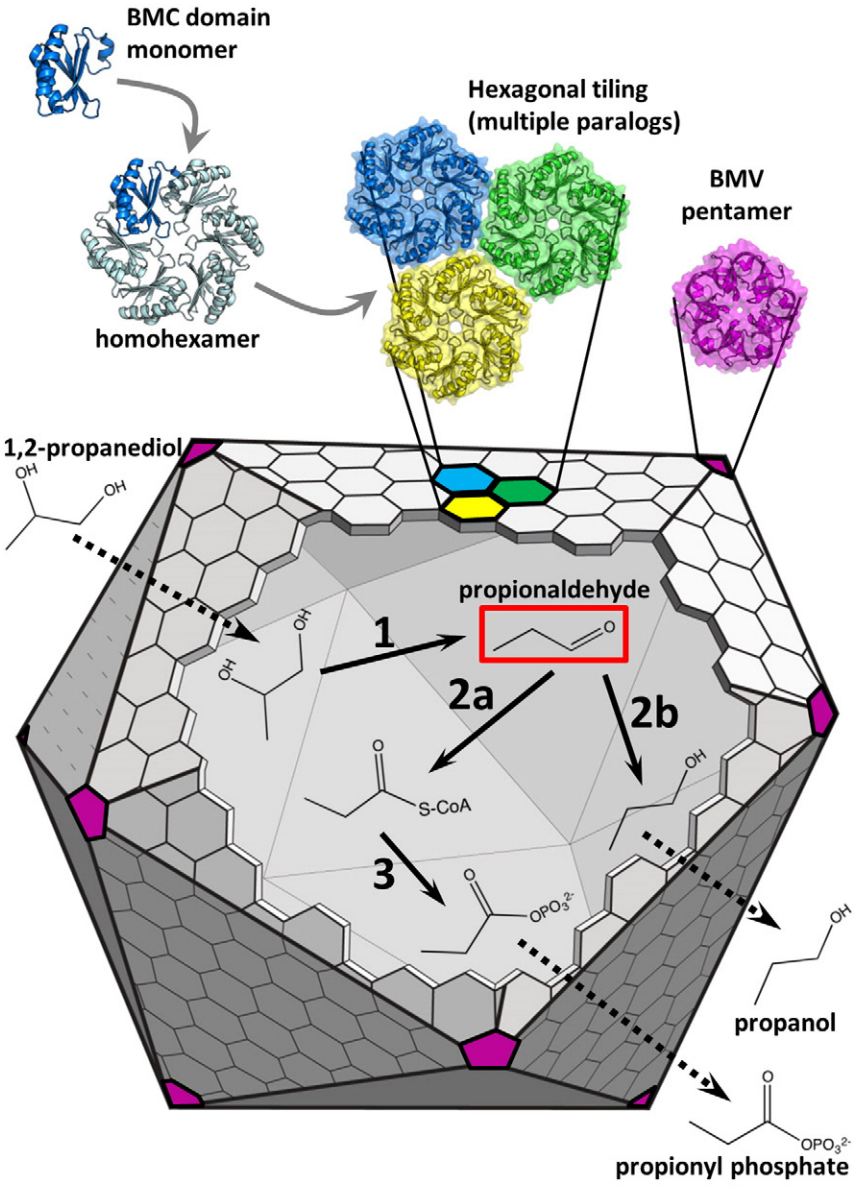
***Correspondence to Todd O. Yeates:** Department of Chemistry and Biochemistry, University of California, Los Angeles, 611 Charles E. Young Drive East, Los Angeles, CA 90095, USA. yeates@mbi.ucla.edu*
http://dx.doi.org/10.1016/j.jmb.2014.07.018
Edited by J. H. Naismith

Abstract

Recently, progress has been made toward understanding the functional diversity of bacterial microcompartment (MCP) systems, which serve as protein-based metabolic organelles in diverse microbes. New types of MCPs have been identified, including the glycyl-radical propanediol (Grp) MCP. Within these elaborate protein complexes, BMC-domain shell proteins [bacterial microcompartment (in reference to the shell protein domain)] assemble to form a polyhedral barrier that encapsulates the enzymatic contents of the MCP. Interestingly, the Grp MCP contains a number of shell proteins with unusual sequence features. GrpU is one such shell protein whose amino acid sequence is particularly divergent from other members of the BMC-domain superfamily of proteins that effectively defines all MCPs. Expression, purification, and subsequent characterization of the protein showed, unexpectedly, that it binds an iron-sulfur cluster. We determined X-ray crystal structures of two GrpU orthologs, providing the first structural insight into the homohexameric BMC-domain shell proteins of the Grp system. The X-ray structures of GrpU, both obtained in the apo form, combined with spectroscopic analyses and computational modeling, show that the metal cluster resides in the central pore of the BMC shell protein at a position of broken 6-fold symmetry. The result is a structurally polymorphic iron-sulfur cluster binding site that appears to be unique among metalloproteins studied to date.

Introduction

Prokaryotic organisms have evolved remarkable metabolic diversity through adaptation to highly varied environments. Many organisms have specialized metabolic pathways that allow them to utilize small molecules that are abundant within their specific niches as key carbon sources. In a number of intriguing cases, however, these specialized metabolic pathways produce intermediate compounds that are cytotoxic, that are mutagenic, and/or that diffuse freely across the cell membrane at physiological pressure and temperature [1–3]. In order to confine these mischievous intermediates and to increase metabolic flux, many bacteria encapsulate sequential enzymes from such metabolic pathways within giant protein complexes called bacterial microcompartments (MCPs) [4–7]. Although they are completely proteinaceous and lack any lipid membrane, MCPs are functionally analogous to eukaryotic organelles based on their ability to sequester specific subcellular processes. To date, MCPs have been associated with



	Pdu	Grp
Reaction 1	B_{12}-dependent diol dehydratase	glycyl-radical diol dehydratase
Reaction 2a	phosphotransacylase	
Reaction 2b	propionaldehyde dehydrogenase	
Reaction 3	alcohol dehydrogenase	

Fig. 1 (*legend on next page*)

at least 7 specific metabolic pathways [8,9], and MCP operons have been detected in approximately 17% of sequenced bacterial genomes.

The ability of MCPs to carry out their highly sophisticated functions can be attributed to the unique structure of their outer shells. This shell must act as a semi-permeable diffusion barrier, restricting the efflux of small, nonpolar intermediates while allowing the passage of more polar, but similarly sized substrates and products [1,6,10,11]. The semi-permeability of the shell is the result of its assembly from thousands of BMC-domain shell proteins [bacterial microcompartment (in reference to the shell protein domain)] (Fig. 1). These small (~100-residue) protein domains oligomerize into cyclic hexamers, which further assemble into tightly packed, two-dimensional molecular sheets [12,13]. Integration of pentameric vertex proteins [BMV proteins (bacterial microcompartment vertex, in reference to the protein domain)] allows these molecular sheets to form a closed polyhedral shell, which is perforated only by small pores found at the centers of some shell protein hexamers [14–16]. An additional level of complexity is added by the presence of multiple BMC protein paralogs simultaneously within a single MCP shell [17–20]. Based on X-ray crystal structures of paralogous BMC proteins [13,20,21] and phenotypic analysis of MCPs with varied BMC protein mutations [22–24], it has become clear that each paralog serves a unique role within the context of the shell. The putative functions of various types of BMC protein paralogs have been reviewed extensively [6,10,11].

In a number of saccharolytic pathogens, the initial steps of 1,2-propanediol catabolism are sequestered within an MCP (Fig. 1) [3,17]. When these pathogenic bacteria infect their hosts, they are capable of hydrolyzing cell-surface glycans from the host tissue, releasing monosaccharides, which become a primary source of metabolic energy for the proliferating bacteria [25]. These carbohydrate molecules, notably fucose and rhamnose, are first degraded to 1,2-propanediol [26–28]. Propanediol is then metabolized to propionate, which enters the methylcitrate cycle, providing a source of metabolic energy [3]. The conversion of 1,2-propanediol to propionate proceeds through a reactive propionaldehyde intermediate, which is cytotoxic and mutagenic if released into the cytoplasm [3]. Organisms that are capable of propanediol catabolism, therefore, must encapsulate several steps of the pathway within an MCP to sequester the aldehyde intermediate.

Two distinct types of 1,2-propanediol catabolic MCPs have been identified and studied previously (Fig. 1). The first of these two types, the Pdu MCP (for *p*ropane*d*iol *u*tilization), has been studied for a number of years and is relatively well understood. Within the Pdu MCP, 1,2-propanediol is first converted to propionaldehyde by a vitamin-B_{12}-dependent diol dehydratase enzyme (PduCDE) [3,26]. The aldehyde is then converted to propionyl-CoA by an aldehyde dehydrogenase enzyme or to propanol by an alcohol dehydrogenase enzyme, which facilitates recycling of the NAD^+ cofactor within the MCP [29]. In the final enzymatic step that occurs within the MCP lumen, propionyl-CoA is converted to propionyl-phosphate. The second type of propanediol catabolic MCP has been implicated only recently through microarray analysis [27] and a comparative genomic analysis [8]. This MCP degrades 1,2-propanediol using a series of enzymatic reactions similar to those found in the Pdu MCP, except that the initial dehydration reaction is catalyzed by a glycyl-radical diol dehydratase enzyme, rather than a vitamin-B_{12}-dependent enzyme. Consequently, this MCP has been named *g*lycyl-*r*adical *p*ropanediol (Grp). In the case of the Grp MCP, it has been suggested that, in addition to retaining the propionaldehyde intermediate, the MCP shell might also maintain an anaerobic environment for the glycyl-radical enzyme, whose activated state is sensitive to oxygen-mediated cleavage of the polypeptide backbone [30].

In contrast to the Pdu MCP, which has been studied in detail, only limited experimental work has been performed to characterize the Grp MCP. Along with a genomic analysis that highlighted the Grp MCP operon, Jorda *et al.* provided initial biochemical analysis of a BMC-domain shell protein from the Grp MCP [8]. Almost simultaneously, Petit *et al.* demonstrated that *Clostridium phytofermentans* expresses Grp-type MCPs during the fermentation of fucose and rhamnose [27]. Shortly after, Wheatley *et al.* determined the structure of a pentameric vertex protein, GrpN, from the Grp MCP of *Rhodospirillum rubrum* [15]. These three reports provide the only experimental characterization of the Grp MCP to date. Because of its overall similarity to the Pdu MCP, some of the functional details of the Grp MCP can be inferred by analogy. However, the difference in the first enzymatic step between these two systems also implies a certain degree of functional divergence. For example, the Pdu MCP contains a system of enzymes (PduGHOS) that regenerates inactivated B_{12} cofactors [17,31,31,32], while the

Fig. 1. Schematic of propanediol catabolic MCPs. Degradation of 1,2-propanediol occurs in an MCP. Homohexameric BMC-domain shell protein paralogs assemble into a proteinaceous shell surrounding a series of internalized enzymes that convert the substrate 1,2-propanediol into the products propanol and propionyl-phosphate. The chemistry involves a toxic intermediate (propionaldehyde), which is highlighted by a red box. The first of these enzymatic steps differs in Pdu-type *versus* Grp-type MCPs. Black broken arrows represent molecular transport events, while black continuous arrows represent enzymatic reactions.

Grp MCP instead contains a glycyl-radical enzyme activase [8]. Each of these two biochemical systems requires access to a different set of substrates, cofactors, and reducing equivalents, which indicates that the protein shells of the Pdu and Grp systems must differ to some extent with respect to their function as a diffusion barrier. Therefore, further experimental characterization of the protein shell of the Grp MCP should highlight the similarities and differences between the Pdu and Grp MCP systems and will provide insight into how the Grp MCP supports the activation of an encapsulated glycyl-radical enzyme.

Some Grp operons contain a gene encoding a particularly divergent BMC-domain shell protein. Analysis of various MCP operons revealed that it is fairly common for a particular type of MCP to vary in gene composition from one species to another [7,8]. These species-to-species variations in MCP composition are presumed to represent functional adaptations that modify the "core" MCP machinery, which is invariant between species. In the genomic analysis of the Grp MCP, Jorda *et al.* counted that 9 out of the 23 Grp operons they examined contain a gene corresponding to a divergent BMC-domain shell protein [8]. The closest homolog of this polypeptide in the Pdu MCP system is the PduU shell protein, and consequently, we have named this divergent Grp shell protein "GrpU". The GrpU protein is an interesting subject of inquiry for several reasons. First, because it is not universally conserved within Grp operons, it represents a functional variation on the core Grp MCP system. Second, the GrpU sequence is highly divergent from other BMC-domain proteins, and therefore, the degree of structural similarity between this BMC-domain protein and other members of the protein family is unclear.

Here we present a structural study of two GrpU orthologs from phylogenetically distant bacterial species. The first of these GrpU proteins is from *Clostridiales bacterium* 1_7_47FAA and is referred to hereafter as "*Clost*_GrpU". While little information is available about this particular species, the *Clost*_GrpU sequence is closely related to GrpU sequences from other *Clostridia* that are human pathogens, including *Clostridium botulinum* and *Clostridium tetani*, the causative agents of botulism and tetanus. The other GrpU sequence used in this work is from *Pectobacterium wasabiae*, a plant pathogen that causes stem rot. This sequence is hereafter referred to as "*Pecwa*_GrpU". Our work combines X-ray crystallography, spectroscopic methods, and molecular modeling to demonstrate that the GrpU shell protein is a novel type of BMC-domain metalloprotein that binds an iron-sulfur (Fe-S) cluster using a conserved, but uniquely polymorphic, binding motif that has not been observed before.

Results

Spectroscopic evidence of Fe-S clusters in GrpU

Solutions of purified GrpU proteins appear brownish in color and have spectral features consistent with the presence of an iron-sulfur (Fe-S) cluster. Electronic absorption spectra (UV-visible) of purified GrpU solutions reveal broad peaks near 420 nm (Fig. 2a). Specifically, the absorption maxima of these peaks occur at 411 nm for *Clost*_GrpU and at 415 nm for *Pecwa*_GrpU. Absorbance in this region of the electromagnetic spectrum is known to result from Fe-S charge-transfer bands [33]. Additionally, electron paramagnetic resonance (EPR) spectra of untreated and dithionite-reduced *Clost*_GrpU samples provide further evidence for the presence of an Fe-S cluster (Fig. 2b). The continuous-wave, X-band EPR spectrum of the untreated sample is relatively featureless. However, chemical reduction of the

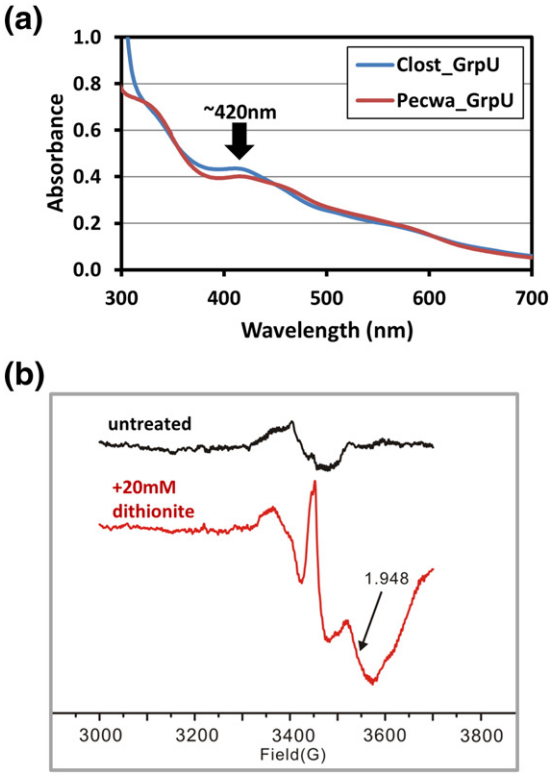


Fig. 2. Spectroscopic characterization of GrpU proteins. Absorbance spectra (a) are shown for *Clost*_GrpU and *Pecwa*_GrpU, demonstrating broad peaks at 411 nm and 415 nm, respectively. Absorption maxima in this region, near 420 nm, are characteristic of Fe-S charge-transfer bands [33]. Continuous-wave EPR spectra for *Clost*_GrpU (b) reveal the appearance of a prominent feature upon reduction of the protein with dithionite, which is characteristic of a bound Fe-S cluster [33,34].

sample with dithionite produces a distinct signal with a *g*-factor of 1.948. Our spectroscopic analysis reveals that GrpU likely binds to a [4Fe-4S] cluster, though line-broadening in the EPR spectrum makes it impossible to completely rule out the possibility that the ligand is a less common [2Fe-2S] cluster [33,34].

We note that similar spectroscopic features have been observed for the PduT BMC protein, which were demonstrated to result from the presence of a [4Fe-4S] cluster [24]. Based on the similarities of the Grp and Pdu MCP systems, it is likely that GrpU and PduT coordinate similar metal clusters. Furthermore, we observed a disappearance of spectroscopic features after storage of the samples for several days at 4 °C, indicating that the Fe-S cluster is somewhat labile and prone to dissociation upon protein oxidation.

In order to confirm that the observed spectroscopic phenomena do indeed result from an Fe-S cluster bound to GrpU and not from some other metal species, we also performed size-exclusion chromatography followed by inductively coupled plasma mass spectrometry (ICP-MS) to analyze the elemental composition of *Clost*_GrpU. Elution of the protein from size-exclusion chromatography was monitored by measurement of UV absorbance, followed immediately by ICP-MS detection. Comparison of the UV and ion chromatograms clearly demonstrates the co-elution of the *Clost*_GrpU hexamer with iron (^{56}Fe) (Fig. S1). In contrast, ion chromatograms for manganese (^{55}Mn), nickel (^{58}Ni), cobalt (^{59}Co), copper (^{63}Cu), zinc (^{66}Zn), and molybdenum (^{95}Mo) show no significant peaks co-eluting with the GrpU hexamer.

GrpU structure determination

We determined X-ray crystal structures of two GrpU homologs, *Clost*_GrpU and *Pecwa*_GrpU, at 2.5 Å and 2.8 Å resolution, respectively. In both cases, structure determination presented considerable difficulty due to the high level of symmetry and high Wilson *B*-factors for the X-ray data. Despite these difficulties, solutions were found in space group $P2_13$ for *Clost*_GrpU and space group $H3$ for *Pecwa*_GrpU. In both crystal structures, GrpU forms homohexamers that sit upon 3-fold crystallographic symmetry axes. Additionally, 2-fold noncrystallographic symmetry (NCS) axes lie perpendicular to the crystallographic 3-folds, resulting in what appear to be homododecameric arrangements of monomers with apparent 622 point group symmetry. Following molecular replacement, we performed restrained refinement of the atomic coordinates, which converged to yield final models that agreed reasonably well with the observed X-ray data and had excellent molecular geometry (Table 1).

In general, the *Clost*_GrpU and *Pecwa*_GrpU structures have similar features (Fig. 3). Likelihood-weighted electron density maps ($2mF_o - DF_c$ and $mF_o - DF_c$) used to generate our structural models were somewhat noisy, likely due to the presence of disordered segments in the crystal lattice (Fig. S2). Nevertheless, we were able to reliably model residues 1–19, 32–64, and 70–102 of *Clost*_GrpU, as well as residues 2–67 and 72–94 of *Pecwa*_GrpU. The *Clost*_GrpU and *Pecwa*_GrpU sequences are 40% identical, and consequently, our two structures have a low coordinate RMSD; any of the four independent protein molecules in our *Clost*_GrpU structure can be superimposed upon any of the four molecules in our *Pecwa*_GrpU structure with coordinate RMSD values of 1–2 Å calculated on C^{α} atoms. GrpU molecules that are part of the same hexamer, but are crystallographically independent, can be superimposed on each other with coordinate RMSD values of less than 1 Å.

Unfortunately, neither of our electron density maps contains features corresponding to the presumptive Fe-S clusters. A number of phenomena could be responsible for the instability and subsequent loss of the cluster over the course of our crystallization and diffraction experiments. First, the Fe-S cluster

Table 1. Diffraction data and refinement statistics

Parameter	*Clost*_GrpU	*Pecwa*_GrpU
X-ray wavelength (Å)	0.9791	0.9792
Nominal resolution range (Å)	19.8–2.50	61.0–2.79
Unit cell dimensions		
a, *b*, *c* (Å)	130.10, 130.10, 130.10	117.85, 117.85, 76.02
α, β, γ (°)	90, 90, 90	90, 90, 120
Space group	$P2_13$	$H3$
Total reflections	25,255	9715
Multiplicity	4.1	12.9
Completeness (%)	98.3 (99.8)	99.5 (97.5)
$\langle I/\sigma I \rangle$	17.6 (1.1)	10.2 (2.8)
Wilson *B*-factor (Å^2)	80.94	94.70
$CC_{½}$[a]	0.999 (0.446)	0.997 (0.897)
CC^*[a]	1.00 (0.768)	0.999 (0.972)
CC_{work}[a]	0.958 (0.664)	0.925 (0.759)
CC_{free}[a]	0.945 (0.650)	0.948 (0.461)
R_{work}[b]	0.190	0.215
R_{free}[b]	0.221	0.267
No. of atoms	2536	2637
Protein residues	302	317
Solvent molecules	80	4
Average atomic *B*-factor (Å^2)	87.1	76.4
RMSD (bonds) (Å)	0.01	0.01
RMSD (angles) (°)	1.26	1.60
Ramachandran plot (%)		
Favored	99.0	97.5
Allowed	1.0	2.5
Outliers	0.0	0.0
MolProbity Clashscore[c]	4.07	5.20

[a] Calculated according to Karplus and Diederichs [65].

[b] R_{work} and R_{free} are given by the following equation, computed for the working and test sets of reflections, respectively: $R = \frac{\sum_{hkl} ||F_o| - |F_c||}{\sum_{hkl} |F_o|}$.

[c] As described by Chen, et al. [64].

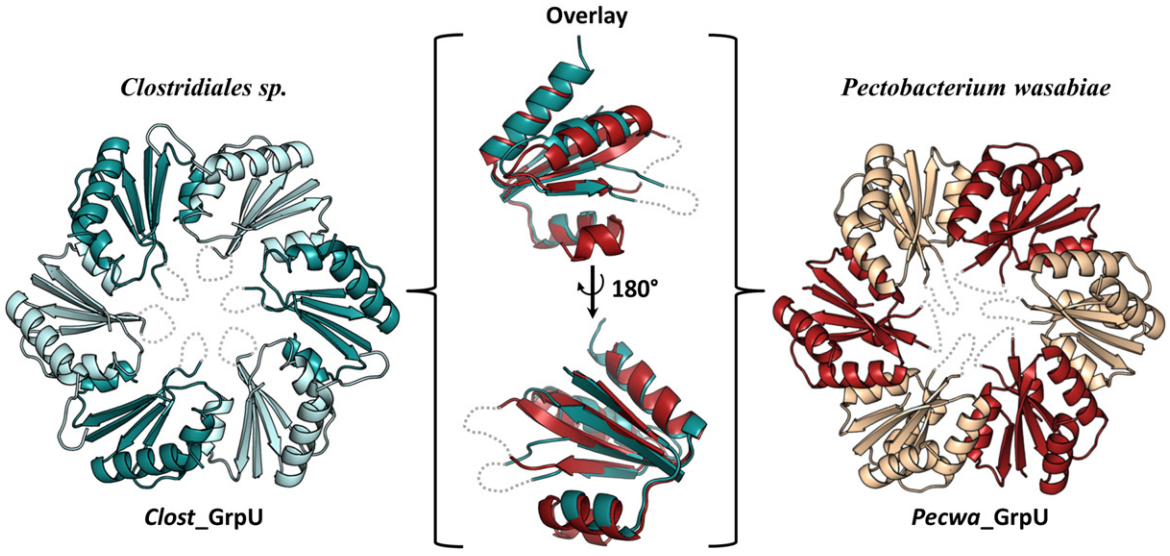


Fig. 3. X-ray crystal structures of GrpU proteins. Hexameric assemblies of Clost_GrpU (blue, left) and Pecwa_GrpU (red, right) are shown. Monomers that are crystallographically independent are colored in different shades, and the disordered β3-β4 loops are represented as gray broken lines. In the center, an overlay of *Clost*_GrpU and *Pecwa*_GrpU monomers illustrates their structural similarity.

binding site is highly solvent exposed in GrpU, and we observed spectroscopically that the bound Fe-S clusters are unstable over several days. In an effort to circumvent this issue, we attempted to purify and crystallize the GrpU proteins anaerobically. However, the crystals obtained anaerobically were similar to those obtained aerobically, and no additional electron density was visible in the vicinity of the cluster binding site. Second, it is known that Fe-S clusters tend to be acid labile, and acid treatment has been used to remove Fe-S clusters from their associated proteins [35]. In our experiments, *Clost*_GrpU and *Pecwa*_GrpU only yielded high-quality crystals suitable for diffraction experiments when crystallized under acidic conditions, thereby introducing another potential source of cluster instability. Finally, the coincidence of a pseudo-6-fold oligomeric symmetry axes with a true 3-fold crystallographic symmetry axes in both of our GrpU crystals has the potential to produce rotational disorder that could further obscure visualization of the Fe-S cluster binding site.

Comparison of GrpU with other BMC-domain proteins

Although GrpU has a highly divergent sequence, its structure is quite similar to other BMC proteins. Our structures of GrpU reveal that the polypeptide assumes a typical BMC-domain fold, with the "circularly permuted" topology previously observed in a number of other BMC-domain proteins (Fig. 3) [20,36,37]. Furthermore, GrpU monomers oligomerize to form homohexamers, consistent with previous knowledge about BMC protein assembly [10,11,13]. Interestingly, GrpU structures have an average RMSD of 2.1 Å (calculated on C^{α} atoms), when compared to a representative set of 17 other BMC-domain protein structures from the Protein Data Bank (PDB), though the average sequence identities between these proteins are only 24%, and the alignments typically have poor coverage due to sequence gaps and circular permutation. Furthermore, a BLAST (*b*asic *l*ocal *a*lignment *s*earch *t*ool) search [38] against sequences in the PDB using a GrpU query sequence does not find a single hit from the BMC-domain family with a significant *E*-value ($E < 10$), indicating that the GrpU sequence is highly divergent from other members of the BMC protein family. A plot of C^{α} RMSD *versus* sequence identity for the pairwise comparison of 19 BMC-domain structures, including *Clost*_GrpU and *Pecwa*_GrpU, illustrates that, while GrpU sequences have a low level of sequence similarity to other BMC-domain proteins, their structures are not particularly different (Fig. 4).

Disorder in GrpU crystal structures

Our GrpU structures show a substantial percentage of disordered residues. The individual monomers represented in our crystal structures are missing electron density for 13–23% of residues, depending on the specific molecule, the majority of which are not located at either terminus of the protein. These disordered regions include the loops that occupy the centers of the individual hexamers (Fig. 3). Residues 65–69 in each of the four chains

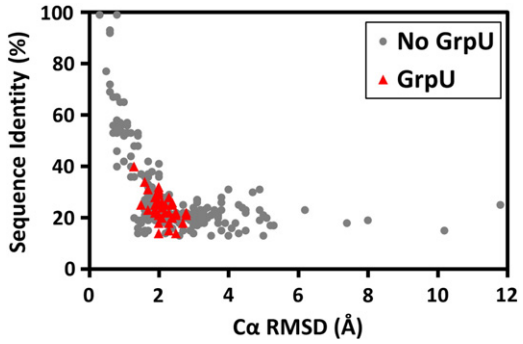


Fig. 4. Comparison of BMC-domain sequences and structures. A plot of sequence identity *versus* C^{α} RMSD for pairwise alignment of 24 BMC-domain proteins, including *Clost*_GrpU and *Pecwa*_GrpU, demonstrates that GrpU proteins are structurally similar to the rest of the BMC superfamily despite having highly divergent sequences. Comparisons involving GrpU are shown in red. Note that the pairwise sequence identities reported tend to overestimate the degree of similarity between proteins when there are substantial gaps in alignment, as is the case for many of the GpuU alignments.

within the asymmetric unit of our *Clost*_GrpU structure, which occupy the central pore region of the hexamer, lack interpretable electron density. Likewise, in all four chains of the asymmetric unit of our *Pecwa*_GrpU structure, the segment covering residues 68–71 at the center of the hexamers also lacks interpretable electron density. In addition to the pore residues described above, our GrpU structures suffer from poor or absent electron density for a segment of residues that lies on the luminal side of the hexamer. As a result, our structures have additional unmodeled segments corresponding to residues 20–31 in *Clost*_GrpU and residues 29–32 (chains A and D only) in *Pecwa*_GrpU.

A conserved sequence motif in GrpU

The disordered loop, six copies of which occupy the centers of the GrpU hexamers, contains a conserved sequence motif. We performed a multiple sequence alignment of 15 nonredundant GrpU sequences, taken from Grp operons analyzed by Jorda *et al.* in their comparative genomic study [8]. We observed that, within this set of representative GrpU sequences, there is a perfectly conserved GXCPQ sequence motif (Fig. 5a). This conserved motif contains Cys67 in Clost_GrpU and Cys69 in Pecwa_GrpU. By mapping the position of the conserved GXCPQ motif onto our structural models of GrpU, we determined that this motif is located in the loop connecting β-strands 3 and 4. In the context of the GrpU hexamer, six copies of this loop, one per monomer, protrude toward the central pore region of the oligomer. In both of our GrpU crystal structures,

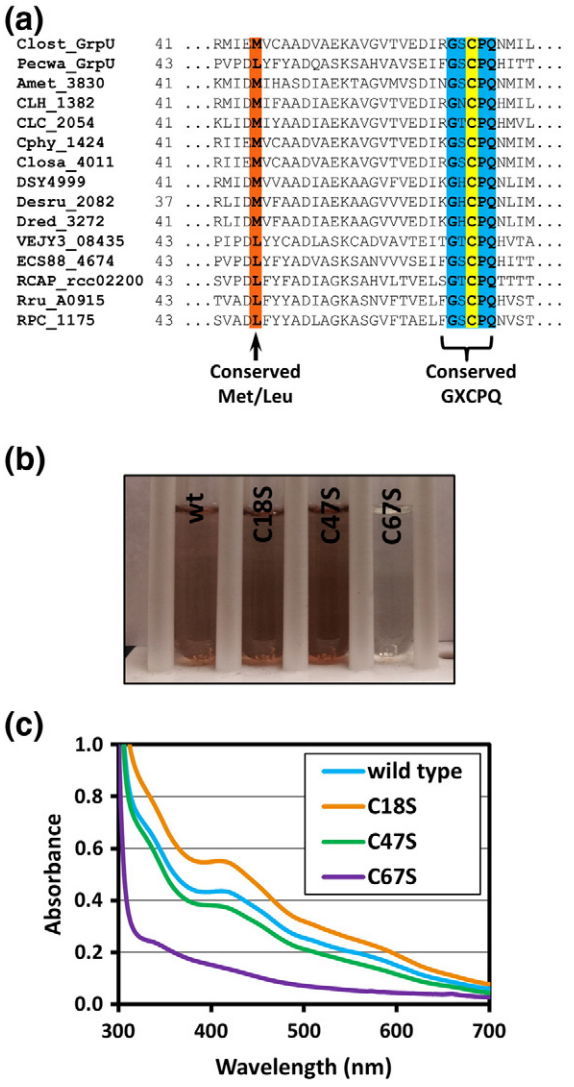


Fig. 5. A conserved cysteine coordinates the Fe-S cluster. A sequence alignment of 15 GrpU orthologs (a) shows perfect conservation of a GXCPQ sequence motif, as well as a semi-conserved methionine or leucine at another sequence position. The GXCPQ motif spans residues 65–69 in *Clost*_GrpU and residues 67–71 in *Pecwa*_GrpU. Solutions of the purified *Clost*_GrpU C67S mutant lack the characteristic brownish color (b), and absorption spectra of *Clost*_GrpU and three point mutants (C18S, C47S, C67S) demonstrate that only the C67S mutation eliminates the signal attributed to the Fe-S cluster (c).

this loop region lacks interpretable electron density and appears to be disordered, as described above.

Identification of cysteine residues that bind the Fe-S ligand in GrpU

We used site-directed mutagenesis to confirm that the conserved cysteine in the GXCPQ motif (Cys67 in *Clost*_GrpU and Cys69 in *Pecwa*_GrpU) coordinates

the putative iron-sulfur cluster. We systematically mutated each of the three cysteine residues in the sequence of *Clost*_GrpU to serine, in order to determine the effects of these mutations on the ability of GrpU to bind to its Fe-S ligand. We immediately observed that the purified C67S mutant lacked the brownish appearance characteristic of the wild-type protein (Fig. 5b). Cysteine 67 is the conserved cysteine residue found in the GXCPQ motif of *Clost*_GrpU. In contrast, the other cysteine mutants, C18S and C47S, retained the brownish color seen for the wild-type protein. Absorbance spectra of these mutants confirmed our visual observation, showing that the C18S and C47S mutants absorb light at 411 nm, while the C67S mutant does not (Fig. 5c). This result demonstrates that, in *Clost*_GrpU, Cys67 is required for Fe-S cluster binding. Furthermore, we note that the *Pecwa*_GrpU sequence contains only one cysteine (C69), which is found in the GXCPQ motif, and therefore, the Fe-S cluster must be ligated by this residue.

Computational modeling of metal-bound GrpU

In the absence of direct visualization of the Fe-S cluster by X-ray crystallography, we considered it important to evaluate whether the proposed binding loop could adopt a conformation compatible with ligation of a central 4Fe-4S cluster. Using our crystal structure of *Clost*_GrpU as a starting point, we were able to prepare a computational model of a *Clost*_GrpU hexamer with a cubic [4Fe-4S] cluster bound at its central pore. We selected the [4Fe-4S] cluster over other types of Fe-S clusters based on the EPR data and on the probable functional analogy between GrpU and PduT, a BMC-domain protein from the Pdu MCP that has been shown to coordinate a [4Fe-4S] cluster [24]. Our model demonstrates that a GrpU hexamer is indeed capable of accommodating its presumptive metal cluster at the proposed binding site, without substantially altering the portions of the structure that were well resolved in the crystal structure.

Our modeling protocol began with a 6-fold symmetric model of a hexamer, in which the loops absent from our crystal structures were modeled by comparison with homologous structures. We subsequently performed iterative, full-atom energy minimization while relaxing the oligomeric symmetry from 6-fold to 3-fold with ligand restraints between alternating Cys67 residues. We note that the Fe-S cluster could be bound in two alternate orientations related by 180° flipping; that is, the fourth Fe atom, which is not ligated to a cysteine residue, could point either up or down. We considered both possibilities in the computational modeling, but the true orientation could not be clearly distinguished from the qualities of the two resulting models or their calculated energetics.

Despite being unable to confidently establish the binding orientation of the Fe-S cluster, one thing that was clear was that conformational differences must exist between the cluster of β3-β4-binding loops on adjacent protein subunits of the hexamer. Otherwise, collisions would occur near the center (Fig. 6). In order to avoid steric clashes, adjacent subunits must adopt alternating configurations. Another observation is that the Fe-S cluster is positioned very low in the pore, making the luminal side of the GrpU hexamer appear as a flat surface, in contrast to the concave feature that is typically found on the luminal side of most BMC hexamers. Because of its position close to the luminal side of the hexamer, the Fe-S cluster appears to lie at the bottom of a funnel-shaped cavity when viewed from the cytosolic face of the hexamer.

A GrpU-like protein from a Pdu MCP

Somewhat surprisingly, an analysis of Pdu operons (i.e., those that encode B_{12}-dependent propanediol-utilizing MCPs rather than the glycyl-radical type) revealed that at least one organism, *Anaerobaculum mobile*, contains a shell protein sequence closely resembling GrpU. This operon can be identified as a true Pdu operon based on the presence of a B_{12}-dependent diol dehydratase enzyme (PduCDE), and we note that, in addition to the GrpU-like shell protein, this operon contains the normal repertoire of BMC-domain shell proteins typically found in the Pdu MCP (PduABKJTU) [17]. The GrpU-like sequence from this Pdu operon bears some similarity to the shell protein PduU, including a slightly extended N-terminus with the capacity to form a small, six-stranded β-barrel in the context of a hexamer. However, a pairwise alignment of the sequence in question with known PduU and GrpU sequences demonstrates that it is more similar to GrpU (average pairwise sequence identity of 26%) than it is to PduU (average pairwise sequence identity of 13%). Construction of a phylogenetic tree using known GrpU and PduU sequences also confirms that this sequence is more closely related to GrpU homologs than it is to PduU homologs (Fig. 7). Furthermore, sequence alignment demonstrates that this polypeptide contains the conserved GXCPQ sequence motif that we define here as characteristic of GrpU proteins.

Discussion

Here we provide the first structural insight into the homohexameric BMC-domain shell proteins from the recently discovered Grp MCP. BMC-domain proteins are the major component of MCP shells, and previous studies of other MCP systems have highlighted the mechanistic understanding that can

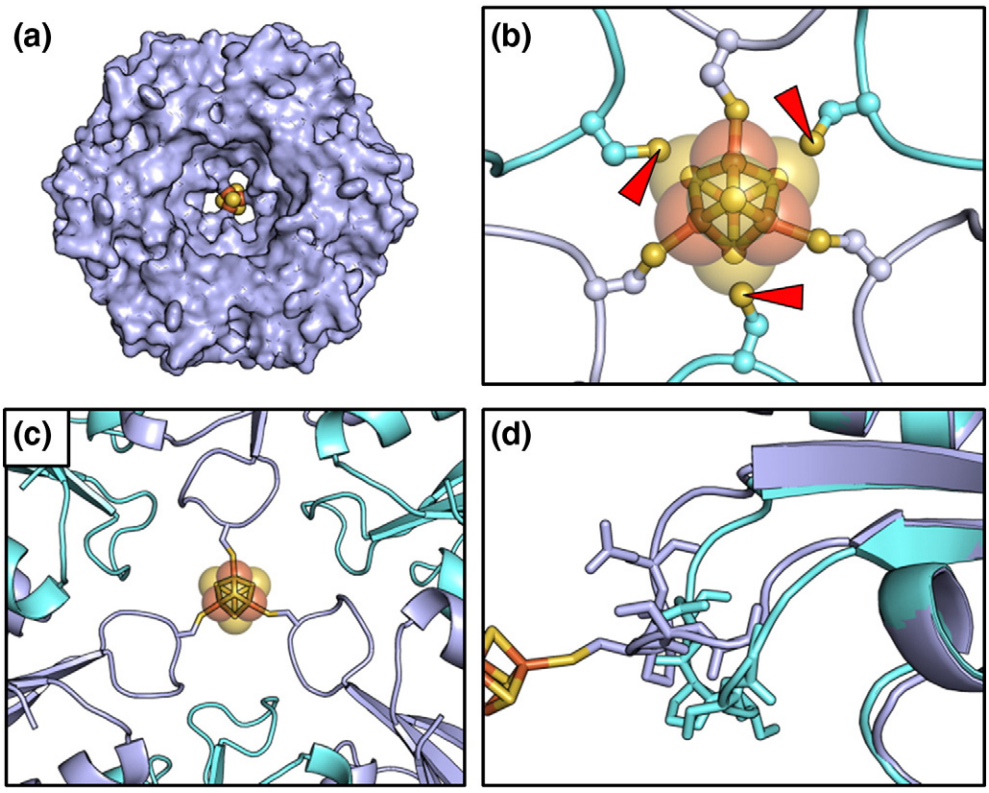


Fig. 6. Computational modeling of *Clost*_GrpU bound to a [4Fe-4S] cluster. Using our crystal structure as a starting point, we prepared a computational model of *Clost*_GrpU bound to a [4Fe-4S] cluster (a). A model having perfect 6-fold symmetry was judged to be impossible, based on a steric clash involving the Fe-S cluster and the S^{γ}-atoms of alternating Cys67 residues at the center of the hexamer (red arrowheads) (b). Our model demonstrates that this clash can be relieved if the symmetry of the hexamer is relaxed from 6-fold to 3-fold (c) via alternating conformations of the conserved GXCPQ motif. (d) One of multiple allowable configurations of the binding loops that would be compatible with Fe-S cluster binding, with amino acid side chains shown as sticks.

be gained from their structural elucidation [14,20,21]. Toward the goal of extracting similar types of insights into glycyl-radical type (Grp) MCPs, we have determined X-ray crystal structures of two GrpU homologs (Fig. 3). One of these homologs is taken from an organism whose genomic sequence is annotated only as *C. bacterium* 1_7_47FAA. This sequence is 82% identical with the GrpU sequence from *Clostridium saccharolyticum*, and GrpU sequences with high levels of sequence similarity (54–67% identity) are found in species of *Clostridia* that are human pathogens, including *C. botulinum* and *C. tetani*. The other GrpU homolog used in this study originates from the plant pathogen *P. wasabiae.* It is interesting to note that although both genera *Clostridium* and *Pectobacterium* contain a number of pathogenic organisms with Grp operons, they are evolutionarily distant from one another; *Pectobacteria* are Gram-negative Proteobacteria, while *Clostridia* are Gram-positive Firmicutes. The conservation of the Grp MCP in such distantly related pathogens suggests its potential importance in infection or other host–pathogen interactions.

Structural overview of GrpU

The structure of GrpU is quite similar to those of other BMC-domain proteins despite its highly divergent sequence. One feature of GrpU that initially attracted our attention was its apparent dissimilarity to other BMC-domain proteins. BLAST searches do not identify any alignments between GrpU sequences and other members of the BMC-domain superfamily as having significant *E*-values ($E < 10$), which we initially interpreted as evidence that GrpU structures might be quite different from other known BMC-domain structures. Surprisingly, our X-ray crystal structures of *Clost*_GrpU and *Pecwa*_GrpU reveal that these proteins exhibit only minor, though important, structural variation relative to the other members of the BMC superfamily. In order to quantify the similarities between GrpU and other, less-divergent members of the BMC superfamily, we calculated sequence identities and coordinate RMSD values (calculated on C^{α} atoms) for pairwise comparisons of 24 BMC proteins of known structure, including *Clost*_GrpU

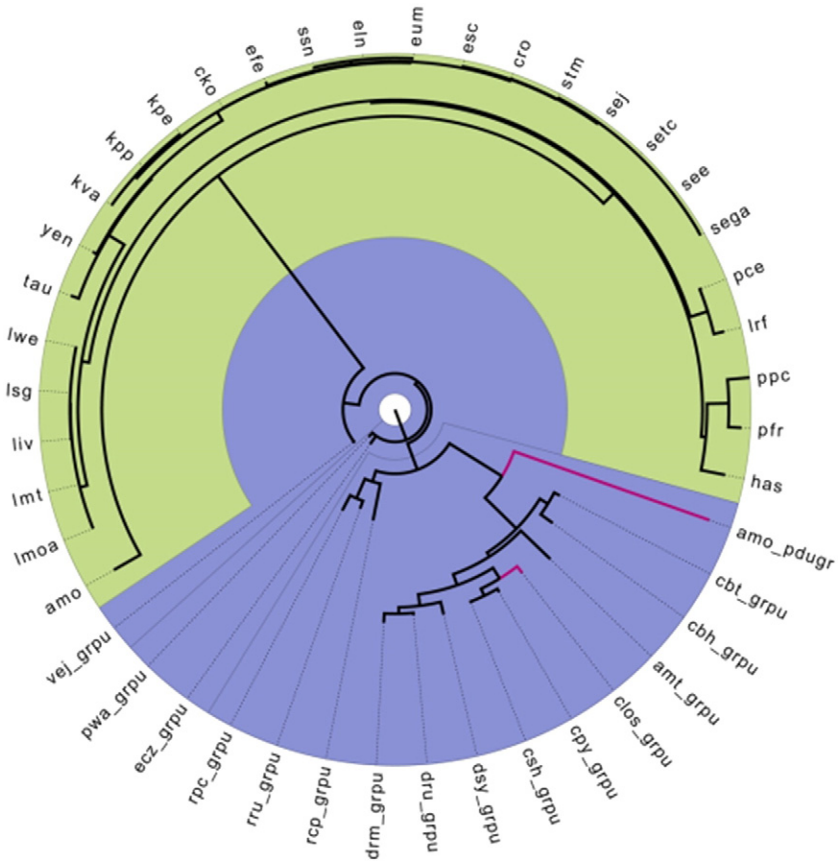


Fig. 7. A GrpU ortholog from a Pdu operon. A phylogenetic tree (a) demonstrates the relationship of GrpU (blue background) and PduU (green background) protein sequences. Pink lines highlight *Clost*_GrpU and another GrpU homolog that is found in a true Pdu-type operon but that clearly clusters with GrpU sequences.

and *Pecwa*_GrpU. A plot of sequence identity *versus* C^{α} RMSD for the 276 pairwise comparisons demonstrates that, while GrpU sequences typically have low sequence identities when compared to other BMC proteins, they also tend to have low C^{α} RMSD values for the corresponding structural alignments (Fig. 4), emphasizing that, although GrpU proteins have relatively divergent sequences, they are structurally similar to other members of the BMC superfamily.

In both of our GrpU crystal structures, hexamers appear to pack within the lattice as dodecameric assemblies with local D_6 symmetry. The possibility that MCP shells may consist of a double layer of oppositely faced BMC-domain shell proteins has been raised in previous studies [39–41]. The dodecameric assemblies we observed in crystals of *Pecwa*_GrpU and *Clost*_GrpU are distinct from each other, and neither appears to be arranged in a way that would be compatible with formation of extended double-layer sheets, though this does not discount the possible relevance of local dodecameric assemblies.

Identification of a Fe-S cluster binding site in GrpU

GrpU likely binds a cubic iron-sulfur cluster at its central pore, similar to the PduT shell protein from the Pdu MCP. Absorbance spectra of GrpU shell protein samples show features that are consistent with the presence of an iron-sulfur cluster (Fig. 2a), particularly a broad absorbance peak near 420 nm [33]. Additionally, the EPR spectra obtained for *Clost*_GrpU are diagnostic of a bound Fe-S cluster. Specifically, the EPR spectrum collected from an untreated *Clost*_GrpU sample is relatively flat, whereas reduction of the same sample with dithionite produces a signal with a *g*-factor of 1.948 (Fig. 2b). Both the *g*-factor and redox response of the observed signal are characteristic of either a [4Fe-4S] or a [2Fe-2S] cluster [33,34]; the features we observe in the EPR spectra are somewhat broad, possibly due to intrinsic flexibility of the cluster binding site. Thus, our data cannot clearly differentiate between these two possibilities. As a final measure, to ensure that the spectroscopic features

we observe do indeed result from the presence of an Fe-S cluster and not from some other type of bound metal, we performed liquid chromatography coupled to ICP-MS, which confirmed that iron is the only metal showing significant co-elution with the GrpU hexamer (Fig. S1).

The spectroscopic signatures we observed for GrpU proteins are nearly identical with those observed by Parsons *et al.* in their study of the PduT BMC-domain shell protein, which were shown to arise from a cubic [4Fe-4S] cluster [24]. Additional structural characterization of PduT revealed that, in contrast to GrpU, it is a tandem BMC-domain polypeptide, which assembles into a pseudohexameric trimer and binds a cubic [4Fe-4S] cluster at its central pore [21,24,42]. In order to bind the Fe-S cluster at the center of a symmetric trimer, each of the three PduT monomers contributes one cysteine, whose sulfur atoms ligate the Fe-S cluster at the 3-fold axis of oligomeric symmetry. In the GrpU shell protein, mutation of a single, perfectly conserved cysteine residue ablates the spectroscopic signal assigned to the Fe-S cluster (Fig. 5b and c), indicating that GrpU binds its Fe-S ligand in a manner similar to PduT, in which multiple monomers each contribute a single cysteine residue to a binding site that lies at the center of a homo-oligomer. The key difference is that, in GrpU, only three of the six chemically identical subunits contribute their cysteines to ligate the metal cluster. Mapping of these critical cysteine residues onto our structural models confirms their location within loops (which appear disordered in the absence of the bound metal cluster) near the oligomeric symmetry axis.

Structural features of the metal cluster binding site

Disorder at the center of the GrpU hexamer, in the β3-β4 loop, breaks the 6-fold oligomeric symmetry, as required for binding of the Fe-S cluster at this position. There is no known example of a biological Fe-S cluster with 6-fold symmetry [43], and to our knowledge, there have been no previous observations of any type of Fe-S cluster present at a coordination site with 6-fold symmetry. However, because the cubic Fe-S cluster does have 3-fold symmetry, which is also contained within the 6-fold oligomeric symmetry of GrpU, the GrpU hexamer could bind the cluster (in one of two orientations—up *versus* down) using a coordination site where the 6-fold symmetry is broken down to 3-fold symmetry. In order to explore whether the loops (which were not visible in the crystal structures) could adopt configurations compatible with cluster binding, we performed a computational modeling exercise. This exercise showed that a binding site with cysteine residues obeying a 6-fold symmetric arrangement is not possible because the S^{γ} atoms of the cysteine residues that do not participate in the Fe-S^{γ} bonds would clash with the sulfur atoms that are part of the Fe-S cluster itself (Fig. 6b). Consequently, we concluded that a GrpU hexamer cannot retain strict 6-fold oligomeric symmetry when bound to its presumptive [4Fe-4S] ligand.

By relaxing the oligomeric symmetry of our computational model of *Clost*_GrpU (from 6-fold to 3-fold) while constraining alternating Cys67 residues to form a [4Fe-4S] coordination site, we were able to show that the cysteine-containing loops can adopt conformations that would allow binding to the Fe-S cluster without steric overlap, as long as modest structural differences between alternating chains in the cyclic hexamer are allowed (Fig. 6c and d). Our modeling exercise was unable to resolve an ambiguity regarding the up versus down orientation of the cluster. But in either case, the ability of the central loop of GrpU to break its 6-fold symmetry and adopt alternate conformations is important for its capacity to bind the central cluster without steric clash.

In GrpU, a structural feature of the β-sheet (within each monomer) appears to promote access to alternate protein configurations near the center of the oligomer. In particular, three successive strands, β2, β3, and β4, have an unusual geometry that leads to conformational variation of the β3-β4 loop, which occupies the center of the oligomer. The outer two of these three strands, β2 and β3, appear to splay unusually far apart from one another at the edge of the sheet. This arrangement of β2 and β3 creates a situation in which the middle strand, β4, cannot simultaneously form hydrogen bonds with both of its neighboring strands in the sheet, leading to two alternative conformations (Fig. 8a and b) and implying the existence of dynamic disorder in the absence of the Fe-S cluster. Evidence for both of these conformations can be found in our *Clost*_GrpU model and in corresponding electron density maps. During the late stages of model building and refinement, after modeling the major conformation of the β4 strand in *Clost*_GrpU, a strong positive density feature (6.7σ) in the $mF_o - DF_c$ difference electron density map indicated a second, minor conformation (Fig. 8c). This difference electron density was present near all four crystallographically independent molecules in the *Clost*_GrpU structure. We were able to successfully model the second conformation in one of the four crystallographically independent molecules, although our attempt to model a second conformation of the β4 strand in the other three *Clost*_GrpU molecules led to a drop in R_{work}, but not in R_{free}, and consequently these conformers were omitted from the final coordinates. Additionally, in *Clost*_GrpU, a methionine residue (Met45) adjacent to the β4 strand also occupies two alternative rotamers in a fashion that correlates with the alternative positions of the strand (Fig. 8c). Failure to model either of these alternate side-chain

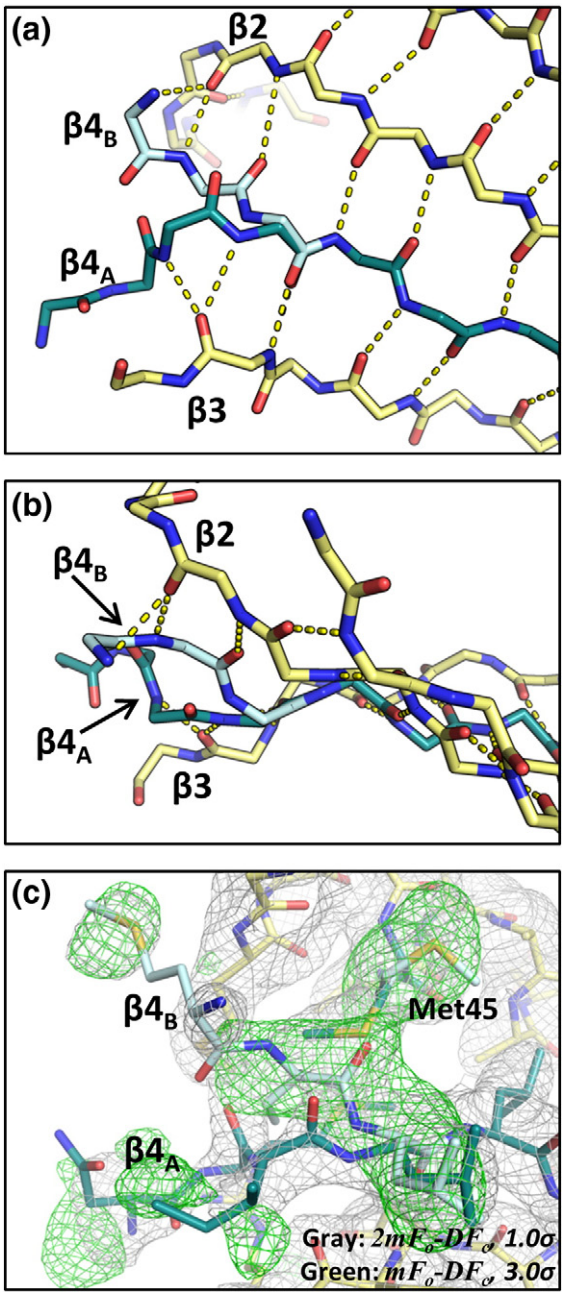


Fig. 8. Multiple conformations of the β3-β4 loop. A distinctive splaying of the β2 and β3 strands in GrpU leads to a situation in which the strand between them, β4, cannot simultaneously form hydrogen bonds with both of its partners, leading to two alternative conformations [(a and b) side and top views, respectively; side chains omitted for clarity]. Positive electron density features in both $2mF_o - DF_c$ and $mF_o - DF_c$ electron density maps calculated during the late stages of model building and refinement (c) indicate that the β4 strand and Met45 occupy correlated alternate conformations.

conformations results in strong positive peaks (8.1σ) in the $mF_o - DF_c$ difference electron density map. We note that this methionine residue is well conserved in GrpU; all GrpU sequences have either methionine or leucine at this position, indicating that motion of this residue may also play an important role in promoting structural heterogeneity at the metal binding site.

Insights into GrpU function

Although we have provided detailed structural characterization of GrpU, our work reveals little information about the function of this shell protein within the Grp MCP. Based on the presence of the Fe-S ligand in GrpU, we have developed several hypotheses about its function. Our first hypothesis is that GrpU facilitates electron transport through the shell of the Grp MCP in order to help generate the glycyl-radical at the active site of the diol dehydratase enzyme. The glycyl-radical within the diol dehydratase enzyme is generated with the help of a secondary activase enzyme, which requires *S*-adenosylmethionine and a single electron donor [8,30,44]. It is possible that the single electron required for this reaction is generated by oxidation of a cytoplasmic reductant and then transported through the MCP shell via GrpU. An alternative hypothesis is that GrpU functions to transport intact Fe-S clusters through the MCP shell. The activase in the Grp system requires a [4Fe-4S] cluster for its enzymatic activity [30], and it is possible that GrpU transports [4Fe-4S] clusters to supply the activase enzyme with fresh cofactor. GrpU might be a well-suited transport protein for a cubic Fe-S cluster because it would bind the cluster with only three cysteine ligands. This might lead to a weaker affinity for the Fe-S cluster relative to another protein that binds the same cluster with four cysteine residues, leading to an affinity gradient compatible with directional transport. A third hypothesis is that GrpU might act as a sensor of electrochemical environment and/or intracellular oxygen levels, in order to regulate the function of the Grp MCP.

Genetic variability between Grp operons creates further confusion about the function of GrpU. That not every Grp operon contains GrpU indicates that its specialized function may represent a modification to the core Grp MCP [8]. On the other hand, there are Grp operons that contain GrpU alongside other presumptive Fe-S-containing BMC shell proteins, namely orthologs of PduT, indicating some functional redundancy in these systems. One such Grp operon that contains GrpU and a PduT homolog is from the organism *Desulfitobacterium halfniense*. Additionally, we have identified at least one true Pdu operon (i.e., encoding PduCDE enzyme), from *A. mobile*, that contains a GrpU-like shell protein; it is more similar to GrpU than it is to PduU, another closely related shell protein from the Pdu MCP system (Fig. 7a). The anomalous distribution of GrpU across MCP operons is evidence that GrpU plays a complicated role in these MCP systems. Hopefully, future studies of GrpU

will integrate our structural characterization with the results of genetic and biochemical assays designed to unravel the function of this unusual shell protein.

Conclusions

The results presented here provide unexpected evidence that the previously uncharacterized GrpU MCP shell protein binds a metal cluster at a unique site of coordination. We used absorption spectroscopy to demonstrate the presence of a Fe-S cluster in GrpU and confirmed that a conserved sequence motif ligates the cluster at the center of a GrpU hexamer. Two X-ray crystal structures of GrpU proteins at moderate resolutions reveal a binding motif that is disordered in the absence of the Fe-S cluster. A subsequent computational study of GrpU demonstrated that 6-fold oligomeric symmetry must break down in order to accommodate the presumptive [4Fe-4S] cluster. Our work with GrpU provides the first insight into the BMC-domain shell proteins of the Grp MCP. Based on the chemistry associated with the Grp MCP, the Fe-S cluster in this shell protein probably facilitates the transport of either reducing equivalents or intact metal clusters across the shell, although more experiments will be required to develop a full understanding of GrpU function.

In addition to being the first BMC-domain homohexamer from the Grp MCP to be experimentally characterized, GrpU represents a new variation on BMC-domain metalloproteins. The PduT shell protein is another known Fe-S binding BMC-domain protein. The functions of GrpU and PduT are likely similar, yet the two proteins have dramatic structural differences. GrpU and PduT have different topology (by virtue of circular permutation), and GrpU is a single-domain homohexamer, while PduT is a tandem-domain homotrimer. This difference in oligomerization reflects a difference in how these two proteins break the 6-fold oligomeric symmetry typical of the BMC-domain proteins in order to accommodate their Fe-S ligands; PduT breaks 6-fold symmetry by domain duplication and subsequent sequence divergence (i.e., sequence polymorphism), while 6-fold symmetry is broken by conformational polymorphism between otherwise identical protein subunits in GrpU. As far as we are aware, the symmetry-breaking mechanism by which GrpU binds its Fe-S ligand is unique among metalloproteins studied to date.

Materials and Methods

Cloning of recombinant GrpU proteins

The codon-optimized DNA sequences corresponding to amino acids 1–101 of the GrpU protein from *C. bacterium* 1_7_47FAA (*Clost*_GrpU) and amino acids 1–99 of the GrpU protein from *P. wasabiae* (*Pecwa*_GrpU) were designed using the DNAWorks Web server [45]. Segments of linear DNA containing the designed sequences were independently assembled from small oligonucleotides (Integrated DNA Technologies), as suggested by DNAWorks, using recursive PCR [46]. The PCR products were digested with NdeI and XhoI restriction enzymes (New England Biolabs), purified by agarose gel electrophoresis, and ligated into the pET-22b+ expression vector, which had also been digested with NdeI and XhoI restriction enzymes, using Quick Ligase (New England Biolabs). The restriction sites selected for incorporation of the GrpU sequences into pET-22b+ append a hexahistadine purification tag (sequence: –LEHHHHHH) at the C-terminus of the polypeptides and place the GrpU genes under the control of the T7 promoter region. Point mutations (C18S, C47S, and C67S) were introduced into the *Clostridiales* GrpU sequence using the QuikChange method (Stratagene). All DNA constructs were verified by dideoxy chain termination sequencing [47]. Protein sequences of *Clost*_GrpU and *Pecwa*_GrpU are provided in the supplemental information.

Protein overexpression and purification

Expression plasmids (pET-22b+) bearing the GrpU genes were independently transformed into *Escherichia coli* BL21 (DE3) chemically competent cells (New England Biolabs) for protein expression. For expression of both protein constructs (*Clost*_GrpU and *Pecwa*_GrpU), we added 1 mM isopropyl-β-D-thiogalactopyranoside to cell cultures grown in selective Luria–Bertani broth during the exponential phase of growth. Growth media was supplemented with 1 mM L-cysteine and 50 mg/L ferric ammonium citrate when protein overexpression was induced, in order to supply the cells with additional iron and sulfur. After 4 h of protein overexpression at 37 °C, the cells were harvested by centrifugation for 15 min at 5000*g* and stored at −20 °C.

Cell pellets containing overexpressed *Clost*_GrpU protein were resuspended in buffer containing 20 mM Tris, 300 mM sodium chloride, and 0.03% polysorbate 20 at pH 8.0 [supplemented with cOmplete Protease Inhibitor Cocktail (Roche), 10 mM $MgCl_2$, 1 mg/mL lysozyme, and 100 units/mL of both DNase and RNase] and then lysed using a high-pressure emulsifier (EmulsiFlex C3; Avestin). We centrifuged the cell lysate at 30,000*g* for 30 min to remove insoluble material and then used a HisTrap nickel affinity column (GE Healthcare) to purify the protein from clarified lysate. The bound protein was eluted with lysis buffer containing 300 mM imidazole and then dialyzed against a buffer containing 20 mM Tris and 50 mM sodium chloride at pH 8.0. The dialyzed sample was then loaded onto a HiTrapQ anion-exchange column (GE Healthcare) that was pre-equilibrated with the dialysis buffer and subsequently eluted with a linear gradient of buffer containing 20 mM Tris and 1 M sodium chloride at pH 8.0. Finally, the eluted protein sample was dialyzed once again against a buffer consisting of 20 mM Tris and 50 mM sodium chloride at pH 8.0.

*Pecwa*_GrpU protein was purified from cell pellets by first resuspending the cells in buffer containing 50 mM Tris and 300 mM sodium chloride at pH 8.0, supplemented with cOmplete Protease Inhibitor Cocktail (Roche). Cells

were lysed by ultrasonic disruption (Vibra-Cell VCX; Sonics & Materials, Inc.), and the cell lysate was clarified by centrifugation at 30,000*g* for 30 min. The lysate was loaded onto a HisTrap nickel affinity column and the bound protein was eluted with lysis buffer containing 350 mM imidazole. Finally, the *Pecwa*_GrpU sample was subjected to gel filtration using a Superdex75 column (GE Healthcare), which provided a pure protein sample in buffer consisting of 18 mM Tris and 100 mM sodium chloride at pH 7.6.

Absorption spectroscopy

Absorption spectra of *Clost*_GrpU and corresponding point mutants (C18S, C47S, and C67S) were measured using a Cary-60 spectrophotometer (Agilent Technologies). We performed a wavelength scan from 250 nm to 800 nm using a 0.5-nm step size and a 0.1-s integration time. Spectra were collected immediately after protein purification and dialysis, in order to minimize loss of the Fe-S cluster upon exposure to oxygen in the atmosphere. For comparison of the wild-type protein with the various point mutants, individual spectra of the mutant constructs were scaled to the wild-type spectrum according to protein concentration.

The absorption spectrum of *Pecwa*_GrpU was measured using a Cary-300Bio spectrophotometer (Agilent Technologies). We performed a wavelength scan from 240 nm to 700 nm using a 1-nm step size and a 0.2-s integration time. Prior to measuring the absorption spectrum, we added 5 mM dithiothreitol in order to minimize loss of the Fe-S cluster due to protein oxidation.

EPR spectroscopy

We prepared *Clost*_GrpU protein for spectroscopic analysis by adding 25% (v/v) glycerol as a cryoprotectant, loading approximately 100 µL into a quartz EPR tube with an inner diameter of 3 mm and flash-cooling with liquid nitrogen before loading into the sample cavity of the spectrometer. Two samples, one untreated and one reduced with 20 mM dithionite, were prepared at identical protein concentrations. Continuous-wave EPR data for *Clost*_GrpU were collected with a Bruker ELEXSYS E580 X-band spectrometer equipped with a dielectric resonator (ER 4118X-MD5). The experimental temperature was equilibrated at 10 K using a temperature controller (Oxford) cooled with liquid helium. The resonance frequency was 9.676 GHz. Data were acquired using a field scan of 700G with an incident power of 2.0 mW, a modulation frequency of 100 kHz, and a modulation amplitude of 5G. Typical signal acquisition time was approximately 10 min for each sample. The reported EPR spectra are after baseline correction using a sample containing buffer only.

Inductively coupled plasma mass spectrometry

Elemental analysis of *Clost*_GrpU samples was performed using an inductively coupled plasma mass spectrometer (ICP-MS) (Agilent 7500ce) operating in line with a liquid chromatography system (Agilent 1200 series) equipped with a Superdex200 5/150GL size-exclusion column (GE Healthcare). Following equilibration of the column in buffer containing 25 mM potassium phosphate (pH 8.0) and 25 mM sodium chloride, we injected GrpU samples into the liquid chromatography system and analyzed the eluate by UV absorbance (214 nm and 280 nm) and then directed it to the ICP-MS using the high matrix interface. The ICP-MS was set to analyze the samples using the Octopole Reaction System running in the helium gas mode with time-resolved detection. We measured the signal intensities for the ions with mass-to-charge ratio of 55, 56, 58, 59, 63, 66, and 95, representing the metals manganese, iron, nickel, cobalt, copper, zinc, and molybdenum, respectively.

Protein crystallization

Purified *Clost*_GrpU protein was concentrated to approximately 12.5 mg/mL and crystallized by hanging-drop vapor diffusion in mother liquor consisting of 2.0 M ammonium sulfate and 0.1 M sodium acetate at pH 4.5. Crystallization drops were prepared by mixing 1 µL of protein solution with 1 µL of mother liquor, sealing the drop above a 0.5-mL reservoir (Hampton VDX crystallization plate) and allowing the system to equilibrate at 296 K. The highest-quality Clost_GrpU crystals were obtained when 10 mM oxidized glutathione was added to the protein sample at least 30 min prior to setting up the crystallization experiment.

Purified *Pecwa*_GrpU protein was crystallized using a hanging-drop protocol similar to the one used for crystallization of *Clost*_GrpU. In the case of *Pecwa*_GrpU, the protein sample was concentrated to approximately 60 mg/mL and crystallized in mother liquor consisting of 30% 2-methyl-2,4--pentanediol and 0.025 M sodium potassium phosphate at pH 5.8.

We attempted to purify and crystallize both *Clost*_GrpU and *Pecwa*_GrpU anaerobically, in order to preserve the Fe-S clusters. Unfortunately, crystals obtained under anaerobic conditions were identical with similar crystals grown in aerobic conditions and did not provide additional information about the Fe-S ligands.

X-ray data collection and processing

Prior to X-ray data collection, crystals of *Clost*_GrpU were harvested, cryoprotected using a 1:1 mixture of crystallization mother liquor and 4.0 M trimethylamine-*N*-oxide, and flash-cooled in a liquid nitrogen cryostream at 100 K. We collected single-crystal X-ray diffraction data at the Advanced Photon Source using beamline 24-ID-C equipped with a Pilatus 6M-F detector and operating at 12,663 eV. Crystals were maintained at cryogenic temperature (100 K) throughout the course of data collection. We indexed and integrated reflections to 2.5 Å resolution using XDS, performed scaling and merging with *XSCALE*, and converted intensities to structure factors using *XDSCONV*. The free set of reflections (10%) was designated by *XDSCONV* [48].

For X-ray data collection from *Pecwa*_GrpU crystals, harvested samples were cryoprotected using a mixture of 30% glycerol and 70% crystallization mother liquor and then flash-cooled in a liquid nitrogen cryostream at 100 K. We collected single-crystal X-ray diffraction data at the

Advanced Photon Source using beamline 24-ID-C equipped with a Pilatus 6M-F detector and operating at 12,662 eV. Crystals were maintained at cryogenic temperature (100 K) throughout the course of data collection. We indexed and integrated reflections to 2.8 Å resolution using *DENZO* and performed scaling and merging with *SCALEPACK* [49]. Intensities were converted to structure factors using *Ctruncate* and the free set of reflections (5%) was assigned using *Uniqueify*, both within the CCP4 utility *sca2mtz* [50].

Information regarding data collection and processing for both *Clost*_GrpU and *Pecwa*_GrpU is provided in Table 1.

Phase calculation by molecular replacement

Initial phase determination for the *Clost*_GrpU diffraction data was performed using the *phenix.mr_rosetta* software [51], using an initial, lower-resolution (3.2 Å) data set. Fragment files and alignment information required for *phenix.mr_rosetta* were obtained from the *Robetta* and *hhpred* Web servers, respectively. The solution obtained by *phenix.mr_rosetta* at 3.2 Å was subsequently used as a search model in a straightforward molecular replacement procedure with the 2.5-Å data set, implemented using the Phaser software [52].

In order to find an acceptable molecular replacement solution for *Pecwa*_GrpU, we first had to expand our data from space group *H*3 to the lower-symmetry space group *P*1. Using the *P*1 data, we performed a molecular replacement search using the Phaser software [52], with a search model consisting of a full BMC hexamer constructed from PDB ID 3IA0. Once a solution was found in *P*1, we selected four polypeptide chains corresponding to an asymmetric unit in space group *H*3 and proceeded with a model corresponding to the trigonal space group.

Crystallographic model building and refinement

Following molecular replacement, atomic models of GrpU proteins were prepared by iterative steps of model building and automated refinement of the model coordinates against the observed X-ray data. Model building was performed with Coot v0.7 [53]. The final model of *Clost*_GrpU was refined using BUSTER (version 2.10.0) with automatic NCS restraints and automatic refinement of TLS parameters [54]. The final model of *Pecwa*_GrpU was generated by restrained refinement using *REFMAC* (version 5.8.0049) with automatic NCS restraints and a single TLS group for each chain [55]. The final models were deposited in the PDB [56] under the following accession codes: 4OLO and 4OLP for *Clost*_GrpU and *Pecwa*_GrpU, respectively. Refinement and model validation statistics are provided in Table 1.

Sequence and structural alignment of BMC-domain proteins

Pairwise sequence and structure alignments of 24 BMC-domain proteins, including both *Clost*_GrpU and *Pecwa*_GrpU, were performed using the Dali Web server [57]. The PDB accession codes for the 22 non-GrpU proteins used in the comparison are as follows: 3BN4, 2A1B, 2A10, 2EWH, 3H8Y, 3FCH, 3I96, 4AXI, 3I6P, 4AXJ, 3N79, 3PAC, 3NWG, 3I82, 4EDI, 3U27, 4FAY, 3NGK, 3CGI, 4HT5, 3IO0, and 3IA0. These 22 BMC-domain homologs were selected in order to represent a variety of different species and MCP systems.

Multiple sequence alignment of GrpU homologs

For multiple sequence analysis of GrpU homologs, we used a set of 15 nonredundant GrpU sequences. Fourteen of these sequences, including the *Pecwa*_GrpU sequence, were taken from the original bioinformatics analysis of the Grp operon performed by Jorda *et al.* [8], and the 15th sequence is that of *Clost*_GrpU. We used the MUSCLE software to perform the multiple sequence alignment [58].

Computational modeling of metal-bound GrpU

Computational models of *Clost*_GrpU bound to a cubic [4Fe-4S] cluster were prepared in several steps. First, we used MODELLER (version 9.10) to model the disordered loop regions for a single monomer from the *Clost*_GrpU crystal structure [59]. The model of the complete monomer was then assembled back into a hexamer by applying a 6-fold symmetry operator, followed by two cycles of rotamer optimization using Rosetta [60] and then symmetry-restrained, full-atom relaxation using *CNS* (version 1.2) [61,62], which performs conformational sampling using the CHARMM force field [63] to evaluate energetics. This hexameric model was then subjected to structural relaxation, allowing backbone flexibility, using MODELLER (version 9.10) [59]. During this relaxation step, we constrained the hexamer to have 3-fold symmetry (thereby modeling two independent chains), and we harmonically restrained the distances of alternating cysteine S^{γ} atoms so that they would form an equilateral triangle with a side length of approximately 6.5 Å. This distance restraint places the sulfur atoms in positions that can accommodate ligation of an Fe-S cluster. After obtaining this 3-fold symmetric model, we attached the Fe-S group to the corresponding cysteine residues and performed 1000 cycles of symmetry-restrained, full-atom conformational sampling with energy minimization, again using *CNS* (version 1.2) [61,62] in the presence of the ligated Fe-S cluster. Two models were prepared, with the Fe-S cluster in each of two nonequivalent orientations (up *versus* down) as described in Results.

The computational models were judged to be geometrically plausible, based on satisfactory energies after minimization and on above-average scores from the MolProbity geometry analysis Web server [64].

Analysis of Pdu operons

A reliable set of 28 Pdu operons was defined by computationally identifying operons containing both PduCDE genes and genes for BMC-domain shell proteins. Computationally identified operons were confirmed by manual inspection. By carefully analyzing shell protein sequences within this manually curated set of Pdu operons, we identified a GrpU homolog in the Pdu operon of *A. mobile* (Kyoto Encyclopedia of Genes and Genomes: Anamo_0124) [65].

Calculation of a phylogenetic tree comparing GrpU and PduU

As a first step toward generating a phylogenetic tree comparing GrpU and PduU sequences, we created a multiple sequence alignment including 28 PduU sequences, 15 GrpU sequences, and the GrpU-like sequence from the Pdu operon of *A. mobile*. The phylogenetic tree was generated from this multiple sequence alignment using the PhyML software [66], and the results were visualized using *FigTree*.

Accession numbers

Coordinates and structure factors have been deposited in the PDB with accession numbers 4OLO and 4OLP.

Acknowledgements

The authors thank Drs. Thomas Bobik and Duilio Cascio for helpful discussions, Michael Collazo for assistance with crystallization experiments, and the beamline scientists at Northeastern Collaborative Access Team for assistance with X-ray diffraction experiments. We also thank Dr. Robert Gunsalus and his research group for allowing the use of their anaerobic chamber, Dr. Wayne Hubbell for access to EPR instrumentation and expertise, and Sara Bassilian for technical assistance with collection of ICP-MS data. This work was supported by National Institutes of Health Grant R01AI081146 (T.O.Y.) and by a Ruth L. Kirschstein National Research Service Award (grant GM007185, awarded to M.C.T.). X-ray data collection was supported by Department of Energy Grant DE-FC02-02ER63421 and the Northeastern Collaborative Access Team beamlines of the Advanced Photon Source, which is supported by National Institutes of Health Grant RR-15301(NCRR). Use of the Advanced Photon Source is supported by the Department of Energy, Office of Basic Energy Sciences, under Contract DE-AC02-06CH11357.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2014.07.018.

Received 19 March 2014;
Received in revised form 1 July 2014;
Accepted 20 July 2014
Available online 4 August 2014

Keywords:
metalloprotein;
BMC-domain;
glycyl-radical propanediol (Grp);
conformational polymorphism;
symmetry-breaking

Abbreviations used:
MCP, microcompartment; Grp, glycyl-radical propanediol; PDB, Protein Data Bank; EPR, electron paramagnetic resonance; ICP-MS, inductively coupled plasma mass spectrometry; NCS, noncrystallographic symmetry.

References

[1] Cheng S, Liu Y, Crowley CS, Yeates TO, Bobik TA. Bacterial microcompartments: their properties and paradoxes. BioEssays News Rev Mol Cell Dev Biol 2008;30:1084–95.
[2] Penrod JT, Roth JR. Conserving a volatile metabolite: a role for carboxysome-like organelles in *Salmonella enterica*. J Bacteriol 2006;188:2865–74.
[3] Sampson EM, Bobik TA. Microcompartments for B12-dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate. J Bacteriol 2008;190:2966–71.
[4] Kerfeld CA, Heinhorst S, Cannon GC. Bacterial microcompartments. Annu Rev Microbiol 2010;64:391–408.
[5] Shively JM, Ball F, Brown DH, Saunders RE. Functional organelles in prokaryotes: polyhedral inclusions (carboxysomes) of *Thiobacillus neapolitanus*. Science 1973;182:584–6.
[6] Yeates TO, Crowley CS, Tanaka S. Bacterial microcompartment organelles: protein shell structure and evolution. Annu Rev Biophys 2010;39:185–205.
[7] Yeates TO, Kerfeld CA, Heinhorst S, Cannon GC, Shively JM. Protein-based organelles in bacteria: carboxysomes and related microcompartments. Nat Rev Microbiol 2008;6:681–91.
[8] Jorda J, Lopez D, Wheatley NM, Yeates TO. Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. Protein Sci Publ Protein Soc 2013;22:179–95.
[9] Erbilgin O, McDonald KL, Kerfeld CA. Characterization of a planctomycetal organelle: a novel bacterial microcompartment for the aerobic degradation of plant saccharides. Appl Environ Microbiol 2014;80:2193–205.
[10] Yeates TO, Jorda J, Bobik TA. The shells of BMC-type microcompartment organelles in bacteria. J Mol Microbiol Biotechnol 2013;23:290–9.
[11] Yeates TO, Thompson MC, Bobik TA. The protein shells of bacterial microcompartment organelles. Curr Opin Struct Biol 2011;21:223–31.
[12] Dryden KA, Crowley CS, Tanaka S, Yeates TO, Yeager M. Two-dimensional crystals of carboxysome shell proteins recapitulate the hexagonal packing of three-dimensional crystals. Protein Sci 2009;18:2629–35.
[13] Kerfeld CA, Sawaya MR, Tanaka S, Nguyen CV, Phillips M, Beeby M, et al. Protein structures forming the shell of primitive bacterial organelles. Science 2005;309:936–8.
[14] Tanaka S, Kerfeld CA, Sawaya MR, Cai F, Heinhorst S, Cannon GC, et al. Atomic-level models of the bacterial carboxysome shell. Science 2008;319:1083–6.
[15] Wheatley NM, Gidaniyan SD, Liu Y, Cascio D, Yeates TO. Bacterial microcompartment shells of diverse functional types possess pentameric vertex proteins. Protein Sci Publ Protein Soc 2013;22:660–5.
[16] Sutter M, Wilson SC, Deutsch S, Kerfeld CA. Two new high-resolution crystal structures of carboxysome pentamer proteins reveal high structural conservation of CcmL orthologs among distantly related cyanobacterial species. Photosynth Res 2013;118:9–16.

[17] Bobik TA, Havemann GD, Busch RJ, Williams DS, Aldrich HC. The propanediol utilization (*pdu*) operon of *Salmonella enterica* serovar *typhimurium* LT2 includes genes necessary for formation of polyhedral organelles involved in coenzyme B12-dependent 1,2-propanediol degradation. J Bacteriol 1999;181:5967–75.
[18] Havemann GD, Bobik TA. Protein content of polyhedral organelles involved in coenzyme B12-dependent degradation of 1,2-propanediol in *Salmonella enterica* serovar *typhimurium* LT2. J Bacteriol 2003;185:5086–95.
[19] Kofoid E, Rappleye C, Stojiljkovic I, Roth J. The 17-gene ethanolamine (*eut*) operon of *Salmonella typhimurium* encodes five homologues of carboxysome shell proteins. J Bacteriol 1999;181:5317–29.
[20] Tanaka S, Sawaya MR, Yeates TO. Structure and mechanisms of a protein-based organelle in *Escherichia coli*. Science 2010;327:81–4.
[21] Crowley CS, Cascio D, Sawaya MR, Kopstein JS, Bobik TA, Yeates TO. Structural insight into the mechanisms of transport across the *Salmonella enterica* Pdu microcompartment shell. J Biol Chem 2010;285:37838–46.
[22] Cheng S, Sinha S, Fan C, Liu Y, Bobik TA. Genetic analysis of the protein shell of the microcompartments involved in coenzyme B12-dependent 1,2-propanediol degradation by *Salmonella*. J Bacteriol 2011;193:1385–92.
[23] Parsons JB, Frank S, Bhella D, Liang M, Prentice MB, Mulvihill DP, et al. Synthesis of empty bacterial microcompartments, directed organelle protein incorporation, and evidence of filament-associated organelle movement. Mol Cell 2010;38:305–15.
[24] Parsons JB, Dinesh SD, Deery E, Leech HK, Brindley AA, Heldt D, et al. Biochemical and structural insights into bacterial organelle form and biogenesis. J Biol Chem 2008; 283:14366–75.
[25] Vercellotti JR, Salyers AA, Bullard WS, Wilkins D. Breakdown of mucin and plant polysaccharides in the human colon. Can J Biochem 1977;55:1190–6.
[26] Obradors N, Badía J, Baldomà L, Aguilar J. Anaerobic metabolism of the L-rhamnose fermentation product 1,2-propanediol in *Salmonella typhimurium*. J Bacteriol 1988; 170:2159–62.
[27] Petit E, LaTouf WG, Coppi MV, Warnick TA, Currie D, Romashko I, et al. Involvement of a bacterial microcompartment in the metabolism of fucose and rhamnose by *Clostridium phytofermentans*. PLoS One 2013;8:e54337.
[28] Saxena RK, Anand P, Saran S, Isar J, Agarwal L. Microbial production and applications of 1,2-propanediol. Indian J Microbiol 2010;50:2–11.
[29] Cheng S, Fan C, Sinha S, Bobik TA. The PduQ enzyme is an alcohol dehydrogenase used to recycle NAD^+ internally within the Pdu microcompartment of *Salmonella enterica*. PLoS One 2012;7:e47144.
[30] Frey PA. Radical mechanisms of enzymatic catalysis 1. Annu Rev Biochem 2001;70:121–48.
[31] Cheng S, Bobik TA. Characterization of the PduS cobalamin reductase of *Salmonella enterica* and its role in the Pdu microcompartment. J Bacteriol 2010;192:5071–80.
[32] Sampson EM, Johnson CLV, Bobik TA. Biochemical evidence that the *pduS* gene encodes a bifunctional cobalamin reductase. Microbiol Read Engl 2005;151:1169–77.
[33] Sweeney WV, Rabinowitz JC. Proteins containing 4Fe-4S clusters: an overview. Annu Rev Biochem 1980;49:139–61.
[34] Hanson G, Berliner L, editors. Metals in Biology, vol. 29. New York, NY: Springer New York; 2010.
[35] Tsibris JCM, Namtvedt MJ, Gunsalus IC. Selenium as an acid labile sulfur replacement in putidaredoxin. Biochem Biophys Res Commun 1968;30:323–7.
[36] Crowley CS, Sawaya MR, Bobik TA, Yeates TO. Structure of the PduU shell protein from the Pdu microcompartment of *Salmonella*. Struct Lond Engl 1993 2008;16:1324–32.
[37] Pang A, Liang M, Prentice MB, Pickersgill RW. Substrate channels revealed in the trimeric *Lactobacillus reuteri* bacterial microcompartment shell protein PduB. Acta Crystallogr Sect D Biol Crystallogr 2012;68:1642–52.
[38] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–10.
[39] Cai F, Sutter M, Cameron JC, Stanley DN, Kinney JN, Kerfeld CA. The structure of CcmP, a tandem bacterial microcompartment domain protein from the β-carboxysome, forms a subcompartment within a microcompartment. J Biol Chem 2013;288:16055–63.
[40] Klein MG, Zwart P, Bagby SC, Cai F, Chisholm SW, Heinhorst S, et al. Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport. J Mol Biol 2009;392:319–33.
[41] Samborska B, Kimber MS. A dodecameric CcmK2 structure suggests β-carboxysomal shell facets have a double-layered organization. Structure 2012;20:1353–62.
[42] Pang A, Warren MJ, Pickersgill RW. Structure of PduT, a trimeric bacterial microcompartment protein with a 4Fe-4S cluster-binding site. Acta Crystallogr Sect D Biol Crystallogr 2011;67:91–6.
[43] Beinert H, Holm RH, Münck E. Iron-sulfur clusters: nature's modular, multipurpose structures. Science 1997;277:653–9.
[44] Frey M, Rothe M, Wagner AF, Knappe J. Adenosylmethionine-dependent synthesis of the glycyl radical in pyruvate formate-lyase by abstraction of the glycine C-2 pro-S hydrogen atom. Studies of [2H]glycine-substituted enzyme and peptides homologous to the glycine 734 site. J Biol Chem 1994;269:12432–7.
[45] Hoover DM, Lubkowski J. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. Nucleic Acids Res 2002;30:e43.
[46] Prodromou C, Pearl LH. Recursive PCR: a novel technique for total gene synthesis. Protein Eng 1992;5:827–9.
[47] Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 1977;74:5463–7.
[48] Kabsch W. XDS. Acta Crystallogr Sect D Biol Crystallogr 2010;66:125–32.
[49] Otwinowski Z, Minor W. [20] Processing of X-ray diffraction data collected in oscillation mode. Macromol Crystallogr Part A, 276. Academic Press; 1997 307–26.
[50] Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr Sect D Biol Crystallogr 2011; 67:235–42.
[51] Terwilliger TC, DiMaio F, Read RJ, Baker D, Bunkoczi G, Adams PD, et al. phenix.mr_rosetta: molecular replacement and model rebuilding with Phenix and Rosetta. J Struct Funct Genomics 2012;13:81–90.
[52] McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr 2007;40:658–74.
[53] Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr Sect D Biol Crystallogr 2010;66:486–501.
[54] Blanc E, Roversi P, Vonrhein C, Flensburg C, Lea SM, Bricogne G. Refinement of severely incomplete structures

with maximum likelihood in BUSTER-TNT. Acta Crystallogr Sect D Biol Crystallogr 2004;60:2210–21.
[55] Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. *REFMAC* 5 for the refinement of macromolecular crystal structures. Acta Crystallogr Sect D Biol Crystallogr 2011;67:355–67.
[56] Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, et al. The protein data bank: a computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535–42.
[57] Holm L, Rosenström P. Dali server: conservation mapping in 3D. Nucleic Acids Res 2010;38:W545–9.
[58] Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004;32:1792–7.
[59] Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen M-Y, et al. Comparative protein structure modeling using MODELLER. Curr Protoc Protein Sci Editor Board John E Coligan Al 2007; 2.9.1–2.9.3.1 [Chapter 2:Unit 2.9].
[60] Simons KT, Ruczinski I, Kooperberg C, Fox BA, Bystroff C, Baker D. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. Proteins 1999;34:82–95.
[61] Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, et al. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr Sect D Biol Crystallogr 1998;54:905–21.
[62] Brunger AT. Version 1.2 of the crystallography and NMR system. Nat Protoc 2007;2:2728–33.
[63] Brooks BR, Brooks CL, MacKerell AD, Nilsson L, Petrella RJ, Roux B, et al. CHARMM: the biomolecular simulation program. J Comput Chem 2009;30:1545–614.
[64] Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr Sect D Biol Crystallogr 2009;66:12–21.
[65] Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. Nucleic Acids Res 2004;32:D277–80.
[66] Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 2010;59:307–21.