Citation | Born, David A. 2018. Crystallographic Snapshots of Functional Motions in Cobalamin Maintenance and Methylphosphonate Production. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

Citable link | http://nrs.harvard.edu/urn-3:HUL.InstRepos:41129203

Terms of Use | This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA
Crystallographic Snapshots of Functional Motions in Cobalamin

Maintenance and Methylphosphonate Production

A dissertation presented

by

David A. Born

to

The Committee on Higher Degrees in Biophysics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biophysics

Harvard University

Cambridge, Massachusetts

March 2018
Crystallographic Snapshots of Functional Motions in Cobalamin Maintenance and Methylphosphonate Production

Abstract

Biochemical experiments provide key insight into the molecular logic and mechanisms underlying the biology surrounding us. Supplementing this biochemistry with detailed molecular models helps confirm, refine, and create hypotheses regarding complex biological systems. This thesis leverages crystallography to contextualize complex biochemical data. In Part I, we fill a major gap in the understanding of chaperone-mediated cobalamin delivery and repair. The molecular snapshots presented open new doors in the study of cobalamin transfer and help explain the molecular basis of human disease. In Part II, a set of crystal structures are used to compare and contrast two related enzymes involved in phosphonate metabolism. A subtle rearrangement in the active site of one enzyme hints at the molecular basis for methylphosphonate metabolism, and this hypothesis is confirmed by structure-guided engineering. Although diverse, both stories showcase the value of atomic resolution in clarifying complex biochemical processes.
Abstract................................................................................................................................. iii

Acknowledgements................................................................................................................... vi

Part 1 ........................................................................................................................................ 1

Chapter I: Introduction to Chaperone-Mediated Adenosylcobalamin Delivery and Repair .... 1

G-protein chaperones in adenosylcobalamin delivery .......................................................... 1
NTPase chaperones in precious cofactor transport pathways ................................................. 3
The role of G-proteins in AdoCbl delivery and repair ............................................................ 6
Isobutyryl-CoA mutase fused depicts the relative position of the G-protein and mutase ..... 7
The Switch III region of the AdoCbl G-protein chaperone .................................................... 8
References .............................................................................................................................. 9

Chapter II: IcmF Establishes an Interdimer Interface to Facilitate Cbl Delivery and Repair .... 12

Summary .................................................................................................................................... 12
Introduction ................................................................................................................................. 13
Results: ...................................................................................................................................... 15
  Evidence of an interdimer interface in IcmF ........................................................................... 15
  Residues at the interdimer interface participate in GTP hydrolysis ......................................... 20
  IcmF forms multimers when bound to a GTP analog .............................................................. 21
  Multimer formation depends on the cofactor state of IcmF .................................................... 21
Discussion: .................................................................................................................................. 24
Materials and Methods .............................................................................................................. 28
  Mutagenesis ............................................................................................................................. 28
  Protein expression and purification ......................................................................................... 28
  Electron microscopy ................................................................................................................. 29
  Crystallization ......................................................................................................................... 30
  Data collection and processing ................................................................................................. 31
  Structure determination and refinement .................................................................................. 31
  GTPase assays ......................................................................................................................... 31
References ................................................................................................................................... 32

Chapter III: ............................................................................................................................... 34

A G-protein Chaperone Dimer Rearrangement Underlies Interaction with its B12-dependent Enzyme ..................................................................................................................... 34

Summary .................................................................................................................................... 34
Introduction .................................................................................................................................. 35
Results ......................................................................................................................................... 37
  Cbl-binding domains are GTPase activating proteins .............................................................. 37
  G-protein:Cbl-binding domain complexes are stable ............................................................... 38
  MeaB:meMCM<sub>cbl</sub> crystal structure reveals an active chaperone arrangement ............... 39
  The Switch III region is poised for GTP hydrolysis across the dimer interface of MeaB:MeMCM<sub>cbl</sub> ................................................................. 44
Discussion ................................................................................................................................... 46
Materials and methods ............................................................................................................ 51
  Preparation of Cbl-binding domain constructs ...................................................................... 51
  Protein purification ................................................................................................................... 51
Acknowledgements

I am a strong believer in balancing my work and my happiness. When choosing my graduate laboratory, the people were what sold me on the Drennan lab. It was a great joy to work in an environment with some of the greatest scientists in our field, but it was truly amazing to be able to do the same while surrounded by friends. Whether we were relaxing in the lab kitchen or crowding behind Tsehai’s computer, a joke and a hypothesis were never far apart. This community kept me motivated each day, and it taught me how to be a strong scientist.

Cathy’s role in assembling this great group of people is to be applauded. I am grateful for Cathy taking the odd step of allowing a Harvard student to work in her lab and jump over the administrative hurtles required. Having a supportive adviser has been one of the greatest parts of graduate school. I had the freedom to take my science in the directions I chose, explore career paths outside of the “track”, and know that Cathy would support my choices.

Every member of the Drennan lab has contributed to my scientific and personal growth in their own ways. I’m grateful for the variety of scientific perspectives within our lab. Whether it is arguing about the details of an enzymatic mechanism or demanding that “C”-fonts are not allowed, I continue to enjoy the constructive criticism and intellectual challenge everyone provides.

In particular, I thank Marco for teaching me the way of things during my first two years in the lab. His scientific rigor is next to none, and he fostered a Drennan lab social community that is difficult to rival.

Outside of the lab I have many people to thank. Beginning, of course with my parents, who have always been keen to listen to my scientific ramblings, even when I utterly fail at
“laymanizing” my research. Growing up in a household where a sound argument could sway opinions contributed in no small part to my critical thinking as a scientist.

Finally, I would like to thank my partner, Emily. Without her unwavering support and patience, I don’t know where I would be today. Together with our dog Bea, an endless source of happiness, we’ve gone on many adventures and continue to challenge each other to improve every day.
G-protein chaperones in adenosylcobalamin delivery

Cobalamin (Cbl, B12) is a complex cobalt-containing molecule and an essential human vitamin. Two human enzymes, methionine synthase (MetH) and methylmalonyl-CoA mutase (Mut or MCM), utilize Cbl derivatives to perform key functions in cellular metabolism (Figure I.1). A complex series of proteins bind, transport, modify, deliver, and repair Cbl and Cbl-derivatives within the human body. Dysfunction in proteins involved in these processes results in human disease, primarily in the forms of homocysteinuria and methylmalonic aciduria (1-3). Genetic characterization of these disorders of Cbl metabolism have informed much of the biochemistry of human Cbl utilization (1, 2).

The following chapters investigate the final step in Cbl delivery within the mitochondria, where AdoCbl is formed and transferred to Mut (Figure I.2). Once Cob(II)alamin has been transported into the mitochondria, it is bound to ATP:Cob(I)alamin adenosyltransferase (ATR). ATR catalyzes a one electron reduction of Cob(II)alamin to Cob(I)alamin, enabling nucleophilic attack of the Co(I) onto C5 of ATP, yielding AdoCbl and triphosphate (4, 5). ATR then directly transfers AdoCbl to Mut (6) where it serves as an essential cofactor for the conversion of methylmalonyl-CoA to succinyl-CoA (Figure I.1B). This final step in AdoCbl transfer is gated by the G-nucleotide state of the G-protein chaperone, methylmalonic aciduria type A protein (MMAA) (7). During catalysis, the Mut enzyme will eventually inactivate due to decoupling of the 5’-dAdo radical from Cob(II)alamin, yielding inactive Cob(II)alamin (8, 9). At this stage, the
ATR rebinds Cob(II)alamin in another G-protein-gated transfer and repairs the inactivated cofactor to AdoCbl \((8, 10)\). This work seeks to investigate the precise structural role of the G-protein chaperone in both AdoCbl transfer and inactive cofactor repair.

**Figure I.1.** Cobalamin derivatives are used as cofactors in two human enzymes. (A) Chemical structure of cobalamin. The upper axial coordination position of the cobalt varies. (B) The two reactions catalyzed by Cbl-dependent enzymes in humans. Top, methionine synthase utilizes MeCbl to catalyze methylation of homocysteine to form methionine. Bottom, methylmalonyl-CoA mutase catalyzes a carbon skeleton rearrangement on L-methylmalonyl-CoA to form succinyl-CoA using AdoCbl as a cofactor. The reaction arrows denote the direction of metabolic flux for both reactions.
Figure 1.2. Schematic for AdoCbl delivery within the mitochondria. AdoCbl is formed from Cob(II)alamin (Cbl(II)) and ATP by ATP:Cob(I)alamin adenosyltransferase (ATR) following reduction by an electron transfer protein. ATR then transfers AdoCbl to Mut in a process mediated by the G-protein chaperone MMAA. The nomenclature used for human and bacterial AdoCbl systems are presented on the right of the dividing line.

NTPase chaperones in precious cofactor transport pathways

The G-protein chaperones participating in AdoCbl delivery belong to the SIMIBI (signal recognition particle, MinD, and BioD) family of P-loop NTPases that includes chaperones involved in nickel utilization such as, HypB, UreG, and CooC (11-15). These SIMIBI family members play important roles in the activation of the complex metalloenzymes [Ni-Fe]-hydrogenase, urease, and carbon monoxide dehydrogenase respectively (12, 13, 16). The mechanism by which these chaperones assist in nickel delivery is still contested (17), however,
structural and biochemical data suggest a nucleotide-induced dimerization event enables interaction with accessory proteins required for Ni(II) insertion (18-22). Whether or not the dimer arrangements require complex formation with accessory proteins is a key area of research (17). The active dimeric arrangement of the NTPase dimer has been directly observed in HypB (22) and UreG (21) (Figure I.3A,B), and modeled for CooC (23). All three systems share a dimeric arrangement that is conserved across many of the SIMIBI family members, suggesting a widespread role of this particular quaternary structure (24).

The relationship between the AdoCbl G-protein chaperones MeaB and MMAA and the Ni(II) chaperones is an intriguing parallel. Both Ni(II) and Cbl can be considered rare and potentially toxic molecules requiring specific protein machinery for their efficient usage. MMAA and the bacterial ortholog, MeaB, share the same overall fold as HypB, UreG, and CooC (21-23). However, the dimeric arrangement observed in crystal structures of MeaB and MMAA is distinct from the Ni(II) chaperones (Figure I.3), largely due to an extended C-terminal dimerization region (25, 26). A number of loop regions at the dimer interface of both characterized AdoCbl G-protein chaperones are also disordered in crystal structures, including the so-called Switch III region that is critical for chaperone function (26, 27). Taken together, the structural differences between the AdoCbl chaperones and the Ni(II) chaperones along with the disordered nature of key signaling residues suggests there is a second, GTP-hydrolysis active conformation of MeaB and MMAA that is yet to be observed directly (25).
Figure 1.3. Structures of SIMIBI NTPase chaperone family members. (A) HypB (PDB 2HF8) (24) bound to GTPγS. (B) UreG (PDB 5XKT) (16) bound to GTPγS. (C) MeaB (PDB 2QM7) (25) bound to GDP. (D) MMAA (PDB 2WWW) (26) bound to GDP. All structures are superimposed by the upper protomer within the dimer and show bound nucleotide in yellow ball-and-stick.
The role of G-proteins in AdoCbl delivery and repair

The biochemical logic of AdoCbl G-protein chaperones has been extensively studied (7). Although most characterization has been on the ATR, MeaB, and MCM derived from *Methylobacterium extorquens*, recent work has also incorporated characterization of a fused protein system from *Cupriavidus metallidurans* termed isobutyryl-CoA mutase fused (IcmF), a natural bacterial fusion protein whereby a single G-protein chaperone domain has been inserted between the Cbl-binding domain and substrate binding domain of an AdoCbl-dependent mutase (28-30). Studies directly evaluating the human system have also been performed, and the data are consistent with studies from *M. extorquens* (26).

The G-protein chaperone MeaB appears to perform a number of different roles relating to activation, protection, and repair of the AdoCbl-bound state of MCM. Firstly, MeaB gates delivery of AdoCbl from ATR in a GTP-hydrolysis dependent manner (10, 31). After AdoCbl delivery, MeaB remains bound to MCM and reduces the rate of MCM cofactor inactivation (31). Once the AdoCbl is damaged, resulting in Cob(II)alamin, MeaB uses the binding energy of GTP to release inactive cofactor from MCM for subsequent repair by ATR (10).

In the fused IcmF system, the biochemical role of the G-domain is somewhat more convoluted, as experimental setups in which high concentrations of GTP or GTP analogs are added lead to aggregation of the protein (unpublished data). IcmF appears to not require GTP binding or hydrolysis for initial AdoCbl loading (29), but IcmF requires GTP binding and hydrolysis for inactive cofactor repair by ATR (28). The detailed nucleotide-dependence of delivery and repair processes has yet to be elucidated in the human system.
For the G-protein chaperone to exert its effect on the mutase, the two proteins must come together as a complex. MeaB binds to \textit{meMCM} in the absence of nucleotide \((32)\) but this binding is enhanced in the presence of a non-hydrolyzable GTP analog \((33)\). MMAA only binds \textit{hsMCM} in the presence of a non-hydrolyzable GTP analog and in the absence of AdoCbl \((26)\).

The dynamics of GTP hydrolysis in the presence/absence of MCM have also been assessed in both systems. Both MeaB and MMAA are significantly more active for GTP hydrolysis when in complex with their cognate MCM \((26, 32)\). The activity of both MeaB:MCM and MMAA:MCM is similar to IcmF \((29)\). The GTP-activating protein (GAP) properties of the mutase suggest significant structural changes at the GTP binding site of the chaperone upon complex formation.

\textbf{Isobutyryl-CoA mutase fused depicts the relative position of the G-protein and mutase}

Crystal structures of IcmF provided much needed insight into the molecular interaction underlying complex formation between the G-proteins and mutases \((30)\). These structures demonstrate a domain arrangement whereby the G-domain binds directly to the Cbl-binding domain and the two domains move as a rigid unit to open and close the AdoCbl binding site \((30)\). However, the structures of IcmF have not provided conclusive evidence for the structural changes that occur upon GTP binding, as IcmF-GMPPNP exhibits no large-scale structural changes that would imply “switch-like” signaling between the G-domain and the mutase domains (Chapter II). Even with a structural depiction of a chaperone:Cbl-binding domain interface, creating a structural model of MeaB:MCM or MMAA:MCM has been impeded by the dimeric arrangements of MeaB and MMAA. The IcmF G-protein is a single domain, whereas MeaB and MMAA are homodimers. Modeling these homodimers using IcmF has not yet been
informative, likely due to conformational rearrangements in MeaB and MMAA upon complex formation with MCM.

**The Switch III region of the AdoCbl G-protein chaperone**

Mutations in numerous locations on the MMAA protein cause methylmalonic aciduria (34, 35). A subset of these mutations has been investigated in the orthologous MeaB:MCM system (27). Intriguingly, mutations in a conserved region termed Switch III decrease the GTPase activity of MeaB only when in complex with MCM (27). The Switch III region is structurally variable in published crystal structures of MeaB (25, 27) and is disordered in the crystal structure of MMAA (35). The dependence on complex formation for Switch III function implies that the region undergoes a significant structural rearrangement following complex formation with MCM such that it can participate in GTP hydrolysis. The Switch III region is a central focus in the following chapters as its dramatic rearrangement appears to underlie GTPase activity.
References


Chapter II:  
IcmF Establishes an Interdimer Interface to Facilitate 
Cbl Delivery and Repair

Summary
The valuable adenosylcobalamin (AdoCbl) cofactor is used by one human enzyme, methylmalonyl-CoA mutase (Mut). Ensuring that Mut utilizes intact AdoCbl is a complex and tightly regulated process involving three proteins: an ATP:cob(I)alamin adenosyltransferase (ATR), a G-protein chaperone, and Mut itself. In concert, these three proteins are able to produce AdoCbl from cob(II)alamin, deliver AdoCbl to Mut, and repair the damaged cofactor that is formed during catalysis. Dysfunction in any of these proteins leads to methyldalonic aciduria, an inborn error of metabolism. Here we investigate the molecular interactions between the G-protein chaperone and mutase proteins using a natural bacterial fusion protein, isobutyryl-CoA mutase fused (IcmF). Leveraging crystallography and electron microscopy, we propose and validate a protein:protein interface between G-domains of IcmF that depicts an active conformation of the G-domain. In this conformation, IcmF G-domains are poised to structurally wedge open the Cbl-binding domain enabling either AdoCbl delivery or inactive cofactor repair. Electron microscopy additionally demonstrates the molecular requirements for G-domain function. Our data suggest a unified model between bacterial and human systems whereby AdoCbl G-protein chaperones undergo a conformational rearrangement upon binding their cognate mutase and GTP, wedging the Cbl-binding domain open.

Contributions: This chapter was written with Professor Catherine L. Drennan. Dallas Fonseca and Francesca Vaccaro assisted in electron microscopy experiments. Marco Jost contributed to crystallography and data analysis.
Introduction

Cobalamin (Cbl) derivatives are used by two human enzymes, methionine synthase and methylmalonyl-CoA mutase (Mut), in the forms of methylcobalamin (MeCbl) and adenosylcobalamin (AdoCbl), respectively. Mut is located within the mitochondria and requires AdoCbl to function (1). Dysfunction of Mut, whether from disrupted Cbl trafficking or from mutation within the mut gene results in methylmalonic aciduria, an inborn error of metabolism (2).

For successful Mut function, Cbl must be taken up from the diet, transported into the mitochondria, converted to AdoCbl by ATP:Cob(I)alamin adenosyltransferase (ATR), and delivered directly to Mut. Much work has been done in recent years to better understand the molecular mechanisms underlying transport, modification, and delivery of Cbl on pathway to Mut (1). The critical final step in AdoCbl delivery to Mut involves three proteins: ATR, methylmalonic aciduria type A protein (MMAA), and Mut (1). ATR directly delivers AdoCbl to Mut while MMAA is a G-protein chaperone that mediates the transfer without binding AdoCbl (3). The biochemical requirements for successful AdoCbl transfer to Mut have been extensively characterized in multiple model systems for Cbl-delivery including human (4), bacterial (3), and a natural bacterial fusion protein in which the chaperone and mutase proteins are contained on a single polypeptide, isobutyryl-CoA mutase fused (IcmF) (5, 6). In addition to AdoCbl delivery, the same three-protein system has been implicated in repair of inactivated Cbl following cofactor damage (7, 8).

Although extensive biochemical and structural data exist on the proteins involved in AdoCbl delivery to Mut, a molecular explanation of how the G-protein chaperone mediates AdoCbl delivery and Cbl repair has remained elusive. For example, a region of the human and
bacterial G-protein chaperones MMAA (9, 10) and MeaB (11), respectively, called Switch III (residues 265-276 of MMAA and 177-188 of MeaB) was identified as being necessary for productive signaling between the G-protein and its cognate mutase, but how Switch III functioned was unclear. Crystal structures of IcmF depict the relative arrangement of G-protein chaperone and mutase domains and demonstrate that the G-protein domain and Cbl-binding domain move together as a rigid unit upon binding AdoCbl (5). However, the relevance of comparing IcmF to a human-like system in which the G-protein chaperone MMAA exists as a homodimer has been questioned (7). Additionally, IcmF residues that correspond to the Switch III region (333-344) appeared far from the intramolecular GTP-binding site (5).

Here we evaluate whether a previously unrecognized feature of the IcmF crystal lattice structure, an interdimer interface composed of the G-protein domains, is relevant to the GTPase activity that is associated with AdoCbl loading. In particular, we establish that this interface forms outside of the crystal; the interdimer interface can be propagated to form a long, multimeric structure in which consecutive IcmF dimers bind end to end at G-protein domains. Structural investigations, reported here and previously (5), provide a molecular explanation of how this interdimer interface could play a role in AdoCbl delivery, with structures showing that the interdimer interface physically wedges open the Cbl-binding domain and substrate-binding domains of IcmF for AdoCbl loading. Additionally, the interdimer interface explains the reported relevance of the Switch III region for GTP hydrolysis, with the structure reported here showing the first direct interaction of Switch III to bound guanine nucleotides. We further consider whether the relative orientation of G-protein domains observed in the IcmF interdimer interface may be an accessible, and physiologically relevant conformation in human and bacterial systems in which the G-protein and mutase are on separate polypeptides.
Results:

Evidence of an interdimer interface in IcmF

The crystal structure of IcmF bound to the non-hydrolyzable GTP analog, GMPPNP (Table II.1) is similar to previously reported IcmF crystal structures (5). The asymmetric unit of the crystal lattice is composed of one IcmF dimer, and each protomer of the dimer demonstrates a unique conformation of IcmF. One protomer, Chain A, has sufficient freedom within the crystal lattice to alternate between open and closed conformations of the Cbl-binding domain (5). The second protomer, Chain B, demonstrates no flexibility and is structurally restrained in an inactive open conformation by a large (2400 Å² as determined using PISA (12)) crystal lattice contact (Figure II.1)(5). In this open conformation, the Cbl-binding domain is held away from the substrate-binding domain, preventing AdoCbl-dependent turnover (Figure II.2). Instead of being catalytically active, this open conformation appears relevant for Cbl insertion and/or removal processes, which are mediated by IcmF’s G-protein domains. In Chain B of IcmF’s G-protein domain, the local environment surrounding the third GMPPNP phosphate is comprised of two residues (Gln341 and Lys344) from a symmetry-related crystal contact that directly interact with oxygens of the third GMPPNP phosphate (Figure II.3A,B). The direct interaction of the third phosphate with these two residues from a symmetry-related dimer was the first indication that an interdimer interface may be relevant for IcmF function. Notably, these two residues are part of the Switch III region (333-344 in IcmF) (Figure II.3C).
### Table II.1. Crystallographic data collection and refinement statistics.

<table>
<thead>
<tr>
<th>Data collection</th>
<th>IcmF-GMPPNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>H32</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
</tr>
<tr>
<td>$a$, $b$, $c$ (Å)</td>
<td>314.4, 314.4, 343.3</td>
</tr>
<tr>
<td>$\alpha$, $\beta$, $\gamma$ (º)</td>
<td>120, 120, 120</td>
</tr>
<tr>
<td>Resolution† (Å)</td>
<td>100-3.25 (3.37-3.25)</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.979</td>
</tr>
<tr>
<td>$R_{max}$ (%)‡</td>
<td>16.6 (100.0)</td>
</tr>
<tr>
<td>$R_{max}$ (%)‡</td>
<td>6.6 (42.1)</td>
</tr>
<tr>
<td>$CC_{1/2}$‡</td>
<td>(71.9)</td>
</tr>
<tr>
<td>$I/\sigma_{I}$‡</td>
<td>11.3 (1.9)</td>
</tr>
<tr>
<td>Completeness (%)‡</td>
<td>100 (100)</td>
</tr>
<tr>
<td>Redundancy‡</td>
<td>6.8 (6.7)</td>
</tr>
<tr>
<td>No. of reflections‡</td>
<td>101331 (5027)</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>3.25</td>
</tr>
<tr>
<td>No. of reflections used</td>
<td>101298</td>
</tr>
<tr>
<td>Average B-factors (Å²)</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>80.6</td>
</tr>
<tr>
<td>nb-CoA</td>
<td>103.3</td>
</tr>
<tr>
<td>GMPPNP</td>
<td>70.7</td>
</tr>
<tr>
<td>Mg²⁺</td>
<td>63.3</td>
</tr>
<tr>
<td>Cl⁻</td>
<td>63.3</td>
</tr>
<tr>
<td>$R_{work}/R_{free}$ (%)§</td>
<td>24.1/25.9</td>
</tr>
<tr>
<td>No. atoms</td>
<td></td>
</tr>
<tr>
<td>protein</td>
<td>16001</td>
</tr>
<tr>
<td>GMPPNP</td>
<td>64</td>
</tr>
<tr>
<td>nb-CoA</td>
<td>86</td>
</tr>
<tr>
<td>Mg²⁺</td>
<td>2</td>
</tr>
<tr>
<td>Cl⁻</td>
<td>1</td>
</tr>
<tr>
<td>R.m.s. deviations</td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.005</td>
</tr>
<tr>
<td>Bond angles (º)</td>
<td>0.747</td>
</tr>
<tr>
<td>Rotamer outliers (%)</td>
<td>4.0</td>
</tr>
<tr>
<td>Ramachandran angles</td>
<td></td>
</tr>
<tr>
<td>Outlier (%)</td>
<td>0.86</td>
</tr>
<tr>
<td>Allowed (%)</td>
<td>5.09</td>
</tr>
<tr>
<td>Favored (%)</td>
<td>94.05</td>
</tr>
</tbody>
</table>

† Values in parentheses indicate highest resolution bin.
§ Reflections used to calculate $R_{free}$ comprise 5% of observed reflections.
Figure II.1. Interdimer interface in the IcmF crystal lattice. (A) Two IcmF homodimers form a symmetric chaperone:chaperone interface in chain B of IcmF crystal structures. Chaperone domains are shown in teal, Cbl-binding domains are shown in yellow and orange, substrate-binding domains are shown in green, and linker regions are shown in magenta. Chain A is shown in gray. GMPPNP is shown in yellow ball-and-stick representation. (B) Close-up view of the interdimer interface. (C) As in (B) and rotated 90° towards the viewer. The relative chaperone domain arrangement is reminiscent of structures of HypB (13) and UreG (14).
Figure II.2. Difference in Cbl-binding domain positioning in open and closed states. (A) Open state of IcmF-GMPPNP, colored as in Figure II.1. The G-domain forming the interdimer interface is shown in light blue. (B) Closed state of IcmF-GDP bound to AdoCbl (PDB 5CJT) (5), positioned and colored as in (A). Red lines denote the hinge-like motion of Cbl-binding domain closure upon AdoCbl binding (5).
Figure II.3. Two residues interact with the bound GMPPNP across the interdimer interface. (A) Ribbon representation of the interdimer interface at the G-nucleotide binding site. The chaperone domains are colored teal and light blue, Cbl-binding domain colored orange, substrate-binding domain green, and GMPPNP yellow. The 22 Å distance between K344’ and GMPPNP bound within the same domain is shown as a violet dotted line. (B) Ball-and-stick representation of the chaperone:nucleotide interactions with ribbon representation only for the crystal lattice-contributed chaperone domain. Gln341 and Lys344 from the lattice-contributed chaperone domain directly interact with the GMPPNP gamma phosphate. Mg\(^{2+}\) is shown as a green sphere. Coloring as in Figure II.1 (C) Sequence alignment of the Switch III region of IcmF, MeaB from *Methylobacterium extorquens*, and human MMAA. Residues investigated in this work are highlighted in red.
Residues at the interdimer interface participate in GTP hydrolysis

To validate the involvement of the interdimer interface in GTP hydrolysis, two key residues, Gln341 and Lys344, were mutated individually to Ala. IcmF-K344A is inactive for GTP hydrolysis and IcmF-Q341A has substantially reduced GTPase activity (Figure II.4). The distance between Lys344 and the third phosphate of GMPPNP bound within the same protomer is ~22 Å, too far to contribute directly to hydrolysis (Figure II.3). However, Lys344 directly interacts with the third phosphate of the GMPPNP molecule located across the interdimer interface in IcmF-GMPPNP, supporting the involvement of the interdimer interface in GTP hydrolysis (Figure II.4).

Figure II.4. GTPase activity of IcmF variants. All assays were performed in triplicate. Error bars represent ± standard deviation.
**IcmF forms multimers when bound to a GTP analog**

The IcmF homodimer is assumed to be symmetric in solution. If an interdimer interface were to exist upon binding GTP, it could be extended on both sides of the IcmF homodimer to produce an elongated multimer where IcmF dimers bind end to end through G-protein domain interactions. We assayed for oligomerization using negative stain electron microscopy of purified IcmF in the presence or absence of the non-hydrolyzable GTP analog GMPPCP. Electron microscopy images reveal that addition of GMPPCP induces an elongated multimeric conformation in IcmF that is not seen in IcmF without nucleotide (Figure II.5).

**Figure II.5.** Negative stain electron microscopy of IcmF. (A) IcmF in the absence of nucleotide forms single homodimers. (B) IcmF in the presence of 1 mM GMPPCP forms long multimers. Scale bars represent 100 nm.

**Multimer formation depends on the cofactor state of IcmF**

GTP-induced opening of the Cbl-binding domain by multimerization would be unproductive for catalysis when IcmF is in an active state bound to AdoCbl. In this state, opening the Cbl-binding
domain would inactivate the enzyme and potentially release valuable AdoCbl or radical intermediates into solution. Negative stain EM reveals that indeed IcmF-AdoCbl does not form a multimer upon addition of GMPPNP (Figure II.6A), suggesting that multimer formation is blocked by the presence of an active cofactor. Consistently, enzyme containing inactive cofactor, achieved through photolysis of the AdoCbl Co-C bond or through direct addition of hydroxocobalamin (OHCbl), does form multimers upon addition of GMPPNP (Figure II.6B,C), consistent with the chaperone domain enabling cofactor repair.
**Figure II.6.** Negative stain electron microscopy of IcmF in different cofactor states. (A) IcmF incubated with AdoCbl followed by incubation with GMPPNP. IcmF is seen as almost exclusively monodisperse homodimers. (B) Sample as in (A) but exposed to light for 2 min prior to staining. IcmF forms multimers following light exposure. (C) IcmF incubated with OHCbl followed by incubation with GMPPNP. IcmF-OHCbl forms multimers similarly to light-exposed IcmF-AdoCbl, consistent with photolysis-induced, and G-nucleotide triphosphate-dependent multimer formation. Red dotted circles in (B) and (C) highlight exemplary multimers.
**Discussion:**

Although a wealth of structural and biochemical data exists on MMAA, Mut (4) and their bacterial homologs MeaB (15) and MCM (16), a high-resolution structure of the complex between a dimeric chaperone and its cognate mutase has remained elusive (17). Findings that MMAA and MeaB have greatly improved GTPase activity when in complex with their cognate mutases (4, 18) suggests an active conformation that has not been characterized through crystal structures of the individual G-proteins.

Prior to this work, the fused IcmF system appeared to behave differently from MMAA and MeaB due to the separation of G-domains on opposite sides of the IcmF homodimer (5, 17). The structure of IcmF-GMPPNP reveals that Switch III residues in IcmF interact directly with the GTP-analog of a neighboring G-domain across a crystal lattice contact. Electron microscopy of IcmF-GMPPCP confirms that the chaperone interface occurs in solution. Furthermore, residues in the Switch III region of the IcmF G-domain are involved in GTP hydrolysis, demonstrating a physiological role for IcmF multimerization. IcmF discriminates between active and inactive cofactor states and selectively multimerizes in the presence of GTP analogs only when inactive cofactor is bound. Taken together these data lead us to propose a concise molecular model for chaperone-mediated cofactor delivery and repair in the IcmF model system (Figure II.7).
**Figure II.7.** Molecular model for chaperone-mediated cofactor delivery and repair. IcmF, shown as one half of the homodimer for clarity, with GDP bound equilibrates between open and closed conformations of the Cbl-binding domain, with the G-domain and Cbl-binding domain moving together as a rigid unit (5). Upon GTP binding, IcmF homodimers assemble at the interdimer interface, structurally locking the Cbl-binding domains into an open conformation. In the open conformation of mutase domains, ATR is able to deliver AdoCbl. The binding energy of AdoCbl and the hydrolysis of GTP release the interdimer interface and allow Cbl-binding domain closure. In the catalytically competent homodimeric state (upper right) IcmF is prone to occasional inactivation (6). Following damage to the AdoCbl cofactor and formation of Cob(II)alamin, the interdimer interface is able to reform in a GTP-dependent manner, opening the Cbl-binding domain for Cob(II)alamin release to ATR (7). Once damaged cofactor has been removed, ATR can again deliver AdoCbl to IcmF and repeat the cycle.
The relative arrangement of G-domains in the IcmF multimer is comparable to the proposed quaternary arrangement of the human and bacterial systems MMAA:Mut and MeaB:MCM, respectively (17). However, the orientations of G-domains at the interdimer interfaces in the IcmF multimer are distinct from crystal structures of either MMAA (4) or MeaB (15). In these structures of dimeric chaperones, the G-nucleotide-binding regions are farther apart than in IcmF, and the Switch III regions do not interact with the nucleotide from the neighboring protomer within the chaperone dimer (Figure II.8A,B). The loss of GTPase activity due to Switch III region mutations cannot be explained by the crystallographic structures of MMAA or MeaB. In contrast, when monomers of MMAA and MeaB are superimposed onto the IcmF interdomain interface structure, the loss in GTPase activity following Switch III mutation has a molecular rationale. As is the case with IcmF, Switch III residues are now juxtaposed to the nucleotide-binding site (Figure II.8C,D). The fact that mutations of Switch III in IcmF, MMAA:Mut, and MeaB:MCM elicit a similar response is consistent with G-domains of all three systems achieving a similar conformation when in an active GTP-bound and mutase-bound state. Although further work will be required to provide structural support for the modeled states of MeaB and MMAA shown in Figure II.8C,D, we now have a molecular proposal for the role of Switch III residues in the AdoCbl G-protein chaperones.
Figure II.8. Conformational changes required for MeaB (PDB 4JYB) (11) and MMAA (PDB 2WWW) (4) to form an IcmF-like dimeric arrangement. Switch III regions (residues 177-188 of MeaB and 265-276 of MMAA), although largely disordered in both structures, are highlighted red and circled red. (A) MeaB bound to GMPPNP in the crystallographic dimer arrangement. GMPPNP is shown in yellow ball-and-stick. (B) MMAA bound to GDP in the crystallographic dimer arrangement with the upper protomer at the same orientation as MeaB in (A). GDP is shown in yellow ball-and-stick. (C) MeaB superimposed onto the arrangement of G-domains at the IcmF interdimer interface (Figure II.1). (D) MMAA superimposed onto the arrangement of G-domains at the IcmF interdimer interface. Switch III regions in both MeaB and MMAA are far from the neighboring nucleotide in both crystallographic conformations (A, B) and approach the neighboring nucleotide in the IcmF-modeled conformations (C, D).
Materials and Methods

Mutagenesis

IcmF-Q341A (Forward Primer: 5’-GCGCGGCCAGCGCTTCGAGAAGATCGAC-3’; Reverse Primer: 5’-GTCGATCTTCTCGAGCGCTGGCCGCGC-3’) and IcmF-K344A (Forward Primer: 5’-CCAGCCAGCTCGAGGCGATCGACATGCTCGAC-3’; Reverse Primer: 5’-GTCGAAGCCTGATCGACTGCTGCCTGACTGCTCGAC-3’) pET28a vectors were engineered from wild type IcmF-pET28a (5) using the QuikChange Site-Directed Mutagenesis Kit following manufacturer’s instructions. Sequences were verified by Sanger sequencing.

Protein expression and purification

IcmF with an N-terminal His-tag from Cupriavidus metallidurans and all variants were expressed and purified as follows. A 10 mL starter culture of BL21(DE3)pLysS cells (Invitrogen) transformed with pET28a-IcmF was grown in LB medium supplemented with 50 mg/L kanamycin and 34 mg/L chloramphenicol at 37 °C with shaking. After 20 h, 2 mL of starter culture was used to inoculate 1 L of LB supplemented with 50 mg/L kanamycin and 34 mg/L chloramphenicol and grown at 37 °C with shaking. When the optical density at 600 nm (OD$_{600}$) was between 0.5 and 0.6, the temperature was reduced to 16 °C. After 2 h, cells were induced with 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). Cells were harvested by centrifugation after 10-12 h induction at 16 °C. Cell pellets were flash-frozen in liquid N$_2$ and stored at -80 °C.
Cell pellet from 1 L of culture was resuspended in 80 mL of lysis buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 20 mM imidazole) supplemented with 1 mM PMSF, 1 mM benzamidine-HCl, 5 µL Benzonase (Novagen), and 1 Roche “cOmplete” EDTA-free protease inhibitor cocktail tablet. Cells were lysed by sonication. Cell lysate was clarified by centrifugation, and the supernatant was filtered through a 0.22 µm filter. Filtered lysate was loaded onto a 5 mL HisTrap column (GE Healthcare) and eluted with a gradient of elution buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 500 mM imidazole). Elution fractions corresponding to IcmF were concentrated in 50 kDa molecular-weight cutoff (MWCO) concentrators and buffer exchanged twice with buffer A (50 mM HEPES pH 7.5) prior to loading onto a MonoQ ion-exchange column. IcmF was eluted from the MonoQ column with a linear gradient of buffer B (50 mM HEPES pH 7.5, 500 mM NaCl). Elution fractions from the MonoQ column were concentrated in 50 kDa MWCO concentrators and loaded onto an S200 16/60 size-exclusion column equilibrated in buffer C (20 mM HEPES pH 8, 50 mM NaCl). IcmF eluted from size-exclusion in a single peak with an estimated molecular weight corresponding to a dimer. Elution fractions were concentrated in 50 kDa MWCO concentrators to 10-20 mg/mL and flash frozen in liquid N₂ before being stored at -80 °C.

**Electron microscopy**

Negative-stain electron micrograph grids for 3D reconstruction of IcmF multimers were prepared with protein solutions of 20 µg/mL wild type IcmF, 200 µM GMPPCP (Sigma-Aldrich), and 300 µM MgCl₂ in buffer C. The solution was applied to a 300 mesh continuous carbon grid (EMS) and stained three times with a 2% uranyl acetate solution.
Electron micrograph grids for IcmF-K344A were prepared identically to wild type IcmF: 20 μg/mL IcmF-K344A, 200 μM GMPPNP, and 300 μM MgCl₂ in buffer C.

Samples and grids containing AdoCbl were prepared in the dark under red light to prevent Co-C bond cleavage. Light-cleaved IcmF-AdoCbl samples were prepared by exposing IcmF-AdoCbl samples to white light for 15 min.

Preliminary images were collected on an FEI Tecnai Spirit electron microscope at 80 kV at the W.M. Keck Microscopy Facility at the Whitehead Institute. The grids were imaged at 49,000x magnification with a pixel size of 2.55 Å at the specimen level.

**Crystallization**

IcmF protein crystals were prepared as described previously (5). The hanging-drop vapor diffusion technique was used with a precipitant solution containing 700 mM potassium sodium tartrate, 100 mM Tris pH 8.5, and 0.5% (w/v) PEG MME 5000. IcmF protein solution (1 μL) containing 7 mg/mL IcmF and 5 mM n-butyryl-CoA (Sigma-Aldrich) in buffer C was mixed with 1 μL precipitant solution on a glass cover slide and sealed over 500 μL precipitant solution. Crystals of IcmF formed within two weeks at 22 ºC, contained bound substrate, and were amenable to soaking experiments with G-nucleotides. Crystals of IcmF-GMPPNP were generated by transfer of apo-IcmF crystals into a precipitant solution supplemented with 5 mM n-butyryl-CoA, 1.5 mM MgCl₂, and 1 mM GMPPNP. All crystals were cryoprotected by step-wise transfer through four solutions of increasing glycerol concentration to the final cryo solution composed of soak solution supplemented with 25% (v/v) glycerol before being flash-frozen in liquid N₂.
**Data collection and processing**

The data set of IcmF-GMPPNP was collected at the Advanced Photon Source (APS) on beamline 24ID-C with a Pilatus 6M pixel detector at 100 K and 12664 eV. IcmF-GMPPNP was collected in a 120° wedge and processed with HKL2000 (19).

**Structure determination and refinement**

IcmF-GMPPNP was solved through rigid body refinement with IcmF/GDP (PDB ID 4XC8) as a starting model in Phenix (20). The asymmetric unit is composed of one IcmF homodimer. Initial refinements included positional refinement, domain-specific non-crystallographic symmetry restraints, group B-factor refinement, and simulated annealing to remove model bias. Nucleotides, AdoCbl, and Mg$^{2+}$ ions were added and refined with bond length restraints in Phenix. Refinement restraints for GMPPNP were generated using the Grade Web Server (http://grade.globalphasing.org) (21). Iterative rounds of model building and refinement in Coot (22) and Phenix, respectively, yielded R<sub>work</sub> and R<sub>free</sub> of 24.2% and 26.0%, respectively. The final model contains residues 23-1093 (of 1093) in chain A and chain B. The model was confirmed by use of composite omit electron density maps.

**GTPase assays**

The $k_{\text{cat}}$ values for GTP hydrolysis by wild type and mutant IcmF variants were determined using the EnzCheck GTPase assay kit following manufacturer’s instructions. Assays contained 2.5 µM IcmF (concentration determined using A<sub>280</sub> with an extinction coefficient of 84690 M<sup>-1</sup>cm<sup>-1</sup> and molecular weight of 121 kDa) and a starting GTP concentration of 200 µM as higher
concentrations of GTP resulted in decreased GTPase activity, likely due to multimerization of IcmF. Reported $k_{cat} \pm$ s.d. values are the results of at least three independent experiments.

References:


7. Z. Li et al., Cofactor editing by the G-protein metallochaperone domain regulates the radical B12 enzyme IcmF. Journal of Biological Chemistry 292, 3977-3987 (2017).


Chapter III:
A G-protein Chaperone Dimer Rearrangement Underlies Interaction with its B12-dependent Enzyme

Summary
The delivery of adenosylcobalamin to methylmalonyl-CoA mutase (MCM) and repair of inactive cofactor formed during mutase enzymatic activity are both mediated by a dimeric G-protein chaperone. Although there is extensive biochemical data detailing the specific molecular requirements governing cofactor delivery and repair, the structural underpinnings of G-protein signaling are unclear. Here we use a minimal model system composed of a dimeric G-protein chaperone (MeaB) and the Cbl-binding domain of MCM from *Methylobacterium extorquens*, MCM_{cbl}, to investigate GTP-induced conformational changes within MeaB. The MeaB:MCM_{cbl} complex is stable in solution when MeaB is bound to a GTP analog and MCM_{cbl} enhances GTP hydrolysis by MeaB. An orthologous model system composed of the human G-protein chaperone, MMAA and human MCM_{cbl} displays similar biochemical properties. The crystal structure of MeaB:MCM_{cbl} reveals large rearrangements in quaternary structure leading to a GTP-hydrolysis competent conformation of the MeaB dimer. The MeaB:MCM_{cbl} structure also shares significant similarity with a crystal lattice interface in crystal structures of IcmF, supporting a unified model for G-protein chaperone function across all model systems.

Contributions: This chapter was written with Professor Catherine L. Drennan. Francesca Vaccaro assisted in crystallography, structural modeling, and analysis.
Introduction

The trafficking and modification of complex metallocofactors is essential for numerous metabolic processes. Vitamin B12, or cobalamin (Cbl), is an essential human nutrient taken up through the diet, transported, and modified through a complex series of protein-mediated steps.

Two human enzymes utilize Cbl cofactors to perform key metabolic functions: methionine synthase uses methylcobalamin (MeCbl) to catalyze the formation of methionine from homocysteine \( \text{(1)} \), and methylmalonyl-CoA mutase (MCM or Mut) uses adenosylcobalamin (AdoCbl) to catalyze a carbon skeleton rearrangement of methylmalonyl-CoA to succinyl-CoA \( \text{(2)} \). MCM activity is necessary for efficient metabolism of certain amino acids, fatty acids, and other metabolites \( \text{(2)} \). MCM activity in humans can be disrupted directly by mutations within the enzyme itself or indirectly by dysfunction in any protein necessary for transport and modification of the essential AdoCbl enzyme cofactor \( \text{(3)} \). Deficiency of MCM activity leads to methylmalonic aciduria (MMA), an inborn error of propionate metabolism \( \text{(4)} \).

The final step in MCM activation is the direct transfer of AdoCbl from ATP:cob(I)alamin adenosyltransferase (ATR) to the Cbl-binding domain of MCM \( \text{(5)} \). Cofactor transfer is gated by a G-protein chaperone, methylmalonic aciduria type A protein (MMAA), through a GTP-hydrolysis dependent mechanism \( \text{(6)} \). MMAA also participates in the repair of damaged Cbl formed during MCM catalysis \( \text{(6)} \).

Two bacterial model systems for the study of AdoCbl transfer to MCM have yielded extensive biochemical data on the interplay between ATR, MMAA, and MCM. The orthologous three-protein system from Methylobacterium extorquens behaves similarly to the human system with the distinction that \textit{M. extorquens} MCM is a heterodimer composed of one active and one inactive subunit whereas human MCM is a homodimer. The isobutyryl-CoA mutase fused
(IcmF) system from *Cupriavidus metallidurans*, a natural fusion protein of a single G-protein chaperone domain with an AdoCbl-dependent mutase, has also been informative for the study of AdoCbl transfer to AdoCbl-dependent mutases (7-9). The crystal structure of IcmF demonstrates the relative positioning of G-protein and mutase domains (8). Recent work has revealed a chaperone-to-chaperone interface between IcmF homodimers that appears to be directly involved in AdoCbl delivery and repair to/from the mutase Cbl-binding domain (Chapter II). The relative orientation of G-protein chaperone domains observed in IcmF is distinct from the dimeric arrangement of MMAA (10) and the G-protein chaperone from *M. extorquens* (MeaB) (11). Given the structural discrepancies between IcmF and the homodimeric MeaB and MMAA proteins, a structure-based mechanistic model for G-protein chaperone function has not been proposed.

Here we find that MeaB and MMAA interact directly with an isolated Cbl-binding domain of their respective cognate MCM (MCM_{cbl}). The crystal structure of MeaB bound to β,γ-methyleneguanosine 5′-triphosphate (GMPPCP) in complex with MCM_{cbl} demonstrates a large conformational change in MeaB upon binding MCM_{cbl}, creating a dimeric arrangement of MeaB protomers distinct from previously published crystal structures (11, 12). The resulting dimeric arrangement is induced by binding GMPPCP and forming a protein:protein complex with the Cbl-binding domain of MCM. Restructuring of a loop region termed Switch III, that is necessary for MCM-mediated GTPase activation (12), explains the molecular mechanism of MCM GTPase activation. The “active” conformation of MeaB is analogous to the relative orientation of chaperone domains in the IcmF interdimer interface (Chapter II) and supports a unified mechanism for chaperone-mediated AdoCbl delivery to B12-dependent mutases.
Results

*Cbl*-binding domains are GTPase activating proteins

Full length MCMs demonstrate GTPase activating protein (GAP) properties for their cognate G-protein chaperones in human (10) and bacterial (13) systems. The necessary interactions between chaperone and mutase that facilitate activation are unclear. Structures of IcmF imply that a majority (~70% of the surface area) of the chaperone:mutase interface occurs on the mutase Cbl-binding domain (8). To investigate whether the Cbl-binding domain is sufficient to act as a GAP for its cognate G-protein chaperone, single Cbl-binding domains from MCM of *Methylobacterium extorquens* (*meMCM*<sub>cbl</sub>) and MCM from humans (*hsMCM*<sub>cbl</sub>) were engineered, expressed, and purified in *E. coli*. GTP hydrolysis activities of MeaB and MMAA are increased upon addition of the cognate Cbl-binding domain *meMCM*<sub>cbl</sub> and *hsMCM*<sub>cbl</sub> (Figure III.1), respectively, demonstrating that the Cbl-binding domains exhibit GAP properties.

![GTPase activity of MMAA and MeaB with and without cognate MCM<sub>cbl</sub>](image)

**Figure III.1.** GTPase activity of MMAA and MeaB with and without cognate MCM<sub>cbl</sub>. GTPase activity is relative to the G-protein alone. Assays performed in duplicate (MMAA) and triplicate (MeaB). Error bars present ± standard deviation.
**G-protein:Cbl-binding domain complexes are stable**

For the Cbl-binding domains to activate their cognate G-proteins, a direct molecular interaction must occur. Biochemical evidence indicates that stable G-protein:mutase complexes form in the presence of non-hydrolyzable GTP analogs \((10, 13)\), suggesting a method for trapping this protein:protein complex. Incubation of either MeaB or MMAA with GMPPCP and the cognate MCM\(_{cbl}\) results in a stable complex that elutes from SEC with a predicted 2:2 stoichiometry between G-protein and Cbl-binding domain (Figure III.2). The protein complex can be purified from the free G-protein dimer by SEC (data not shown).

**Figure III.2.** Size-exclusion chromatography trace of MeaB with and without \(me\text{MCM}_{cbl}\) and various nucleotides. MeaB only forms a complex with \(me\text{MCM}_{cbl}\) in the presence of the non-hydrolyzable GTP analog GMPPCP. MMAA:\(hs\text{MCM}_{cbl}\) also forms a complex in the presence of GMPPCP. Elution peaks at 14.3 mL corresponds to approximately 70 kDa in molecular weight, and 13.5 mL corresponds to approximately 110 kDa in molecular weight. The MMAA dimer is \(~80\text{ kDa, slightly larger than MeaB. MMAA and MMAA:}\text{MCM}_{cbl}\) elute at \(~13.8\text{ mL and 13.1 mL, respectively.}\)
**MeaB:meMCM\textsubscript{cbl} crystal structure reveals an active chaperone arrangement**

Using purified MeaB:meMCM\textsubscript{cbl} bound to GMPPCP, we were able to obtain diffraction quality crystals of the MeaB:meMCM\textsubscript{cbl} complex. The 3.1 Å-resolution crystal structure of purified MeaB:meMCM\textsubscript{cbl} with bound GMPPCP reveals a second conformation of MeaB distinct from the previously observed crystallographic conformations of MeaB (11, 12) and MMAA (10) (Table III.1, Figure III.3). The quaternary arrangement of protomers is similar to the interdimer interface previously reported in crystal structures of IcmF (Figure III.3C) (Chapter II).

SIMIBI G3E GTPases utilize conserved GTP binding motifs (14, 15). The Switch I, Switch II, Walker A, Walker B, and P-loop motifs are all resolved in MeaB:MCM\textsubscript{cbl} (Figure III.4). The Switch III region of the AdoCbl-associated G-protein chaperones has only been recently identified (12). Although residues of Switch III have displayed structural heterogeneity in published structures of MeaB (11, 12), in MeaB:MCM\textsubscript{cbl} Switch III is well ordered (Figure III.4). The ordering of Switch III situates residues for direct interaction with nucleotide bound in the opposite protomer of the MeaB dimer (Figure III.5). The positioning of MeaB Switch III residues Gln185 and Lys188 is consistent with their involvement in GTP hydrolysis only when in complex with MCM (12) (Figure III.5).

Intriguingly, the arrangement of MeaB protomers and the structure of key loop regions at the chaperone interface are consistent with the chaperone:chaperone crystal lattice interface in crystal structures of IcmF (Figure III.3C and Figure III.5) (Chapter II) (8).
Table III.1. Crystallographic data collection and refinement statistics.

<table>
<thead>
<tr>
<th>Data set</th>
<th>MeaB:MCM_{cbl}</th>
</tr>
</thead>
</table>

**Data collection**
- **Space group**: \( P2_12_12_1 \)
- **Cell dimensions**: \( a, b, c (\text{Å}) = 66.1, 81.0, 166.2 \)
- **Wavelength (Å)**: 1.54
- **Resolution (Å)**: 3.10
- **\( R_{\text{meas}} \) (%)**: 27.4 (103.7)
- **CC1/2**: (80.1)
- **\( I/\sigma_I \)**: 7.5 (2.2)
- **Completeness (%)**: 94.2 (100)
- **Redundancy**: 6.7 (7.3)
- **No. of reflections**: 112637 (8942)

**Refinement**
- **Resolution (Å)**: 3.10
- **No. of reflections used**: To be determined*
- **Average B-factors (Å\(^2\))**
  - **Protein**: To be determined*
  - **Non-protein**: To be determined*
  - **\( R_{\text{work}}/R_{\text{free}} \) (%)**: To be determined*
- **No. atoms**
  - **protein**: To be determined*
  - **GMPPCP**: To be determined*
  - **Mg\(^{2+}\)**: To be determined*
- **R.m.s. deviations**
  - **Bond lengths (Å)**: To be determined*
  - **Bond angles (')**: To be determined*
  - **Rotamer outliers (%)**: To be determined*
  - **Ramachandran angles**
    - **Outlier (%)**: To be determined*
    - **Allowed (%)**: To be determined*
    - **Favored (%)**: To be determined*

*Structure is currently being refined. This table will be updated with final statistics.
† Values in parentheses indicate highest resolution bin.
§ Reflections used to calculate \( R_{\text{free}} \) comprise 5% of observed reflections.
Figure III.3. MeaB:MCM_cbl demonstrates a new conformation of the MeaB dimer. (A) MeaB:MCM_cbl shown in ribbon representation. The MeaB homodimer is colored in dark and light green, and the MCM_cbl protomers are colored in dark and light purple. GMPPCP is shown in yellow ball-and-stick representation. (B) The MeaB homodimer from (A) shown without MCM_cbl. (C) The interdimer interface of IcmF-GMPPNP (Chapter II) contains the same quaternary domain arrangement as MeaB:MCM_cbl. IcmF is shown in ribbon representation with G-domain (residues 161-418) and Cbl-binding domain (residues 21-160) of one protomer in dark blue and pink and the same domains from the second, symmetry-related protomer in light blue and pink. GMPPNP is shown in yellow ball-and-stick representation. (D) Structure of MeaB alone (brown and orange, PDB 2QM7) (11) with the upper MeaB protomer oriented as in MeaB:MCM_cbl in (B). GDP is shown in yellow ball-and-stick representation. (E) Structure of MMAA (dark and light purple, PDB 2WWW) (10) with the upper protomer in the same orientation as the upper MeaB protomer from MeaB:MCM_cbl in (B). GDP is shown in yellow ball-and-stick representation. (F) MeaB dimer arrangement from MeaB:MCM_cbl oriented as in (B) with the upper protomer shown as a gray surface and the lower protomer in green ribbon representation. The lower protomer of MeaB alone (orange, PDB 2QM7) (11) is superimposed in the same orientation as (D). The upper protomer of the MeaB dimer is very similar to the upper protomer of MeaB:MCM_cbl and is not shown for clarity. A red arrow denotes the conformational change required to form the MeaB:MCM_cbl complex.
Figure III.3 (Continued).
Figure III.4. Structure of nucleotide-binding regions of MeaB in MeaB:MCM<sub>cbl</sub>. (A) Key nucleotide-binding regions: the Switch I, Switch II, Switch III, base specificity loop, and the P-loop are colored orange, gray, pink, dark blue, and purple, respectively. (B) Labeled representation as in (A) without ribbon representation. Primes indicate that residues are contributed by second protomer.
**Figure III.5.** Structure of the Switch III regions of MeaB:MCM_{cbl} and IcmF. (A) Zoomed in view of the Switch III region (residues 177-188, pink) of MeaB:MCM_{cbl} (non-Switch III region in green). GMPPCP is shown in yellow. (B) Zoomed in view of the Switch III region (residues 333-344, pink) of IcmF at the symmetry-related interdimer interface (non-Switch III region of G-domains colored dark and light blue and Cbl-binding domain colored dark pink). Nucleotides from the other side of the dimer interface can be seen in the background. (C) Sequence alignment of the Switch III regions of MeaB, MMAA, and IcmF. Key residues interacting directly with G-nucleotide across the dimer interface are highlighted in red.

The Switch III region is poised for GTP hydrolysis across the dimer interface of MeaB:MCM_{cbl}

The conformational change in the MeaB dimer allows the Switch III region to come into direct contact with bound nucleotide from across the dimer interface (Figure III.5). Previous work has
demonstrated that the Switch III region is required for MCM-mediated GTPase activation and mutations to Switch III in MMAA cause MMA (12). In crystal structures of MeaB or MMAA alone, the Switch III region is highly variable or disordered (10-12). The ordering of this key region of MeaB upon complex formation with MCM may be the molecular signal required for GTPase activation. As was observed in IcmF (Chapter II), a lysine residue from Switch III (K188 in MeaB, K344 in IcmF) directly interacts with an oxygen from the gamma phosphate of GMPPCP. Lys188 is required for MCM-induced GTPase activation (12). A lysine participating in dimer-mediated NTP hydrolysis is reminiscent of other structures from the SIMIBI G3E family including HypB (16) and UreG (17), although the Switch III region is not conserved between MeaB and HypB or UreG.
**Discussion**

In this chapter, we demonstrate a rearrangement of the MeaB dimer in response to a bound GTP analog and MCM\textsubscript{cbl}. The result of this conformational change is a dimeric arrangement consistent with structures of IcmF (Chapter II). The high degree of structural similarity between IcmF mutase domains and MCM from humans (10) and from bacteria (18) allows us to produce a molecular model for the complete MeaB:MCM structure (Figure III.6). This model provides a visualization of our proposed functional role for the G-protein in which it establishes a physical wedge between Cbl-binding and substrate-binding domains of the mutase enzyme (Chapter II and Figure III.6).

The human model system, MMAA:MCM\textsubscript{cbl}, displays analogous GTPase activation and solution complex formation behavior to MeaB:MCM\textsubscript{cbl}. Taken together with previous results indicating the importance of Switch III in both systems (12) and the tertiary structure similarities between MeaB and MMAA (10, 11) we proposed a consensus mechanism for G-protein-mediated AdoCbl delivery and cofactor repair (Figure III.7).
Figure III.6. Molecular model of the full-length MeaB:MCM complex. MeaB:MCM\textsubscript{cbl} was used to superimpose the domain arrangement of Chain B from IcmF-GMPPNP (Chapter II) onto a single MeaB:MCM\textsubscript{cbl} interaction. Using IcmF to model the tertiary arrangement of MCM domains in an open conformation, the substrate-binding domain of MCM (MCM\textsubscript{substrate}, blue) and the inactive MCM subunit (MCM\textsubscript{inactive}, yellow) from \textit{P. freudenreichii} (PDB 4REQ)\textsuperscript{(18)}, were placed by structural superimposition of MCM\textsubscript{substrate} onto the IcmF substrate-binding domain. \textit{M. extorquens} MCM and \textit{P. freudenreichii} MCM share 30\% sequence identity in the active subunit (MCM\textsubscript{cbl} and MCM\textsubscript{substrate}). The resulting complex demonstrates a domain arrangement with an open MCM\textsubscript{cbl}, and the overall architecture is consistent with negative stain electron microscopy of the MeaB:MCM complex\textsuperscript{(19)}. Red lines indicate the relative angle of MCM\textsubscript{cbl} and MCM\textsubscript{substrate} compared to an active, closed conformation with bound AdoCbl.
Figure III.7. Molecular model for G-protein-mediated AdoCbl delivery and cofactor repair. GTP binding induces a conformational change in the G-protein dimer, structurally wedging open the Cbl-binding domain of the mutase (Chapter II). With the mutase Cbl-binding domain fixed in the open conformation, ATR can transfer AdoCbl to the apo-mutase. AdoCbl binding and GTP hydrolysis release the G-protein hydrolysis-active conformation, causing a change in dimeric configuration, and allowing the mutase Cbl-binding domain to close. Following cofactor damage during catalysis, GTP binding can again induce a wedge-like conformation of the G-protein dimer and reopen the Cbl-binding domain for subsequent repair of the damaged cofactor by ATR. Only one substrate-binding domain and one Cbl-binding domain of the mutase is shown for simplicity.
Central to the dimer rearrangement observed in our crystal structure of MeaB:MCM_{cbl} is the ordering of key loop regions at the dimer interface, specifically the Switch III region. Mutations to Switch III cause MMA (20), and Switch III variants of MeaB are not activated for GTP hydrolysis upon binding MCM (12). The ordering of Switch III to interact directly with the nucleotide from across the dimer interface solidifies its role in hydrolysis only when in an active conformation, bound to MCM or MCM_{cbl}. Conformational switching between the crystallographic conformation of MeaB alone and the mutase-bound conformation elegantly explains the biochemical data reported to date on MeaB and MMAA.

An intriguing consideration is the apparent ability of MeaB to detect when cofactor damage has occurred (13, 21). For this to be the case, MeaB-induced Cbl-binding domain opening must be triggered by formation of Cob(II)alamin following AdoCbl degradation. Direct evidence supporting the chaperone’s ability to discriminate between cofactor states has been observed in IcmF using electron microscopy (Chapter II). IcmF establishes chaperone domain interactions and forms multimers in the presence of non-hydrolyzable GTP analogs and inactive cofactor but does not form multimers when AdoCbl is bound, even in the presence of a non-hydrolyzable GTP analog (Chapter II). These multimer structures appear to open the Cbl-binding domain and allow for cofactor repair by ATR. A plausible mechanism for triggering cofactor removal is that the binding energy between the Cbl-binding domain and substrate-binding domains of MCM is too large for MeaB to open the Cbl binding site when MCM is bound to AdoCbl. The Ado moiety of the AdoCbl extends into the substrate-binding TIM barrel domain, increasing the contact region by ~200 Å² (as determined using PISA (22)). Following formation of Cob(II)alamin, fewer contacts are made between domains, lowering the energetic barrier for
domain opening and Cob(II)alamin release. It remains to be determined how an open state of the Cbl-binding domain triggers Cob(II)alamin uptake by ATR.

In summary, we have provided the first structural evidence that the homodimeric G-protein chaperones involved in AdoCbl trafficking undergo a large conformational change upon binding GTP and the Cbl-binding domain of their target B12-dependent mutase. This conformational change enables switch-like behavior of the G-protein in response to G-nucleotides and situates the G-protein as a structural wedge between the Cbl-binding and substrate-binding domain of the mutase, facilitating cofactor delivery or repair.
Materials and methods

Preparation of Cbl-binding domain constructs

The DNA encoding MeaB and the Cbl-binding domain of methlymalonyl-CoA mutase (WP_015857646.1) from Methyllobacterium extorquens (meMCM_{cbl}) were synthesized individually by GenScript and inserted into pET28a expression vectors at the NdeI and XhoI restriction sites. Both vectors contained N-terminal hexahistidine tag. The Cbl-binding domain of MCM comprised residues 545-712 of the full-length protein.

Preliminary data indicated the meMCM_{cbl} with an N-terminal hexahistidine has low solubility. A pET28a vector containing meMCM_{cbl} with a C-terminal hexahistidine affinity tag was constructed using Gibson assembly (23) at the NcoI and XhoI restriction sites to improve meMCM_{cbl} solubility.

Genes encoding human MMAA (AAI01182.1) and human MCM (NP_000246.2) without the mitochondrial leader peptides were synthesized individually by GenScript and inserted into pET28a expression vectors at the NdeI and XhoI restriction sites using Gibson assembly. The Cbl-binding domain of human MCM, hsMCM_{cbl}, residues 576-750, was cloned into a pET28a expression vector through Gibson assembly at the NdeI and XhoI restriction sites with an N-terminal hexahistidine affinity tag.

Protein purification

Expression and purifications of MeaB, MMAA, meMCM_{cbl}, and hsMCM_{cbl} were conducted following the same procedure. An overnight starter culture of 100 mL LB supplemented with 50 mg/L kanamycin was inoculated from a single colony of T7 express cells (NEB) transformed with the appropriate gene and grown at 37 °C with shaking. The starter culture was used to
inoculate 1 L of LB supplemented with 50 mg/L kanamycin at 37 °C with shaking. The 1 L culture was induced with 0.5 mM IPTG when OD_{600} reached ~0.4-0.6 and grown for 16 h at 18 °C with shaking. Cells were harvested by centrifugation and flash frozen in liquid N\textsubscript{2} before being stored in a -80 °C freezer.

A cell pellet from 1 L of cells was resuspended in 80 mL lysis buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 25 mM imidazole) supplemented with 1 mM PMSF and 1 Roche EDTA-free protease inhibitor cocktail. Cells were lysed by ultrasonication and cell lysates were clarified by centrifugation at 25,000 x g for 30 min at 4 °C. Clarified lysate was passed through a 0.2 µm filter before being loaded onto a 1 mL HisTrap column (GE Healthcare). The column was washed with lysis buffer and protein was eluted with a linear gradient of elution buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 500 mM imidazole). Elution fractions were concentrated in either a 30 kDa molecular-weight cutoff (MWCO) concentrator (for MeaB and MMAA) or a 10 kDa MWCO concentrator (for meMCM_{cbl} and hsMCM_{cbl}) before being loaded onto an S200 16/60 size-exclusion column (SEC) equilibrated with SEC buffer (50 mM HEPES pH 7.5, 500 mM NaCl). Elution fractions from SEC were concentrated in a 30 kDa or 10 kDa MWCO concentrator for MeaB and MMAA or meMCM_{cbl} and hsMCM_{cbl} respectively. Final concentrations were ~35 mg/mL for MeaB, ~10 mg/mL for MMAA, ~6 mg/mL for meMCM_{cbl}, and ~5 mg/mL for hsMCM_{cbl}, as determined by A_{280} for MeaB and MMAA or Bradford for meMCM_{cbl} and hsMCM_{cbl}. After concentration, protein samples were flash frozen in liquid N\textsubscript{2} and stored in a -80 °C freezer.
**GTPase assays**

The EnzCheck phosphate assay kit (Molecular Probes) was used for all GTPase assays following the manufacturer’s instructions. Direct comparison assays were used to validate the GTPase-activating protein properties of meMCM<sub>cbl</sub> and hsMCM<sub>cbl</sub>. MeaB (2.5 µM) or MMAA (5 µM) was supplemented with the appropriate MCM<sub>cbl</sub> (5 µM meMCM<sub>cbl</sub> or 10 µM hsMCM<sub>cbl</sub>) in an assay solution containing 200 µM GTP (MeaB:meMCM<sub>cbl</sub>) or 500 µM GTP (MMAA:hsMCM<sub>cbl</sub>). Activity was normalized to MeaB (2.5 µM) or MMAA (5 µM) without addition of MCM<sub>cbl</sub>.

**Complex formation as determined by analytical size-exclusion chromatography**

A S200 10/300 GL SEC column (GE Healthcare) was used to assess complex formation between MeaB and meMCM<sub>cbl</sub> or MMAA and hsMCM<sub>cbl</sub>. A 300 µL volume of sample containing 50 µM MeaB or MMAA and 100 µM meMCM<sub>cbl</sub> or hsMCM<sub>cbl</sub> supplemented with 500 µM of the appropriate nucleotide was injected onto the SEC column. The estimated molecular weights of elution peaks were calculated by comparison to a set of gel-filtration standards (Bio-Rad).

**Crystallographic structure determination of MeaB:meMCM<sub>cbl</sub> complex**

The complex between MeaB and meMCM<sub>cbl</sub> was prepared by incubating 207 µM MeaB (in SEC buffer) and meMCM<sub>cbl</sub> (in SEC buffer) with 1 mM GMPPCP (Sigma) for 1 h. MeaB:meMCM<sub>cbl</sub> complex was purified by size exclusion chromatography (SEC) on a S200 16/60 column (GE Healthcare) equilibrated with SEC buffer. The complex eluted with a molecular weight of approximately 110 kDa. Following SEC, the complex was concentrated to 182 µM MeaB in a 30 kDa MWCO concentrator.
Crystals were formed by the sitting-drop vapor diffusion method with a precipitant solution composed of 20% (w/v) PEG 3350 and 200 mM LiCl. Briefly, 150 nL of protein solution (182 µM complex in SEC buffer supplemented with 182 µM GMPPCP), was mixed with 230 nL precipitant solution and incubated over 70 µL precipitant solution using a Phoenix liquid handling robot (Art Robbins Instruments). Crystals grew within 2 wk at 20 ºC. Crystals were looped by step-wise transfer through three drops of cryo solution to a final solution containing precipitant solution supplemented with 20% glycerol (v/v) and flash frozen in liquid N2.

Data collection and processing

Data for MeaB:MCMcbl were collected on an in-house Cu-Kα rotating anode source (Rigaku) with a Saturn 944 CCD detector in a single 180º wedge with 0.5º per image. Data were processed in XDS and scaled in XSCALE (24) (Table III.1).

Structure determination and refinement

A structure of MeaB:MCMcbl was solved by molecular replacement (MR) with Phaser (25) using data to 3.1 Å-resolution. The final model contained two MeaB protomers and two meMCMcbl protomers in the asymmetric unit. Two MeaB protomers (PDB 2QM7) (11), with the C-terminal dimerization helices removed, were used for the MR. Solutions were identified for each protomer of the MeaB dimer in separate MR searches. The MR partial solution containing placed MeaB protomers was used as a starting point for MR to find the MCMcbl domains. The search model used was a homology model of meMCMcbl that was created using Sculptor on the Cbl-binding domain of methylmalonyl-CoA mutase from Propionibacterium freudenreichii subsp. 
schermyii (PDB 4REQ) (18). Following MR, iterative rounds of model building and refinement in Coot and Phenix, respectively, yielded a final model with an $R_{\text{work}}$ of 28.0% and $R_{\text{free}}$ of 31.5%.
References


9. Z. Li et al., Cofactor editing by the G-protein metallochaperone domain regulates the radical B12 enzyme IcmF. Journal of Biological Chemistry 292, 3977-3987 (2017).


Chapter IV:

Concluding Thoughts on Chaperone-Mediated Cobalamin Delivery and Repair

The G-protein chaperones involved in AdoCbl delivery and repair present an intriguing set of mechanistic questions. The heart of these questions can be summarized as follows: how does a G-protein chaperone affect Cbl mobilization to and from an AdoCbl-dependent mutase without binding Cbl itself? This question has proved difficult to answer, even with a wealth of biochemical and structural data in three orthologous systems.

The goal of this work was to provide a structure-based mechanism for AdoCbl chaperone function. Leveraging our previous structural data depicting the relative arrangement of chaperone and mutase domains in IcmF, we extend the structure-based model to include snapshots of the chaperone in action as it wedges open the Cbl-binding and substrate-binding domains of the mutase. Structural characterization of the two-component MeaB:MCM_{cbl} system and the fusion protein IcmF demonstrate a conserved mechanism for AdoCbl chaperone function.

IcmF proved a serendipitous starting point for this work. The symmetry related interdimer interface within our crystal structures had been observed previously (1). However, the physiological relevance of a dimer-of-dimers or a multimeric IcmF was unclear. Identification of the conserved Switch III region at the IcmF interdimer interface and its involvement in GTP hydrolysis solidified the hypothesis that our crystal lattice reproduces a relevant interface. Pursuing this hypothesis further, and leveraging previous observations that IcmF is prone to precipitate or form aggregates at high concentrations of G-nucleotides, led us to electron microscopy (EM) as a tool to validate formation of an interdimer interface in solution.
The utility of EM extends beyond simply observing the multimerization process. Because of the relatively simple experimental setup and low concentrations of protein and ligands required, EM has proved an ideal tool to directly monitor the molecular requirements for chaperone function during cofactor delivery and repair in IcmF. We have made first strides in understanding the molecular logic of the Cbl repair process whereby the chaperone domain enables release of inactive cofactor. The same experimental setup has great potential in clearly depicting each step in our proposed molecular model (Figure II.7).

Progress in the IcmF system, and recently published work in the MeaB:MCM system (2) highlight the complex multimeric nature of chaperone-mediated Cbl mobilization. To unite our structural model of IcmF with the biochemical data on MeaB and MMAA, direct characterization of a homodimeric chaperone in complex with its cognate mutase was required. However, these systems demonstrate heterogeneity in quaternary conformation (2) and in multimeric state (3) due to the symmetric nature of the G-protein chaperone homodimer and the size and multimeric structures of MCM in bacteria and humans. Our engineered minimal model system, whereby the Cbl-binding domain is removed from the substrate binding domain of the mutase, was created to minimize the multimer complexity in the chaperone:mutase system while maintaining the key interactions directing chaperone function. The chaperone:MCM\textsubscript{cbl} system may even be a naturally occurring arrangement in certain archaeal species that harbor separate genes for the Cbl-binding and substrate-binding domains of AdoCbl-dependent mutases.

Our crystal structure of MeaB bound to MCM\textsubscript{cbl} reveals the missing piece in understanding the role of G-protein chaperones in AdoCbl delivery and repair. The activated conformation of MeaB whereby the Switch III region directly interacts with nucleotide bound across the dimer interface is a key conformation for chaperone activity. These structural data,
combined with our insights from IcmF, lead to a structure-based model whereby binding GTP and MCM causes a conformational change in the chaperone that physically wedges open the Cbl-binding and substrate-binding domains of MCM (Figure III.7).

The work presented here clarifies our understanding of the chaperone:mutase interaction. Further research is required to fully understand the role of ATR in delivering and repairing AdoCbl. Stable protein complexes of ATR with MCM or MeaB:MCM have not been observed, although data indicate cofactor transfer occurs directly between ATR and MCM (4, 5). Some of the techniques used in this work, including EM of IcmF multimers and the MeaB:MCM<sub>cbl</sub> model system, may be useful starting points in understanding the structural underpinnings of the ATR:MCM interaction.
References


Part 2

Chapter V:

Structural Basis for Methylphosphonate Biosynthesis

Summary

Methylphosphonate synthase (MPnS) is a mononuclear iron enzyme that catalyzes the formation of methylphosphonate, a metabolic precursor to methane in the upper ocean. Here we determine a 2.35-Å resolution structure of MPnS and discover that it has an unusual 2-His-1-Gln iron-coordinating triad. We further solve the structure of a related enzyme, hydroxyethylphosphonate dioxygenase from Streptomyces albus (SaHEPD), and find that it displays the same motif. SaHEPD can be converted into an MPnS by mutation of Gln-adjacent residues, identifying the molecular requirements for methylphosphonate synthesis. Using these sequence markers, we find numerous putative MPnSs in marine microbiomes and confirm that MPnS is present in the abundant Pelagibacter ubique. Ubiquity of MPnS-containing microbes supports the proposal that methylphosphonate is a source of methane in the upper, aerobic ocean, where phosphorus-starved microbes catabolize methylphosphonate for its phosphorus.

Contributions: Emily Ulrich and Spencer Peck purified and characterized enzymes and all enzyme variants. Kou-San Ju performed phylogenetic analysis of MPnS genes. Wilfred van der Donk helped interpret and discuss results.

This research was originally published in Science.

**Main text:**

Methane concentrations in oxygenated surface ocean waters are super-saturated relative to the atmosphere (1). This methane has been shown to derive from a biological source (2), however, canonical methanogenesis only occurs in obligate anaerobic archaea (3). Although anoxic microenvironments could provide habitats for methanogenic archaea (4), *in situ* methanogenesis has not been directly observed. Thus, the source of biological methane from the aerobic upper ocean is unknown, a conundrum referred to as the “oceanic methane paradox” (5). Recently, the C-P lyase pathway was identified as a biogenic source of methane in aerobic organisms prevalent in surface ocean waters (6, 7). Under phosphorus stress, these organisms use the C-P lyase pathway to cleave the C-P bond in methylphosphonate (MPn), releasing methane and sequestering the phosphorus in the form of phosphate (6, 7). The release of methane through metabolism of MPn has been proposed as a resolution to the oceanic methane paradox (8). Indeed, MPn was recently shown to be ubiquitous in dissolved organic matter from the ocean (9), and an MPn biosynthetic enzyme was recently identified but thus far only shown to be present in one marine microbe, the abundant archaeon *Nitrosopumilus maritimus* (10). The key enzyme discovered, MPn synthase (MPnS), uses molecular oxygen and a non-heme mononuclear Fe(II) center to convert 2-hydroxyethylphosphonate (2-HEP) into MPn and CO$_2$ (10) (Figure V.1). MPnS is similar in sequence to hydroxyethylphosphonate dioxygenase (HEPD) (11) and hydroxypropylphosphonate epoxidase (HppE) (10, 12), which catalyze the formation of hydroxymethylphosphonate (HMP) from 2-HEP (Figure V.1) (11) and the formation of the antibiotic fosfomycin from 2-S-hydroxypropylphosphonate (13), respectively.
Figure V.1: The reactions of MPnS and HEPD have been proposed to proceed through a common radical intermediate (14). The hydrogen derived from the C2 pro-R position in 2-HEP is highlighted red throughout. The radical recombination in MPnS or HEPD is colored red or blue, respectively. For an alternative mechanism involving a cationic intermediate, see Figure V.11.

To establish the molecular determinants of MPnS activity and thereby establish sequence markers to employ in the search for additional MPnSs, we determined X-ray structures of MPnS from *N. maritimus* in both the substrate-free and the Fe(II)- and substrate-bound states to 2.37 Å and 2.35 Å resolution, respectively (Table V.1). As expected based on sequence (Figure V.2), the overall structure of MPnS is highly similar to the HEPD from *Streptomyces viridochromogenes* (*SvHEPD*) (Figure V.3) (11). Both enzymes are dimeric and contain the same domain structure: two β-sheet domains (β1 and β2) which together comprise a bicupin fold and two entirely α-helical domains (α1 and α2) (Figure V.3A). The active site is situated in the first of the two cupin folds and completed by contributions from the N-terminal α-helical domain and the C-terminal tail of the second protomer. As noted previously (11), the dimeric architectures of *SvHEPD*, and now MPnS, are analogous to the tetramer structure of HppE from *Streptomyces wedmorensis*
(12) with each protomer of HppE corresponding to a single $\alpha$-helical domain and a single $\beta$ cupin fold.
Table V.1. Crystallographic data collection and refinement statistics.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Apo-SohEHPD*</th>
<th>Datasets used in phasing</th>
<th>Cd(II)-SohEHPD*</th>
<th>Cd(II)-SohEHPD anomalous*</th>
<th>2-HEP-bound Cd(II)-SohEHPD</th>
<th>Substrate-free Fe(II)-SohEHPD</th>
<th>2-HEP-bound Fe(II)-MPnS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P2₁</td>
<td>P2₁</td>
<td>P2₁</td>
<td>P2₁</td>
<td>P₂₁</td>
<td>C₂</td>
<td>P₂₁</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td>β (°)</td>
<td>89.7, 101.4, 99.6</td>
<td>89.6, 101.2, 99.1</td>
<td>89.4, 101.1, 99.1</td>
<td>89.2, 100.7, 99.1</td>
<td>352.4, 53.2, 260.1</td>
<td>70.9, 351.3, 76.9</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9791</td>
<td>0.9791</td>
<td>1.7401</td>
<td>0.9791</td>
<td>0.9791</td>
<td>0.9791</td>
<td>0.9791</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>100 – 171</td>
<td>100 – 162</td>
<td>100 – 191</td>
<td>100 – 180</td>
<td>100 – 237</td>
<td>100 – 235</td>
<td></td>
</tr>
<tr>
<td>Rmeas (%)</td>
<td>11.2 (97.0)</td>
<td>9.3 (95.9)</td>
<td>11.6 (58.7)</td>
<td>12.1 (62.1)</td>
<td>15.7 (93.4)</td>
<td>10.4 (97.4)</td>
<td></td>
</tr>
<tr>
<td>Rfree (%)</td>
<td>3.1 (28.3)</td>
<td>4.5 (49.2)</td>
<td>3.9 (26.3)</td>
<td>5.6 (34.0)</td>
<td>6.1 (47.4)</td>
<td>4.0 (37.9)</td>
<td></td>
</tr>
<tr>
<td>CC1/2</td>
<td>99.9 (84.0)</td>
<td>99.3 (49.0)</td>
<td>99.8 (77.0)</td>
<td>99.3 (74.0)</td>
<td>99.4 (42.0)</td>
<td>99.9 (75.0)</td>
<td></td>
</tr>
<tr>
<td>No. of atoms</td>
<td>23.8 (3.3)</td>
<td>14.4 (1.4)</td>
<td>15.8 (3.2)</td>
<td>11.3 (2.1)</td>
<td>11.9 (1.2)</td>
<td>15.2 (2.0)</td>
<td></td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97.4 (95.4)</td>
<td>93.7 (76.7)</td>
<td>98.6 (96.1)</td>
<td>97.7 (92.7)</td>
<td>96.7 (90.7)</td>
<td>89.8 (99.3)</td>
<td></td>
</tr>
<tr>
<td>Redundancy</td>
<td>12.6 (11.3)</td>
<td>3.9 (3.1)</td>
<td>7.7 (4.5)</td>
<td>4.1 (2.7)</td>
<td>6.2 (3.3)</td>
<td>6.9 (7.0)</td>
<td></td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>186,162 (9,054)</td>
<td>206,521 (8,385)</td>
<td>264,384 (12,938)</td>
<td>156,962 (7,478)</td>
<td>160,320 (7,560)</td>
<td>135,649 (11,087)</td>
<td></td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>1.8</td>
<td>2.37</td>
<td>2.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of reflections used</td>
<td>156,919</td>
<td>160,314</td>
<td>135,624</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-factors (Å²)</td>
<td>Average</td>
<td>25.6</td>
<td>49.0</td>
<td>46.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>protein</td>
<td>23.7</td>
<td>49.1</td>
<td>46.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>water</td>
<td>37.5</td>
<td>47.2</td>
<td>41.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-HEP</td>
<td>32.2</td>
<td>-</td>
<td>44.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fe</td>
<td>37.8</td>
<td>61.8</td>
<td>43.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>glycerol</td>
<td>40.9</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>formate</td>
<td>-</td>
<td>-</td>
<td>37.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rwork/Rfree (%)</td>
<td>15.3/18.9</td>
<td>20.3/24.1</td>
<td>19.1/23.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. atoms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>protein</td>
<td>13601</td>
<td>27049</td>
<td>28307</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>water</td>
<td>1951</td>
<td>1351</td>
<td>987</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-HEP</td>
<td>28</td>
<td>-</td>
<td>56</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fe</td>
<td>4</td>
<td>5</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>glycerol</td>
<td>60</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>formate</td>
<td>-</td>
<td>-</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R.m.s. deviations</td>
<td>Bond lengths (Å)</td>
<td>0.006</td>
<td>0.003</td>
<td>0.004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>0.84</td>
<td>0.59</td>
<td>0.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rotamer outliers</td>
<td>6</td>
<td>29</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ramachandran angles</td>
<td>Outlier (%)</td>
<td>0</td>
<td>0.1</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allowed (%)</td>
<td>0.9</td>
<td>1.9</td>
<td>1.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favored (%)</td>
<td>99.1</td>
<td>98.0</td>
<td>98.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Structure was not refined to completion.
† Bijvoet pairs were not merged during data processing.
‡ Values in parentheses indicate highest resolution bin.
§ Reflections used to calculate Rfree comprise 5% of observed reflections.

\[
R_{\text{meas}} = \frac{\sum_{hkl} \sqrt{n} \Sigma_{i=1}^{n} |I_i(hklt) - \langle I(hklt)\rangle|}{\sum_{hkl} \Sigma_{i=1}^{n} I_i(hklt)}
\]

\[
R_{\text{pim}} = \frac{\sum_{hkl} \frac{1}{n} \Sigma_{i=1}^{n} |I_i(hklt) - \langle I(hklt)\rangle|}{\sum_{hkl} \Sigma_{i=1}^{n} I_i(hklt)}
\]

\[
CC1/2 = \frac{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2 \sum_{i=1}^{n} (Y_i - \bar{Y})^2}}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2 \sum_{i=1}^{n} (Y_i - \bar{Y})^2} \sqrt{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}}, \text{ the Pearson’s correlation coefficient between two half-data sets}
\]

\[
X \text{ and } Y.
\]

\[
R_{\text{work/free}} = \frac{\sum_{hkl} [F_{\text{obs}}(hklt) - F_{\text{calc}}(hklt)]}{\sum_{hkl} F_{\text{obs}}(hklt)}
\]
Figure V.2: Sequence alignment of the N-terminal region that contains a cupin fold and the active site residues of MPnS-related proteins. Iron-coordinating residues are highlighted red. Substrate-binding residues are highlighted blue. For enzymes with structurally similar active sites, the Gly-Tyr or Ile-Phe pairs that determine reaction pathway are highlighted in yellow (see Text for more information). SwHppE and SvHEPD have no yellow highlighting since they have different active site geometry. SwHppE, HppE from *Streptomyces wedmorensis*; SvHEPD, HEPD from *Streptomyces viridochromogenes*; SaHEPD, HEPD from *Streptomyces albus*; NmMPnS, MPnS from *Nitrosopumilus maritimus*; PuMPnS, MPnS from *Pelagibacter ubique* strain HTCC7217.
**Figure V.3:** Overall structure of *NmMPnS* and similar enzymes. (A) An *NmMPnS* monomer consists of the following domains: α1, α2, β1, and β2. Both the N- and C-termini of *NmMPnS* are near the α1-domain. The protein residues ligating iron (H148, Q152, and H190) and 2-HEP are shown in ball-and-stick and the iron is shown as an orange sphere. (B) Structure of the active sites of *NmMPnS* (cyan), *SvHEPD* (yellow; PDB 3GBF), and HppE (green; PDB 3SCG). Structures are aligned by the two conserved His residues for each protein. The iron-coordinating residues are shown in ball-and-stick representation, and part of the cupin barrel is shown in ribbon representation. The MPnS Fe(II) is depicted as an orange sphere. (C) *NmMPnS* (cyan) superimposed onto *SvHEPD* (yellow) (r.m.s.d. of 3.4 Å over 384 out of 443 Cα atoms as calculated with PyMOL). (D) *NmMPnS* (cyan) superimposed onto *SaHEPD* (yellow). (E) The *NmMPnS* dimer (cyan and pink) shows an extensive interface formed by the two α-helical domains and linker regions from each protomer. The average dimerization interface is 7670 Å² as calculated using PISA (15)). (F) An HEPD dimer from *S. viridochromogenes* (green and yellow (11)).
Figure V.3: (continued).
Mononuclear non-heme iron enzymes are traditionally characterized by a three-residue iron-binding motif composed of two His residues and one carboxylate-bearing residue (Asp or Glu), termed the facial triad (16). In the active site of MPnS, a Gln residue (Q152) coordinates the Fe(II) along with two His residues (H148 and H190), forming an unusual 2-His-1-Gln facial triad (Figure V.4A, Figure V.5A). The position of Gln152 in MPnS is distinct from that of the iron coordinating Glu (E176) in SvHEPD (11), originating from an adjacent β-strand at the analogous position of the iron-coordinating Glu in HppE (12) (Figure V.3B). The substrate 2-HEP coordinates Fe(II) in the active site of MPnS through its hydroxyl oxygen and one phosphonate oxygen (Figure V.4A, Figure V.5A, Figure V.6A). The phosphonate moiety is further stabilized by interaction with Arg102, Tyr110, Asn145 from the cupin barrel, Lys28′ from the α-helical domain of neighboring protomer, and Trp449′ from the C-terminal tail of the neighboring protomer. Interestingly, the Trp449-containing C-terminal tail in MPnS is mobile; in the crystal structure of substrate-free MPnS, residues Ala442-Ser457 have an alternative position where Trp449 is no longer in the active site and a channel is open from the active site to bulk solvent (Figure V.7A,B). Thus, movement of the C-terminal tail and Trp449 appears to open and close the active site in MPnS. Although Trp449 is conserved between MPnS and SvHEPD, the residue has not been visualized in the SvHEPD structures (11). Three other residues of the cupin fold in MPnS (Y108, I126, F192) further contribute to substrate binding through hydrophobic interactions (Figure V.4A, Figure V.5A).
Figure V.4: Active site structure of NmMPnS and SaHEPD. (A) The NmMPnS active site is composed of residues from the $\beta$1-domain (cyan) and the $\alpha$1'-domain (pink) with two key residues discussed in the text in light brown. Fe(II) (orange sphere) is ligated by substrate, Gln152, His148, and His190. Although electron density cannot be used to distinguish oxygen from nitrogen, we are showing the oxygen of the Gln side chain coordinating Fe(II). (B) SaHEPD active site composed of residues from the $\beta$1-domain (yellow) and the $\alpha$1'-domain (green) for two protomers of the asymmetric unit in which Gln153 is in the iron-binding conformation. (C) Alternative conformation of SaHEPD where Gln153 is displaced from the Fe(II).
Figure V.5: Stereo view of *NmMPnS* and *SaHEPD* active sites. (A) The *NmMPnS* active site is composed of residues from the β1-domain (cyan) and the α 1’-domain (pink) is shown with key residues in ball-and-stick and with iron shown as an orange sphere. Residue positions distinguishing the two enzymes are shown in light brown. (B) *SaHEPD* active site composed of residues from the β1-domain (yellow) and the α1’-domain (green) for two protomers of the asymmetric unit in which Gln153 is in the iron-binding conformation. (C) Alternative conformation of *SaHEPD* where Gln153 is displaced from the Fe(II). For clarity, additional waters within the active site are omitted.
Figure V.5. (continued).
Figure V.6: Electron density within the active sites of *Nm*MPnS and *Sa*HEPD. (A) Substrate 2-HEP bound in *Nm*MPnS active site with difference omit electron density contoured at 2.5 σ, shown in blue mesh, surrounding the substrate. (B) Substrate 2-HEP bound in the *Sa*HEPD active site with omit map as described in (A). (C) Composite omit electron density map contoured at 1.3 σ (blue mesh) surrounding the iron-coordinating ligand Gln152 in *Nm*MPnS. (D) Composite omit electron density (contoured at 1.5 σ and in blue mesh) surrounding Gln153 and a water molecule in the active site of *Sa*HEPD in which Gln153 is acting as a ligand to Fe(II). (E) Composite omit electron density (contoured at 1.5 σ and in blue mesh) surrounding Gln153 and water molecules within another active site of *Sa*HEPD in which Gln153 is displaced as a ligand to Fe(II). Instead, a water molecule is serving as a ligand to Fe(II).
Figure V.6. (continued).
**Figure V.7:** The MPnS C-terminus gates access to the active site. Ribbon representation of the substrate-free MPnS active site in the (A) closed and (B) open conformations. Red dotted lines depict the solvent channel to the active site that opens following rearrangement of the C-terminus. The 2-His-1-Gln triad of the MPnS active site and Trp449 is shown in ball-and-stick with $2F_o-F_c$ electron density contoured at 1.0 $\sigma$ in blue mesh. The two protomers in the dimer are shown in cyan and violet and residues 443-450 are highlighted in red.
To determine if the presence of a 2-His-1-Gln motif is predictive of MPnS activity, we selected to clone, express and characterize a gene product from *Streptomyces albus* (strain NRRL B-16041) that displays a 2-His-1-Gln motif and 34% sequence identity to MPnS. Using a previously described phosphorus NMR spectroscopy assay (11), this enzyme was shown to produce only HMP when 2-HEP is provided as substrate (Figure V.8), classifying this enzyme as an HEPD. HEPDs containing a 2-His-1-Gln facial triad will hereafter be referred to as Class II whereas the previously characterized 2-His-1-Glu HEPDs will be notated Class I. Identification of Class II HEPDs that share the same iron-ligating ligand set as MPnS demonstrates that the 2-His-1-Gln motif is not diagnostic of enzyme function, raising the question, what sequence motifs distinguish an MPnS from a Class II HEPD?
**Figure V.8:** $^{31}$P NMR spectroscopic analysis of the reactions with wild-type and double mutant enzyme variants and 2-HEP as substrate. Peak assignments were confirmed by spiking with authentic material. The chemical shifts of phosphonates in $^{31}$P NMR are very sensitive to pH near their pK$_a$ values, accounting for the small changes in chemical shifts between experiments.
To investigate the structural differences between MPnS and a Class II HEPD, we determined a crystal structure of HEPD from *S. albus* (*SaHEPD*) to 1.8 Å resolution with 2-HEP and Fe(II) bound (Table V.1). *SaHEPD* has the same fold and oligomeric state as MPnS, and is highly similar (r.m.s.d. of 1.9 Å over 427 out of 450 C\textalpha{} atoms as calculated with PyMOL (17)) (Figure V.2D). Consistent with the sequence alignment (Figure V.2), all three iron-coordinating residues (H149, Q153, and H190 in *SaHEPD*) and all residues forming the substrate-binding pocket (R103, Y109, Y111, I127, N146, F192, K28’ and W449’ in *SaHEPD*) are conserved between *SaHEPD* and MPnS (Figure V.4A,B, Figure V.5A,B, Figure V.6B). The crystal lattice affords four independent views of this active site, and in two of the four active sites in the asymmetric unit, the electron density for Gln153 is consistent with an iron-coordinating conformation. However, in the other two active sites, the electron density indicates that a water molecule has replaced Gln153 as the ligand to Fe(II) and that Gln153 is now pointing away from the iron (Figure V.6D,E). This Gln is positioned such that it can hydrogen bond to Tyr163 that packs against Fe(II) ligand His190 (Figure V.4C, Figure V.6E). Interestingly, Tyr163 is one of two residues, Gly184 is the other, that is close to the ligand triad and not identical between *SaHEPD* (Y163/G184) and MPnS (F162/I184). In MPnS, Phe162 and Ile184 are in van der Waals contact with each other and fill in the area under Gln152 such that movement of Gln away from the Fe(II) does not appear possible (Figure V.4A). In contrast, the two residues (Y163 and G184) in *SaHEPD* form a cavity that can accommodate a water molecule when Gln153 is a ligand to Fe(II) (Figure V.4B) and can also accommodate an alternative (non-coordinating) conformation of Gln153 (Figure V.4C).
To investigate the importance of the alternative conformation of Gln153 in SaHEPD, we mutated both Tyr163 and Gly184 in SaHEPD to the corresponding residues in MPnS, Phe and Ile, respectively. Our structures suggest that in this double mutant the solvent cavity beneath Gln153 would be filled, restricting Gln153 to an iron-binding configuration as observed in MPnS. Indeed, the purified SaHEPD-Y163F/G184I variant catalyzes the conversion of 2-HEP to MPn with no detectable side products and with kinetic parameters ($k_{cat}: 1.2 \pm 0.1 \, \text{s}^{-1}$, $K_{m,HEP} 15 \pm 2 \, \mu \text{M}$) very similar to those of purified wild-type SaHEPD (Figure V.8, Figure V.9, Table V.2). The catalytic efficiency is in fact slightly better than for NmMPnS ($k_{cat}: 0.25 \, \text{s}^{-1}$; $K_{m,HEP} 11 \, \mu \text{M}$) (18). These results demonstrate that switching the identity of only two non-coordinating active site residues is sufficient to transform an HEPD into an MPnS. We then investigated which of these two residues is most important; SaHEPD-Y163F produced only HMP, whereas SaHEPD-G184I only produced MPn (Figure V.10). Thus, disruption of hydrogen bonding between Tyr163 and Gln153 is insufficient to alter the reaction outcome, but preventing Gln153 from occupying the alternative conformation does result in the switch in product formation. We also replaced the unusual Gln iron-coordinating ligand with Glu or Ala in both SaHEPD and NmMPnS. SaHEPD-Q153E was inactive, whereas the Q153A mutant still produced HMP, albeit with reduced efficiency (Figure V.10). NmMPnS-Q152A produced HMP with a very low level of activity, whereas NmMPnS-Q152E was inactive under the same condition (Figure V.10). Thus, removal of the Gln ligand is strongly deleterious for both MPnS and Class II HEPD enzymes (for discussion, see Figure V.10). Replacing the Glu ligand in Class I HEPD with Gln had no effect on activity (14).
Table V.2. Kinetic Parameters.

<table>
<thead>
<tr>
<th></th>
<th>SaHEPD</th>
<th>SaHEPD Y163F/G184I</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_m$, 2-HEP apparent (µM)</td>
<td>$12 \pm 2$</td>
<td>$15 \pm 2$</td>
</tr>
<tr>
<td>$K_m$, O₂ (µM)</td>
<td>$61 \pm 6$</td>
<td>$95 \pm 15$</td>
</tr>
<tr>
<td>$k_{cat}$ (s⁻¹)</td>
<td>$1.1 \pm 0.1$</td>
<td>$1.2 \pm 0.1$</td>
</tr>
<tr>
<td>$k_{cat}/K_m$, 2-HEP (M⁻¹s⁻¹)</td>
<td>$(9.2 \pm 1.7) \times 10^4$</td>
<td>$(8.0 \pm 1.3) \times 10^4$</td>
</tr>
<tr>
<td>$k_{cat}/K_m$, O₂ (M⁻¹s⁻¹)</td>
<td>$(1.8 \pm 0.2) \times 10^4$</td>
<td>$(1.3 \pm 0.2) \times 10^4$</td>
</tr>
</tbody>
</table>
Figure V.9: Steady-state Michaelis-Menten kinetics of SaHEPD (red trace) and SaHEPD Y163F/G184I (blue trace) with respect to 2-HEP (top) and oxygen (bottom). Error bars represent the standard deviation calculated from triplicate reactions. For kinetic parameters, see Table V.2.
**Figure V.10:** $^{31}$P NMR spectroscopic analysis of the reactions with single mutant enzyme variants and 2-HEP as substrate. Peak assignments were confirmed by spiking with authentic material. The chemical shifts of phosphonates in $^{31}$P NMR are very sensitive to pH near their pK$_a$ values, accounting for the small changes in chemical shifts between experiments. All experiments were performed with 10 µM enzyme, 2 mM 2-HEP, for 2 h. *SaHEPD-G184I* produces MPn whereas *SaHEPD-Y163F* produces HMP. Substitution of Ile for Gly in *SaHEPD-G184I*, which places a bulky residue directly below iron-ligand Gln153, converts the enzyme activity to that of an MPnS. This result indicates the necessity of an open cavity below Gln153 for HEPD activity and the importance of a filled cavity for MPnS activity. In contrast, a hydrogen bond between Gln153 and Tyr does not appear important for HEPD activity given that *SaHEPD-Y163F* is active as a HEPD. Replacing Gln153 in *SaHEPD* with Ala results in a less active HEPD enzyme, but does not convert enzyme activity to that of an MPnS. This result is consistent with the proposal that MPnS activity requires Gln ligation of the iron, presumably to position the formate intermediate appropriately. Replacing Gln153 with Glu in *SaHEPD* resulted in an inactive protein. Like *SaHEPD*, *NmMPnS-Q152E* was inactive and *NmMPnS-Q152A* produced a small amount of HMP (and thus was converted to an HEPD). The finding that *NmMPnS-Q152A* has HEPD activity is consistent with the idea that iron-ligation is required for MPn production. Although we can rationalize the conversion of enzyme activity between HEPD and MPnS in these Gln-substituted mutant proteins, we note that the decreases observed in overall enzyme activity could be due to a number of different factors, including mutation-induced deviations in oxygen binding to iron and/or activation and/or substrate binding among other factors.
Figure V.10: (continued)
We next considered why restricting Gln movement would alter product formation. An iron-bound MPn-based radical and an iron-bound formate are the proposed reaction intermediates for both enzymes (Figure V.1, Figure V.11) (19); in MPnS, the MPn-based radical abstracts the formate hydrogen atom to yield MPn and CO₂, and in HEPD the MPn-based radical reacts with the iron-bound hydroxyl yielding HMP and HCO₂⁻ (14, 18). Therefore, the fate of the intermediate appears to depend directly on the geometric arrangement of formate relative to the MPn-based radical. MPnS requires the formate hydrogen atom to be in close proximity to the radical (Figure V.1). Inspection of the substrate-bound MPnS active site reveals that the positioning of an iron-bound formate intermediate would be most directly affected by the side chain of Gln152 (Figure V.4A, Figure V.5A). With the side chain oxygen of the Gln coordinating the iron, the side chain amide nitrogen would be positioned appropriately to interact with the iron-bound formate, securing its position near the iron-bound MPn-radical (Figure V.12A). Thus, residues that fix the position of Gln, like a juxtaposed Ile184, favor MPn production. In contrast, residues that permit Gln to adopt an alternative (non iron-ligating) position, like a juxtaposed Gly184, should support HMP production; Gln movement allows formate movement, potentially increasing the distance between the formyl hydrogen atom and the iron-bound MPn radical, disfavoring hydrogen-atom abstraction (Figure V.12B). The same arguments can be made if the reaction were to proceed via an ionic mechanism (Figure V.11), except that formate would be appropriately positioned by the Gln for hydride rather than hydrogen atom transfer in MPnS. This model also explains the outcome of all mutant enzymes.
Figure V.11: Proposed reaction mechanism for Class II HEPD and MPnS. Both enzymes are believed to follow the same mechanism to generate a radical intermediate that can either abstract a hydrogen atom from formate to generate MPn (MPnS, red arrows) or combine with a ferric hydroxide to form HMP (HEPD, blue arrows). Alternatively (pathway b), the radical might be oxidized to a cation that can either be trapped by hydride transfer from formate (MPnS; red arrows) or hydroxide (HEPD; blue arrows) (20). Regardless of whether the last step occurs by hydrogen atom transfer or hydride transfer, both pathways require formate to be appropriately positioned for MPnS catalysis. Previous studies have provided strong evidence for the proposal that the two enzymes only differ in the last step and not at earlier steps (14).
Figure V.12: Proposed intermediates for MPnS and SaHEPD. (A) Gln152 in MPnS coordinates the formate intermediate, enabling H-atom abstraction by the MPn radical. (B) Formate displaces Gln153 from its iron-coordinating position in SaHEPD, preventing H-atom abstraction by the MPn-based radical and promoting reaction with the iron-bound hydroxide.

Next, we searched genomes for additional MPnSs. Two homologs that contain Ile at position 184 were identified in the NCBI non-redundant protein sequences database. One of these enzymes, a hypothetical protein from Candidatus Pelagibacter ubique strain HTCC7217 (WP_029455269.1), was expressed as a maltose-binding protein fusion enzyme in E. coli. Indeed, this enzyme converts 2-HEP to MPn without detectable side products, consistent with it being an MPnS (Figure V.8). Altogether, these results establish a sequence signature that defines MPnS enzymes: a 2-His-1-Gln facial triad with an Ile at position 184 to pack against the Gln. Phylogenetic analysis suggests that Val at position 184 may also play the same role as Ile (Figure V.13).

Extending the MPnS sequence signature to a phylogenetic analysis of genomes, metagenomes, and transcriptome shotgun assemblies reveals additional candidate MPnS-coding
genes in uncultivated marine microbes and even eukaryotic marine dinoflagellates including
dinoflagellates of the genera *Symbiodinium* (Figure V.13, Figure V.14). In contrast to HEPD-
coding genes, which are present in microbial strains and metagenomes of both marine and
terrestrial origin, all candidate MPnS-coding genes identified to date derive from marine
environments, suggesting a relevant role for MPn in marine ecosystems. Importantly, *P. ubique*
belongs to the SAR11 clade of α-proteobacteria, and the SAR11 clade, along with
thaumarchaeota (e.g. *N. maritimus*), are two of the most abundant marine microorganisms (21,
22). Thus, the identification of MPnS enzymes in *P. ubique* and *N. maritimus* strongly supports a
widespread role of MPn in marine carbon and phosphorus cycling, further supporting the MPn
hypothesis for oceanic methane production and potentially resolving the “oceanic methane
paradox.”
Figure V.13: Phylogeny of MPnS, HEPD, and HppE-like proteins. HEPD proteins are colored based on the iron-ligating facial triad: Class I HEPD, 2-His-1-Glu (green); or Class II HEPD, 2-His-1-Gln (brown). MPnS proteins are colored based on the amino acid at position 184, either Ile (blue) or Val (magenta). Complete labeling of sequence names is provided in Figure V.14.
Figure V.14: Phylogeny of MPnS, HEPD, and HppE-like proteins structures as in Figure V.13 with complete labeling of sequence names.
Materials and Methods

Materials

Restriction enzymes (NdeI, XhoI, HindIII, DpnI), Phusion® HF polymerase, and the Gibson Assembly® kit were purchased from New England Biolabs (NEB). Fail Safe polymerase and buffers were purchased from Epicentre. Plasmids were purified using the QIAprep Spin Miniprep Kit (Qiagen).

LB medium components were purchased from Fisher Scientific. Autoinduction media components were purchased from Sigma Aldrich, Fisher Scientific, Alfa Aesar, and Acros Organics. Ampicillin was purchased from Fisher and GoldBio and used at a concentration of 100 µg/mL. Chloramphenicol was purchased from Sigma Aldrich and used at a concentration of 12.5 µg/mL. Isopropyl-β-D-thiogalactopyranoside (IPTG) was purchased from IBI Scientific. Protein purification buffer components HEPES, glycerol, and imidazole were purchased from Fisher Scientific, and KCl and maltose were purchased from Sigma Aldrich. Lysozyme from chicken egg white and DNase were purchased from Sigma Aldrich. Ni-NTA resin was purchased from Qiagen. Amylose resin was purchased from New England Biolabs. PD10 desalting columns were purchased from GE Healthcare. 2-Hydroxyethylphosphonate (2-HEP) was prepared as described previously (19). Iron ammonium sulfate, chelex 100 sodium, and methylphosphonate were purchased from Sigma Aldrich. D₂O was purchased from Cambridge Isotope Labs.
**Primers for cloning and site-directed mutagenesis**

All primers were purchased from Integrated DNA Technologies, Inc.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>pET15-XhoI Fw (amplification</td>
<td>CTCGAGGATCCGGCTGCTAAACAAAGCCCGAAAGG</td>
</tr>
<tr>
<td>of linear pET-15b vector)</td>
<td></td>
</tr>
<tr>
<td>pET15-NdeI Rv (amplification</td>
<td>CATATGGCTGCCGCGCGCAACCAGGCGCGCTG</td>
</tr>
<tr>
<td>of linear pET-15b vector)</td>
<td></td>
</tr>
<tr>
<td>pMAL-HindIII Fw (amplification</td>
<td>AAGCTTGACCTGCGCGTCGTTTTTCAACACGTCG</td>
</tr>
<tr>
<td>of linear pMAL-c2x vector)</td>
<td></td>
</tr>
<tr>
<td>pMAL-EcoRI Rv (amplification</td>
<td>GAAATTCTGAATAACCTTTTCGATCCGGAGG</td>
</tr>
<tr>
<td>of linear pMAL-c2x vector)</td>
<td></td>
</tr>
<tr>
<td><em>S. albus</em> (SaHEPD) Fw</td>
<td>GCAGCGGCCCTGTGCCGCCGCCGCCAGCATTAGACGACGCC</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><em>S. albus</em> (SaHEPD) Rv</td>
<td>CTTTCGGGGCTTTTGGTTAGCAGCAGCAGCTCAGGTACGATC</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><em>P. ubique</em> Fw</td>
<td>GCAGCGGCCTGTTGCCGCCGCCGCCATATGAGCACG</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><em>P. ubique</em> Rv</td>
<td>ATCAAAATAATACAAATTCAG</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><em>P. ubique</em>_pMAL Fw (amplification</td>
<td>CTTTCGGGGCTTTTGGTTAGCAGCAGCAGCTCAGGTACGATC</td>
</tr>
<tr>
<td>from pET-15b into pMAL-c2x)</td>
<td></td>
</tr>
<tr>
<td><em>P. ubique</em>_pMAL Rv (amplification</td>
<td>CTTTCGGGGCTTTTGGTTAGCAGCAGCAGCTCAGGTACGATC</td>
</tr>
<tr>
<td>from pET-15b into pMAL-c2x)</td>
<td></td>
</tr>
<tr>
<td><em>SaHEPD</em> Fw Y163F</td>
<td>CTTTCGGGGCTTTTGGTTAGCAGCAGCAGCTCAGGTACGATC</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><em>SaHEPD</em> Rv Y163F</td>
<td>CTTTCGGGGCTTTTGGTTAGCAGCAGCAGCTCAGGTACGATC</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><em>SaHEPD</em> Fw G184I</td>
<td>CTTTCGGGGCTTTTGGTTAGCAGCAGCAGCTCAGGTACGATC</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><em>SaHEPD</em> Rv G184I</td>
<td>CTTTCGGGGCTTTTGGTTAGCAGCAGCAGCTCAGGTACGATC</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><em>SaHEPD</em> Fw Q153E</td>
<td>CTTTCGGGGCTTTTGGTTAGCAGCAGCAGCTCAGGTACGATC</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><em>SaHEPD</em> Rv Q153E</td>
<td>CTTTCGGGGCTTTTGGTTAGCAGCAGCAGCTCAGGTACGATC</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><em>SaHEPD</em> Fw Q153A</td>
<td>CTTTCGGGGCTTTTGGTTAGCAGCAGCAGCTCAGGTACGATC</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><em>SaHEPD</em> Rv Q153A</td>
<td>CTTTCGGGGCTTTTGGTTAGCAGCAGCAGCTCAGGTACGATC</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><em>NmMPnS</em> Fw Q152A</td>
<td>CTTTCGGGGCTTTTGGTTAGCAGCAGCAGCTCAGGTACGATC</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><em>NmMPnS</em> Rv Q152A</td>
<td>CTTTCGGGGCTTTTGGTTAGCAGCAGCAGCTCAGGTACGATC</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><em>NmMPnS</em> Fw Q152E</td>
<td>CTTTCGGGGCTTTTGGTTAGCAGCAGCAGCTCAGGTACGATC</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><em>NmMPnS</em> Rv Q152E</td>
<td>CTTTCGGGGCTTTTGGTTAGCAGCAGCAGCTCAGGTACGATC</td>
</tr>
</tbody>
</table>
**Strains and Plasmids**

<table>
<thead>
<tr>
<th>Name</th>
<th>Features</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. coli</em> DH5α λpir</td>
<td>Sup E44, ΔlacU169 (φlacZΔM15), recA1, endA1, hsdR17, thi-1, gyrA96, relA1, λpir phage lysogen</td>
<td>(23)</td>
</tr>
<tr>
<td><em>E. coli</em> 5-α</td>
<td></td>
<td>NEB</td>
</tr>
<tr>
<td>Rosetta 2 (DE3) pLYsS</td>
<td></td>
<td>Novagen</td>
</tr>
<tr>
<td>pET-15b</td>
<td>AMP&lt;sup&gt;R&lt;/sup&gt;, encodes an N terminal 6xHis-tag® of the protein of interest, includes a thrombin recognition sequence</td>
<td>Novagen</td>
</tr>
<tr>
<td>pMAL-c2x</td>
<td>AMP&lt;sup&gt;R&lt;/sup&gt;, encodes an N terminal maltose binding protein (MBP) tag of the protein of interest, includes a factor Xa recognition sequence</td>
<td>NEB</td>
</tr>
<tr>
<td>pET15b-SaHEPD</td>
<td>AMP&lt;sup&gt;R&lt;/sup&gt;, NdeI and XhoI restriction sites</td>
<td>This study</td>
</tr>
<tr>
<td>pET15b-SaHEPD-YFGI</td>
<td>AMP&lt;sup&gt;R&lt;/sup&gt;, NdeI and XhoI restriction sites</td>
<td>This study</td>
</tr>
<tr>
<td>pET15b-NmMPnS</td>
<td>AMP&lt;sup&gt;R&lt;/sup&gt;, NdeI and BamHI restriction sites</td>
<td>(10)</td>
</tr>
<tr>
<td>pMALc2x-PuMPnS</td>
<td>AMP&lt;sup&gt;R&lt;/sup&gt;, EcoRI and HindIII restriction sites</td>
<td>This study</td>
</tr>
<tr>
<td>pET15b-NmMPnS-Q152A</td>
<td>AMP&lt;sup&gt;R&lt;/sup&gt;, NdeI and BamHI restriction sites</td>
<td>This study</td>
</tr>
<tr>
<td>pET15b-NmMPnS-Q152E</td>
<td>AMP&lt;sup&gt;R&lt;/sup&gt;, NdeI and BamHI restriction sites</td>
<td>This study</td>
</tr>
<tr>
<td>pET15b-SaHEPD-G184I</td>
<td>AMP&lt;sup&gt;R&lt;/sup&gt;, NdeI and XhoI restriction sites</td>
<td>This study</td>
</tr>
<tr>
<td>pET15b-SaHEPD-Y163F</td>
<td>AMP&lt;sup&gt;R&lt;/sup&gt;, NdeI and XhoI restriction sites</td>
<td>This study</td>
</tr>
<tr>
<td>pET15b-SaHEPD-Q153E</td>
<td>AMP&lt;sup&gt;R&lt;/sup&gt;, NdeI and XhoI restriction sites</td>
<td>This study</td>
</tr>
<tr>
<td>pET15b-SaHEPD-Q153A</td>
<td>AMP&lt;sup&gt;R&lt;/sup&gt;, NdeI and XhoI restriction sites</td>
<td>This study</td>
</tr>
</tbody>
</table>

**Molecular cloning and QuikChange mutagenesis**

Cloning of the pET15b-NmMPnS vector encoding for *Nitrosopumilus maritimus* MPnS was described previously (10).

pET15b-SaHEPD was prepared by amplification of the gene from *Streptomyces albus* subsp. albus strain NRRL B-16041 genomic DNA. The 50 µL reaction contained 1X Fail Safe buffer G, 1 µM of each primer, 250 ng gDNA, and 0.05 U/µL Phusion polymerase. The samples were run through the following program on a BioRad thermocycler: 98 °C 2 min, 35 cycles of 98
°C 15 s, 60 °C 15 s, 72 °C 1 min, then 72 °C 2 min. pET15b-PuMPnS (MPnS from *Pelagibacter ubique*) was prepared by amplification of the gene from *P. ubique* HTCC7217 genomic DNA. The 50 µL reaction contained 1X Fail Safe buffer F, 0.5 µM of each primer, 40 ng gDNA, and 0.05 U/µL Phusion polymerase. The following PCR program was run: 98 °C 2 min, 42 cycles of 98 °C 15 s, 50 °C 15 s, 68 °C 1.5 min, then 68 °C 2 min.

The PCR samples were analyzed on an 1% agarose gel and purified using the Wizard® SV kit (Promega) or the QiaQuick PCR purification kit (Qiagen), and the products (200 ng, except 100 ng was used for *P. ubique*) were used directly for Gibson assembly (24) with 100 ng linear pET-15b DNA (see below for preparation). The 20 µL ligation reactions contained 1X NEB Gibson master mix. The ligation reactions were incubated at 50 °C for 1 h (45 °C for *P. ubique*). An aliquot (~2 µL) was used to transform chemically competent *E. coli* DH5α cells (NEB). For *P. ubique*, a three-fold diluted aliquot (1 µL) was used to transform electrocompetent *E. coli* DH5α cells. The cells recovered in SOC media for 1 h at 37 °C before being plated on LB/amp100 plates and grown overnight at 37 °C. Colonies were picked for plasmid preparation, and the sequence was confirmed by sequencing at ACGT, Inc.

Linear pET-15b was prepared by digesting the circular plasmid with NdeI, and PCR amplifying the resulting product. The reaction (50 µL) contained 1X Fail Safe buffer G, 50 µM of each primer, 20-40 ng of linear plasmid, and 0.03 U/µL Fail Safe polymerase. The sample was run through the following program on a BioRad thermocycler: 98 °C 3 min, 30 cycles of 98 °C 10 s, 60 °C 1 min, 72 °C 5 min, then 72 °C 8 min. The product was treated with DpnI and purified with the Wizard® SV kit (Promega).

For amplifying the *P. ubique* gene from pET-15b-PuMPnS to make pMAL-c2x-PuMPnS, a 50 µL reaction was run containing 1X Fail Safe buffer A, 1 ng template (pET-15b-
PuMPnS), 0.5 µM of each primer, and 0.05 U/µL Phusion polymerase. The PCR program was the same as above for cloning from the gDNA except 30 cycles were performed. The PCR products were purified using the Qiagen PCR purification kit and used for Gibson assembly as described above (performed at 50 °C using linear pMAL-c2x prepared by digestion with HindIII). The ligation product (1 µL) was used to transform electrocompetent E. coli DH5α cells. After recovery in SOC media (37 °C for 1 h), the culture was plated on LB/amp100 plates, and colonies were picked for plasmid preparation. The correct sequence was confirmed by sequencing at ACGT, Inc.

Mutagenesis reactions were performed using the 10X reaction buffer and polymerase provided in the Quikchange II Site-Directed Mutagenesis Kit (Agilent). A typical reaction (50 µL) contained 1X Pfu buffer, 50 ng template plasmid DNA (pET15b- SaHEPD or pET15b-NmMPnS), 0.25 µM of each primer, 0.04 mM dNTPs (Novagen), 5% v/v DMSO, and 0.05 U/µL Pfu Ultra HF polymerase. Samples were run through the following program on a BioRad thermocycler: 95 ºC 1 min, 20 cycles of 95 ºC 1 min, 60 ºC or 50 ºC 1.5 min, 72 ºC 15 min, then 72 ºC 5 min Samples were then treated with DpnI and purified using the Qiagen PCR Purification Kit. The Quikchange product (6 µL) was used to transform chemically competent (NEB) or electrocompetent E. coli DH5α cells. After recovery in LB (37 ºC for 1 h), the culture was plated on LB/amp100 plates, and colonies were picked for plasmid preparation. A plasmid harboring the correct mutation was confirmed by sequencing either by ACGT, Inc. or the University of Illinois-Urbana Champaign Biotechnology Center. The entire process was repeated using plasmids harboring the single mutations as templates to make the double mutations (G184I followed by Y163F for SaHEPD).
**SaHEPD protein expression and purification for crystallography.**

Wild-type SaHEPD was expressed and purified following a similar protocol to that previously described for HEPD from *Streptomyces viridochromogenes* (11). The pET15b expression vectors were transformed into *E. coli* BL21 Rosetta(DE3) cells. Cells were grown in LB media supplemented with ampicillin (100 μg/mL) and chloramphenicol (25 μg/mL) at 37 °C until OD$_{600}$ reached 0.6, at which time the temperature was lowered to 18 °C and expression was induced by addition of 0.3 mM IPTG. After 14-22 h, cells were harvested by centrifugation (10,000 x g, 5 min, 4 °C). Cells were flash frozen in liquid N$_2$ and stored at −80 °C.

Cells from 1 L of LB were resuspended in 25 mL of buffer C (50 mM HEPES pH 7.5, 200 mM KCl, 10% (v/v) glycerol) supplemented with 20 mM imidazole, 0.4 U/mL DNase, and 1 mg/mL lysozyme at room temperature. Cells were lysed by two passages through a French pressure cell (15,000 psi) and the lysates were clarified via centrifugation (35,000 x g, 1 h, 4 °C). Clarified lysates were affinity purified using a 5 mL Ni-NTA column previously equilibrated with buffer C supplemented with 20 mM imidazole. Protein was eluted with buffer C supplemented with 250 mM imidazole. Elution fractions were combined and concentrated using a 30 kDa MWCO centrifugal filter and then buffer exchanged into buffer C with either a NAP-25 column or dialysis tubing following the manufacturer’s instructions. Protein was flash frozen in liquid N$_2$ and stored at −80 °C for future use.

**MPnS protein expression and purification.**

For the purification of MPnS from *N. maritimus*, typically, *E. coli* Rosetta 2 (DE3) pLysS cells (Novagen) harboring pET15b-NmMPnS grown in 1 L of LB media were resuspended in 30 mL of buffer A (50 mM HEPES pH 7.5, 300 mM NaCl, 10% (v/v) glycerol) supplemented with 20
mM imidazole, 0.3 U/mL DNase, 1 mM PMSF, 1 Roche Protease Inhibitor Cocktail tablet, and 1 mg/mL lysozyme at 4 °C. Cells were lysed by sonic disruption. The lysates were clarified via centrifugation (25,000 × g, 30 min, 4 °C). Clarified lysates were affinity purified using a 5 mL HisTrap column previously equilibrated with buffer A. Protein was eluted with a linear gradient of 20-250 mM imidazole in buffer A. Elution fractions were combined and concentrated using a 30 kDa molecular-weight cutoff (MWCO) centrifugal filter. MPnS was further purified on a SuperDex 200 pg 16/60 gel filtration column pre-equilibrated with buffer B (50 mM HEPES pH 7.5, 100 mM NaCl). The His6 affinity tag was removed by thrombin cleavage (15 µg/mL thrombin protease) for 5-10 h at 20 °C. Tag-cleaved MPnS was separated from the His6 affinity tag cleavage products using a 5 mL HisTrap column. Tag-cleaved MPnS was liberated from nonspecific binding to the Ni-NTA column with 50 mM HEPES pH 7.5 and 300 mM NaCl. MPnS was then concentrated and buffer exchanged into buffer B using a 30 kDa MWCO centrifugal filter to a concentration of ~20 mg/mL (as determined by UV/Vis absorbance at 280 nm using an extinction coefficient of 64,540 cm⁻¹M⁻¹), flash frozen in liquid N2, and stored at −80 °C for future use.

NmMPnS and SaHEPD eluted from gel filtration chromatography in single peaks with estimated molecular weights of approximately 100 kDa representing the dimeric form of each enzyme.

Obtaining crystals of apo- and Cd(II)-SaHEPD for use in phasing

Obtaining data quality crystals of SaHEPD required the use of both microseeding and metal soaks. Data quality crystals were not observed when SaHEPD was co-crystallized with Fe(II). Metal-free crystals were soaked with either Cd(II) or Fe(II) to obtain metal-bound structures.
Seed crystals of apo-SaHEPD were obtained by the sitting drop vapor diffusion technique at 18 °C. An aliquot (120 nL) of protein solution (6.8 mg/mL SaHEPD in buffer C) was added to 120 nL of precipitant solution (240 mM sodium malonate, 20% (w/v) PEG 3350) in a 96-well sitting drop tray using a PHOENIX liquid handling robot (Art Robbins Instruments). Seed crystals were transferred to a microcentrifuge tube, diluted 1:40 in precipitant solution, and vortexed to produce a microcrystal stock.

Optimized crystals of apo-SaHEPD were obtained by the hanging drop vapor diffusion technique with microseeding at 22 °C. On a glass coverslip, an aliquot (0.75 µL) of protein solution (6.8 mg/mL SaHEPD in buffer C) was mixed with 1 µL microcrystal stock. The coverslip was sealed with grease over a reservoir containing 500 µL reservoir solution (300 mM sodium malonate, 22% (w/v) PEG 3350). Crystals formed within 2 d and reached a maximum size after approximately 14 d.

Apo-SaHEPD crystals grown in malonate solution were transferred stepwise in three steps of increasing glycerol concentration into a cryogenic solution containing 300 mM sodium malonate, 22% (w/v) PEG 3350, and 20% (v/v) glycerol, soaked in that solution for 15 s, and then flash-frozen in liquid N₂.

To generate crystals with Cd(II) bound in the active site, apo-SaHEPD crystals were transferred to a 2 µL drop of reservoir solution supplemented with 10 mM CdCl₂. Crystals were soaked overnight over reservoir solution before being transferred stepwise in three steps of increasing glycerol concentration into a cryogenic solution containing 300 mM sodium malonate, 22% (w/v) PEG 3350, 10 mM CdCl₂, and 20% (v/v) glycerol, soaked in that solution for 15 s, and then flash-frozen in liquid N₂. Both apo and Cd(II)-bound SaHEPD crystals contain malonate bound in the active site from the crystallization solution.
**Obtaining crystals of 2-HEP-bound Fe(II)-SaHEPD**

Seed crystals of SaHEPD bound to substrate (2-HEP, synthesized as described previously (25)) were obtained similarly to those of apo-SaHEPD using hanging drop vapor diffusion technique at 20 °C with an altered protein solution (6.8 mg/mL SaHEPD in buffer C, 4 mM 2-HEP) and an altered precipitant solution (200 mM sodium malate, 24% (w/v) PEG 3350) in a 96-well sitting drop tray using a PHOENIX liquid handling robot (Art Robbins Instruments). A microcrystal seed stock was created using the seed crystals as described above.

Optimized crystals of SaHEPD bound to 2-HEP were obtained by the hanging drop vapor diffusion technique at 22 °C in an anaerobic environment (95% Ar, 5% H₂; Coy Laboratory Products, Inc.). On a glass coverslip, an aliquot (0.75 µL) of protein solution (6.8 mg/mL SaHEPD in buffer C supplemented with 4 mM 2-HEP) was mixed with 1 µL microcrystal stock. The coverslip was sealed with grease over a reservoir containing 500 µL reservoir solution (255 mM sodium malate, 24% (w/v) PEG 3350). Crystals formed within 2 d and reached a maximum size after approximately 14 d. These crystals were used for Fe(II) soaking experiments.

To soak Fe(II) into the crystals, 2-HEP-bound SaHEPD crystals were transferred to a reservoir solution supplemented with 2 mM Fe(NH₄)₂(SO₄)₂. Crystals were soaked for 30 s before being transferred stepwise in three steps of increasing glycerol concentration to a cryogenic solution containing 255 mM sodium malate, 24% (w/v) PEG 3350, 2 mM 2-HEP, 2 mM Fe(NH₄)₂(SO₄)₂, and 20% (v/v) glycerol, soaked in that solution for 15 s, and then flash-frozen in liquid N₂.
**Data collection for SaHEPD**

All SaHEPD data sets were collected at the Advanced Photon Source (Argonne, Illinois, USA) on beam line 24ID-C using a Pilatus 6M pixel detector at a temperature of 100 K. SaHEPD crystals belonged to space group $P2_1$.

Data were collected on crystals of Cd(II)-SaHEPD at a wavelength of 0.9791 Å (12663 eV) in two wedges of 180° and at a wavelength of 1.7401 Å (7125 eV) using the inverse beam method with 18 sets of paired 20° wedges to capture Cd anomalous signal. An isomorphous dataset on apo-SaHEPD was collected to facilitate single isomorphous replacement with anomalous scattering (SIRAS). Data for 2-HEP-bound Fe(II)-SaHEPD, were collected at 0.9791 Å (12662 eV) in a single wedge of 300°.

Data for Cd(II)-SaHEPD, apo-SaHEPD, and 2-HEP-bound Fe(II)-SaHEPD were integrated in HKL2000 and scaled in Scalepack (26) (Table S1).

**Structure determination of SaHEPD**

The phase problem for SaHEPD data was solved by single isomorphous replacement with anomalous scattering (SIRAS) using the apo-SaHEPD dataset and a Cd(II)-SaHEPD dataset collected at $\lambda=0.9791$ Å for isomorphous replacement and a Cd(II)-SaHEPD dataset collected at $\lambda=1.7401$ Å for Cd-anomalous scattering (Table S1). Sixteen initial cadmium positions were determined using the Cd(II)-SaHEPD SAD dataset in ShelxD (27). Experimental maps extending to 1.9 Å resolution and an initial polyalanine model of Cd(II)-SaHEPD were generated by SAD in ShelxE (27). The experimental maps and ShelxE-generated polyalanine model were used to build a polyalanine model of four protomers in the asymmetric unit (two dimers) to facilitate solvent flattening in SOLOMON (28). Maps were improved by further refinement of the 16
cadmium sites in SHARP/autoSHARP (29), followed by SIRAS phase calculation using both the apo-SaHEPD dataset and the $\lambda=1.7401$ Å Cd(II)-SaHEPD dataset (Table S1). Solvent flattening was performed in SHARP using SOLOMON with the previous polyalanine model to define solvent boundaries. The overall figure of merit (acentric) was calculated by SHARP to be 0.41 to 3.5 Å resolution and the 3.5 Å resolution cutoff was used in generating experimental electron density maps. Using these experimental electron density maps, an initial model of Cd(II)-SaHEPD was built by manual adjustment in Coot until rigid body refinement in Phenix yielded $R$-factors of 31.7% and 31.5% for the working $R$-factor and the free $R$-factor (5% of reflections), respectively. Subsequent rounds of model building and refinement were performed in Coot (30) and Phenix (31), respectively, using native Cd(II)-SaHEPD data extending to 1.62 Å resolution. Sidechains were added to residues with clear electron density before waters were placed automatically using Phenix. Subsequent iterative refinement included positional refinement, non-crystallographic symmetry restraints, and individual $B$-factor refinement until the $R$-factors were 17.5% and 20.5% for working $R$ and free $R$, respectively. This model was used only to build the model of apo-SaHEPD.

A model of Cd(II)-SaHEPD with Cd(II) and waters removed was used as an initial model for refinement of apo-SaHEPD. Initial refinement of apo-SaHEPD included positional refinement, non-crystallographic symmetry restraints, individual $B$-factor refinement, and simulated annealing to remove model bias. Apo-SaHEPD was refined by iterative rounds of model building and refinement in Coot and Phenix, respectively, until the working $R$-factor was 17.1% and the free $R$-factor was 19.1%. This model was used only to solve the structures of 2-HEP-bound Fe(II)-SaHEPD and MPnS by molecular replacement (see below) and was not refined to completion.
The structure of 2-HEP-bound Fe(II)-SαHEPD was determined to 1.8 Å resolution by molecular replacement in Phaser (32) with apo-SαHEPD as the search model. Following molecular replacement, ten cycles of simulated annealing were performed in Phenix to minimize model bias. Fe(II) and 2-HEP were added to each active site, and the Fe(II) was fixed using bond restraints in Phenix. The parameter file for 2-HEP was sourced from the CCP4 monomer library (33). Waters were placed automatically using Phenix. The model was refined by cycles of manual adjustment in Coot followed by positional refinement, individual B-factor refinement, and TLS parameterization with non-crystallographic symmetry restraints in Phenix (Table S1). Simulated annealing composite omit maps were generated in Phenix and used to validate the final 2-HEP-bound Fe(II)-SαHEPD model. The 2-HEP-bound Fe(II)-SαHEPD structure contains, of 450 residues, residues 9-450, residues 9-450, residues 5-450, and residues 9-450, for chains A-D, respectively.

**Crystallization of MPnS**

Crystals of substrate-free MPnS were obtained by the hanging drop vapor diffusion technique at 24 °C in an anaerobic environment (95% Ar, 5% H₂; Coy Laboratory Products, Inc.). On a glass coverslip, an aliquot (1 µL) of protein solution (20 mg/mL MPnS in buffer B) was mixed with 1 µL reservoir solution (100 mM tri-ammonium citrate, 100 mM citric acid pH 4.0, 100 mM NH₄Cl, 18% (w/v) PEG 3350, 2 mM Fe(NH₄)₂(SO₄)₂). The coverslip was sealed with grease over a reservoir containing 500 µL reservoir solution. Crystals formed within 7 d and reached a maximum size after approximately 14 d. Crystals were transferred stepwise in three steps of increasing glycerol concentration into a cryogenic solution containing reservoir solution.
supplemented with 20% (v/v) glycerol, soaked in that solution for 15 s, and then flash-frozen in liquid N₂.

Microseeding was used to obtain data quality crystals of 2-HEP-bound Fe(II)-MPnS. Seed crystals of 2-HEP-bound MPnS were obtained by the hanging drop vapor diffusion technique at 24 °C in an anaerobic environment (95% Ar, 5% H₂; Coy Laboratory Products, Inc.). An initial microcrystal stock of 2-HEP-bound MPnS was prepared by mixing an aliquot (1 µL) of protein solution (10 mg/mL MPnS in buffer B supplemented with 2.5 mM 2-HEP) with 1 µL reservoir solution (190-220 mM NH₄Cl, 20-22% (w/v) PEG 3350) and sealing with grease over 500 µL reservoir solution. Crystals were harvested after 2 wk, diluted 1:2500 in reservoir solution, and vortexed to produce a microcrystal stock. To prepare optimized crystals, 1 µL microcrystal stock was mixed with an aliquot (1 µL) of protein solution (10 mg/mL MPnS in buffer B supplemented with 4 mM 2-HEP) and sealed above 500 µL of reservoir solution (220 mM NH₄Cl, 21% (w/v) PEG 3350). Crystals formed within 14 d and reached a maximum size after approximately 30 d. Fe(II) in the form of FeCl₂ was soaked into 2-HEP-bound MPnS crystals as MPnS did not produce high-quality crystals when cocrystallized with Fe(II) salts and 2-HEP. In particular, MPnS crystals grown with 2-HEP were manually transferred to a soak solution (220 mM NH₄HCO₃, 21% (w/v) PEG 3350), soaked for 30 s, and then transferred to soak solution supplemented with 2 mM FeCl₂. Crystals were soaked for 30 s before being transferred stepwise in three steps of increasing glycerol concentration into a cryogenic solution containing soak solution supplemented with 2 mM FeCl₂, 2.5 mM 2-HEP, and 20% (v/v) glycerol, soaked in that solution for 15 s, and then flash-frozen in liquid nitrogen.
**Data collection for MPnS**

Data for the substrate-free MPnS structure were collected at the Advanced Photon Photon Source (Argonne, Illinois, USA) on beam line 24ID-C using a Pilatus 6M pixel detector at a temperature of 100 K and at 0.9791 Å (12662 eV) in a single wedge of 360°. Data for the 2-HEP-bound Fe(II)-MPnS structure were collected at Stanford Synchrotron Radiation Laboratory (Menlo Park, California, USA) on beam line BL9-2 using a Pilatus 6M pixel detector at a temperature of 100 K and at 0.9791 Å (12662 eV) in a single wedge of 360°. Crystals of substrate-free MPnS belong to space group C2, and crystals of 2-HEP-bound Fe(II)-MPnS belong to space group P2₁.

The substrate-free MPnS dataset was integrated in HKL2000 and scaled in Scalepack (26) (Table V.1). The 2-HEP-bound Fe(II)-MPnS dataset was processed in XDS and scaled in XSCALE (34) (Table V.1).

**Structure determination of MPnS**

The structure of substrate-free MPnS was solved by molecular replacement in Phaser (32) using data trimmed to 3 Å resolution. A search model based on the structure of apo-SαHEPD was generated using Sculptor (35), which removes sidechains and regions of divergent sequence. For successful molecular replacement, the MPnS search model was further divided into three domains: residues 13-75, residues 238-300, and residues 79-237 together with residues 301-420. These domains correspond to the α₁, α₂, and β₁/β₂ domains, respectively. Eight protomers were identified in the asymmetric unit corresponding to four dimers. Following molecular replacement, data were extended to 2.37 Å resolution for initial refinement in Phenix. Loop regions were modeled in Coot and sidechains were added to residues with clear electron density. Fe(II) was modeled in 5 protomers where electron density indicated metal occupancy in the
active site. Low metal-occupancy may be due to Fe(II) chelation by citrate in the crystallization condition. Waters were placed automatically using Phenix. Iterative rounds of model building were performed in Coot and positional refinement and individual B-factor refinement with non-crystallographic symmetry restraints was performed in Phenix with data extending to 2.37 Å resolution (Table V.1). The substrate-free MPnS structure contains, of 457 residues, residues 9-450, residues 10-450, residues 10-448, residues 11-451, residues 11-451, residues 20-441, residues 10-451, and residues 11-439, for chains A-H, respectively.

The structure of 2-HEP-bound Fe(II)-MPnS was solved by molecular replacement in Phaser using substrate-free MPnS as a starting model with data extending to 2.35 Å resolution. Fe(II) and 2-HEP were added manually in Coot, and waters were added automatically in Phenix. The parameter file for 2-HEP was sourced from the CCP4 monomer library (33). The model was refined by iterative rounds of model building in Coot and positional refinement and individual B-factor refinement with non-crystallographic symmetry restraints in Phenix (Table V.1). Simulated annealing composite omit maps were generated in Phenix and used to validate the final 2-HEP-bound Fe(II)-MPnS model. The 2-HEP-bound Fe(II)-MPnS structure contains, of 457 residues, residues 10-456, residues 10-455, residues 10-456, residues 4-456, residues 10-453, residues 10-456, residues 10-455, and residues 10-451, for chains A-H, respectively. Crystallographic software packages were compiled by SBGrid (36).

SaHEPD/NmMPnS (and variants) protein expression and purification for enzymology

*E. coli* Rosetta 2 (DE3) pLysS cells (Novagen) harboring the designated plasmid were grown in LB/amp100/cam12.5 media while shaking at 37 ºC until an optical density at 600 nm of 0.6-0.8 was reached. The culture flasks were then placed in ice water for 10 min before induction with
IPTG (0.1 mM final concentration). The cultures were incubated for 10-12 h while shaking at 18 °C. The cells were harvested by centrifugation, and the pellets were washed with 25 mM HEPES pH 7.5 before resuspension in lysis buffer (50 mM HEPES pH 7.5, 200 mM KCl, 20 mM imidazole, 10% (v/v) glycerol). The resuspended cells were treated with lysozyme (1 mg/mL) and DNase (1000 U/30 mL lysis buffer) and incubated at 4 °C for 30 min before lysis by two passages through a chilled French pressure cell (Thermo Electron Corporation). The cell debris was removed by centrifugation at 35,000 × g for 1 h at 4 °C. The supernatant was then incubated with equilibrated Ni-NTA resin (2.5 mL of resin per L of overexpression culture) for 30 min at 4 °C before the flow-through was collected. The resin was washed twice with 10 column volumes of wash buffer (50 mM HEPES pH 7.5, 200 mM KCl, 50 mM imidazole, 10% (v/v) glycerol). Protein was eluted with elution buffer (50 mM HEPES pH 7.5, 200 mM KCl, 250 mM imidazole, 10% (v/v) glycerol), and fractions were pooled based on absorbance at 280 nm measured by a NanoDrop spectrophotometer (Thermo Scientific). The protein was concentrated to ~2 mL using an Amicon Ultra centrifugal filter with a 30 kDa molecular weight cut off. The protein solution was desalted using a PD10 column equilibrated with storage buffer (50 mM HEPES pH 7.5, 200 mM KCl, 10% (v/v) glycerol) and the eluent was stored at -80 °C in aliquots. For kinetics, the protein was additionally purified by size exclusion chromatography (Superdex 200 preparative grade resin, GE Healthcare, on an AKTA FPLC). Purity was assessed by SDS-PAGE, and the concentration was measured by absorbance at 280 nm using a NanoDrop spectrophotometer and a theoretical extinction coefficient (M⁻¹ cm⁻¹) of 80,790 for SaHEPD WT, G184I, Q153E, and Q153A, 79,300 for SaHEPD Y163F/G184I and Y163F, and 64,540 for NmMPnS WT, Q152A and Q152E (calculated using the SIB ExPASy Bioinformatics Resources Portal, http://www.expasy.org (37)).
**PuMPnS expression and purification**

Attempts to obtain a purified His-tagged enzyme were hampered by low expression. Similarly, a codon-optimized gene expressed with a His-tag resulted in low expression. As a result, a maltose binding protein-fusion was chosen for expression, purification, and analysis of activity. *E. coli* Rosetta 2 (DE3) pLysS cells (Novagen) harboring the designated plasmid (pMAL-c2x-*PuMPnS*) were grown in ZYM-5052 autoinduction media (Studier2005)/amp100/cam12.5 (with tryptone substituted for N-Z amine) while shaking at 37 °C until an optical density at 600 nm of 0.6-0.8 was reached. The cultures were incubated for 14 h while shaking at 18 °C. The cells were harvested by centrifugation, and the pellets were washed with 25 mM HEPES pH 7.5 before resuspension in lysis/wash buffer (50 mM HEPES pH 7.5, 200 mM KCl, 10% (v/v) glycerol). The resuspended cells were treated with lysozyme (1 mg/mL) and DNase (1000 U/30 mL lysis buffer) and incubated at 4 °C for 30 min before lysis by two passages through a chilled French pressure cell (Thermo Electron Corporation). The cell debris was removed by centrifugation at 35,000 × g for 1 h at 4 °C. The supernatant was then incubated with equilibrated amylose resin (1.25 mL of resin per L of overexpression culture) for 30 min at 4 °C before the flow-through was collected. The resin was washed twice with 10 column volumes of lysis/wash buffer. Protein was eluted with elution buffer (50 mM HEPES pH 7.5, 200 mM KCl, 10 mM maltose, 10% (v/v) glycerol), and fractions were pooled based on absorbance at 280 nm measured by a NanoDrop spectrophotometer (Thermo Scientific). The protein was concentrated to ~2 mL using an Amicon Ultra centrifugal filter with a 30 kDa molecular weight cut off. The protein solution was desalted using a PD10 column equilibrated with storage buffer (50 mM HEPES pH 7.5, 200 mM KCl, 10% (v/v) glycerol). The protein was additionally purified by size exclusion chromatography and stored at -80 °C in aliquots. Purity was assessed by SDS-PAGE, and the concentration was
measured by absorbance at 280 nm using a NanoDrop spectrophotometer and a theoretical extinction coefficient (M⁻¹cm⁻¹) of 121,130.

**Activity assay by ³¹P NMR spectroscopy**

NMR experiments were performed on an Agilent 600 MHz spectrometer equipped with an OneNMR probe. Enzymes (100 µM final concentration) were reconstituted in an anaerobic chamber (Coy Laboratory Products) with an atmosphere of 97%/3% N₂/H₂ with 1 molar equivalent of Fe(II)(NH₄)₂(SO₄)₂ in 50 mM HEPES pH 7.5 with incubation for 10 min on ice. A typical assay (500 µL) contained 10 µM enzyme and 2 mM 2-HEP in 50 mM HEPES pH 7.5, whereas 20 µM of PuMPnS was used to form additional product. After incubation for 2 h at room temperature, the reaction was treated with Chelex (~100 µL added) for 20 min while shaking, and the protein was removed using a Millipore centrifuge filter with a 30 kDa molecular weight cut off. D₂O (100 µL) was added (20% final v/v) before analysis by ³¹P NMR spectroscopy. Spectra were analyzed using MestReNova software version 8.0.0.

**Kinetic characterization of SaHEPD**

Kinetic characterization was performed on a Clark-type O₂ electrode (Hansatech Instruments). SaHEPD (WT or Y163F/G184I) was anaerobically reconstituted with varying amounts of Fe(II)(NH₄)₂(SO₄)₂ as described above. The amount of Fe(II) for optimal activity (measured by rate of oxygen consumption in air-saturated 50 mM HEPES pH 7.5 at 20 ºC with 1 µM enzyme and 250 µM 2-HEP) was determined to be 6 molar equivalents. This value was used to prepare reconstituted enzyme for determination of steady-state Michaelis-Menten parameters. For determination of $K_{m \text{ apparent}, 2-\text{HEP}}$, the rate of oxygen consumption was measured in air-saturated
50 mM HEPES pH 7.5 at 20 °C (280 µM O₂) with 1 µM enzyme and varying amounts of 2-HEP (10-1000 µM) in a final volume of 1 mL. For determination of \( K_{m,oxygen} \), the rate of oxygen consumption was measured in 50 mM HEPES pH 7.5 at 20 °C with varying oxygen content (25-525 µM, obtained by either sparging with O₂ or N₂ until the desired concentration was obtained), 1 µM enzyme, and 800 µM 2-HEP in a final volume of 1 mL. Each reaction was performed in triplicate. Rates were measured using the OxyGraph Plus software version 1.02, and data were fit to the Michaelis-Menten equation using Igor Pro version 6.32A.

Bionformatics

The JGI-IMG and NCBI data repositories (accessed April 10, 2016) (non-redundant protein (nr), metagenomic protein (env_nr), whole-genome shotgun contigs (wgs), and transcriptome shotgun assembly (tsa) databases) were mined for additional MPnS, HEPD, and HppE-like sequences using BLAST. Biochemically characterized homologs were used to seed queries (A9A1T2, MPnS, \textit{N. maritimus} strain SCM1; AAU00079, PhpD, \textit{S. viridochromogenes} strain DSM 40736; WP_033205181.1, HppE, \textit{S. wedmorensis} strain NRRL 3426).

A series of criteria was applied to produce a list of high-quality sequences for subsequent analyses. For hits deriving from isolate microbial strains, the genome neighborhood was inspected for phosphoenolpyruvate (PEP) mutase encoding genes, a rigorously established method for discovering biosynthetic loci for phosphonic acids (38). Those lacking a gene PEP mutase within a 20 kb window were eliminated. Hits from metagenome and metatranscriptome datasets that were less than 70% the length of \textit{NmMPnS} were eliminated. Finally, sequences were manually de-replicated so that identical entries returned from queries of the various databases are only represented once.
Alignments were performed using MAFFT (39). Approximated maximum-likelihood phylogenies were constructed using FastTree 2 with gamma 20 correction (40) and visualized with FigTree. A separate alignment composed of only HEPD and MPnS was analyzed to identify amino acids that coordinate iron and mediate specificity.
References


Chapter VI:

Methylphosphonate production in perspective

The oceanic methane paradox is an intriguing molecular mystery with global implications. Methane is a key greenhouse gas, and the size of the ocean demands we pay close attention to its role in methane production. Small variations in methane flux from the ocean surface, potentially occurring as a result of climate change, may have a direct impact on global methane levels. Our work in identifying sequence signatures of methylphosphonate (MPn) production \(^1\) is a key step in further studying oceanic methane production.

Marine microbiology is a complex field. Many of the organisms playing key roles in ocean biogeochemistry are unculturable microbes. As sequencing-based techniques have accelerated forward in recent years, marine microbiologists have begun to rely heavily on metagenomic, metatranscriptomic, and metaproteomic techniques to study the complex biochemistries occurring \textit{in situ}. The analyses of data generated from these techniques are limited by the quality of available annotations. In the context of MPn production, until this work, there was no sequence annotation to accurately identify genes encoding MPnS in marine metagenomics sequences, limiting research into the role of MPn in marine methane production.

With a sequence signature for MPnS in hand, researchers can now directly investigate whether the methylphosphonate hypothesis \(^2\) is contributing to methane production at their study site. Although published only recently, we have already been in contact with a research team looking to explain large methane blooms over isolated Arctic ocean sea ice. The researchers noted that these methane blooms have been seen before, but were often set aside as artifacts with no reasonable source. Although we cannot provide direct evidence, there are numerous correlations suggesting MPn-containing organisms are abundant in Arctic and
Antarctic sea ice, providing a plausible explanation for the aforementioned methane blooms (3-5). We envision our sequence signatures, and the biochemistry and structural biology supporting them, will give researchers the confidence to develop experimental strategies to assess directly the role MPn plays in marine methane production.
References


