STRUCTURE-GUIDED STUDIES OF BACTERIAL COMPETITION MECHANISMS

by

Kateryna Podzelinska

A thesis submitted to the Department of Biochemistry

In conformity with the requirements for

the degree of Doctor of Philosophy

Queen’s University

Kingston, Ontario, Canada

(September, 2010)

Copyright © Kateryna Podzelinska, 2010
Abstract

Microorganisms have evolved a stunning array of strategies for nutrient competition, ranging from concerted effort of antibiotic release to kill off competing species, to evolving complex enzymatic pathways that are capable of scavenging nutrients from sources not utilizable by other organisms. The carbon-phosphorous (C-P) lyase pathway is a survival mechanism that is activated during phosphate limitation in certain species of bacteria and enables cleavage of the extremely stable C-P bond in order to obtain phosphorous from organophosphonates. The structure and biochemical characterization of PhnP, a critical accessory protein from C-P lyase pathway of *E. coli* is presented in this thesis. The structure of PhnP revealed a conserved metal-dependent hydrolase active site with two Mn$^{2+}$ ions, and another unique mononuclear Zn$^{2+}$ site that appears to have a structural role. A non-physiological ligand that fortuitously co-crystallized with the enzyme provided insights into the catalytic features of the active site. We were able to demonstrate hydrolytic activity towards a number of phosphodiesterase substrates, and propose a plausible physiological role for PhnP. These results contribute to deciphering the mechanism of phosphonate utilization, which would allow design of bioremediation programs to remove toxic phosphonates from the environment. Antibiotic production is another mechanism for resource competition. Structural characterization of CmlS, a halogenase from the chloramphenicol biosynthesis pathway of *S. venezuelae* is presented. The crystal structure revealed a novel covalent modification of its FAD cofactor, which was confirmed through ESI-MS and chemical denaturation studies. The unique C-terminal domain, active site architecture, and the position of the C-terminus suggest that halogenation mechanism of CmlS may differ from the currently proposed mechanism for structurally related halogenases. This work provides early steps towards understanding mechanisms of enzymatic halogenations, which is of great scientific, as well as pharmaceutical interest.
Acknowledgements

First of all, I would like to express my deep gratitude to my supervisor Dr. Zongchao Jia for giving me the opportunity to study in his laboratory. I greatly appreciate his guidance through the tricky waters of crystallization and structure solution. I also cannot thank him enough for his understanding, support and encouragement that allowed me to successfully get through the personal difficulties during my studies. I would also like to extend most sincere appreciation of all the advice and help from the past and present members of the laboratory. In particular I would like to acknowledge Dr. Gour Pal for teaching me the foundation skills for crystallography, Dr. Qilu Ye for coming to the rescue of the most finicky crystals and for sharing crystallographic tricks with me, Dr. Jimin Zhang and Dr. Michael Suits for their guidance through structure solution and assistance with countless hurdles that computer programs threw at me, and Laura van Staalduinen and Dr. Mark Currie for their friendship and continual emotional support.

I would like to thank Dr. Steven Smith, Dr. Andrew Craig and Dr. John Allingham for serving as members of my committee. I greatly appreciate guidance and helpful discussions with my collaborators, Dr. David Zechel and Dr. Bjarne Hove-Jensen, who have helped me tremendously throughout the projects. I would like to acknowledge hard work and dedication of the two M.Sc. students from Dr. Zechel’s lab: Shumei He, who has collaborated with me on PhnP project, and Ryan Latimer, who worked with me on CmlS project.

I would like to acknowledge Kim Munro for his extensive technical expertise with biophysical protein characterization and his cheerful approach to life, which always had a positive effect on me when the experiments were not working. Many thanks to Alexei Soares of NSLS for X-ray data collection and sharing his expertise during the RapiData course of 2007 and remote data collection. I will never forget the kindness and helpfulness of Scott Crawley and Rob Eves,
who had helped me to overcome numerous problems encountered with molecular biology and protein work.

Finally, I would like to thank my family for their support. I could always count on Alex, my dear brother, and Patrick, my partner of many years, for help and support in the worst of times.

The author of this thesis has been supported by the Ontario Graduate Scholarship Award, as well as number of Queen’s internal fellowships.
Table of Contents

Abstract............................................................................................................................................ii
Acknowledgements.........................................................................................................................iii
Table of Contents.............................................................................................................................v
List of Figures ......................................................................................................................................viii
List of Abbreviations .........................................................................................................................x
Chapter 1 General introduction........................................................................................................ 1
  1.1 Mechanisms of bacterial competition .................................................................................... 1
  1.2 Phosphorous utilization mechanisms in bacteria ................................................................. 3
    1.2.1 Occurrence, uptake and degradation of phosphorous compounds......................... 3
    1.2.2 Phosphonate utilization pathways .......................................................................... 6
    1.2.3 Characterization of Pn utilization operon of E. coli .................................................. 12
  1.3 Antimicrobial mechanisms of competition .......................................................................... 15
    1.3.1 The role of antibiotic production in resource competition ....................................... 15
    1.3.2 Natural halogenated compounds ........................................................................... 16
    1.3.3 Halogenase overview ......................................................................................... 18
    1.3.4 Chloramphenicol biosynthesis pathway .................................................................. 21
Chapter 2 Expression, purification and preliminary diffraction studies of PhnP....................... 26
  2.1 Abstract ............................................................................................................................... 27
  2.2 Introduction ......................................................................................................................... 28
  2.3 Materials and methods ....................................................................................................... 29
    2.3.1 Cloning, expression and purification .................................................................... 29
    2.3.2 Crystallization ....................................................................................................... 30
    2.3.3 Data collection and diffraction measurements ...................................................... 31
  2.4 Results and discussion ......................................................................................................... 32
Chapter 3 Structure of PhnP, a phosphodiesterase of the carbon-phosphorus lyase pathway for
  phosphonate degradation ............................................................................................................ 37
  3.1 Abstract ............................................................................................................................... 38
  3.2 Introduction ......................................................................................................................... 39
  3.3 Experimental procedures .................................................................................................... 40
    3.3.1 Materials ................................................................................................................ 40
    3.3.2 Expression, purification, and size exclusion chromatography of PhnP and mutants... 41
3.3.3 Substrate screening ....................................................................................................... 41
3.3.4 Steady-state kinetic analysis ......................................................................................... 41
3.3.5 RNase activity assay ..................................................................................................... 42
3.3.6 ICP-MS analysis ........................................................................................................... 43
3.3.7 Metal dependence of PhnP activity............................................................................... 43
3.3.8 TLC analysis ................................................................................................................. 43
3.3.9 PhnP structure solution and refinement ........................................................................ 44
3.4 Results.................................................................................................................................. 44
3.4.1 Substrate specificity and product analysis .................................................................... 44
3.4.2 Metal ion dependence of PhnP ...................................................................................... 50
3.4.3 Kinetic analysis of wild type PhnP ............................................................................... 54
3.4.4 The crystal structure of PhnP ........................................................................................ 54
3.4.5 Structural homology to tRNase Z endonucleases ......................................................... 59
3.4.6 The Zn^{2+} binding site............................................................................................... 59
3.4.7 Structural homology to PqqB........................................................................................ 62
3.4.8 The PhnP active site...................................................................................................... 63
3.5 Discussion............................................................................................................................ 67

Chapter 4 Expression, purification and preliminary diffraction studies of CmlS................. 72
4.1 Abstract................................................................................................................................ 73
4.2 Introduction.......................................................................................................................... 74
4.3 Materials and methods ......................................................................................................... 75
4.3.1 Cloning, expression and purification ............................................................................ 75
4.3.2 Crystallization............................................................................................................... 79
4.3.3 Data collection and diffraction measurements.............................................................. 79
4.4 Results and discussion ......................................................................................................... 80

Chapter 5 Chloramphenicol biosynthesis: the structure of CmlS, a flavin-dependent halogenase showing a covalent flavin–aspartate bond ........................................................................ 85
5.1 Abstract................................................................................................................................ 86
5.2 Introduction.......................................................................................................................... 87
5.3 Materials and methods ......................................................................................................... 91
5.3.1 Expression, purification, and crystallization................................................................. 91
5.3.2 Data collection, structure determination, and refinement ............................................. 91
5.3.3 Electrospray ionization mass spectrometry analysis..................................................... 92
5.3.4 Chemical denaturation and flavin content analysis of wild type CmlS and D277N mutant ............................................................... 92
5.4 Results......................................................................................................................... 93
  5.4.1 The overall structure of CmlS .............................................................................. 93
  5.4.2 The C-terminal domain ....................................................................................... 98
  5.4.3 CmlS structural homologues .............................................................................. 98
  5.4.4 The FAD binding site: a covalently bound FAD cofactor ................................. 99
  5.4.5 Electrospray ionization mass spectrometry analysis of CmlS ......................... 103
  5.4.6 Spectroscopic characterization and denaturation of wild type CmlS and D277N mutant ............................................................................................................. 105
  5.4.7 Interaction of the C-terminus with the active site ............................................. 107
  5.4.8 The halogenation active site ............................................................................. 109
  5.4.9 A potential general acid catalyst ...................................................................... 111
5.5 Discussion.................................................................................................................. 111
5.6 Conclusions............................................................................................................... 117

Chapter 6 Discussion, summary, and conclusions ......................................................... 119
  6.1 Importance of studying microbial competition mechanisms .............................. 119
  6.2 Insights into the mechanism of the C-P lyase pathway ........................................ 120
  6.3 Insights into the chloramphenicol biosynthesis pathway .................................... 128
  6.4 Conclusion ............................................................................................................. 135

Appendix A Additional data ............................................................................................ 137
  1. Purification of PhnP.................................................................................................. 138
  2. Purification of CmlS ............................................................................................... 140

References.................................................................................................................... 142
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1-1</td>
<td>Pathways of phosphorous acquisition in bacteria</td>
<td>5</td>
</tr>
<tr>
<td>Figure 1-2</td>
<td>Pathways for the degradation of phosphonate</td>
<td>7</td>
</tr>
<tr>
<td>Figure 1-3</td>
<td>Examples of phosphonate cleavage products</td>
<td>9</td>
</tr>
<tr>
<td>Figure 1-4</td>
<td>Presumed mechanism of the C–P bond cleavage</td>
<td>11</td>
</tr>
<tr>
<td>Figure 1-5</td>
<td>Examples of five classes of halogenases</td>
<td>19</td>
</tr>
<tr>
<td>Figure 1-6</td>
<td>Gene cluster and pathway for Cm biosynthesis</td>
<td>24</td>
</tr>
<tr>
<td>Figure 2-1</td>
<td>PhnP crystallization</td>
<td>33</td>
</tr>
<tr>
<td>Figure 2-2</td>
<td>Diffraction of SeMet PhnP crystals</td>
<td>35</td>
</tr>
<tr>
<td>Figure 3-1</td>
<td>PhnP sequence alignment</td>
<td>46</td>
</tr>
<tr>
<td>Figure 3-2</td>
<td>Screening PhnP hydrolytic activity</td>
<td>48</td>
</tr>
<tr>
<td>Figure 3-3</td>
<td>TLC analysis of PhnP reaction product</td>
<td>49</td>
</tr>
<tr>
<td>Figure 3-4</td>
<td>The crystal structure of PhnP</td>
<td>55</td>
</tr>
<tr>
<td>Figure 3-5</td>
<td>Size exclusion chromatogram of PhnP</td>
<td>56</td>
</tr>
<tr>
<td>Figure 3-6</td>
<td>A comparison of PhnP with close structural homologues</td>
<td>60</td>
</tr>
<tr>
<td>Figure 3-7</td>
<td>Active site of PhnP</td>
<td>65</td>
</tr>
<tr>
<td>Figure 4-1</td>
<td>CmlS sequence alignment</td>
<td>76</td>
</tr>
<tr>
<td>Figure 4-2</td>
<td>CmlS crystallization</td>
<td>81</td>
</tr>
<tr>
<td>Figure 4-3</td>
<td>Diffraction of SeMet CmlS crystals</td>
<td>82</td>
</tr>
<tr>
<td>Figure 5-1</td>
<td>CmlS sequence alignment</td>
<td>88</td>
</tr>
<tr>
<td>Figure 5-2</td>
<td>The structure of CmlS</td>
<td>96</td>
</tr>
<tr>
<td>Figure 5-3</td>
<td>Covalent attachment of FAD cofactor</td>
<td>101</td>
</tr>
<tr>
<td>Figure 5-4</td>
<td>Alignment of CmlS and CndH</td>
<td>102</td>
</tr>
<tr>
<td>Figure 5-5</td>
<td>ESI mass spectrum of CmlS</td>
<td>104</td>
</tr>
<tr>
<td>Figure 5-6</td>
<td>Absorbance spectra of CmlS and its D277N mutant</td>
<td>106</td>
</tr>
<tr>
<td>Figure 5-7</td>
<td>The halogenation active site of CmlS</td>
<td>108</td>
</tr>
<tr>
<td>Figure 5-8</td>
<td>The HOCl tunnel of CmlS</td>
<td>110</td>
</tr>
<tr>
<td>Figure 5-9</td>
<td>Proposed mechanism for halogenation by CmlS</td>
<td>113</td>
</tr>
<tr>
<td>Figure 6-1</td>
<td>A proposed model for Pn utilization</td>
<td>127</td>
</tr>
<tr>
<td>Figure 6-2</td>
<td>Possible pathways for halogenation of Cm</td>
<td>130</td>
</tr>
</tbody>
</table>
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2-1</td>
<td>Diffraction data for SeMet PhnP crystals</td>
<td>36</td>
</tr>
<tr>
<td>Table 3-1</td>
<td>ICP-MS analysis of PhnP</td>
<td>51</td>
</tr>
<tr>
<td>Table 3-2</td>
<td>Metal ion dependence of PhnP</td>
<td>52</td>
</tr>
<tr>
<td>Table 3-4</td>
<td>Kinetic parameters of PhnP</td>
<td>53</td>
</tr>
<tr>
<td>Table 3-5</td>
<td>PhnP data collection statistics</td>
<td>58</td>
</tr>
<tr>
<td>Table 3-6</td>
<td>PhnP active site distances</td>
<td>64</td>
</tr>
<tr>
<td>Table 4-1</td>
<td>Diffraction data for SeMet CmlS crystals</td>
<td>84</td>
</tr>
<tr>
<td>Table 5-1</td>
<td>CmlS data collection statistics</td>
<td>94</td>
</tr>
</tbody>
</table>
List of Abbreviations

The abbreviations used are:

5PR1,2cP  5-phosphoribose-1,2-cyclic phosphate
amu         atomic mass unit
ABC         ATP-binding cassette
AEPn        2-aminoethylphosphonate
Alkylphosphonate  AlPn
BAP         bacterial alkaline phosphatase
bpNPP       bis(p-nitrophenoxy) phosphate
CHES        2-(cyclohexylamino) ethanesulfonic acid
Cm          chloramphenicol
CmlS        *Streptomyces venezuelae* chloramphenicol halogenase
CndH        *Chondromyces croatus* tyrosyl halogenase
C-O-P       carbon-oxygen-phosphorous
C-P         carbon-phosphorus
EAS         electrophilic aromatic substitution
EDTA        ethylenediaminetetraacetic acid
ESI         electrospray ionization
EtPn        ethylphosphonate
EtPnR       α-1-(ethylphosphono)ribose
FAD         flavin adenine dinucleotide
FDH         flavin-dependent halogenase
GR          glutathione reductase
GdHCl       guanidine hydrochloride
HPLC        high-performance liquid chromatography
ICP         inductively coupled plasma
IPTG        isopropyl-β-D-1-thiogalactopyranoside
LB          Luria Bertani Broth
MALDI       matrix-assisted laser desorption ionization
MePn        methylphosphonate
MeOH  methanol
MS  mass spectrometry
NADH  reduced β-nicotinamide adenine dinucleotide
Ni-NTA  nickel-nitriloacetic acid
NSLS  National Synchrotron Light Source
PAPA  p-aminophenylalanine
PDB  Protein Data Bank
PEG  polyethylene glycol
PHBH  *Pseudomonas fluorescens* para-hydroxybenzoate hydroxylase
PhnP  *E. coli* protein from C-P lyase pathway
Pi  inorganic phosphate
Pit  inorganic phosphate transport system
Pn  phosphonate
pNP-TMP  thymidine-5’-monophosphate p-nitrophenyl ester
PnP  phosphonopyruvate hydrolase
PolyP  polyphosphate
ppG2’,3’p  guanosine 5’-bispyrophosphate-2’,3’-cyclic phosphate
ppGpp  guanosine 3’,5’-bispyrophosphate
pppGpp  guanosine 3’-diphosphate-5’-triphosphate
PrnA  *Pseudomonas fluorescens* tryptophan halogenase
PRPP  5-phospho-D-riboyl α-1-diphosphate
ψi  phosphate-starvation-inducible
Pst  Pi-specific transport system
Pt  phosphite
R1P  ribose-1-phosphate
R1,5P  ribose-1,5-bisphosphate
RFU  relative fluorescence units
r.m.s.d  root mean squared deviation
SAD  single anomalous dispersion
SDS-PAGE  sodium dodecyl sulphate polyacrylamide gel electrophoresis
SeMet  selenomethionine
TLC  
thin layer chromatography

Tricine  
N-[2-hydroxy-1,1-bis(hydroxymethyl) ethyl]glycine

WT  
wild type
Chapter 1
General introduction

1.1 Mechanisms of bacterial competition

Microorganisms are essential components of every ecosystem and are recognized as a dominant life form on earth, comprising the major living biomass on the planet. Every environment is populated by numerous species of microorganisms, requiring common elements, such as light, oxygen, nutrients and space for surviving and passing on their genes. The limited nature of these resources results in fierce competition between different species and also between members of the same species. The outcome of this competition depends on the relative ability of the organisms to utilize these resources. Natural selection drives the emergence of new adaptations and survival strategies between competing organisms that allow specialization for colonizing various environmental niches.

Nutritional resources are a focal point of microbial competition. All organisms require an energy source and various forms of carbon, nitrogen, oxygen, phosphorous, sulfur, as well as a number of metal ions in order to live. Limitation in the amount of any of these nutrients presents a survival challenge. Jacques Monod was a pioneer of the study of bacterial growth kinetics and was the first one to demonstrate the relationship between limiting nutrient concentrations and bacterial growth. In a medium where all but one nutrient were provided in excess, Monod showed that the bacterial growth yield has a linear dependence on the initial concentration of the limiting nutrient. This relationship was then incorporated into the equation for exponential bacterial growth - an equation that is similar to the Michaelis–Menten representation of enzyme kinetics. This model describes a relationship between growth rate and the concentration of a limiting nutrient.
In natural environments, nutritional limitation imposes a major restriction on bacterial growth. Shortage in any of the essential nutrients forces microorganisms to undergo numerous metabolic changes to survive in the nutrient-limiting conditions. Microbes adapt to shifts in environmental factors by means of a wide variety of genotypic and phenotypic changes. These include appropriate modification of enzyme synthesis to take up the growth-limiting nutrient, modulation of uptake rates for nutrients available in excess, rerouting of metabolic pathways to avoid possible blockages due to specific nutrient limitation, and coordination of synthetic rates to maintain balanced growth.

Microorganisms have evolved numerous strategies to augment their acquisition of nutritional resources. Motility, space competition, rapid growth rate, coordinated behaviour, predation, and antimicrobial production can all be interpreted as a way for one organism to maximize its nutrient uptake at the expense of another. There are two general types of the nutritional competition: scramble and contest. The former is a passive mechanism where one organism is better adapted to consuming a particular resource, such as taking a limiting nutrient or colonizing a niche, thereby depleting that resource and depriving another organism of that resource. The latter involves one competitor actively harming another through predation or release of antimicrobials.

Despite the existence of intra-species competition, there is emerging evidence that bacteria engage in cooperative behaviour, which can be used to enhance acquisition of nutritional resources. Bacteria utilize the quorum sensing mechanism to achieve effects that are only feasible when carried out by a group of organisms. Quorum sensing involves production of diffusible signal molecules that regulate cell-density-dependent gene expression. For example, iron is one of the major limiting factors of the \textit{in vivo} growth of parasitic bacteria. This occurs because under aerobic conditions, iron exists in the largely insoluble ferric form (Fe(III)). Additionally, many
host species defend against infection by actively withholding iron from pathogenic bacteria using proteins with high iron affinity. In response to iron limitation, bacteria have evolved a number of mechanisms for scavenging iron from their hosts. One mechanism involves the cooperative release of siderophores, which bind to the insoluble and host-bound iron with high affinity, making it available for utilization by parasitic bacteria.

Several mechanisms of competition actively restrict or remove a nutrient from one organism and supply it to another. For example, some species of bacteria found in wastewater treatment plants upregulate a scavenging pathway for phosphorous sequestration when phosphate is readily available. Phosphate is then incorporated into polyphosphate chains that are stored in the cells for use under phosphate-limiting conditions. Such accumulation of an essential nutrient facilitates the dominance of these organisms over other species.

The work presented in this thesis adds to the growing body of research aimed at deducing the biochemical mechanisms involved in two strategies of bacterial competition: the antibiotic production pathway and the nutrient-scavenging pathway, in particular phosphorous acquisition.

1.2 Phosphorous utilization mechanisms in bacteria

1.2.1 Occurrence, uptake and degradation of phosphorous compounds

Phosphorous is the 11th most common element on earth. It is never found in its elemental form in nature due to high reactivity. Instead it usually occurs in its most oxidized form of +5, such as inorganic minerals, free phosphate and phosphate esters, like phosphoethanolamine. The common mineral form of phosphorous is calcium phosphate contained in rocks that over time became available to organisms as phosphate due to the weathering of rock surfaces. Phosphate is a nutrient that most often limits growth of microorganisms in aquatic environments.
Phosphorous also occurs in its reduced forms of +3 (phosphite (Pt) and phosphonate (Pn)) and +1 (hypophosphite) in anaerobic environments, such as lake sediments. Pn are a large class of organophosphorus molecules that contain a carbon-phosphorous (C-P) bond, which is much more stable than the labile carbon-oxygen-phosphorous (C-O-P) bond found in phosphate esters. The early evidence for natural Pn was based upon finding organophosphorus that was released during combustion, which resisted strong-acid and strong-base hydrolysis. The first natural C-P compound was isolated from ciliates and identified as 2-aminoethylphosphonate (AEPn) in 1959. AEPn is a structural analog of the ethanolamine phosphate moiety in phospholipids. Lipids with C-P bonds (phosphonolipids) are the corresponding Pn analogs of phosphatidylethanolamine, phosphatidylcholine, and phosphatidylserine, and are found in various organisms from protozoa to mammals. Pn also exist covalently bound to proteins as well as other structural components of the cell.

Bacteria use inorganic phosphate (Pi) as the preferred source of phosphorous (Figure 1.1). The components of the inorganic phosphate transport system (Pit) are expressed constitutively and are responsible for uptake of Pi when it is available in excess. Under bountiful nutritional conditions bacteria can incorporate excess Pi into polyphosphate (polyP) chains of 1000 residues or more for storage. PolyP can serve as a phosphorous source for the biosynthesis of nucleic acids and phospholipids under Pi starvation conditions. Bacteria have evolved several complex systems to survive under Pi starvation conditions. When Pi concentrations fall below 4 µM, the pho regulon is activated and transcription of several sets of phosphate-starvation-inducible (psi) genes is induced up to 1500-fold. One such gene set encodes a high-affinity Pi transport system (Pst), which serves as a major scavenger of Pi containing molecules. In E. coli the Pst system is encoded by the pstS, pstA, pstB, and pstC genes, producing a periplasmic protein-dependent transporter similar to those of histidine, maltose, and ribose. It belongs to the superfamily of ATP-binding cassette (ABC) transporters. When Pi is not available,
Figure 1.1 Pathways of phosphorous acquisition in bacteria (Figure adapted from Hirota et al.\textsuperscript{17}).
bacteria are forced to utilize alternative phosphorous sources, such as organophosphates (Pi esters), Pt, and Pn\textsuperscript{18}. Since most Pi esters are not transportable, Pi must be cleaved off before being taken up. This is achieved by a variety of hydrolases, including bacterial alkaline phosphatase (BAP), that use a nucleophilic attack by an activated serine residue upon the phosphate ester to form a phosphoserine ester intermediate. The ester is subsequently hydrolyzed by hydroxide, while an alkoxide leaving group acquires a proton from a general acid residue. Compounds like Pt and Pn are taken up through the activity of the C-P lyase transporter, encoded by \textit{phnCDE} genes. Pt is thought to be enzymatically oxidized to Pi before being used as a phosphorous source by a mechanism similar to phosphate ester with formation of hydride anion as a leaving group\textsuperscript{19}. Utilization of Pn compounds is limited to certain species of Gram-negative, and a few Gram-positive bacteria and requires cleavage of the highly stable C-P bond through a C-P lyase pathway\textsuperscript{20}. Pathways for Pn degradation are described in the next section. Marine cyanobacteria that grow on the surface of oligotrophic sea, use Pn extensively because Pi concentrations are extremely low, but a considerable amount of Pn is present\textsuperscript{21}. Additionally, motile bacteria can also exhibit chemotactic responses toward Pi for their survival during Pi starvation. Pi taxis allows bacteria to adjust rapidly to sudden nutrient concentration changes, which provides a further selective advantage.

### 1.2.2 Phosphonate utilization pathways

Pn are utilized by four pathways found in various species of bacteria: (a) phosphonopyruvate hydrolase pathway (PnP\textsubscript{y} hydrolase); (b) phosphonoacetate hydrolase pathway; (c) phosphonoacetaldehyde hydrolase (phosphonatase) pathway; and (d) carbon-phosphorus lyase (C-P lyase) pathway (Figure 1.2). The main distinction between these Pn utilization systems is that the first three pathways act upon chemically activated C-P compounds that contain an electron-withdrawing \(\beta\)-carbonyl group next to the scissile bond. On the other
Figure 1-2 Pathways for the degradation of Pn compounds: (a) Phosphonopyruvate hydrolase pathway; (b) Phosphonoacetate hydrolase pathway; (c) Phosphonoacetaldehyde (phosphonatase) pathway; (d) C-P lyase pathway; (Figure adapted from White and Metcalf).
hand C-P lyase is believed to catalyze a direct scission of the C-P bond using a radical mechanism.

The phosphonopyruvate hydrolase pathway genes are expressed only in the presence of the substrate, regardless of the level of Pi in the medium. This is consistent with the primary role of phosphonopyruvate as a carbon and energy source, rather than a phosphorous source. PnPy hydrolase catalyzes the cleavage of the C-P bond of PnPy to yield Pi and pyruvate (Figure 1.2a), while phosphonoacetate hydrolase cleaves the C-P bond of phosphonoacetate to yield Pi and acetate (Figure 1.2b). Thus, both the Pi and the C compound liberated from these reactions can be directly used as nutrient sources. Importantly, activating groups are already present in PnPy and phosphonoacetate, allowing their direct hydrolysis without the need for prior rearrangement.

The phosphonatase pathway is under the control of the pho regulon in most bacteria. It consists of two steps. The first step involves the substrate deamination, where the amine group from 2AEPn is transferred onto the pyruvate, resulting in the formation of 2-phosphonoacetaldehyde and alanine, which allows 2AEP to be used as a nitrogen source. This reaction introduces an activating carbonyl moiety onto the β-carbon and results in destabilization of the C-P bond. In the second step, 2-phosphonoacetaldehyde is hydrolyzed to acetaldehyde and Pi (Figure 1.2c). The reaction proceeds via the formation of a covalently bound imine intermediate with the carbonyl group and the side chain of a lysine residue in the enzyme. In many bacteria AEPn is degraded thorough the C-P lyase pathway as well.

The C-P lyase pathway demonstrates the broadest substrate specificity, acting on alkylphosphonates (AlPn), AEPn, phenylphosphonate, and in some cases phosphonomethylglycine (glyphosate) for use as a sole phosphorous source (Figure 1.2d). It must be noted that unactivated AlPn are degraded only through C-P lyase pathway. Some examples of known reactions are shown in Figure 1.3. It is also the most widespread pathway of Pn utilization.
<table>
<thead>
<tr>
<th>Organophosphonate</th>
<th>Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{CH}_3\text{PO}_3\text{H}_2$</td>
<td>$\text{CH}_4$</td>
</tr>
<tr>
<td>$^{13}\text{CH}_3\text{PO}_3\text{H}_2$</td>
<td>$^{13}\text{CH}_4$</td>
</tr>
<tr>
<td>$\text{CD}_3\text{PO}_3\text{H}_2$</td>
<td>$\text{CD}_3\text{H}$</td>
</tr>
<tr>
<td>$\overset{\text{O}}{\overset{\text{P}}{\text{O}}-\text{OH}}$</td>
<td>$\overset{\text{O}}{\overset{\text{P}}{\text{O}}-\text{OH}}$ (0.5%)</td>
</tr>
<tr>
<td>$\overset{\text{P}}{\text{O}}_3\text{H}_2$</td>
<td>$\overset{\text{P}}{\text{O}}_3\text{H}_2$ (1%)</td>
</tr>
<tr>
<td>$\overset{\text{P}}{\text{O}}_3\text{H}_2$</td>
<td>$\overset{\text{P}}{\text{O}}_3\text{H}_2$ (2%)</td>
</tr>
<tr>
<td>$\overset{\text{P}}{\text{O}}_3\text{H}_2$</td>
<td>$\overset{\text{P}}{\text{O}}_3\text{H}_2$</td>
</tr>
<tr>
<td>$\overset{\text{P}}{\text{O}}_3\text{H}_2$</td>
<td>$\overset{\text{P}}{\text{O}}_3\text{H}_2$</td>
</tr>
</tbody>
</table>

Figure 1-3 Examples of known products of Pn cleavage by C-P lyase pathway (Figure from Frost et al.\textsuperscript{23}).
among bacteria, and is the only means of Pn catabolism in *E.coli*. Based on the degradation products of various radiolabeled AlPn, a radical mechanism was proposed (Figure 1.4)\(^{23}\). The process is initiated by generation of a phosphonyl radical. Subsequent fragmentation of this reactive intermediate would lead to a metaphosphate and alkyl radical. Abstraction of a hydrogen atom by the alkyl radical would yield the corresponding alkane. The finding of not only alkanes but also alkenes among degradation products is evidence of the radical-based mechanism of degradation\(^{23}\). The electrophilic metaphosphate group would be quickly attacked by a nucleophile, such as water or a hydroxyl group, to produce a phosphate. It is important to note that the phosphorous product has not been conclusively identified, though it is suggested that nucleotides act as acceptors of the phosphoryl group of AlPn\(^{24}\). This hypothesis is supported by the finding of Avila *et al.*, where \(\alpha\)-1-ethylphosphono-ribose was identified as an intermediate of ethyphosphonate (EthPn) degradation\(^{25}\).

Understanding the mechanism of Pn utilization has important environmental implications. Large amounts of man-made Pn enter natural ecosystems through the extensive use of herbicides, like glyphosate, pesticides, and flame retardants. Fertilizers are used extensively around the world, accounting for 80% of the total phosphorous manufacturing. The extreme stability of the C-P bond results in accumulation of these compounds in the environment, and presents a removal challenge. Since certain microorganisms are able to cleave the C-P bond cleavage, they present environmentally friendly options for removal of these harmful compounds from the environment.
Figure 1-4 Presumed mechanism of the C–P bond cleavage by the C-P lyase pathway (Figure from Frost et al.\textsuperscript{23}).
1.2.3 Characterization of Pn utilization operon of *E. coli*

Despite the importance of understanding the reaction mechanisms involved in Pn utilization, little biochemical information is available about the C-P lyase pathway. To date, studies of the C-P lyase function have been limited to intact bacterial cells\(^{16,25}\), since numerous attempts to reconstitute of the C-P lyase activity in cell-free systems were unsuccessful. However, the repression of Pn utilization by inorganic phosphate and the limited cell growth in a minimal medium with Pn as the sole phosphorus source imposed serious restrictions on studies of Pn degradation in intact bacterial cells. These issues were eventually remedied by engineering an *E. coli* strain with a mutation in the *pstS* gene, resulting in a constitutive expression of Pn operon regardless of the amount of Pi present\(^{26}\). The cell growth is also improved due to presence of Pi, allowing for a metabolite accumulation studies in *phn* operon mutants.

Despite numerous attempts, no group could reliably and reproducibly demonstrate C-P lyase activity in lysed cells. Two claims of C-P lyase activity in cell extracts have been made thus far. Murata *et al.* have demonstrated the release of phosphate in presence of phosphonoacetate, but the fate of organic moiety remained undetermined\(^{27}\). Later this report was proven to be incorrect, because phosphate was released even in the absence of phosphonoacetate\(^{28}\). Kononova *et al.* took a different approach and monitored C-P bond cleavage by the amount of methane gas released when methylphosphonate (MePn) was added to the cell lysate. This approach is considered to be a much more reliable indicator of Pn cleavage than release of phosphate, since phosphate can be produced by a number of other irrelevant reactions. Despite a multitude of alklylation and dealkylation reactions occuring in the cell, the C-P lyase activity is unique in producing a free alkane. The group has tested the addition of various sugars, nucleotides and cofactors to the lysate, but in the end they managed to reconstitute only 1% of C-P lyase activity in cell extracts, compared to that observed in intact cells\(^{29}\). While it is entirely possible that a true C-P lyase activity was observed, it may have been very inefficient due to disruption of cellular
compartmentalization, dilution of possible cofactors, enzyme inhibition though byproducts of unrelated reactions, or requirement for other intact pathways that would provide intermediates to the C-P lyase pathway.

The mutational analysis involved two distinct types of experiments. The ability to cleave a C-P bond was assessed by growing cells on LB media to a high density, washing them with Pi-free buffer and transferring them to a minimal media with MePn as the sole source of phosphorous. The air from the headspace of the culture was then monitored for the release of a methane gas. In another type of experiment Pn utilization was measured by cell’s ability to grow on minimal solid or liquid media supplemented with organophosphonates as a sole source of phosphorus. Thus, some mutants were able to cleave the C-P bond due to the presence of a functional set of core enzymes, yet were unable to grow on the resulting product due to a defect in the genes required for transport or assimilation of the phosphate product.

The enzymes required for Pn utilization are encoded by a 14-gene operon \textit{phnCDEFGHIJKLMNOP} and are collectively known as the C-P lyase pathway \textsuperscript{30}. Sequence analysis of \textit{phnCDE} gene products suggested that they encode a Pn transporter. PhnE has high hydrophobicity content and shows sequence homology to integral membrane proteins. It is the prime candidate for a transmembrane protein, and it likely forms a dimer. PhnD was shown to be a soluble periplasmic protein with nanomolar affinity for 2AEPn and lower affinity for a number of other Pn \textsuperscript{31}. PhnC has sequence homology to nucleotide-binding domains of ABC transporters, and is possibly the cytoplasmic component of the Pn transporter. Mutational studies have shown that the \textit{phnCDE} genes are not required for C-P bond cleavage, yet are essential for cellular growth on Pn \textsuperscript{32}. \textit{PhnF} encodes a transcriptional regulator protein \textsuperscript{33}. Mutational studies have shown that \textit{phnGHIJKLM} genes are essential for C-P bond cleavage, and disruption of any of them abolished methane production. It is currently not known whether some or all of these
proteins form an enzymatic complex, and whether this complex is cytoplasmic, or if it is tethered to the membrane through PhnM protein, which has a hydrophobic sequence in the N-terminus. PhnK and PhnL proteins also show sequence homology to ATP-binding proteins, and may form an enzyme complex together with other core members. PhnN has been characterized as an ATP dependent kinase which phosphorylates 5-phospho-α-D-ribofuranosyl-1,5-diphosphate to produce 5-phospho-α-D-ribofuranosyl pyrophosphate (PRPP)\(^34\). PRPP is a precursor in a redundant pathway for NADH biosynthesis, and is also used to synthesize purine and pyrimidine nucleosides and aromatic amino acids histidine and tryptophan. \textit{phnN} gene disruption did not eliminate C-P cleavage ability, although \textit{phnN} mutants grew poorly on Pn\(^35\), suggesting an accessory role for PhnN in the metabolic pathway that utilizes a phosphate-containing product. The \textit{phnO} gene was shown to be non-essential for C-P bond cleavage or Pn utilization as a source of phosphorous. Its gene product was characterized \textit{in vitro} as an acetyl-CoA dependent N-acetyltransferase that is responsible for acetylating aminoAlPn\(^36\). Finally, PhnP was annotated as an accessory protein since its gene disruption does not affect C-P bond cleavage, yet completely prevents cellular growth on Pn, suggesting a critical function in the phosphate product assimilation pathway\(^35\).

The members of this pathway have resisted structural characterization due to the difficulty of purifying these enzymes in soluble form, possibly due to the fact that an entire complex or some of its members may be membrane-associated. Out of the 14 proteins only PhnF, PhnH and PhnP have been structurally characterized to date, with the latter two structures produced by Dr. Jia’s lab. In agreement with sequence alignment data, the structure of PhnF revealed a transcriptional regulator motif\(^37\). PhnH was shown to be an essential enzyme for C-P bond cleavage through mutational analysis. Structural homology searches, cleft analysis and screening multiple libraries of potential ligands did not make the substrate or the catalyzed reaction immediately apparent\(^38\), reflecting the challenges associated with structure-guided
biochemical characterization. PhnP displayed highest sequence homology to tRNases, as well as hydrolytic activity towards a number of phosphodiester substrates. As it was unclear how a tRNase might contribute to bacterial utilization of organophosphonates, we set out to characterize PhnP structurally. The structural information offers subtle clues to its biological activity, though determination of the in vivo reaction still presents a challenge, as described in Chapter 3 and discussed in Chapter 6. The work presented in the first part of this thesis is focused on deciphering the mechanism of Pn utilization through the structural and functional characterization of PhnP, a protein from the Pn utilization pathway of E.coli K12.

1.3 Antimicrobial mechanisms of competition

1.3.1 The role of antibiotic production in resource competition

Natural antimicrobial compounds comprise a wide range of antibiotics and secondary metabolites, such as streptomycin, methylenomycin, acyl homoserine lactones, colicin, antimicrobial peptides, such as nisin and subtilisin\(^4\). Many of these molecules are potent, highly specific toxins that are usually produced during stressful conditions, such as nutrient limitation, and result in the rapid elimination of neighboring cells that are not immune or resistant to their effect. Release of antibiotics as a means of competition for resources is not a straightforward process. If a single microbial cell were to release antimicrobial compounds at random, it would likely be a futile process, since the local extracellular concentrations of these compounds would not be high enough to kill. Basal production of such factors could actually be damaging to the producer, because subinhibitory levels of an antimicrobial could induce a resistance in target species. Therefore, antibiotic-producing organisms face a challenge of delivering a killing dose of antibiotic, yet preventing the acquisition of tolerance in competing species. Bacteria have overcome this dilemma by evolving a social behavior called quorum sensing\(^39\). It involves
delaying the release of antimicrobial agents until a local quorum is achieved, ensuring the presence of sufficient cell numbers. In this case synchronous release of these compounds by a large number of bacteria produces fully inhibitory antimicrobial levels. Consequently, many species of bacteria use quorum sensing for regulation of antimicrobial functions.

Recently, there has been some controversy regarding the ecological role of the compounds that are currently defined as antimicrobial agents. Several groups have observed that subinhibitory concentrations of antimicrobials can trigger or alter the transcriptional responses of sensitive bacteria\textsuperscript{40,41}. It was proposed that the true function of many of these molecules in nature is to act as signal molecules within and between species. These molecules might serve many functions, like altering central metabolic pathways, contributing to nutrient scavenging or participating in developmental pathways. For example, several antimicrobial molecules that can cause K\textsuperscript{+} leakage, have recently been shown to upregulate the formation of extracellular matrix components in \textit{Bacillus subtilis} and thus stimulate biofilm formation\textsuperscript{42}. Similarly, \textit{Pseudomonas aeruginosa} have also been shown to influence gene expression in several bacterial species through production of small redox-active antimicrobial agents\textsuperscript{43}. Many of these naturally produced antimicrobial agents are organohalogenated compounds, and new compounds with antimicrobial activities are still being discovered in marine and terrestrial environments.

\subsection*{1.3.2 Natural halogenated compounds}

The first halometabolite, 3,5-diiodotyrosine, was isolated from the marine eukaryote \textit{Gorgonia cavolinii} in 1896. Later, this compound was also found to be produced in the thyroid gland of mammals\textsuperscript{44}. Only about 30 naturally produced halogenated compounds were known by 1960, but as of the beginning of the 21\textsuperscript{st} century, more than 3800 had been discovered. Chlorinated compounds comprise about 50\% of currently known organohalogens, and are usually
produced by terrestrial plants, fungi, lichen, bacteria, insects, and even some higher animals, including humans. Brominated compounds comprise about 45% of known compounds, owing their production to numerous marine plants, animals and bacteria. Iodine- and fluorine-containing compounds occur less frequently and account for the remaining 5%\textsuperscript{45}. Several hundred natural marine products contain both chlorine and bromine.

Many potent antibiotics have been discovered amongst natural organohalogens. These include chloramphenicol, pyrrolnitrin, chondrochloren, rebecamycin and vancomycin, to name a few. The bacterium *Amycolatopsis orientalis* produces the life-saving glycopeptide antibiotic vancomycin, which has been used for nearly 50 years to treat penicillin-resistant infections\textsuperscript{46}. The potency and selectivity of natural antibiotics and hormones often depends on the key structural features provided by simple halogen substituents. For example, replacing a single chlorine atom on the antibiotic vancomycin with hydrogen results in a 70% decrease of its antibacterial activity\textsuperscript{47}. Unfortunately, the widespread use of these antibiotics results in emergence of bacterial resistance, and modified or new drugs are needed to overcome this problem\textsuperscript{48}. Natural organohalogens are emerging as a rich resource for new drug development.

Integration of halogens into complex molecules through traditional synthetic methods is challenging due to the lack of control over reaction specificity and regioselectivity. In the natural world, this issue is circumvented through evolution of numerous halogenating enzymes and reaction mechanisms that result in halogen incorporation into specific substrates with remarkable regio- and stereoselectivity. Studying halogenating enzymes both from a structural and mechanistic perspective will pave the road for development of novel synthetic methods and production of novel antibiotics with enhanced functions. Such research is vital for staying ahead of the continually emerging resistance to drugs. An example of utilizing enzymes for medical use
includes a fluorinase from the microorganism *Streptomyces cattleya*, which produces $^{18}$F-labeled molecules for use in positron emission tomography (PET).

### 1.3.3 Halogenase overview

Halogenating enzymes can be divided into five classes based on the form of halogenating agent and the cofactor or another group required to produce such an agent. There are hydrogen peroxide (H$_2$O$_2$)-requiring haloperoxidases (heme-dependent and or vanadium-dependent), the oxygen-dependent halogenases (flavin-dependent or non-heme iron-dependent), and the nucleophilic halogenases. Structural information was essential in elucidating the reaction mechanism for all classes of halogenases. Examples of reactions catalyzed by each class of halogenases are shown in Figure 1.5.

Though heme-dependent and vanadium-dependent haloperoxidases are different in structure, both form a metal-bound hydrogen peroxide, which reacts with halide ions to produce a metal-bound hypohalite ion. After dissociation from the metal, the hypohalous acid reacts with substrate in solution. Such halogenation lacks regioselectivity and substrate specificity.

The heme-dependent haloperoxidases catalyze halogenation of aromatic, electron-rich substrates using H$_2$O$_2$ and halide$^{50}$. These enzymes have been found in organisms from bacteria to mammals, and differ in sequence, structure, and the attachment of their heme group. Mammalian peroxidases display covalent attachment of the heme group$^{51}$, while in enzymes from other organisms it is not attached. The general mechanism of heme-dependent haloperoxidases involves oxidation of heme-Fe(III) by H$_2$O$_2$ to produce a $^{\cdot}$heme-Fe(IV)=O species$^{52}$. This compound then oxidizes the halide by two electrons to produce an enzyme-bound heme-Fe(III)-OX intermediate (where X is a halide) (Fig 1.5A). It is not clear whether this electrophilic
Figure 1-5 Examples of halogenation reactions catalyzed by five classes of halogenases. (A) Heme-dependent thyroid haloperoxidase halogenating thyroid hormone; (B) Vanadium-dependent halogenating a generic substrate monochlorodimedone; (C) Flavin-dependent halogenase RebH halogenating tryptophan; (D) Non-heme iron-dependent halogenase SyrB2 acting on peptidyl carrier-bound threonine; (E) Nucleophilic fluorinase acting on AdoMet (Figure from Blasiak and Drennan\textsuperscript{53}).
intermediate is attacked by an electron-rich substrate, or if it forms a free hypohalous acid (HOX or XO) that reacts with the substrate in solution.

Vanadium haloperoxidases have been found in many organisms, including fungi, marine algae, and bacteria. They also catalyze halogenation of aromatic, electron-rich substrates using a vanadate cofactor, H₂O₂, and a halide (Fig 1.5B). H₂O₂ binding to vanadate produces an activated peroxy intermediate, which is then attacked by a halide ion. As with heme-dependent halogenases, there is still some debate as to the nature of halogenating species. It is not clear whether the substrate reacts predominantly with free hypohalous acid or with some form of enzyme-bound V-OX intermediate. The elucidation of the reaction mechanism is hindered by the fact that the exact physiological roles and substrates of most vanadium-dependent haloperoxidases are unknown, and the enzyme activity is often assayed with a generic monochlorodimedone (MCD) substrate.

The first evidence for FADH₂-dependent halogenases emerged in the late 1990’s, and they collectively stand out for their ability to halogenate a wide array of substrates, including indole, pyrrole, phenyl, quinone and alkynyl groups. Several flavin-dependent halogenases have now been characterized, and they can be divided into two main groups: those that catalyze chlorination of free small-molecule substrates (RebH, PrnA), and those that react with substrates tethered to a thiolation domain in a nonribosomal polypeptide synthetase (NRPS) system (PltA, SgcC3, CndH). To date only four halogenases of this class have been characterized structurally. PrnA from Pseudomonas fluorescens, RebH from Lechevalieria aerocolonigenes, and PyrH from Streptomyces rugosporous all catalyze the chlorination of the free amino acid L-tryptophan at two different positions: C7 for PrnA and RebH, and C5 for the PyrH. The structure of CndH, a tyrosyl halogenase from chondrochloren biosynthesis pathway of Chondromyces crocatus, was recently solved, but the substrate for CndH is unknown, as
biosynthesis proceeds through a peptide-carrier bound intermediate. Flavin-dependent halogenases require reduced flavin (FADH$_2$), which is provided by a separate NADH-dependent reductase (Fig 1.5C). Through spectroscopic studies the oxygen has been observed to react with reduced flavin to produce a peroxide intermediate at C4α of the flavin ring$^{61}$. It is then attacked by a Cl$^-$ ion bound near the ring to produce hypohalous acid that diffuses through a hydrophobic tunnel connecting the FAD-binding site with the active site, where a strictly conserved lysine is located. The HOCl proceeds to forms a long-lived chloramine intermediate with this lysine, which then directs a regiospecific chlorination of the substrate$^{62}$.

Non-heme iron and O$_2$-dependent halogenases act on a phosphopantetheine-tethered substrate during natural product biosynthesis by NRPS machinery (Fig 1.5D). These enzymes coordinate Fe (II) using two amino acid side chains, a chloride ion, a water molecule and a bidentate interaction with α-ketogluterate (αKG). Oxygen binding to the iron center leads to oxidative decarboxylation of αKG and formation of an Fe(IV)-oxo intermediate, which abstracts a hydrogen from the substrate. The resulting substrate radical then combines with Cl$^-$ to give the chlorinated product. The use of a radical halogenation mechanism allows non-heme iron enzymes to regioselectively halogenate substrates at unactivated, aliphatic carbons$^{63}$.

Nucleophilic halogenases are not well characterized. The only structural example to date is a 5′-fluoro-5′-deoxyadenosine synthase (5′-FDAS) from Streptomyces cattleya$^{64}$. This fluorinase catalyzes a nucleophilic attack of F$^-$ on AdoMet in an S$_{N}$2-type reaction, where the L-Met moiety makes as an excellent leaving group (Figure 1.5E)

1.3.4 Chloramphenicol biosynthesis pathway

Chloramphenicol (Cm; original name chloromycetin) was discovered in 1947 in soil-dwelling bacteria, which were subsequently named Streptomyces venezuelae, after the origin of
their discovery. In the same year it was recognized as a powerful therapeutic agent with little side effects, protecting test animals from a number of infections, and showing the greatest potency against rickettsialpox infections (typhus)\textsuperscript{65}. Two years later this “wonderdrug” was used for typhus infection treatment as well as prophylactic treatment in Malaysia and other areas where typhus is endemic\textsuperscript{66}. Currently, it is used as an alternative therapy to treat typhoid fever, some forms of meningitis, and rickettsial infections such as Rocky Mountain spotted fever and typhus.

Cm is effective in controlling the spread of bacterial infections because it inhibits ribosomal protein synthesis. The crystal structure of the 50S ribosomal subunit with the antibiotic revealed that chloramphenicol binds to the 23S ribosomal RNA at the peptidyl transferase cavity of this subunit, thereby inhibiting the peptidyl transferase activity\textsuperscript{67}. Drug production by bacteria is not only energetically costly, but also requires production of the drug resistance enzymes for protection from its own antibiotic. \textit{S. venezuelae} produce an enzyme Cm hydrolase that catalyzes the removal of the dichloroacetyl moiety from the antibiotic, rendering it inactive. The deacylated product, \textit{p}-nitrophenylserinol, is metabolized to \textit{p}-nitrobenzyl alcohol and other compounds\textsuperscript{68}.

The development of drug resistance in target species is a natural concern for antibiotic-producing species, and also has implications for human health, where it is a constant race to develop new antibiotics, as infectious bacteria develop resistance to existing antibiotics. For example, in many Gram-positive and Gram-negative bacteria Cm resistance is mediated by the enzyme chloramphenicol acetyltransferase, which inactivates the antibiotic by catalyzing its \textit{O}-acetylation\textsuperscript{69}.

The operon for Cm biosynthesis contains twelve genes\textsuperscript{70}, and a large number of its encoded enzymes still await functional characterization. The Cm precursor has its origins in the shikimate pathway, which produces aromatic metabolites\textsuperscript{71}. The route to Cm branches at chorismic acid to generate \textit{p}-aminophenylalanine (PAPA), which serves as a precursor of the \textit{p}-
nitrophenylserinol component of the antibiotic (Figure 1.6). These initial steps of the pathway were elucidated from the patterns of isotope incorporation into Cm when bacterial cultures were fed labeled substrates, and from the structures of intermediates accumulated by blocked mutants. However, the sequence of events leading to formation of Cm from PAPA is not clear. The β-hydroxylation reaction of PAPA is believed to be carried out by CmlA, an α-ketoglutarate-dependent non-heme iron dioxygenase, while oxidation of the amino group might be carried out by CmlII, an iron-dependent N-oxidase. Finally, an unusual dichloroacetylation reaction of the α-amino group of PAPA must take place for conversion to Cm. Gene knockout studies in *S. venezuelae* have revealed that in both CmlS and CmlK are necessary for the formation of dichloroacetyl group. CmlK has sequence homology to acyl Co-A synthases and would potentially activate the halogenation substrate or product for transfer onto Cm precursor. CmlS shows sequence homology to FADH$_2$-dependent halogenases. Mutants lacking both *cmlK* and *cmlS* were shown to incorporate a propionyl group in place of the dichloroacetyl group, yielding cornyecin II instead of chloramphenicol. The group that undergoes halogenation reaction remains a mystery due to the contradicting results from radio labeling experiments. When [³H]-dichloroacetate was added to *S. venezuelae* cultures, negligible incorporation of radioactive isotopes into Cm was detected. Therefore Simonsen *et al.* ruled out a possibility of free carboxylic acids acting as substrates for chlorination, and suggested that acylation step should precede chlorination. On the other hand Gottlieb *et al.* showed that [*¹⁴C*] atoms from acetate and several amino acids that stimulate production of Cm, end up being incorporated into the dichloroacetyl portion of Cm. To date, the substrate undergoing dichlorination, and the mechanism of introduction of that moiety into the PAPA precursor remains unidentified.

Finally, CmlP is an important player in the final steps of Cm biosynthesis. Its sequence is homologous to nonribosomal peptide synthases, and it has been proposed to act as a carrier protein for a number of pathway intermediates. To date it is unclear how many enzymatic
Figure I-6 (A) Gene cluster for Cm biosynthesis; (B) Proposed Cm biosynthesis pathway.

Intermediates of the top row have been established, while the bottom two rows represent hypothesized sequence of events (Figure is generously provided by Dr. David Zechel).
tailoring steps occur on the substrate tethered to CmlP, and at which point this intermediate (or
the final product) gets released. It was shown that CmlP first activates PAPA via its adenylation
domain, and then attaches PAPA to its peptidyl carrier domain though a thioester linkage on its
phosphopantetheine arm. Interestingly, despite the presence of the reductase domain, no PAPA
release could be shown in this experiment. Patcholec et al. proposed that CmlP acts as a carrier
protein for a number of Cm intermediates, and a number of enzymatic tailoring steps have to
occur on the carrier-bound PAPA, before it can be released by CmlP reductase domain. CmlP
may play an important role of sequestering PAPA from a general metabolic pool and positioning
this precursor for enzymatic tailoring by other members of the Cm pathway77.

The work presented in the second part of the thesis deals with structural characterization
of halogenase CmlS and lays the foundation for further mechanistic analysis. Besides CmlS, there
are no known examples of enzymes introducing halogen into alkyl groups, so elucidation of the
reaction mechanism of CmlS is of great interest.
Chapter 2

Expression, purification and preliminary diffraction studies of PhnP

Preface:

This chapter was published in the journal Acta Crystallographica F:

Expression, purification and preliminary diffraction studies of PhnP. Acta Cryst. F64, 554-557.

Kateryna Podzelinska performed the native and selenomethionine protein expression, purification, crystallization, crystal harvesting and testing, and diffraction data analysis. Shumei He performed initial native and selenomethionine protein expression and purification. Alexei Soares performed the X-ray data collection and initial processing. Bjarne Hove-Jensen provided the expression plasmid for PhnP. This manuscript was written by Kateryna Podzelinska and Dr. David Zechel, with editorial input from Dr. Zongchao Jia.
2.1 Abstract

PhnP belongs to a 14-gene operon that supports the growth of *E. coli* on AlPn as a sole source of phosphorus; however, the exact biochemistry of Pn degradation by this pathway is poorly understood. The protein was recombinantly expressed in *E. coli* and purified to homogeneity. Sitting-drop vapour diffusion in combination with microseeding was used to obtain crystals that were suitable for X-ray diffraction. Data were collected to 1.3 Å resolution and the crystals exhibit the symmetry space group *C*2, with unit-cell parameters \( a = 111.65, b = 75.41, c = 83.23 \) Å, \( \alpha = \gamma = 90, \beta = 126.3^\circ \).
2.2 Introduction

The availability of free phosphate in nature can be a limiting factor for bacterial growth. Bacteria have adapted by obtaining phosphate from phosphate esters and organophosphonates, the latter of which are characterized by a direct carbon–phosphorus (C–P) bond that is exceptionally stable compared with the carbon–oxygen–phosphorus (C–O–P) bond of phosphate esters. During phosphate limitation, *E. coli* exhibits a many hundred-fold upregulation of the C–P lyase pathway. The pathway has broad substrate specificity and is the only means of Pn breakdown in the K-12 strain.\(^{78}\)

The C–P lyase operon consists of 14 genes (*phnCDEFGHIJKLMNOP*). Based on mutational and sequence analysis, the gene products PhnC–PhnE are thought to be involved in transport of substrates,\(^ {31}\) PhnF is a transcription repressor protein,\(^ {37}\) PhnO is an aminoalkylphosphonate N-acetyltransferase,\(^ {36}\) PhnG–PhnM are enzymes involved in C–P bond cleavage and PhnN is a ribose 1,5-bisphosphokinase,\(^ {34}\) Although the *phnP* gene is conserved in the prokaryotic C–P lyase operon, its precise role has not been determined. Mutational studies have suggested that PhnP is required in the presence of a functional *phnN* product for Pn utilization,\(^ {35}\) Attempts to detect C–P lyase activity in a cell-free system have been inconsistent, possibly owing to the fact that the enzyme complex is predicted to be membrane-associated.\(^ {29,79}\) Consequently, the reaction mechanism of enzymatic C–P bond cleavage is poorly understood.

In order to determine the role of PhnP in the C–P lyase pathway, structural studies have been undertaken. Here, we present the expression, purification, crystallization and preliminary X-ray diffraction analysis of PhnP.
2.3 Materials and methods

2.3.1 Cloning, expression and purification

The phnP gene was amplified directly from *E. coli* HO764 using Vent DNA polymerase (New England Biolabs) and the forward and reverse primers 5’- GAGAATTCAATTAAGAGG-AGAAATTAACTATGAGCCTGACCCTCAGCTCACCAGACGCGCGGCGG-3’ and 5’-TGGT-TGGATCCCCGAGCCCATGGTTATTAATGGTGATGGTGATGGTGCGCCACCCCAATC-TCCATCCCATCAAACCCC-3’, respectively. *Eco*RI and *Bam*HI restriction endonucleases sites are italicized and the starting ATG and hexahistidine codons are shown in bold. The PCR product was digested by *Eco*RI and *Bam*HI and ligated into similarly digested DNA of pUHE23-280. Results were confirmed by DNA sequencing. The plasmid was designated pHO520. The NCBI accession number for the phnP nucleotide and protein sequences is 948600.

PhnP was expressed by co-transformation of pHO520 and pLacI (Novagen) into *E. coli* BL21 (DE3) cells (Novagen). The pLacI vector was necessary to provide sufficient lactose repressor to control transcription from pHO520, which lacks a *lacI* gene. Cells were then grown in Luria-Bertani medium with ampicillin (100 mg ml⁻¹) and chloramphenicol (30 mg ml⁻¹) until the A₆₀₀ of the culture reached 0.6 at 303 K. At this point, the protein expression was induced with 0.5 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG). After further incubation at 288 K for 20 h, the cells were harvested by centrifugation at 3000 g at 277 K for 20 min and stored at 253 K until purification.

Cell pellets were resuspended in 50 mM phosphate pH 7.2, 10 mM imidazole, 300 mM NaCl and lysed with an Emulsiflex-C5 homogenizer (Avestin). The lysed cells were then centrifuged at 40 000g and 277 K for 30 min. The soluble fraction of the lysate was passed through a nickel–nitrilotriacetic acid (Ni–NTA) agarose column (Qiagen) and the captured PhnP was eluted with an imidazole gradient from 10 to 500 mM over ten column volumes at a flow rate
of 5 ml min⁻¹ using an ÄKTA FPLC system (GE Healthcare). Fractions containing >95% pure PhnP, as demonstrated by SDS–PAGE analysis, were pooled, concentrated to 1 ml and further purified by size exclusion using a Superdex 200 column (1 x 30 cm, prep-grade; GE Healthcare) pre-equilibrated with 50 mM HEPES pH 7.2, 150 mM NaCl. SDS-PAGE images following both purification steps and a size exclusion chromatogram are shown in the Section 1 of the Appendix. The pooled fractions from the size exclusion column were concentrated using Millipore Amicon Ultra centrifugal filters (10 000 Da molecular-weight cutoff); this was followed by passage through a 0.22 mm syringe filter. The concentration of purified protein was determined from the absorbance at 280 nm using the calculated extinction coefficient $\varepsilon_{280}^{1\%} = 1.18$ and a molecular weight of 28.67 kDa. The protein was flash-frozen in liquid nitrogen and stored at 193 K.

The selenomethionine derivative of PhnP was produced in the methionine-auxotroph strain DL41 (DE3) grown in M9 SeMET High Yield medium (Medicilon). The selenomethionine-labelled PhnP was purified using the same procedure as described above for the native protein.

### 2.3.2 Crystallization

All crystallization experiments were performed at room temperature. For initial screening, the protein sample was diluted to 360 mM (~ 10 mg ml⁻¹) in protein buffer (20 mM HEPES pH 7.5, 150 mM NaCl). Sitting-drop vapour-diffusion trials were carried out in 96-well plates (Greiner): 1 µl protein solution was mixed with 1 µl crystallization solution and equilibrated against 100 µl well solution. Initial hits were obtained in condition Nos. 38–41 of the PACT Suite (Qiagen) containing 0.1 M MMT buffer pH 5.0–8.0 and 25%(w/v) PEG 1500. MMT is a composite buffer that consists of 37 mM MES, 20 mM L-malic acid and 43 mM Tris.
To further improve the homogeneity of the PhnP preparation, the purified protein was incubated at 315 K for 30 min and precipitate was removed by centrifugation. The protein concentration was adjusted to 280 mM (~8 mg ml\(^{-1}\)); 50 mM L-arginine and 50 mM L-glutamate pH 7.6 were added to the protein sample 1 h prior to crystallization. The protein sample was mixed with crystallization solution in a 1:1 ratio (2 + 2 µl) and placed on a microbridge (Hampton) containing 20 µl Fluorinert (Hampton). The microbridge was placed in a well of a VDX plate containing 1 ml crystallization solution and sealed with a glass cover slip.

The crystallization conditions were optimized using a grid screen of various molecular-weight PEGs versus the pH of the MMT buffer. Addition of PEG 400 was found to prevent crystal cracking upon immersion in cryoprotectant solution prior to flash-freezing. The final crystallization conditions contained 0.1 M MMT pH 5.2–5.8, 8% w/v PEG 8000 and 6% v/v PEG 400. Crystals appeared overnight and reached their maximum size within 48 h. Crystal quality was further improved by the standard streak-seeding technique using a horse hair and a 1/10 dilution of microseed stock. Selenomethionine-derivative crystals of PhnP were obtained using the same procedure as for the native crystals.

2.3.3 Data collection and diffraction measurements

Single-wavelength anomalous dispersion data were collected from SeMet PhnP crystals on the X12B beamline at the National Synchrotron Light Source Brookhaven National Laboratory using an ADSC Quantum-4 CCD detector. All data were collected at 100 K. Prior to flash-freezing in liquid nitrogen, the crystals were sequentially immersed in crystallization solution containing 5%, 15% and 25% v/v 2-methyl-2,4-pentanediol (MPD) as a cryoprotectant. Oscillations of 1°, with an exposure time of 6 s per image, and a crystal-to-detector distance of 200 mm were used. The data were indexed and scaled using DENZO and SCALEPACK\(^2\).
2.4 Results and discussion

The recombinant PhnP protein was purified to homogeneity using a two-step purification protocol. After size-exclusion chromatography, PhnP was found to be approximately 95% pure, as confirmed by SDS–PAGE.

PhnP initially crystallized overnight in condition Nos. 38–41 from the PACT Suite (Qiagen; Figure 2.1a). Expansion of these conditions using the hanging-drop vapour-diffusion method in 24-well screw-cap plates (Qiagen) produced crystals that had split ends and were too small for diffraction studies (less than 0.05 mm in the longest axis; Figure 2.1b). An exhaustive search for compounds that would improve the crystal size was conducted using Additive and Detergent Screens 1, 2 and 3 (Hampton), but no consistent change in crystal size and quality was observed. It has been reported that the addition of 50 mM L-glutamate and L-arginine greatly improves protein solubility and long-term stability. The effect of these amino acids on protein crystallization was assessed by adding them to PhnP samples immediately after thawing and at least 1 h prior to crystallization trials. Surprisingly, the crystal size increased to about 0.2–0.35 mm along the longest axis (Figure 2.1c). Most crystals still possessed internal defects, however, and produced only a few low-resolution spots in X-ray diffraction patterns.

In an attempt to rid the PhnP crystals of internal defects and prevent nucleation along the contact area of the drop with the cover slip, which made it difficult to harvest the crystals, the floating-drop method utilizing microbridges with Fluorinert was used. The crystallization drops did not float on top of the Fluorinert, but sank to the bottom of the microbridge, resembling a microbatch experiment. The use of Fluorinert did not remedy the nucleation problem, but dramatically improved the crystal size and quality (Figure 2.1d). This was attributed to slower equilibration of the drop with the reservoir and hence slower crystal growth. The standard streak-seeding technique was implemented in order to minimize spontaneous nucleation from the
Figure 2-1 (a) Initial crystals obtained from PACT Suite condition Nos. 38–41: 0.1 M MMT buffer pH 5.0–8.0 and 25%(w/v) PEG 1500. Crystals produced during the optimization process in (b) the absence and (c) the presence of 50 mM L-Asp and L-Glu in the protein buffer. (d) Optimized crystals were obtained in 0.1 M MMT pH 5.2–5.8, 8% w/v PEG 8000 and 6% v/v PEG 400.
contact area between the drop and the microbridge.

Short heat treatments can improve protein homogeneity by denaturing partially unfolded molecules\textsuperscript{84}. To further improve the quality of the crystals, purified PhnP was incubated at 315 K for 30 min prior to crystallization in order to precipitate less stable protein. The resulting large and flawless crystals produced diffraction to 1.3 Å resolution (Figure 2.2). The typical crystal had dimensions of 0.3 x 0.2 x 0.2 mm. The crystals exhibited symmetry of the C-centred monoclinic space group \(\text{C}_2\), with unit-cell parameters \(a = 111.65\), \(b = 75.41\), \(c = 83.23\) Å, \(\alpha = \gamma = 90\), \(\beta = 126.3^\circ\). Diffraction data for PhnP were processed in the resolution range 60–1.4 Å. A Matthews coefficient of 2.469 Å\(^3\) Da\(^{-1}\) was obtained\textsuperscript{85}, with a solvent content of 50.1%, representing a dimer in the asymmetric unit. This is consistent with size exclusion results, which indicate that the protein is dimeric in solution (data shown in Chapter 3, Figure 3-5). A summary of the crystal parameters and the statistics of the diffraction data are presented in Table 2.1. Structure determination of PhnP is presented in Chapter 3.
Figure 2-2 Diffraction of SeMet PhnP crystals. The edge of the detector is 1.3 Å; data were processed to 1.4 Å.
Table 2-1 Diffraction data for Se-Met PhnP crystals. Values in parentheses are for the highest resolution shell (1.46 - 1.40 Å).

<table>
<thead>
<tr>
<th>Space Group</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit Cell Parameters (Å, °)</td>
<td>$a = 111.65$, $b = 75.41$, $c = 83.23$ Å, $\beta = 126.3^\circ$</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.97916</td>
</tr>
<tr>
<td>Temperature (K)</td>
<td>100</td>
</tr>
<tr>
<td>Resolution Range (Å)</td>
<td>67.1 - 1.4</td>
</tr>
<tr>
<td>Observed Reflections</td>
<td>1628375</td>
</tr>
<tr>
<td>Unique Reflections</td>
<td>109380</td>
</tr>
<tr>
<td>Data Completeness (%)</td>
<td>94.5 (70.4)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>7.1 (5.1)</td>
</tr>
<tr>
<td>$R_{sym}$ (%)</td>
<td>8.3 (25.7)</td>
</tr>
<tr>
<td>$&lt;I/\sigma I&gt;$</td>
<td>30.5 (5.5)</td>
</tr>
<tr>
<td>Matthews coefficient (Å$^3$ Da$^{-1}$)</td>
<td>2.469</td>
</tr>
<tr>
<td>Solvent content (%)</td>
<td>50.1</td>
</tr>
</tbody>
</table>

*R_{sym} = \Sigma I(k) - <I>/\Sigma I(k), where I(k) and <I> represent the diffraction-intensity values of the individual measurements and the corresponding mean values. The summation is over all measurements.*
Chapter 3
Structure of PhnP, a phosphodiesterase of the carbon-phosphorus lyase pathway for phosphonate degradation

Preface:

This chapter was published in The Journal of Biological Chemistry:


Kateryna Podzelinska was responsible for structure solution, refinement and analysis of PhnP, as well as conducting an RNase assay. ICP-MS analysis was performed independently by Kateryna Podzelinska and Shumei He, though data from the latter source is shown in this manuscript. Shumei He conducted a TLC experiment, metal ion dependence and kinetic characterization of PhnP. Matthew Wathier assisted Shumei He with various parts of her experiments. Screening of PhnP against a phosphodiesterase substrate library was conducted by Michael Proudfoot and Alexander Yakunin. Bjarne Hove-Jensen provided the expression plasmid for PhnP. Also, his expertise and insightful discussions on PhnP involvement in the C-P lyase pathway are greatly appreciated. The manuscript was written by Dr. David Zechel and Kateryna Podzelinska, with editorial input from Dr. Zongchao Jia.
3.1 Abstract

C-P lyase is a multienzyme system encoded by the phn operon that enables bacteria to metabolize organophosphonates when the preferred nutrient, inorganic phosphate, is scarce. One of the enzymes encoded by this operon, PhnP, is predicted by sequence homology to be a metal-dependent hydrolase of the β-lactamase superfamily. Screening with a wide array of hydrolytically sensitive substrates indicated that PhnP is an enzyme with phosphodiesterase activity, having the greatest specificity toward bis(p-nitrophenyl) phosphate and 2’,3’-cyclic nucleotides. No activity was observed toward RNA. The metal ion dependence of PhnP with bis(p-nitrophenyl) phosphate as substrate revealed a distinct preference for Mn\(^{2+}\) and Ni\(^{2+}\) for catalysis, whereas Zn\(^{2+}\) afforded poor activity. The three-dimensional structure of PhnP was solved by X-ray crystallography to 1.4 Å resolution. The overall fold of PhnP is very similar to that of the tRNase Z endonucleases but lacks the long exosite module used by these enzymes to bind their tRNA substrates. The active site of PhnP contains what are probably two Mn\(^{2+}\) ions surrounded by an array of active site residues that are identical to those observed in the tRNase Z enzymes. A second, remote Zn\(^{2+}\) binding site is also observed, composed of a set of cysteine and histidine residues that are strictly conserved in the PhnP family. This second metal ion site appears to stabilize a structural motif.
3.2 Introduction

In many environments inorganic phosphate, an essential nutrient, can fall to extremely low concentrations, forcing microorganisms to utilize other forms of phosphorus to survive. In such cases, organophosphonates can comprise a major fraction of the total phosphorus available to biological systems (e.g. AEPn is a widespread natural product). However, cleavage of the highly stable C-P bond to release inorganic phosphate requires specialized enzymes. One such enzyme activity found widely in bacteria is C-P-lyase\textsuperscript{18}. Cleavage of the C-P bond of organophosphonates by C-P-lyase yields inorganic phosphate and, remarkably, a hydrocarbon. C-P-lyase is actually a multienzyme system, encoded by the \textit{phn} operon (\textit{phnCDEFGHJKLMNOP}), which is induced by low concentrations of phosphate as part of the \textit{pho} regulon. Gene deletion studies in \textit{E. coli} have shown that \textit{phnGHIJKLM} are essential for catalysis of C-P bond cleavage, whereas the remaining genes probably encode transport, regulatory, or accessory functions\textsuperscript{26}. Only a handful of the proteins encoded by the \textit{phn} operon have been characterized to date. PhnD was shown to be a periplasmic binding protein with high affinity for organophosphonates\textsuperscript{31}; the three-dimensional structure of PhnH, one of the proteins essential for C-P-lyase catalysis, was recently solved, but a function has yet to be determined\textsuperscript{38}; PhnN was shown to be an ATP-dependent kinase that provides a redundant pathway to 5-phospho-D-ribofuranosyl-\textalpha{}-1-diphosphate\textsuperscript{34}; and PhnO was demonstrated to be an acetyl-CoA-dependent \textit{N}-acyltransferase with activity toward a wide range of aminoalkylphosphonates\textsuperscript{36}.

Although the \textit{phnP} gene is not essential for C-P bond cleavage by stationary cells in liquid culture\textsuperscript{26}, cell growth on solid media supplemented with MePn or phosphite as the sole phosphorus source is prevented by \textit{phnP} mutations\textsuperscript{35}, suggesting a critical regulatory or accessory role for PhnP. Accordingly, \textit{phnP} appears frequently in the \textit{phn} operon in various species of bacteria, typically following the \textit{phnN} gene\textsuperscript{86}. PhnP is predicted, based on its sequence, to be a
member of the β-lactamase family of metal-dependent hydrolases with greatest homology to enzymes from the tRNase Z (ProDom family PD352433) and ElaC families, the latter erroneously annotated as composed of arylsulfatases but later determined to also belong to the tRNase Z family. The tRNase Z enzymes are endonucleases used by prokaryotes and eukaryotes to cleave a specific phosphodiester bond near the 3’-end of pre-tRNA, yielding a 3’-end that can be coupled to an amino acid. These enzymes typically use two active site bound Zn ions to simultaneously lower the pKₐ of a nucleophilic water molecule and stabilize negative charge development on the phosphodiester linkage undergoing nucleophilic attack. Since it is not clear how a tRNase activity would support cell growth with an organophosphonates as a sole phosphorus source, we set out to characterize the substrate specificity and three-dimensional structure of PhnP to learn more about this critical C-P-lyase enzyme.

### 3.3 Experimental procedures

#### 3.3.1 Materials

All buffers, substrates, and chemicals, unless otherwise noted, were purchased from Sigma-Aldrich Canada Ltd. (Oakville, Canada). DNA oligonucleotides were synthesized by Sigma-Genosys Canada (Oakville, Canada). Microbiology media were purchased from Thermo-Fisher Scientific Canada Ltd. (Ottawa, Canada). DNA sequencing was performed at the Robarts Research Institute (London, Canada). Calf intestinal alkaline phosphatase was obtained from New England Biolabs. RNase A was from Ambion. J. T. Baker, Inc. brand polyethyleneimine-cellulose TLC plates were obtained from Mallinckrodt Baker.
3.3.2 Expression, purification, and size exclusion chromatography of PhnP and mutants

The cloning and expression of *phnP* as well as the purification and crystallization of wild type PhnP were described in Chapter 2. PhnP mutants were created using the QuikChange™ protocol (Stratagene). For the D80A mutant, the primer 5’- ACGCATTATCATATGGCTACGC-TCCAGGGGCTG- 3’ and its complementary sequence were used. For the C21S/C23S/C26S triple mutant, the primer 5’-CGGCATGGGGCTCCGGAGTCTGCGGCCTCCGCCAGAGCGCG-3’ and its complimentary sequence were used. Mutagenic codons are underlined. The mutant genes were verified by sequencing both DNA strands. The mutant alleles were expressed and the mutant proteins were purified in the same fashion as wild type PhnP, with the exception of the triple mutant (C21S/C23S/C26S), which yielded insoluble protein (see below). The oligomeric state of PhnP was determined by size exclusion chromatography on a calibrated Superdex 200 column (GE Healthcare), as described previously.

3.3.3 Substrate screening

PhnP activity was assayed against a panel of naturally occurring phosphodiesters and phosphoanhydrides (all obtained from Sigma), as described previously. Each reaction (80 µl), containing 50 mM Tricine hydrochloride (pH 8.5), 5mM MgCl₂, 0.5mM each MnCl₂, NiCl₂, and CoCl₂, 0.25 mM substrate, and 2 µg of PhnP, was incubated at 37 °C for 20 min and then stopped by the addition of 80 µl of 0.2 M CHES, 10 mM MgCl₂, pH 9.0. The reactions were then incubated with 1 unit of shrimp alkaline phosphatase for 10 min at 37 °C, followed by the addition of 40 µl of malachite green reagent and measurement of the absorbance at 630 nm to determine production of orthophosphate.

3.3.4 Steady-state kinetic analysis

Reactions of PhnP with *bis(p-nitrophenyl) phosphate (bpNPP), thymidine-5’-monophosphate-p-nitrophenyl ester (pNP-TMP)*, or 2’,3’-cyclic nucleotides were performed in 50
mM Tris-HCl (pH 7.2), 150 mM NaCl, divalent metal (0.1 mM ZnCl₂, 1 mM MnCl₂, or 1 mM (NiCl₂) and 1 mg/ml bovine serum albumin at 298 K. Release of p-nitrophenolate was monitored at 405 nm ($\Delta \varepsilon_{405} = 11,500 \text{ M}^{-1} \text{cm}^{-1}$) on a Cary Bio-300 spectrophotometer (Varian). Phosphodiesterase activity toward 2',3'-cyclic nucleotides was measured as follows. A 400-µl solution of substrate, 1 mM MnCl₂, 50mM Tris-HCl (pH 7.2), and 150mM NaCl was reacted with 1.5 µM PhnP for 20 min and then stopped with 400 µl of 0.2 M Tris-HCl, 10 mM MgCl₂, pH9.0. One unit of calf intestinal alkaline phosphatase was added, and the reaction was incubated at 310 K for 10 min. Finally, 200 µl of malachite green reagent was added, and free phosphate was determined colorimetrically at 630 nm using an extinction coefficient of $\Delta \varepsilon_{630} = 90,000 \text{ M}^{-1} \text{cm}^{-1}$. Duplicate reactions containing no PhnP were performed to measure residual inorganic phosphate. The kinetic parameters $k_{\text{cat}}$ and $K_m$ were determined by fitting the dependence of the initial reaction velocities on the substrate concentration with the Michaelis-Menten or Hill equations using GraFit 6 (Erithacus Software Ltd.).

3.3.5 RNase activity assay

The activity of PhnP and the D80A mutant was assayed with the RNaseAlert Lab Test Kit from Ambion (catalogue number 1964). RNase-free pipette tips, plasticware, and water were used for the assay. A 50 µl reaction containing 10 ng of PhnP, 20 mM HEPES (pH 7.5), 150 mM NaCl, 10 μM MnCl₂, and RNA substrate was incubated at 310 K for 1 h. Control reactions contained 0.2 pg of RNase A or buffer only. Cleavage of the RNA oligonucleotide was monitored by fluorescence at $\lambda_{\text{ex}} = 490 \text{ nm}$ and $\lambda_{\text{em}} = 520 \text{ nm}$ using a SpectraMax Gemini XS spectrofluorimeter (Molecular Devices, Sunnyvale, CA).
3.3.6 ICP-MS analysis

The metal content of wild type PhnP was analyzed using a Varian ICP-MS (Ultra Mass). PhnP was dialyzed into 10 mM HEPES, pH 7.2, and diluted to 10 µM for analysis. Metal content was determined by comparison with standard curves of Mn²⁺ and Zn²⁺ ranging from 0 to 1000 ppm. Measurements were performed in triplicate and corrected for residual metal content by analyzing buffer alone.

3.3.7 Metal dependence of PhnP activity

Metal-free “apo-PhnP” was obtained by incubation with 10 mM EDTA for 1 h at 277 K, followed by dialysis into 50mM Tris-HCl, 150mM NaCl, pH 7.2. Apo-PhnP was then incubated with different metal ions by the addition of 0.2 mM ZnCl₂, MnCl₂, NiCl₂, CoCl₂, CaCl₂, FeCl₂, FeCl₃, or MgCl₂. After 1 h of incubation at 277 K, an aliquot of each PhnP sample was added to a reaction containing 2 mM bpNPP in 50 mM Tris-HCl, 150 mM NaCl, 1 mg/ml bovine serum albumin, pH 7.2, equilibrated at 298 K (the final concentration of PhnP was 2 µM). Initial rates were measured by monitoring p-nitrophenolate production as before. PhnP reconstituted with Zn²⁺ was “rescued” by incubation with 10 mM EDTA a second time followed by dialysis. After incubation with 0.2 mM Mn²⁺ as before, PhnP (2 µM) was assayed with bpNPP. The apparent Kₘ values (Kₘ(app)) for Mn²⁺ and Ni²⁺ were obtained by an assay of wild type apo-PhnP or the D80A mutant with 1 mM bpNPP and various concentrations of the metal ion. The resulting curve of initial rates versus metal ion concentration was fit to the Michaelis-Menten equation using GraFit 6.0.

3.3.8 TLC analysis

A reaction containing 700 µl of 50 mM Tris (pH 7.2), 150 mM NaCl, 2 mM MnCl₂, 10 mM 2’,3’-cAMP, and 26 µM wild type PhnP was incubated for 7 h at 294 K. A control reaction
lacking PhnP was incubated under the same conditions. Aliquots (5 µl) of these reactions along with standards of 3’-AMP and 2’-AMP (10 mM each in the above reaction buffer lacking PhnP) were spotted onto a polyethyleneimine-cellulose TLC plate (5 x 20 cm, polyester-backed, with a fluorescent indicator). Once the aliquots had dried, the TLC plate was washed in methanol (MeOH), dried, and then developed in saturated ammonium sulfate (pH 3.5), as described by Bochner. After washing the plate again in MeOH, the dried plate was visualized under UV light (254 nm) and photographed with a standard digital camera.

### 3.3.9 PhnP structure solution and refinement

Diffraction data were collected as described in Chapter 2. Data were indexed, integrated, and scaled using DENZO and SCALEPACK. Initial phases were determined by single wavelength anomalous dispersion at the selenium peak energy using the program autoSHARP, and 10 of the 12 expected selenium atoms in the asymmetric unit were located. The initial model was built automatically using autoSHARP and completed manually in XFIT/XTALVIEW. Structure refinement was performed using REFMAC5. The final model contained a dimer in the asymmetric unit, 882 water molecules, four manganese ions, two zinc ions, and two (S)-malate molecules. The structure factors and atomic coordinates determined for PhnP in this study have been deposited in the Protein Data Bank (PDB) under accession number 3G1P.

### 3.4 Results

#### 3.4.1 Substrate specificity and product analysis

The amino acid sequence of *E. coli* PhnP shows significant homology to metal ion-dependent enzymes that comprise the β-lactamase superfamily (Figure 3.1). The best homology is found with enzymes that hydrolyze phosphodiester bonds, particularly the tRNase Z
endonucleases (21 and 24% sequence identity, respectively, with *E. coli* ZipD and *B. subtilis* tRNase Z), where all residues involved in active site metal ion binding are strictly conserved (Figure 3.1). Accordingly, *E. coli* PhnP showed high activity against the generic phosphodiesterase substrate bpNPP. The pH optimum for this reaction was 7.2 (data not shown). PhnP was inactive against the corresponding monoester, *p*-nitrophenyl phosphate. To identify a potential *in vivo* substrate, PhnP was screened for hydrolytic activity against a wide array of naturally occurring phosphodiesters and phosphoanhydrides in the presence of a mixture of metal ions (Mg$^{2+}$, Mn$^{2+}$, and Ni$^{2+}$) (Figure 3.2). Highest activity was observed with 2’,3’-cyclic nucleotides, whereas virtually no activity was observed with the corresponding 3’,5’-cyclic nucleotides. The reaction with 2’,3’-nucleotides is regiospecific, with only the 3’-nucleotide product being observed by TLC (Figure 3.3). The low apparent activity observed in the screen with FAD, 3’,5’-cIMP, and UDP-GlcA could not be detected by direct initial rate kinetics and is probably due to free phosphate in the substrate preparations.

PhnP also did not display activity against single-stranded RNA beyond background hydrolysis (specific activity of 2.8 x 10$^3$ RFU min$^{-1}$ mg$^{-1}$). This was confirmed when the D80A mutant of PhnP, which is virtually inactive against bpNPP (see Table 3.1), afforded essentially the same low specific activity against the RNA substrate (4.2 x 10$^3$ RFU min$^{-1}$ mg$^{-1}$) as the wild type enzyme. This is consistent with the background hydrolysis arising from another enzyme, probably a constant, trace level of RNase in our PhnP preparations. In contrast, a positive control with RNase A showed specific activity of 2.2 x 1013 RFU min 1 mg 1 toward this single-stranded RNA substrate.
Figure 3-1 Multiple sequence alignment of PhnP homologues. Amino acid sequences are for *E. coli* K-12 PhnP (gi: 536936), *Pseudomonas stutzeri* PhnP (gi: 40804950), *Pyrococcus horikoshii* OT3 PhnP (gi: 14591382), *Marinobacter aquaeolei* VT8 PhnP (gi: 120555210), *Pseudomonas putida* PqqB (gi: 56967240; Protein Data Bank code 1xto), *E. coli* ZipD (gi: 90109091; Protein Data Bank code 2cbn), and *B. subtilis* tRNase Z (gi: 60594108; Protein Data Bank code 1y44). The percentage of sequence identity to *E. coli* PhnP is shown in parentheses. α-Helices and β-strands observed in the *E. coli* PhnP structure are indicated as cylinders and block arrows, respectively. Residues involved in binding the two active site metal ions are highlighted in red. The putative general acid catalyst (GAC) is highlighted in turquoise. Residues involved in binding the structural zinc ion are highlighted in yellow. The exosite of the tRNase Z enzymes involved in binding tRNA is underlined. Strictly, highly and moderately conserved residues are indicated by asterisks, colons, and periods, respectively. The sequence alignment was performed using ClustalW and then edited manually.
Figure 3-2 Screening PhnP hydrolytic activity against a series of phosphate diesters and phosphoanhydrides. PhnP (2 µg) was incubated with each substrate (0.25 mM) in the presence of a mixture of Mg$^{2+}$ (5 mM), Mn$^{2+}$, Ni$^{2+}$, and Co$^{2+}$ (0.5 mM each) in pH 8.5 buffer for 20 min at 37 °C. After treatment with alkaline phosphatase to hydrolyze any monoester product generated by PhnP, the liberated orthophosphate was quantified by absorbance at 630 nm with a malachite green assay. Assay conditions are described under “Experimental Procedures.”
Figure 3-3 TLC (polyethyleneimine-cellulose) analysis of the reaction product of PhnP with 2’,3’-cAMP. Lanes 1 and 2, standards of 2’-AMP and 3’-AMP, respectively (5 µl each from 10 mM solutions in reaction buffer). Lane 3, an aliquot (5 µl) of the PhnP reaction with 2’,3’-cAMP (26 µM PhnP, 10 mM 2’,3’-cAMP in reaction buffer, incubated at 21 °C for 7 h). Lane 4, an aliquot (5 µl) of 2’,3’-cAMP (10 mM in reaction buffer) incubated under the same conditions as the PhnP reaction. Reaction conditions and TLC development are described under “Experimental Procedures.”
3.4.2 Metal ion dependence of PhnP

ICP-MS analysis of PhnP revealed nearly stoichiometric amounts of Zn$^{2+}$ (1.38/monomer) and low levels of Mn$^{2+}$ (0.13/monomer; Table 3.2). To determine which of these metals was essential for catalysis, PhnP was treated with EDTA and then dialyzed against metal free buffer. Although it was possible to reduce the Mn$^{2+}$ content to background levels (as measured by ICP-MS), nearly stoichiometric amounts of Zn$^{2+}$ (0.78/monomer) remained (Table 3.2), suggesting the presence of a high affinity Zn$^{2+}$ binding site on the enzyme.

The Mn$^{2+}$-free form of PhnP was essentially inactive against bpNPP (Table 3.1), whereas incubation of PhnP with 200 $\mu$M Mn$^{2+}$ or Ni$^{2+}$ restored substantial activity (5,000 – 13,000-fold). Zinc ion restored considerably less activity (~200-fold) and in fact appeared to be inhibitory, since equimolar mixtures of Zn$^{2+}$ and Mn$^{2+}$ or Ni$^{2+}$ likewise only restored ~140–190-fold activity. However, substantial activity (1500-fold greater than apo-PhnP) could be restored by treating the Zn$^{2+}$-reconstituted PhnP with EDTA and then reincubating with Mn$^{2+}$ (Table 3.3).

The reaction rate of apo-PhnP with bpNPP as a function of metal ion concentration affords an apparent $K_m$ value ($K_{m(app)}$) of 0.13 ± 0.02 mM for Mn$^{2+}$ and 0.090 ± 0.02 mM for Ni$^{2+}$ (Table 3.1), indicating that at least one metal ion occupies a site of relatively low affinity. Further evidence for Mn$^{2+}$ bound in the active site was obtained by mutating D80, one of the conserved metal ion ligands. The PhnP D80A mutant, in addition to being largely inactive, experienced a 100-fold increase in $K_{m(app)}$ for Mn$^{2+}$ (13 ± 2 mM; Table 3.1). Consistent with this dramatically reduced affinity for the metal ion, the D80A mutant no longer co-purified with Mn$^{2+}$, as indicated by ICP-MS (Table 3.2), whereas the amount of Zn$^{2+}$ bound did not change significantly. More detailed kinetic analysis of PhnP using bpNPP as substrate and saturating metal ion concentrations confirmed that considerably greater $k_{cat}$ and $k_{cat}/K_m$ values are achieved with Ni$^{2+}$ and Mn$^{2+}$ than with Zn$^{2+}$ (Table 3.1). Overall, these data suggest that the binuclear active site of
Table 3-1 Kinetic parameters of PhnP

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Substrate</th>
<th>$k_{cat}$ ($s^{-1}$)</th>
<th>$K_M$ (mM)</th>
<th>$k_{cat}/K_M$ (s$^{-1}$ M$^{-1}$)</th>
<th>$K_{m}^{app}$ metal ion (mM)$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>2',3'-cAMP</td>
<td>1.71 ± 0.06</td>
<td>0.11 ±0.01</td>
<td>15.8 x 10$^3$</td>
<td>ND$^b$</td>
</tr>
<tr>
<td>WT</td>
<td>2',3'-cGMP</td>
<td>1.23 ± 0.07</td>
<td>0.31 ±0.05</td>
<td>3.91 x 10$^3$</td>
<td>ND</td>
</tr>
<tr>
<td>WT</td>
<td>2',3'-cCMP</td>
<td>0.64 ± 0.03</td>
<td>0.14 ±0.03</td>
<td>4.22 x 10$^3$</td>
<td>ND</td>
</tr>
<tr>
<td>WT</td>
<td>pNP-TMP</td>
<td>0.08 ± 0.02</td>
<td>15 ± 5</td>
<td>5.3</td>
<td>ND</td>
</tr>
<tr>
<td>WT</td>
<td>bpNPP</td>
<td>1.19 ± 0.08</td>
<td>2.9 ± 0.5</td>
<td>410</td>
<td>0.13 ± 0.02</td>
</tr>
<tr>
<td>WT</td>
<td>bpNPP (1 mM Ni$^{2+}$)</td>
<td>4.1 ± 0.2</td>
<td>2.3 ± 0.3</td>
<td>1770</td>
<td>0.09 ± 0.02</td>
</tr>
<tr>
<td>WT</td>
<td>bpNPP (0.1 mM Zn$^{2+}$)$^c$</td>
<td>0.030 ± ± 0.001</td>
<td>0.68 ± 0.07</td>
<td>44</td>
<td>ND</td>
</tr>
<tr>
<td>D80A</td>
<td>bpNPP</td>
<td>8.3 x 10$^{-4}$</td>
<td>3.7 ± 0.4</td>
<td>0.22</td>
<td>13 ± 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>± 0.4 x 10$^{-4}$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$Apparent metal ion $K_m$ values for apoenzymes measured with 1 mM bpNPP. Reaction conditions are described under “Experimental Procedures”.

$^b$ND, not determined

$^c$Zn$^{2+}$ concentrations above 0.1 mM inhibited PhnP activity.
Table 3-2 ICP-MS analysis of PhnP.

<table>
<thead>
<tr>
<th>PhnP sample</th>
<th>[Mn$^{2+}$] (µM)</th>
<th>[Zn$^{2+}$] (µM)</th>
<th>Mn$^{2+}$/PhnP monomer</th>
<th>Zn$^{2+}$/PhnP monomer</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT as purified</td>
<td>1.3 ± 0.1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>13.8 ± 1.1</td>
<td>0.13</td>
<td>1.38</td>
</tr>
<tr>
<td>WT EDTA treated</td>
<td>0.06 ± 0.01</td>
<td>7.8 ± 0.03</td>
<td>0.01</td>
<td>0.78</td>
</tr>
<tr>
<td>D80A</td>
<td>ND&lt;sup&gt;b&lt;/sup&gt;</td>
<td>10.5 ± 0.2</td>
<td>0</td>
<td>1.05</td>
</tr>
</tbody>
</table>

<sup>a</sup> S.D. values based on the average of three samples.

<sup>b</sup> Not detectable; value indistinguishable from that obtained with buffer alone.
Table 3-3 Metal ion dependence of PhnP.

<table>
<thead>
<tr>
<th>Sample</th>
<th>( V_o/[E]_T )(min(^{-1}))</th>
<th>Relative rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apo PhnP</td>
<td>0.004</td>
<td>(1)</td>
</tr>
<tr>
<td>Mn(^{2+})</td>
<td>24.3</td>
<td>5600</td>
</tr>
<tr>
<td>Ni(^{2+})</td>
<td>58.7</td>
<td>13500</td>
</tr>
<tr>
<td>Mn(^{2+})/Ni(^{2+})</td>
<td>46.1</td>
<td>10600</td>
</tr>
<tr>
<td>Zn(^{2+})</td>
<td>0.891</td>
<td>205</td>
</tr>
<tr>
<td>Zn(^{2+})/Mn(^{2+})</td>
<td>0.600</td>
<td>138</td>
</tr>
<tr>
<td>Zn(^{2+})/Ni(^{2+})</td>
<td>0.817</td>
<td>188</td>
</tr>
<tr>
<td>(1) Zn(^{2+}), (2) EDTA, (3) Mn(^{2+})</td>
<td>6.57</td>
<td>1510</td>
</tr>
<tr>
<td>Co(^{2+})</td>
<td>0.557</td>
<td>128</td>
</tr>
<tr>
<td>Fe(^{2+})</td>
<td>0.561</td>
<td>129</td>
</tr>
<tr>
<td>Fe(^{3+})</td>
<td>0.261</td>
<td>60</td>
</tr>
<tr>
<td>Cu(^{2+})</td>
<td>0.013</td>
<td>3.0</td>
</tr>
<tr>
<td>Mg(^{2+})</td>
<td>0.022</td>
<td>5.0</td>
</tr>
<tr>
<td>Ca(^{2+})</td>
<td>0.026</td>
<td>6.0</td>
</tr>
</tbody>
</table>

\(^a\) Metal-free apo-PhnP was obtained by incubation with EDTA followed by dialysis. apo-PhnP was incubated with a 0.2 mM concentration of the metal ions listed above then assayed with bpNPP, as described under “Experimental Procedures.”

\(^b\) Specific activities of bpNPP hydrolysis for metal-reconstituted PhnP and 2 mM bpNPP at 25 °C. \( V_o \) = initial rate; \([E]_T\) = total enzyme concentration.

\(^c\) Apo-PhnP incubated with Zn\(^{2+}\) was stripped with EDTA, dialyzed, incubated with Mn\(^{2+}\), and then assayed with bpNPP.
PhnP has greater activity with Mn$^{2+}$ and Ni$^{2+}$ ions and that a single Zn$^{2+}$ ion is bound at a separate, high affinity site.

3.4.3 Kinetic analysis of wild type PhnP

The initial substrate screening results were confirmed by more detailed kinetic analysis (Table 3.1). In the presence of saturating Mn$^{2+}$ (1 mM), PhnP has greatest $k_{cat}/K_m$ values with 2',3'-cAMP, 2',3'-cCMP and 2',3'-cGMP followed by bpNPP. Analysis of the activity at low bpNPP concentrations followed by fitting of the data to the Hill equation revealed modest cooperativity ($n_H = 1.55 \pm 0.04$), similar to that observed with *E. coli* ZipD ($n_H = 1.6$). This cooperativity was not observed with the 2',3'-cyclic nucleotides. Very low activity was also observed with pNP-TMP ($k_{cat}/K_m = 5.3 \text{ M}^{-1}\text{ s}^{-1}$), another general phosphodiesterase substrate. The greater specificity of PhnP toward the 2',3'-cyclic nucleotides is manifested almost entirely by a drop in the value of $K_m$ (110–310 µM) relative to bpNPP ($K_m = 2.9 \pm 0.5 \text{ mM}$), suggesting greater recognition for these substrates in the ground state (in the absence of kinetically significant enzyme-substrate intermediates).

3.4.4 The crystal structure of PhnP

The three-dimensional structure of PhnP from *E. coli* K12 was determined to 1.4 Å resolution using the single-wavelength anomalous dispersion method. The asymmetric unit of the PhnP crystal contained a dimer (see Figure 3.4A). This is probably representative of the oligomeric state of PhnP in solution, as opposed to crystal packing. The PhnP monomer has a predicted molecular mass of 28.67 kDa (confirmed by MALDI-MS; data not shown), whereas PhnP eluted from a size exclusion column with an apparent molecular mass of 44.5 kDa (Figure 3.5), suggesting the formation of a compact dimer in solution. Of 258 residues, molecule A contains residues 2–250, including all of the side chains, whereas molecule B contains residues 3–250. Since only the peptide backbone density was visible for residues 2 and 251 in molecule B,
Figure 3-4 The crystal structure of dimeric PhnP. (A) PhnP dimer. Subunit A is in magenta, and subunit B is in blue. Malate is in yellow, Mn$^{2+}$ ions are in orange, and Zn$^{2+}$ ion is in red. (B) GRASP representation of the PhnP dimer with (S)-malate shown in both active sites. (C) The active site with (S)-malate. Mn$^{2+}$ ions are shown in orange.
Figure 3-5 Size exclusion chromatogram of PhnP (monomer MW = 28.67 kDa). The void volume and elution volumes of selected protein standards (alcohol dehydrogenase, 150 kDa; bovine serum albumin, 66 kDa; carbonic anhydrase, 29 kDa; cytochrome c, 12.4 kDa) are indicated with black triangles (▼). Inset is a plot of log MW for protein standards vs the ratio of elution volume (Ve) to void volume (Vo). The linear fit yields a slope of -0.90 ± 0.06 and a y-intercept of 3.7 ± 0.1 (correlation coefficient = -0.994). The data point marked as an open square (□) corresponds to dimeric PhnP (57.3 kDa) based on the observed Ve/Vo = 2.28.
they were refined as alanine and glycine, respectively. There was no clear density present for the C-terminal hexahistidine tag in either of the subunits (residues 253–258). The model was refined to an $R$-factor of 18.6% and an $R_{\text{free}}$ value of 21.0%, with 882 molecules of water (Table 3.4). The model possesses excellent geometry, where 87.2% of the residues fall in the most favored regions of Ramachandran plot, 10.9% in additional allowed regions, and 1.9% in generously allowed regions, as determined by the PDB validation procedure.

The PhnP fold belongs to the $\alpha/\beta$ class of proteins and falls into the metallo-hydrolase superfamily. The monomer core consists of two mixed $\beta$-sheets that are sandwiched between two layers of $\alpha$-helices. The smaller sheet contains six strands, the first three parallel ($\beta_5$–$\beta_3$) and the next three antiparallel ($\beta_2$, $\beta_1$, and $\beta_{13}$). The larger sheet contains seven strands, the first four of which are parallel ($\beta_{12}$–$\beta_9$); the remaining three sheets are antiparallel ($\beta_{8}$–$\beta_6$). Strands 6 and 7 are joined by a two-residue Type I’ $\beta$-hairpin, whereas strands 2 and 3 are connected by a Type II’ $\beta$-hairpin. The dimerization interface has a buried surface area of 17.4% and is formed by $\alpha$-helices 3 and 4 and portions of loops 1, 3, and 5–10. The interactions are primarily hydrophobic and mainly result from the W19 of loop 1 of each monomer extending far into the hydrophobic pocket of the other monomer. The two H58 side chains from loop 5 of each monomer are in the off-centered parallel orientation, creating a $\pi$-stacking interaction. The dimer is further stabilized through three salt bridges as well as 10 hydrogen-bonding pairs. The dimerization of PhnP results in formation of a deep cleft on the surface of each monomer close to the dimerization interface (Figure 3.4B). Each cleft contains the binuclear metal ion active site (Figure 3.4C). The second monomer contributes R89, W90, D108, D109, and H113 to the cleft formation. The two active site clefts are separated by about 27 Å measured from a more solvent-exposed metal ion of each monomer) and a 90° rotation along the long axis of the dimer. The root mean square deviation value for C-\(\alpha\) atoms the two subunits is 0.3 Å.
Table 3-4 Summary of data collection and refinement statistics (SeMet PhnP)\(^a\).

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space Group</td>
<td>C2</td>
</tr>
<tr>
<td>Unit Cell Parameters</td>
<td>(a = 111.65, b = 75.41, c = 83.23 \text{ Å}, \beta = 126.3^\circ)</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.97916</td>
</tr>
<tr>
<td>Resolution Range (Å)</td>
<td>67.1 - 1.4</td>
</tr>
<tr>
<td>Observed Reflections</td>
<td>1628375</td>
</tr>
<tr>
<td>Unique Reflections</td>
<td>109380</td>
</tr>
<tr>
<td>Data Completeness (%)</td>
<td>94.5 (70.4)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>7.1 (5.1)</td>
</tr>
<tr>
<td>(R_{\text{sym}}) (%)(^b)</td>
<td>8.3 (25.7)</td>
</tr>
<tr>
<td>(&lt;</td>
<td>I</td>
</tr>
<tr>
<td>(R_{\text{work}}) (%)(^c)</td>
<td>18.6 (22.3)</td>
</tr>
<tr>
<td>(R_{\text{free}}) (%)</td>
<td>21.0 (26.1)</td>
</tr>
<tr>
<td>No. of observations (total)</td>
<td>98155 (4914)</td>
</tr>
<tr>
<td>No. of observations for (R_{\text{free}})</td>
<td>5171 (281)</td>
</tr>
<tr>
<td>Root mean square deviations</td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.009</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.32</td>
</tr>
<tr>
<td>Mean temperature factor (Å(^2))</td>
<td>18.8</td>
</tr>
<tr>
<td>No. of protein residues</td>
<td>499</td>
</tr>
<tr>
<td>No. of protein atoms</td>
<td>3883</td>
</tr>
<tr>
<td>No. of water atoms</td>
<td>882</td>
</tr>
<tr>
<td>No. of metal ions</td>
<td>6</td>
</tr>
<tr>
<td>No. of S-malate molecules</td>
<td>2</td>
</tr>
<tr>
<td>Ramachandran statistics (%)</td>
<td></td>
</tr>
<tr>
<td>Most favoured regions</td>
<td>87.2</td>
</tr>
<tr>
<td>Allowed regions</td>
<td>10.9</td>
</tr>
<tr>
<td>Generously allowed regions</td>
<td>1.9</td>
</tr>
<tr>
<td>Disallowed regions</td>
<td>0</td>
</tr>
</tbody>
</table>

\(^a\)The values in parentheses are data for the high-resolution shell (1.435 - 1.399 Å).

\(^b\)\(R_{\text{sym}} = \frac{\Sigma|I(k)| - \langle|I|\rangle\Sigma|I(k)|}{\Sigma|I(k)|}\), where \(I(k)\) and \(\langle|I|\rangle\) represent the diffraction-intensity values of the individual measurements and the corresponding mean values. The summation is over all measurements.

\(^c\)\(R_{\text{work}} = \frac{\Sigma||F_o|| - ||F_c||\Sigma|F_o|}{\Sigma||F_o||\Sigma|F_o|}\) and \(R_{\text{free}} = \frac{\Sigma||F_{o\text{free}}|| - ||F_{c\text{free}}||\Sigma|F_{o\text{free}}|}{\Sigma||F_{o\text{free}}||\Sigma|F_{o\text{free}}|}\), where \(F_o\) is the observed structure factor and \(F_c\) is the calculated structure factor based on the model. No sigma cut-off was applied to the data; 5% of reflections were excluded from the refinement for calculation of \(R_{\text{free}}\).
3.4.5 Structural homology to tRNase Z endonucleases

A search of the Protein Data Bank using DALI revealed that PhnP has high structural homology to metal-dependent hydrolases of the β-lactamase superfamily, particularly the tRNase Z endonucleases (Figure 3.6A). Using root mean square deviation values based on the least squares superimposition of the structurally equivalent C-α atoms, the nearest tRNase Z homologue is the *Bacillus subtilis* enzyme (Protein Data Bank code 1y44) with a root mean square deviation of 2.7 Å (Z score = 23). The tRNases are also homodimers, and the active site residues used for coordination of two active site metal ions are strictly conserved with PhnP (Figure 3.6B). However, several profound structural differences are also observed. The characteristic long exosite used for pre-tRNA binding by *E. coli* ZipD and *B. subtilis* tRNase Z is absent in PhnP (Figure 3.6A). PhnP, like *E. coli* ZipD, possesses fully metal-loaded active sites in both monomers, whereas only one monomer is metal-loaded in the *B. subtilis* enzyme. This is due to a dramatic conformational change between two monomers in the *B. subtilis* tRNase Z, where one monomer has a distorted active site, whereas the other one lacks a resolved exosite but retains a functional active site. H140 and H247 in *B. subtilis* tRNase Z (H143 and H200 in PhnP, respectively) move far out of position in the “inactive” monomer, which prevents metal binding. It has been suggested that tRNA binding to the inactivated monomer causes a conformational change and subsequent activation of the second active site, which would result in cooperative behavior. Another feature that distinguishes PhnP from the tRNase Z hydrolases is the presence of an additional α-helix containing the second metal ion binding site (see below).

3.4.6 The Zn$^{2+}$ binding site

Unlike other members of the β-lactamase family, the PhnP subfamily possesses three strictly conserved cysteines near the N terminus (Figure 3.1). Such sulfur rich binding sites typically bind Zn$^{2+}$ ions, and this is clearly observed in the structure of PhnP (Figure 3.6C). This
Figure 3-6 A comparison of PhnP with close structural homologues. (A) PhnP compared with *P. putida* PqqB (PDB ID: 1xto) and *B. subtilis* tRNase Z (PDB ID: 1y44). (B) Alignment of tRNase Z (magenta) and PhnP (cyan) active site residues. Zinc ions observed in the tRNase active site are shown in red, Mn$^{2+}$ ions observed in the PhnP active site are in orange. PhnP residues are indicated in bold, tRNase Z residues are given in brackets. (C) Alignment of PqqB (slate) and PhnP (cyan) residues comprising the structural Zn$^{2+}$ ion site. PhnP Zn$^{2+}$ ion is in red, PqqB Zn$^{2+}$ ion in raspberry. PhnP residues are indicated in bold, PqqB residues are given in brackets.
metal-binding site is located at the edge of the monomer next to the dimerization interface. A tetrahedral coordination sphere for single Zn$^{2+}$ ion is formed by C19, C21, C23, and H225 residues. The cysteine residues are contributed by loop 1 and α-helix 1, whereas H225 is contributed by α-helix 8. Helices 1 and 2 are flanked by loop 3 and a long extended region of loop 1. This stretch of secondary structure forms a lobe that is tethered to the main body of the protein through hydrophobic interactions with helix 8 and is further stabilized by two intraprotein salt bridges and several hydrogen bonds. Loops 1 and 3 also provide the residues for two of three interprotein salt bridges as well as residues that form the majority of the interprotein hydrogen bonds. As mentioned previously, loop 1 also contains W19, which forms extensive hydrophobic interactions with the deep pocket created by residues of helices 3 and 4 and loops 6 and 8 of the other monomer. Therefore, the area around the Zn$^{2+}$ site is responsible for providing the majority of protein-protein interaction, and its integrity is crucial for the overall stability of the dimer. The structural role of this site is supported by the observation that simultaneous alteration of all three cysteines to serines (C21S/C23S/C26S) produced an insoluble protein.

3.4.7 Structural homology to PqqB

Intriguingly, one of the closest structural homologues to PhnP, based on a DALI search of the Protein Data Bank, is PqqB (Figure 3.6A; Protein Data Bank code 1xto), an enzyme that appears to be involved in the transport of an intermediate in the pyrroloquinoline quinone biosynthetic pathway\textsuperscript{100,101}. Despite the low sequence identity of 22%, PhnP and PqqB monomer structures align with a root mean square deviation of 2.7 Å and Z-score of 24.7. PhnP and PqqB share the Zn$^{2+}$ binding site and structural motif, which is also strictly conserved in the PqqB family of enzymes (Figure 3.1). The arrangement of the scaffold and the three coordinating cysteine residues is virtually identical between the two proteins, whereas the fourth residue is N272 in PqqB rather than histidine as in PhnP (Figure 3.6C). In contrast, there is only moderate
sequence and structural observation between PhnP and PqqB active sites, and the latter does not have metal ions bound (Figure 3.1).

3.4.8 The PhnP active site

The active site in PhnP is located at the loop aggregation area at the edge of the β-sandwich at the dimerization interface (Figure 3.4B). Density was observed for two metal ions in the active site, surrounded by residues that are conserved among PhnP homologues and known to bind metal ions (Figure 3.1). At a resolution of 1.4 Å, the difference density for the metal ions at this site was noticeably smaller than the one found for Zn$^{2+}$ at the cysteine site described above, suggesting the presence of a lighter metal ion. Although 0.13 Mn$^{2+}$/monomer co-purified with PhnP (Table 3.2), this does not necessarily represent the metal ion occupancy in the crystal, since a metal enriched form of the enzyme may have been selectively crystallized. Combining this observation with the co-purification of Mn$^{2+}$ with wild type PhnP, the reduced affinity for Mn$^{2+}$ in the active site mutant D80A (Tables 3.1 and 3.2), and the distinct preference for Mn$^{2+}$ over Zn$^{2+}$ for activity (Tables 3.1 and 3.3), the active site metals were assigned as two Mn$^{2+}$ ions. Distances between metal ions and coordinating residues are summarized in Table 3.5. The metal ions are 3.5 Å apart, and are located about 19 Å away from the Zn$^{2+}$ ion of the same monomer.

Surprisingly, a molecule of (S)-malate was stereoselectively sequestered from 0.1 M racemic malate present in the crystallization buffer (MMT buffer; Qiagen). The (S)-malate molecule binds in a bidendate fashion to the more solvent-exposed MnA ion (Figure 3.7A). Metal ions are labeled as described by Vogel et al.$^{93}$ The MnA ion has octahedral coordination geometry, with axial bonds provided by the (S)-malate α-carboxyl and H76, whereas equatorial bonds are provided by H78, H143, D164, and the (S)-malate hydroxyl. The coordination geometry of less solvent-exposed MnB ion is distorted octahedral. The equatorial bonds are provided by D80, D164, H222, and the (S)-malate hydroxyl, whereas the axial bonds are formed with H81 and a
Table 3-5 Distances between PhnP active site residues, metal ions, and S-malate.

<table>
<thead>
<tr>
<th>Metal ion / malate</th>
<th>Residue / malate (Atom)</th>
<th>Distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MnA</td>
<td>H78 (ND1)</td>
<td>2.2</td>
</tr>
<tr>
<td>MnA</td>
<td>H143 (NE2)</td>
<td>2.2</td>
</tr>
<tr>
<td>MnA</td>
<td>H76 (NE2)</td>
<td>2.2</td>
</tr>
<tr>
<td>MnA</td>
<td>D164 (OD2)</td>
<td>2.3</td>
</tr>
<tr>
<td>MnA</td>
<td>Malate hydroxyl</td>
<td>2.2</td>
</tr>
<tr>
<td>MnA</td>
<td>Malate α-carboxyl</td>
<td>2.2</td>
</tr>
<tr>
<td>MnB</td>
<td>D80 (OD2)</td>
<td>3.1</td>
</tr>
<tr>
<td>MnB</td>
<td>H222 (NE2)</td>
<td>2.5</td>
</tr>
<tr>
<td>MnB</td>
<td>H81 (NE2)</td>
<td>2.3</td>
</tr>
<tr>
<td>MnB</td>
<td>D164 (OD2)</td>
<td>2.0</td>
</tr>
<tr>
<td>MnB</td>
<td>Malate hydroxyl</td>
<td>3.1</td>
</tr>
<tr>
<td>MnB</td>
<td>Water</td>
<td>1.6</td>
</tr>
<tr>
<td>Malate α-carboxyl</td>
<td>H200 (NE2)</td>
<td>2.8 / 3.1</td>
</tr>
<tr>
<td>Malate hydroxyl</td>
<td>D80 (OD1)</td>
<td>2.6</td>
</tr>
<tr>
<td>Malate δ-carboxyl</td>
<td>R89 (NH2)</td>
<td>3.9</td>
</tr>
<tr>
<td>Malate δ-carboxyl</td>
<td>D109 (OD2)</td>
<td>3.7</td>
</tr>
<tr>
<td>Malate δ-carboxyl</td>
<td>D108 (OD2)</td>
<td>4.7</td>
</tr>
</tbody>
</table>
Figure 3-7 Active site of PhnP. (A) Interaction of S-malate with Mn$^{2+}$ ions and D80. A bound water molecule is shown in blue. The difference density for Mn$^{2+}$ ions at 5σ level is shown as red mesh. (B) Interactions of S-malate with H200 and adjacent monomer (residues are shown in gray). The difference density is contoured at 5σ level. Distances between interacting groups are given in Table 3.5.
water molecule, which is located 1.6 Å above the MnB ion. A significant B-factor increase (~ 3 times) was observed for MnB ion compared with the MnA ion in both monomers, indicating higher mobility of the former. This probably reflects differing affinities for the two metal ions, which have been observed by ITC for *E. coli* ZipD and related β-lactamase family enzymes. The hydroxyl group of (S)-malate is not equidistant from each Mn$^{2+}$ due to the bidentate interaction with the MnA ion; the distance to the more solvent exposed MnA ion is 2.2 Å, whereas the distance to the more buried MnB ion is 3.1 Å (Table 5). The malate hydroxyl makes an intriguing interaction with D80, a conserved residue critical for PhnP catalysis (Table 3.1). This hydroxyl remains protonated, despite its Lewis acidic environment and thus is able to form a short 2.6-Å hydrogen bond to OD1 of D80, which in turn maintains a weak ligand interaction with MnA that is 3.1 Å away from OD2. The ionized α-carboxyl group of (S)-malate also forms an ionic hydrogen bond (2.8 Å) with NE2 of a protonated H200 (Figure 3.7B). An ionized D187 in turn stabilizes H200 through an ionic hydrogen bond (2.6 Å). H200 is strictly conserved in tRNase Z endonucleases (His247 of *B. subtilis* tRNase Z and H248 of *E. coli* ZipD) and has been observed to interact with inorganic phosphate bound in the active site of the *B. subtilis* enzyme analogous to the (S)-malate interaction observed with PhnP. Notably, (S)-malate bound in one monomer makes additional contacts with Arg89, D108, and D109 of the other monomer using its second carboxyl group (Figure 3.7B; distances of 3.9, 4.7, and 3.7 Å, respectively). Although probably not a physiological substrate, the interaction of (S)-malate with the residues of the second monomer might be similar to the one provided by the actual substrate, where substrate binding to one monomer may affect a conformational change in the dimer that confers higher affinity for the substrate in the second active site. These interactions could account for the modest cooperativity PhnP showed toward bpNPP.
3.5 Discussion

The hallmark of the β-lactamase family of hydrolases is the use of a pair of active site metal ions as Lewis acid catalysts. The metal ions are thought to simultaneously polarize the P=O or C=O bonds of their respective substrates and lower the pKa of the attacking water molecule, which is typically sandwiched between the two metals. A range of metal ions are utilized in the β-lactamase family, most commonly Zn$^{2+}$, Fe$^{2+}$, and Mn$^{2+}$, with some individual enzymes displaying activity with all three of these metal ions$^{104}$, and mixed metal pairs$^{105}$. PhnP is notable in that it has a distinct preference for Ni$^{2+}$ and Mn$^{2+}$ ions for hydrolysis of bpNPP, whereas Zn$^{2+}$ affords considerably lower activity. In contrast, tRNase Z endonucleases do not show a marked change in activity against bpNPP with Mn$^{2+}$ (a minor 3-fold increase was reported for one enzyme)$^{106}$. Rather, despite sharing the same active site residues as PhnP, Zn$^{2+}$ appears to be the active metal ion for the tRNase Z enzymes$^{93,98,106}$ (Figure 3.6B). Oddly, Mn$^{2+}$ does dramatically enhance activity of these enzymes against more complex tRNA substrates$^{106,107}$. In this case, it is thought that Mn$^{2+}$ mediates RNA folding into a hydrolytically sensitive conformation or mediates binding to the enzyme itself. However, in the case of PhnP, a “chaperone” role between Mn$^{2+}$ and a small substrate like bpNPP is an unlikely reason for enhanced activity with this metal ion. Likewise, the distinct electron density difference observed between the Zn$^{2+}$ site on PhnP (a convenient internal control) and the active site metals argues for a lighter metal, such as Mn$^{2+}$, bound in the active site and supporting catalysis. The possibility of “second shell” side chains that modulate the hardness of the metal binding site has been put forward to account for the metal binding preference of Salmonella typhimurium glyoxylase II$^{104}$, but here too there is considerable conservation between PhnP and the Zn$^{2+}$-dependent tRNase Z enzymes. Evidently there is subtle plasticity that dictates metal ion specificity in the β-lactamase family.

It is certainly possible that in vivo Zn$^{2+}$ may serve as the active metal ion in PhnP, since cytoplasmic Zn$^{2+}$ concentrations are maintained at 45 µM in E. coli$^{108}$, which would afford a low
level of activity (Table 3.1) sufficient for cell growth. However, *E. coli* also has a dedicated Mn\(^{2+}\) transport system\(^{109}\) and can achieve cytoplasmic levels of this metal ion well into the 10\(^{-4}\) M range\(^{110,111}\). This would match or exceed the apparent \(K_m\) for this metal with PhnP (Table 3.1) and would stimulate much greater activity. Interestingly, Mn\(^{2+}\) levels are typically highest in stationary, slowly growing bacteria that are nutrient deprived\(^{111}\), which would be the case when expression of the *phn* operon is increased in response to low phosphate levels. This may provide another level of control of PhnP activity in phosphate-starved cells.

The high resolution structure of PhnP fortuitously complexed with (S)-malate provides excellent insight into the catalytic features of the active site, particularly into the roles of D80 and H200. (S)-malate binds to the more solvent-exposed manganese (MnA) in a bidentate fashion, forcing the hydroxyl closer to MnA than MnB (Figure 3.7A and Table 3.5). Remarkably, the OD2 oxygen of D80 appears to “follow” this hydroxyl in order to maintain a short hydrogen bond (2.6 Å), seemingly in preference to OD1 forming a close ligand interaction to MnB, which is 3.1 Å distant (for comparison, the metal-bridging OD2 of D164 is 2.0 and 2.3 Å from MnB and MnA, respectively). The malate hydroxyl appears to mimic the attacking water molecule (or hydroxide) that distinguishes the \(\beta\)-lactamase family of hydrolases\(^{112}\). The close interaction of D80 with this hydroxyl, even in the presence of a Lewis acidic metal ion (MnB), illustrates its potential to participate in general base catalysis or positioning of a nucleophilic water (or hydroxide). The importance of D80 to catalysis is highlighted by its strict conservation in the \(\beta\)-lactamase family. Studies with \(\beta\)-lactamases have suggested primarily a metal ion binding role for this residue, possibly combined with positioning or general base catalysis\(^{113,114,115}\). Likewise, PhnP D80A shows reduced Mn\(^{2+}\) affinity and a great loss of activity (Tables 3.1 and 3.2). One would expect a close interaction of D80 with the MnB ion to lower its \(pK_a\) and impair its suitability as a general base or hydrogen bond acceptor. The fact that a preferential hydrogen bond is formed by D80 in the presence of a competing metal ion suggests that the Lewis acidity
of MnB is either dampened by its other ligands, such as a closer interaction by D164, or that D80 is physically constrained from approaching MnB more closely.

The (S)-malate complex also reveals a potential role for H200 in stabilizing negative charge on a phosphodiester substrate. The large decreases in $k_{cat}$ upon mutation of this residue in these enzymes$^{93,107,116}$ indicate that most of this interaction with a phosphodiester substrate takes place in the transition state. This might arise from stabilization of increasing negative charge development on the nonbridging oxygens in a phosphorane transition state or general acid-catalyzed proton transfer to the leaving group oxygen.

Although PhnP shares close overall structural and active site homology with tRNase Z endonucleases, there are a number of notable differences. In addition to its metal ion preferences, PhnP lacks the distinctive tRNA binding exosite as well as activity against RNA. However, PhnP does exhibit regiospecific activity against 2',3'-cyclic nucleotides. The *E. coli* tRNase ZipD is not active against 2',3'- or 3',5'-cyclic nucleotides$^{88}$ but will cleave short sequences of unstructured RNA$^{117}$. The noncatalytic Zn$^{2+}$ binding site of PhnP is one of the striking features of this structure, which, intriguingly, is shared by another “accessory” protein, PqqB, of the pyrroloquinoline quinone biosynthetic pathway. PqqB does not appear to play a direct catalytic role in the synthesis of pyrroloquinoline quinone$^{100,118}$. Deletion of the pqqB gene in this pathway does not prevent synthesis of pyrroloquinoline quinone but instead leads to accumulation of a biosynthetic intermediate$^{101}$. For this reason, PqqB is suggested to facilitate transport of the final product pyrroloquinoline quinone or an intermediate across the cytosolic membrane to the periplasm and thereby alleviate product inhibition of PqqC$^{100}$. This echoes observations of the importance of PhnP in organophosphonate degradation. Although disruption of the *phnP* gene in *E. coli* does not prevent C-P bond cleavage by stationary cells in liquid culture$^{26}$, cell growth on solid media supplemented with MePn or phosphite as a sole phosphorus source is prevented$^{35}$. 
Interestingly, simultaneous disruption of *phnN* and *phnP* allows weak growth on solid media, suggesting that PhnP is only essential when active PhnN is present\(^{15}\). The product of the PhnN catalyzed reaction is 5-phospho-D-ribofuranosyl-α-1-diphosphate, a glycosyl donor used in the biosynthesis of purine, pyrimidine, and pyridine nucleotides\(^ {119}\).

It is not clear how the phosphodiesterase activity of PhnP relates to the PhnN-catalyzed reaction. Nevertheless, the degradation of cyclic nucleotides appears to be a highly conserved activity in the C-P-lyase pathway. A survey in the SEED data base (available on the World Wide Web)\(^ {120}\) of 54 bacterial *phn* operons containing *phnM* (an essential gene for C-P bond cleavage) revealed 27 occurrences of *phnP*. Intriguingly, in the 16 operons where *phnP* was absent, the gene *rcsF* was present in its stead (the remaining 11 operons contained neither *phnP* nor *rcsF*). The *rcsF* gene product (DUF1045, pfam06299) belongs to the 2H-phosphodiesterase superfamily and is uniquely associated with *phn* operons\(^ {121}\). This family of phosphodiesterases hydrolyze 2’,3’-cyclic nucleotides or ribosyl-1’,2’-cyclic phosphates as part of tRNA splicing reactions and signal transduction. However, unlike PhnP, these enzymes do not employ active site metals and instead use two histidines as general acid-base catalysts to cleave phosphodiester bonds. It is also noteworthy that the *phnP* and *rcsF* genes almost always occur together with *phnN* (only three *phn* operons of the 54 examined above contained a *phnN* gene without *phnP* or *rcsF*). Analogous to PqqB, PhnP (or RcsF) may be involved in transport or processing of an intermediate of organophosphonate metabolism that contains a cyclic phosphate diester or hydrolysis of a 2’,3’-cyclic nucleotide as part of a signaling pathway. The latter is a distinct possibility, since phosphate starvation in *E. coli* (and other bacteria) leads to the production of the “alarmones” guanosine 3’,5’-bis(diphosphate) (ppGpp) and pppGpp in a RelA ((p)ppGpp synthase)-dependent (but SpoT ((p)ppGpp phosphohydrolase)-independent) fashion as part of the bacterial “stringent response”\(^ {122}\). These alarmones are believed to induce expression of genes of the pho regulon (induction of *phoA* and *pstS* have been directly observed)\(^ {123}\), of which the *phn* operon is a
member. Intriguingly, guanosine 5’-diphosphate 2’,3’-cyclic monophosphate (ppG2’,3’p) was observed in crystal structures of SpoT\textsuperscript{124} and adenylosuccinate synthetase\textsuperscript{125}. In both cases, ppG2’,3’p is observed to bind to these enzymes in an inhibitory fashion. Since adenylosuccinate synthetase is an essential enzyme for the synthesis of AMP, its inhibition results in reduced cell growth, probably as a mechanism for bacterial cells to conserve resources under nutrient-limiting conditions. Likewise, inhibition of SpoT by ppG2’,3’p would prevent the hydrolase activity of this enzyme from degrading ppGpp and halting the stringent response. However, once the expression of the \textit{phn} operon is induced and local organophosphonates are degraded at a rate sufficient to meet the phosphate demands of the cell, it will become necessary to degrade ppG2’,3’p. The 2’,3’-cyclic phosphodiesterase activity of PhnP may provide this mechanism.
Chapter 4
Expression, purification and preliminary diffraction studies of CmlS

Preface:

This chapter was published in the journal Acta Crystallographica F:


Ryan Latimer performed native and selenomethionine protein expression and purification, as well as crystallization under Kateryna Podzelinska’s guidance. Kateryna Podzelinska performed crystal harvesting and testing, as well as diffraction data analysis. Alexei Soares performed X-ray diffraction data collection and initial processing. Anupam Bhattacharya developed the protocol for soluble CmlS purification. Leo Vining provided the initial CmlS construct. This paper was written by Kateryna Podzelinska and Dr. David Zechel, with editorial input from Dr. Zongchao Jia.
4.1 Abstract

CmlS, a flavin-dependent halogenase (FDH) present in the chloramphenicol biosynthetic pathway in *Streptomyces venezuelae*, directs the dichlorination of an acetyl group. The reaction mechanism of CmlS is of considerable interest as it will help to explain how the FDH family can halogenate a wide range of substrates through a common mechanism. The protein has been recombinantly expressed in *E. coli* and purified to homogeneity. The hanging-drop vapour-diffusion method was used to produce crystals that were suitable for X-ray diffraction. Data were collected to 2.0 Å resolution. The crystals exhibit the symmetry of space group $C_2$, with unit cell parameters $a = 208.09$, $b = 57.74$, $c = 59.88$ Å, $\beta = 97.47^\circ$. 
4.2 Introduction

Naturally produced organohalogens often display potent bioactivities and accordingly serve as a rich source of new drugs \(^{45}\). The first enzymes shown to regiospecifically catalyze halogenation were the flavin-dependent halogenases (FDHs) \(^{126}\). Since this seminal discovery, a number of other enzymes that catalyze regiospecific and stereospecific halogenation have been discovered \(^{53,127}\). The FDHs have received particular attention since they are capable of halogenating a diverse array of natural products. The structural characterization of FDHs is still in its infancy, with only four structures known to date: PrnA \(^{128}\), RebH \(^{62}\), *Shewanella frigidimarina* halogenase (PDB code 2pyx) and CndH \(^{60}\). The first three of these enzymes chlorinate tryptophan, yielding 7-chlorotryptophan, whilst the recently characterized CndH chlorinates the ortho position of a phenol ring during the biosynthesis of chondrochloren. Mechanistic studies have shown that the flavin cofactor of FDHs generates HOCl, which is believed to either form a stable chloroamine intermediate \(^{62}\) or hydrogen bond \(^{129}\) to a conserved K residue in the active site (Figure 4.1). The K residue in turn directs regiospecific chlorination of the substrate indole ring through an electrophilic aromatic substitution (EAS) reaction. Interestingly, the residues that are proposed to stabilize the carbocation intermediate \(^{128}\) are not conserved in FDH homologues (Figure 4.1), although many of these enzymes also catalyze EAS reactions \(^{60}\). Therefore, a crucial question is how the FDH family adopts a conserved halogenation machinery to react with a remarkable array of substrates such as indole \(^{128,62}\), pyrrole \(^{130}\), quinone \(^{131}\), phenyl \(^{60}\) and alkynyl groups \(^{132}\).

A unique addition to this list of functional group conversions is provided by CmlS, a FDH that is present in the chloramphenicol biosynthetic pathway found in *S. venezuelae* \(^{73}\). CmlS appears to catalyze what resembles a classical haloform reaction on an acetyl group, with the exception that the reaction stops after two halogenation events, producing the dichloroacetyl moiety on chloramphenicol. A structural view of the CmlS active site is critical to determine the
reaction mechanism. To this end, we report the expression and purification of CmlS and the generation of crystals that currently diffract to 2.0 Å resolution.

4.3 Materials and methods

4.3.1 Cloning, expression and purification

The gene encoding CmlS was PCR-amplified from the plasmid pJV526 with the primers 5’-GC-AGCCATATGACACGATCGAAGGTGGCGA-3’ and 5’-CCGCAAGCTTTTC-AGACCTCGTACTCGAC-3’ (NdeI and HindIII sites, respectively, are in bold). The purified PCR product was digested with NdeI and HindIII and ligated into similarly digested pET-28a (Novagen). The resulting plasmid, pET-28-CmlS, encodes CmlS with an N-terminal hexahistidine tag. Sequencing of both strands of pET-28-CmlS revealed that our cmlS clone differed from the cmlS sequence deposited in GenBank (accession No. AAK08979). The cmlS gene in pET-28-CmlS had two silent mutations (bases 702 and 948) and the DNA sequence from 907–924, which encodes the amino-acid sequence IFRRSV (residues 303–308 of AAK08979), was absent. A sequence alignment performed with our cloned CmlS amino-acid sequence revealed that the IFRRSV insertion would disrupt a highly conserved region shared by the FDH family (Figure 4.1). This suggests that the IFRRSV sequence in GenBank accession No. AAK08979 is incorrect.

To express CmlS, E. coli BL21(DE3) cells (Novagen) were transformed with pET-28-CmlS and grown at 310 K on solid LB medium containing 1% agarose and 50 mg ml⁻¹ kanamycin. A single colony was used to inoculate 5 ml LB medium supplemented with 50 mg ml⁻¹ kanamycin, which was then incubated overnight in an air shaker (225 rev min⁻¹, 310 K) to obtain a saturated culture. The saturated culture (5 ml) was used to inoculate 500 ml LB medium
FAD

CmlS  --- MTRSKVAIQGGDFASVAGLTLHK --- LGHDVTIYERSAFPFRYRVSLLGPTMSI 53
CrpH (27%)  MTSPLNTGQIIQGGPGSSTAATTLLAR --- EGFVDTLLEBEVFPRYHRVHRLPSALRI 57
PrnA (11%)  MNPKIKPQGTAGGMAAGYVLRAQQVQVITLESAAAPIRQVGHEATPSQKLVK 58
RebH (13%)  MSGIKDIIIIVGQSTAGGMAAGYQKLQGTAIDIPLLQAPIDTLGVTETPTNLQTA 58

CmlS  LNRLGQLEK --- IDAQNYVKPSIFTLWG --- QDAQAWPFSFAAPK 94
CrpH  FDGLVREK --- IBAQGFGQRPGVIENW --- TERKWSLNGEFETLDN 98
PrnA  FDFPLGIPERDWMQVQNGAFDAKFKPNWRKSPD --- PGREDYVFYLGSFVNCDC 110
RebH  FDFPLGIPERDWMRQCNATVAKTFINWKTAJGTSARELEDQGDFPFDHPSFLGLKYNE 118

CmlS  APWFDFHAVQVQREE --- FDKLLL 115
CrpH  T --- YSFQVRDRED --- FDHLLL 114
PrnA  GVPLTHYWLKRKREQQF-QQPMAYACQFQPAGALSQKLAPCLUDOTQMS-HAWHFDHALVI 168
RebH  QIPLASHYFWDRSYRQGKTVFPEFHACYKFEVILFDANRSPRRLDKGSQVTN-YAWHFDHALV 177

CmlS  DEARSGKGI - TVHSESETFTVHSVLSD-FDFRVVLTVERGQVESMTVESDFVYDAGG-QGGIPRGR 172
CrpH  EHSKQGG-VKVFOTGKIQLSFDGPEFPSATNSGQNDTTEGIFSDFPMIDASRAGIMATE 173
PrnA  DFLNARVA-CRENGQVYVDDVECLNGYISTLTTKETKRLADLPWIDTHSQMOLLINQ 227
RebH  DFLQRTAFTEKLHRVQRDLHVQDQANQGIRESVATRGRVDADLFVDCSFGQFLLNK 237

Isalofoxazine

CmlS  KLGW --- RQYD-toolbar-AYFVWFKQWKLKDGGDKLGTKYTSIFTEPQJOAHYFECDDLWSVGL 230
CrpH  YLKN --- RHALDFVPQKVSYSYNKKTALSQIGPSDFQYQITFLDESEMSQGV 231
PrnA  ALKSFIPDFDMDSILDSVASSQAYSYFPMDQERVSAYMKSTQDKMLGRPGSV 287
RebH  AMSSPFLMSDSSLHKLQHSAVVQFPDADIVGVFDTSVIAKSMGRKQILMLGRPGTVY 297

CmlS  VVDRSKSAEVRQGAGQDDAQAQQADYFRHVSDEIRVDQIVDWSYDTEVPFADSFRD 290
CrpH  VMHSTYKERRLNLKIDYVEAIATCPLIADLDVALGELDLDVQEDQYDSTDSDFGSAPY 291
PrnA  FSSKFTSRQDATADFLKLW --- LSNNQQLNZKIFVRGNNKRAWVNNCVSI 335
RebH  YSSRFATDEDEVARFECMBWHLW --- P-ETQFLNKKIRFVRGNNKRAWVNNCVSI 346

Cl

CmlS  FLCGDACFCTDFPMHNYQLSLHATQVAVYASAAADIRTRPHGDKADAVHAYNHRREBYEQY 350
CrpH  PISRACFLCPFDRLLSQQHVTLYSTALGSALSASISRTQEGVEGTSQASAFYDSQSYQYALRF 351
PrnA  GLS --- S-CPLFILFSTGILYFYYAALYQLVKHPDTS --- FDPRLDARFNENITYMDCC 388
RebH  GTS --- SCFVRLEFLSTGILYFYYAALYQLVHHPDKS --- INFLVTRFREIEITPMFID 399

CmlS  HQFFLASFYFTPASTEPDFSERFKRRKETESDDDELRTRKWFESLAGNPDQGDPQTVASYFRED 410
CrpH  LVFVQFAPQYDNQVGDYFQARQLSLRDPOSQNLKAFLNLQSGS --- 395
PrnA  RDFQGAHYPTT --- SREDTPFWFLESARHESLDAYQKEQKERYKAGLP 432
RebH  RDFQGAHYFPS --- PRITDTPFWRAN-KERLSAYQKEMQSPKYRMAIN 442

CmlS  RASTMIAIQRHQQRELSPEDSFSEAEBLNRVIRWISDLTRKLSNISRTFKNQGKAVLQKQHR 470
CrpH  --- VEDLEDAGKEDAFVSWNARQQQRSQSHS 423
PrnA  LTTSDSTFFYETFDEYEFKNFWLQNYICYFAILG 468
RebH  --- APASDDAQLYGNFEEEFNFNNWSQNYVCIALDQ 479

CmlS  VEPQGRFLICQREVFVLANGEQLMDAMQFDMDEARQIFQDLAEEEEFYGKTLVERVLEGAVGQEL 530
CrpH  IQDQKALGAIREEIKNEVNCMTAQQPNAVFSGILSAGAAGFDVLWFTTQKQLGQCPQVIPL 483
PrnA  MLDFSRLPLILQPRT-EISIAKAMAPASIRREAINTLSTPLNÝYLSLRLDGQLASLNQH 527
RebH  LVDPASPRLHHM-PQTESDEVFQAVKDQRRNLLETLPSLHFFL --- QOQH 528

CmlS  STQIVVRMLMAGLLTGYDAQKEFVQRYQRLHFGVGEYEVY --- 571
CrpH  QRPSSLHHT --- 491
PrnA  GPTLAAQERQ --- 537
RebH  GRL --- 530
Figure 4-1 Alignment of the sequences of CmlS (as cloned in this work) and the flavin-dependent halogenases CrpH (GenBank accession No. ABM21576), PrnA (PDB code 2apg) and RebH (PDB code 2oa1). The alignment was performed with ClustalW2. Sequence percentage identities relative to CmlS are given in parentheses. Conserved regions of sequence are highlighted, including the FAD diphosphate-binding site (gray), the residues lining the FAD isoalloxazine-ring binding site and the tunnel that guides HOCl (yellow), the chloride-binding site (green) and the K residue which directs chlorination (red). The residues conserved in tryptophan halogenases PrnA and RebH which are thought to stabilize the carbocation intermediate of an EAS reaction are shown in turquoise. Note that CrpH also catalyzes an EAS reaction. The sequence IFRRSV in the CmlS sequence recorded in GenBank (accession No. AAK08979) is located between amino acids Pro302 and L303, which are highlighted with arrows (↓). The very high sequence conservation in this region suggests that the IFRRSV sequence in the CmlS GenBank entry is incorrect.
containing 50 mg ml\(^{-1}\) kanamycin, which was incubated in an air shaker (225 rev min\(^{-1}\), 310 K) until the optical density \((\text{OD}_{600})\) of the culture reached \(\sim 0.6\). The protein expression was induced with 0.5 mM IPTG for 24 h at 288 K, after which time the cells were collected by centrifugation at 3000g for 20 min at 277 K. The resulting cell pellet was stored at 253 K until purification. The cells were resuspended in 50 mM Tris–HCl, 2 mM DTT, 300 mM NaCl, 10 mM imidazole pH 7.5 and lysed using two passes through an EmulsiFlex C5 cell homogenizer (Avestin) at 138 MPa. The lysed cells were then centrifuged at 40 000g for 30 min at 277 K. The supernatant was loaded onto a nickel–nitrilotriacetic acid (Ni–NTA) column (Qiagen). CmlS was then eluted using an imidazole gradient ranging from 10 to 500 mM over ten column volumes at a flow rate of 5 ml min\(^{-1}\) using an ÄKTA FPLC system (GE Healthcare). Fractions containing pure CmlS, as shown by SDS–PAGE analysis (see Appendix), were then pooled and concentrated using a Millipore Amicon Ultra 15 centrifugal filter (30 000 Da molecular-weight cutoff) followed by buffer exchange into 50 mM Tris–HCl, 2 mM DTT pH 7.5 using a PD-10 desalting column (GE Healthcare). SDS-PAGE images following Ni-NTA purification is shown in the Section 2 of the Appendix The concentration of purified CmlS was determined by a Bradford assay. Aliquots of CmlS were then flash-frozen in liquid nitrogen and stored at 193 K. We did not find it necessary to remove the N-terminal hexahistidine tag on recombinant CmlS to obtain high-quality crystals (see below).

The selenomethionine derivative of CmlS was produced in the methionine-auxotroph *E. coli* strain DL41(DE3) grown in M9 SeMET High Yield medium (Medicilon) using the same growth conditions as used for native protein expression. The selenomethionine-labelled CmlS was purified using the same procedure as described above.
4.3.2 Crystallization

All crystallization experiments were performed at room temperature. For initial screening, CmlS was kept at a concentration of 74 mM (~ 5 mg ml\(^{-1}\)) in buffer containing 50 mM Tris–HCl, 2 mM DTT pH 7.5. Sitting-drop vapour-diffusion trials were carried out using Qiagen crystallization screening kits in 96-well plates (Greiner): 1 µl protein solution was mixed with 1 µl crystallization solution and equilibrated against 100 µl well solution. Initial hits were obtained in condition Nos. 54, 64 and 79 of the JCSG+ Suite (Qiagen), all of which contained PEG 3350 together with various salts and buffers.

Crystallization conditions were optimized using a grid screen of PEG 3350 versus the pH of sodium acetate buffer. 1 h prior to crystallization trials, a concentrated solution of L-arginine and L-glutamate at pH 7.6 was added to the CmlS sample to afford a final concentration of 50 mM of each amino acid. Optimization using the hanging-drop vapour-diffusion method was performed in 24-well plates (VDX): 2 µl CmlS solution was mixed with 2 µl crystallization solution on a siliconized glass cover slip and then equilibrated against 1 ml well solution. The final crystallization conditions contained 0.1 M Na HEPES pH 6.8–7.4, 17–22% w/v PEG 3350. Crystals appeared after 3–4 d and reached their maximum size within a further 2 d. Typical crystal dimensions were 0.25 x 0.1 x 0.02 mm. Selenomethionine-derivative (SeMet) crystals of CmlS were obtained using the same procedure as used for the native CmlS crystals.

4.3.3 Data collection and diffraction measurements

Single-wavelength anomalous dispersion data were collected from SeMet CmlS crystals on the X12B beamline at Brookhaven National Laboratory National Synchrotron Light Source using an ADSC Quantum-4 CCD detector. All data were collected at 100 K. Prior to flash-freezing in liquid nitrogen, the CmlS crystals were sequentially immersed in crystallization solution containing 10, 15 and 20% v/v PEG 200 as a cryoprotectant. Oscillations of 1° with an
exposure time of 40 s per image and a crystal-to-detector distance of 200 mm were used. The data were indexed and scaled using DENZO and SCALEPACK.

4.4 Results and discussion

Recombinant CmlS was expressed in soluble form in *E. coli* using a low-temperature protocol. CmlS was then isolated in one step to achieve essentially single-band purity as confirmed by SDS–PAGE (data not shown). The purified yield of CmlS was 32 mg per litre of culture. Concentrated CmlS in pure form had a distinct yellow hue characteristic of a flavin-containing enzyme. Accordingly, the UV–visible spectrum of CmlS revealed absorbance maxima (375 and 458 nm) typical of a bound oxidized flavin.

CmlS initially crystallized in condition Nos. 54, 64 and 79 of the JCSG+ Suite (Qiagen). Expansion of these conditions using the hanging-drop vapour-diffusion method in 24-well plates produced crystals that were large enough for diffraction studies (Figure 4.2), but the initial diffraction was very weak. It has been reported that addition of 50 mM L-glutamate and 50 mM L-arginine greatly improves protein solubility and long-term stability. The effect of these amino acids on crystal quality was assessed by adding them to CmlS samples immediately after thawing and at least 1 h prior to crystallization trials. The crystal morphology and size did not change beyond the normal range of variability; however, the diffraction limit of the crystals increased to 2.3 Å at the home source. The resulting large flawless crystals produced diffraction to 2.0 Å at the synchrotron (Figure 4.3).

The crystals exhibited symmetry of the C-centered monoclinic space group *C*2, with unit-cell parameters \(a = 208.1\), \(b = 57.7\), \(c = 59.9\) Å, \(\beta = 97.5^\circ\). Diffraction data for CmlS were processed in the resolution range 30–2.2 Å. A Matthews coefficient of 2.68 Å³ Da⁻¹ was obtained, with a solvent content of 54.2%, representing the presence of a monomer in the
Figure 4-2 Optimized crystals of CmlS obtained using 0.1 M Na HEPES pH 6.8–7.4, 17–22% w/v PEG 3350.
Figure 4-3 Diffraction of SeMet CmlS crystals. The data were collected to 2.0 Å resolution and processed to 2.2 Å resolution.
asymmetric unit. A summary of the crystal parameters and the statistics of the diffraction data is presented in Table 4.1. The structure solution of CmlS is presented in Chapter 5. In parallel, we are testing various compounds with CmlS (e.g. acetate, chloroacetate, malonate and thioesters of these compounds) in order to determine the optimal substrate for halogenation activity.
Table 4-1 Diffraction data for SeMet CmlS crystals.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Space Group</strong></td>
<td>$C_2$</td>
</tr>
<tr>
<td><strong>Unit Cell Parameters (Å, °)</strong></td>
<td>$a = 208.09$, $b = 57.74$, $c = 59.88$, $\beta = 97.47$</td>
</tr>
<tr>
<td><strong>Wavelength (Å)</strong></td>
<td>0.98160</td>
</tr>
<tr>
<td><strong>Temperature (K)</strong></td>
<td>100</td>
</tr>
<tr>
<td><strong>Resolution Range (Å)</strong></td>
<td>30 - 2.2 (2.28 - 2.20)$^a$</td>
</tr>
<tr>
<td><strong>Observed Reflections</strong></td>
<td>762884</td>
</tr>
<tr>
<td><strong>Unique Reflections</strong></td>
<td>36054</td>
</tr>
<tr>
<td><strong>Data Completeness (%)</strong></td>
<td>99.7 (100)</td>
</tr>
<tr>
<td><strong>Redundancy (%)</strong></td>
<td>7.2 (6.5)</td>
</tr>
<tr>
<td><strong>$R_{sym}^b$ (%)</strong></td>
<td>7.7 (55)</td>
</tr>
<tr>
<td><strong>$&lt;I/\sigma I&gt;$</strong></td>
<td>35.7 (4.1)</td>
</tr>
<tr>
<td><strong>Matthews coefficient (Å$^3$ Da$^{-1}$)</strong></td>
<td>2.68</td>
</tr>
<tr>
<td><strong>Solvent content (%)</strong></td>
<td>54.17</td>
</tr>
</tbody>
</table>

$^a$Values in parentheses are for the highest resolution shell (2.28–2.20 Å).

$^b$R$_{sym} = \Sigma_{hkl} \Sigma_i |I_i(hkl)| - <I(hkl)> | \Sigma_{hkl} \Sigma_i I_i(hkl)$, where $I_i(hkl)$ and $<I(hkl)>$ represent the diffraction-intensity values of the individual measurements and the corresponding mean values. The summation is over all measurements.
Chapter 5
Chloramphenicol biosynthesis: the structure of CmlS, a flavin-dependent halogenase showing a covalent flavin–aspartate bond

Preface:

This chapter was published in the Journal of Molecular Biology:


Kateryna Podzelinska was responsible for the structure solution of CmlS, its refinement and analysis. Ryan Latimer performed the native and selenomethionine protein expression and purification, ESI mass spectrometry, spectroscopic characterization and denaturation experiments, as well as the CmlS crystallization under Kateryna Podzelinska’s guidance. Anupam Bhattacharya developed the protocol for soluble CmlS purification. Dr. Leo Vining provided the initial CmlS construct. This manuscript was written by Dr. David Zechel and Kateryna Podzelinska, with editorial input from Dr. Zongchao Jia.
5.1 Abstract

Chloramphenicol is a halogenated natural product bearing an unusual dichloroacetyl moiety that is critical for its antibiotic activity. The operon for chloramphenicol biosynthesis in \textit{S. venezuelae} encodes the chloramphenicol halogenase CmlS, which belongs to the large and diverse family of flavin-dependent halogenases (FDH’s). CmlS was previously shown to be essential for the formation of the dichloroacetyl group. Here we report the X-ray crystal structure of CmlS determined at 2.2 Å resolution, revealing a flavin monooxygenase domain shared by all FDHs, but also a unique ‘winged-helix’ C-terminal domain that creates a T-shaped tunnel leading to the halogenation active site. Intriguingly, the C-terminal tail of this domain blocks access to the halogenation active site, suggesting a structurally dynamic role during catalysis. The halogenation active site is notably nonpolar and shares nearly identical residues with \textit{C. crocatus} tyrosyl halogenase (CndH), including the conserved K71 that forms the reactive chloramines intermediate. The exception is Y350, which could be used to stabilize enolate formation during substrate halogenation. The strictly conserved residue E44, located near the isoalloxazine ring of the bound flavin adenine dinucleotide (FAD) cofactor, is optimally positioned to function as a remote general acid, through a water-mediated proton relay, which could accelerate the reaction of the chloramine intermediate during substrate halogenation, or the oxidation of chloride by the FAD(C4α)–OOH intermediate. Strikingly, the 8α carbon of the FAD cofactor is observed to be covalently attached to D277 of CmlS, a residue that is highly conserved in the FDH family. In addition to representing a new type of flavin modification, this has intriguing implications for the mechanism of FDHs. Based on the crystal structure and in analogy to known halogenases, we propose a reaction mechanism for CmlS.
5.2 Introduction

Chloramphenicol is a commonly used antibiotic that competitively inhibits the peptidyl transferase activity of bacterial ribosomes. Isolated in 1947 from the soil bacterium *S. venezuelae*, chloramphenicol quickly became a ‘wonder drug’ in the treatment of typhus and is still used for this purpose where typhus is endemic.

Chloramphenicol is also one of the earliest known halogenated natural products, possessing a dichloroacetyl group that is critical for its anti-bacterial activity. Today, over 4000 halogenated natural products have been characterized, and the number and structural diversity exhibited by these compounds continue to expand rapidly. However, despite the long history of halogenated natural product discovery, the identities of many enzymes that install halogen substituents (or halogenases) have emerged only in the past decade, and only a handful have been characterized mechanistically or structurally. The operon encoding chloramphenicol biosynthesis in *S. venezuelae* was reported in 2001 and was found to include a gene, *cmlS*, that is essential for the incorporation of chlorine into an unusual dichloroacetyl moiety. *S. venezuelae* chloramphenicol halogenase (CmlS) has a sequence homology to flavin-dependent halogenases (FDHs), a large and rapidly growing family of enzymes that collectively stand out for their ability to incorporate halogens into an astonishing spectrum of substrates, including as indole, pyrrole, quinone, phenyl, and, in the case of CmlS, alkynyl groups. Thus far, tryptophan halogenases have yielded the most structural and mechanistic details. Structures have been solved for halogenases specific for the synthesis of 7-chlorotryptophan (*P. fluorescens* PrnA and *L. aerocolonigenes* RebH) and 5-chlorotryptophan (*S. rugosporus* PyrH). These structures revealed a shared flavin monooxygenase domain containing a conserved GWxWxIP sequence motif (Figure 5.1) that is unique to FDHs. Mechanistic studies have shown that this conserved domain binds FADH$_2$ (provided externally by an NADH-dependent flavin reductase), which subsequently reacts with molecular oxygen to produce an FAD(C4α)–OOH intermediate. This
Figure 5-1 Structure-based sequence alignment of CmlS (GenBank accession no. AAK08979) with CndH (GenBank accession no. CAQ43074; 34% sequence identity) and PrnA (GenBank accession no. AAB97504; 14% sequence identity). Secondary structure elements observed in the CmlS structure are shown above the alignment. The conserved active site is indicated with a red star. Active-site hydrophobic and hydrophilic residues in CmlS are highlighted with black and red arrows, respectively. Residues E44 and D277 are indicated with blue arrows. Alignment was performed with the Align X module of Vector NTI (Invitrogen) then adjusted by eye. The figure was prepared with ESPript.
species is in turn attacked by a bound chloride ion, generating HOCl. A tunnel some 10 Å long directs the HOCl to a conserved K residue (K79 in PrnA), where substrate chlorination takes place. This K residue is absolutely essential for halogenation activity. Yeh et al. have proposed that the K reacts with HOCl to form a covalent chloramine intermediate, which would be more apt than freely diffusible HOCl to react regioselectively with tryptophan. However, Flecks et al. have argued that a chloramine is not sufficiently electrophilic to react with an indole in an EAS reaction, and they propose that the conserved K directs the more reactive HOCl with a hydrogen bond. The transition state leading to the carbocation intermediate of the EAS reaction is possibly stabilized by negatively charged E346 in PrnA, which is maintained in an ionized state by H101 and H395. However, these three residues are not conserved in the FDH family; instead, hydrophobic F or I residues are found at these same positions in FDH homologues (Figure 5.1), even though many of these enzymes also catalyze EAS reactions. One such example is SgcC3, which chlorinates 3-hydroxy-β-tyrosine ortho to the phenolic oxygen. Unlike the tryptophan halogenases that act on free substrates, SgcC3 activity requires its substrate to be bound as the thioester to a peptidyl carrier protein. Recently, the structure of a similar halogenase, C. crocatus tyrosyl halogenase (CndH), was solved, revealing a considerably more accessible chlorination active site, which is likely necessary to accommodate a peptidyl carrier protein-bound substrate.

Herein, we report the X-ray structure of CmlS, one of the few enzymes known to halogenate an alkyl group. The structure provides a view into the evolution of the FDH family and the versatility of their catalytic machinery. Intriguingly, CmlS is covalently bound to flavin adenine dinucleotide (FAD), raising many questions concerning the dynamics of this cofactor during catalysis.
5.3 Materials and methods

5.3.1 Expression, purification, and crystallization

The expression, purification, and crystallization of CmlS were described in Chapter 4. CmlS concentrations were determined by Bradford assay. The optimized crystallization conditions consisted of 74 µM CmlS in 50 mM Tris–HCl, 2 mM DTT (pH 7.5) mixed 1:1 with 0.1 M Na Hepes (pH 6.8–7.4), and 17–22% w/v polyethylene glycol 3350. Crystals were grown using the hanging-drop method. CmlS crystals formed thick yellow plates with typical dimensions of 0.25 × 0.1 × 0.02 mm.

5.3.2 Data collection, structure determination, and refinement

Diffraction data were collected at the X-12B beamline at the National Synchrotron Light Source (Brookhaven National Laboratory, Upton, NY). The data were collected at 100 K on the ADSC Quantum 4 CCD detector. Data were indexed, integrated, and scaled with DENZO and SCALEPACK, and truncated to 2.2 Å to ensure reasonable data intensity statistics for the highest resolution shell. The structure of CmlS was determined using the single-wavelength anomalous dispersion method and a SeMet derivative of the protein. The heavy atom positions for eight of nine selenium atoms in the asymmetric unit were determined and refined using autoSHARP. The initial model was built automatically using autoSHARP and ARPwARP. Subsequently, substantial manual building was carried out in XFIT/XTALVIEW and Coot. The bulk of the refinement was performed in REFMAC, and simulated annealing refinement was carried out in PHENIX. The final model contained one molecule in the asymmetric unit, with 232 molecules of water and 1 molecule of FAD. No density was observed for the first 19 of 20 residues of the N-terminal histidine purification tag: residues 93–94, 393–407, and 567–571. Residues F390, F408, R409, D499, R559, and L560 were refined as glycine, and residue R481 was refined as alanine, because side-chain density was not observed for these residues. The only
Ramachandran plot outlier (D288) belongs to a solvent-exposed loop and fits well into the electron density. This residue forms a 3.04-Å ionic interaction with K332. The model possesses one cis-peptide bond between F37 and P38. All structure figures presented in this work were generated using PyMOL. The structure factors and atomic coordinates determined for CmlS in this study have been deposited in the PDB under accession number 3I3L.

5.3.3 Electrospray ionization mass spectrometry analysis

CmlS, as purified from *E. coli*, was desalted on a ZipTip C4 microcolumn (Millipore) and eluted with acetonitrile/water (1:1) containing 1% formic acid. The solution was directly infused from a nanospray tip in positive ion mode on an Applied Biosystems MDS QSTAR instrument equipped with a time-of-flight detector set to scan from 100 atomic mass unit (amu) to 2500 amu. The following settings were applied: curtain gas, 30; declustering potential, 80; focusing potential, 250; declustering potential, 2:15; collision gas, 3; ion release delay, 6; ion release width, 5.

5.3.4 Chemical denaturation and flavin content analysis of wild type CmlS and D277N mutant

Absorbance spectra were acquired on a Varian Cary 50 spectrometer at room temperature. A solution of 22 µM CmlS (200 µL) in 6 M GdHCl was used to acquire an initial spectrum from 240 nm to 550 nm. This sample was diluted 20-fold in 6 M GdHCl [containing 50 mM Tris (pH 7.5), 0.5 mM ethylenediaminetetraacetic acid, and 0.125 mM tris(2-carboxyethyl)-phosphine] then concentrated by centrifugation at 277 K in a 4-mL Amicon ultrafiltration device (Millipore; 10 kDa molecular weight cutoff) to a final volume of 200 µL. The spectrum of the retained sample was recorded, then diluted and concentrated as described above before the recording of a third spectrum. The concentration of FAD in CmlS denatured in 6 M GdHCl was calculated from the absorbance at 450 nm using the extinction coefficient for FAD in GdHCl.
(ε_{450}=11,900 \text{ M}^{-1} \text{ cm}^{-1})^{147}. This was then used to calculate the concentration of CmlS from the absorbance of the sample at 280 nm using the corresponding extinction coefficient for FAD at 280 nm (ε_{280}=22,900 \text{ M}^{-1} \text{ cm}^{-1})^{147}, and the calculated extinction coefficient for CmlS in 6 \text{ M GdHCl} (ε_{280}=94,240 \text{ M}^{-1} \text{ cm}^{-1})^{143}. The D277N mutation was introduced by the four-primer PCR method using the mutagenic primers D277N\_forward (5′-GATCGGTGCAGAA-CTGGTCTACGACACC-3′) and D277N\_reverse (5′-CGTAGGACCAGTTCTGCACGATCC-GGAC-3′), along with the flanking primers CmlS\_forward (5′-GCAGCATATGAACACGATCGAAGGTG-3′) and CmlS\_reverse (5′-CCGCAAGCTTTACGACCTCGTACTCG-3′). Mutagenic codons are underlined, and NdeI and HindIII restriction sites are in boldface. The plasmid pET-28-CmlS was used as PCR template. Sequencing of both DNA strands of the resulting pET-28-CmlS-D277N clone confirmed the desired allele. The D277N mutant was expressed and purified to homogeneity as described for the wild-type enzyme in Chapter 4. A spectrum of 4.6 \mu M CmlS D277N in 6 \text{ M GdHCl} solution was recorded, followed by 20-fold dilution in 6 \text{ M GdHCl} and re-concentration to the original volume by ultrafiltration. The spectrum of the retained D277N solution was then recorded.

5.4 Results

5.4.1 The overall structure of CmlS

The crystal structure of CmlS was determined to 2.2 Å resolution using the single-wavelength anomalous dispersion method (Table 5.1). The model was refined to R and R_{free} values of 20.0% and 26.1%, respectively. According to the Ramachandran plot generated during PDB validation, 90.5% of the residues lie on the most favorable regions, with 9.1% of the residues lying in the additionally allowed regions, 0.2% of residues in additionally allowed regions, and 0.2% of the residues lying in the disallowed regions. CmlS belongs to an α/β class of
Table 5-1 Summary of data collection and refinement statistics (SeMet CmlS).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space Group</td>
<td>C2</td>
</tr>
<tr>
<td>Unit Cell Parameters</td>
<td>( a = 208.1 ), ( b = 57.7 ), ( c = 59.9 \ \text{Å}, \beta = 97.5^\circ )</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9798</td>
</tr>
<tr>
<td>Resolution Range (Å)</td>
<td>30.0 - 2.2 (2.28 - 2.20)</td>
</tr>
<tr>
<td>Observed Reflections</td>
<td>762,884</td>
</tr>
<tr>
<td>Unique Reflections</td>
<td>36,054</td>
</tr>
<tr>
<td>Data Completeness (%)</td>
<td>99.4 (94.8)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>7.2 (6.5)</td>
</tr>
<tr>
<td>( R_{\text{sym}} ) (%)</td>
<td>7.7 (55)</td>
</tr>
<tr>
<td>( \langle I/\sigma I \rangle )</td>
<td>35.7 (4.1)</td>
</tr>
<tr>
<td>( R_{\text{work}} ) (%)</td>
<td>20.0 (23.4)</td>
</tr>
<tr>
<td>( R_{\text{free}} ) (%)</td>
<td>26.1 (28.3)</td>
</tr>
<tr>
<td>No. of observations (total)</td>
<td>34,078</td>
</tr>
<tr>
<td>No. of observations for ( R_{\text{free}} )</td>
<td>1786</td>
</tr>
<tr>
<td>Root mean square deviations</td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.021</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.538</td>
</tr>
<tr>
<td>Mean temperature factor (Å^2)</td>
<td>38.2</td>
</tr>
<tr>
<td>No. of protein residues</td>
<td>550</td>
</tr>
<tr>
<td>No. of protein atoms</td>
<td>4356</td>
</tr>
<tr>
<td>No. of water atoms</td>
<td>232</td>
</tr>
<tr>
<td>Ramachandran statistics (%)</td>
<td></td>
</tr>
<tr>
<td>Most favoured regions</td>
<td>90.5</td>
</tr>
<tr>
<td>Allowed regions</td>
<td>9.1</td>
</tr>
<tr>
<td>Generously allowed regions</td>
<td>0.2</td>
</tr>
<tr>
<td>Disallowed regions</td>
<td>0.2</td>
</tr>
</tbody>
</table>

\(^a\)The values in parentheses are data for the high-resolution shell (2.28-2.20 Å).
proteins of the FAD/ NAD(P)-binding domain superfamily, which includes the immediate ancestor of FDHs — flavin-dependent monooxygenases. The overall shape of the protein resembles a triangle, with three lobes arranged above and below the long central α-helix (α10) of 36 residues (Figure 5.2a). The protein contains one molecule of FAD that is bound between the two bottom lobes. The bottom left lobe adopts a characteristic Rossmann fold of glutathione reductases (GR) family members\(^\text{144}\). This lobe possesses the β/β/α layer architecture. The first and outermost solvent-exposed layer is a sheet of three anti-parallel β-strands that form a β-meander motif. The middle layer is a β-sheet composed of five parallel β-strands. The third layer consists of two α-helices that lie parallel with each other and pack against the β-strands of the middle layer, thus burying their hydrophobic side chains in this β-sheet. Despite the overall conservation of topology, the connectivity of the Rossmann fold adopted by CmlS differs from the typical β1–α1–β2–α2–β3 connectivity shared by other GR family members. In CmlS, the first three N-terminal secondary structure elements form a β1–α1–β2 β-motif, while the last two elements are contributed by helix α4 and strand β7. The consensus sequence motif GGGxxG (Figure 5.1) is part of the loop connecting the first β-strand and α-helix in the Rossmann fold, and the N-terminal end of this helix points toward the pyrophosphate moiety of FAD for charge compensation. The two helices of the third layer of the Rossmann fold motif lie in a row and approximately parallel with the two longer helices above them. The topmost helix α10 is the central 36-residue helix that spans the length of the entire protein and begins with a single 3\(_{10}\) turn. On the back side of the Rossmann fold, there are two additional short helices α2 and α3, which form a fourth layer on the bottom right lobe. The bottom right lobe consists of a mixed seven-stranded β-sheet in the core layer and three helices in the surface layer. The first five strands of the sheet (β12–β16) are long and antiparallel with each other. The last two strands (β4 and β5) are short and also anti-parallel, but β4 is parallel with β13. This sheet curves around the isoalloxazine ring of the FAD, and the strand β14 contains another highly conserved FDH motif.
Figure 5-2 The structure of CmlS. (a) The conserved FAD monooxygenase domain is shown in light blue and purple, and the unique C-terminal domain is shown in orange. The final C-terminal residues that lead to the halogenation active site are shown in green. N-termini and C-termini are indicated. FAD is shown in stick format. An unresolved loop between residues 388 and 410 is shown as a broken line. (b) Electrostatic surface representation of CmlS. Blue, positive charge; red, negative charge; gray, neutral charge. (c) CmlS structural homologues: PHBH (PDB ID 1pbe), PrnA (PDB ID 2apg), and CndH (PDB ID 3e1t). The
conserved FAD monooxygenase domain is colored as with CmlS, and the variable C-termini are shown in orange. Bound chloride ions are shown as yellow spheres. In the case of PrnA, the 45-residue insertion unique to the tryptophan halogenases is shown in red. Unresolved loops are shown as broken lines.
GWxWxIP (Figure 5.1). The surface layer consists of three short α-helices, where the first helix begins with a single 3_10 turn. The helices are positioned at right angles to each other and are joined by two-residue and one-residue loops. Finally, the third lobe is made up of the unique C-terminal of CmlS and is discussed in detail in section 1.4.2. The long loop connecting this C-terminal lobe to the rest of the protein is not fully resolved due to disorder.

5.4.2 The C-terminal domain

An intriguing feature of CmlS is the C-terminal domain (residues 380–517), which is mounted directly above the flavin monooxygenase domain (Figure 5.2a). This domain resembles an arch-like structure above the central α-helix, creating a small tunnel just above the middle of that helix (Figure 5.2b). The short helix α12 with a single 3_10 turn initiates this domain, followed by the much longer helix α13, which projects away from the halogenation active site. This is followed by a cluster of three helices (α14, α15, and α16) and two β-strands (β19 and β25), which form the top of the arch. This portion of the C-terminal domain has a weak structural homology to the winged-helix DNA binding domain of RNA polymerase III [Protein Data Bank (PDB) ID 2yu3; Z=4.5, RMSD=2.8 Å] and a domain of unknown function in type II methionine amino peptidases (PDB ID 1xgs; Z=4.3, RMSD=2.4 Å). The primary deviations from the winged-helix domain are the anti-parallel strands β21 and β22, which form a severely twisted β-hairpin structure that projects back down to the central helix α10 of the flavin monooxygenase domain.

5.4.3 CmlS structural homologues

A DALI search of the PDB revealed a strong structural homology between CmlS and CndH (PDB ID 3e1t; Z=44.0, RMSD=2.1 Å, sequence identity=36%), tryptophan halogenases such as PrnA (PDB ID 2apg; Z=31.7, RMSD=4.7 Å, sequence identity=20%), and the flavin-dependent monooxygenase *P. fluorescens* para-hydroxybenzoate hydroxylase (PHBH) (PDB ID 1pbe; Z=31.7, RMSD=3.5 Å, sequence identity=16%). As a dramatic illustration of how a
conserved enzyme fold is adapted to different functions, virtually all of the structural homology between CmlS, CndH, and PrnA is derived from the progenitor flavin monooxygenase domain of PHBH (Figure 5.2c). Some structural differences are observed in the flavin monooxygenase domain of CmlS relative to those of CndH and PrnA. This includes three short helices (α6–α8) at the outer layer of the bottom left lobe of CmlS. PrnA has a single longer helix in this position, as well as a 42-residue insert at position 113. This additional segment (Figure 5.2c, red) consists of a long loop with three short α-helices that project above the long central helix and pack within the C-terminal lobe of PrnA. All three halogenases possess the K residue that forms the proposed chloramine intermediate; however, unlike tryptophan halogenases such as PrnA, both CmlS and CndH are missing a key catalytic E residue (E346 in PrnA), as well as the stabilizing histidine (H395). All three structural homologues also have FAD bound in the elongated conformation and contain the strictly conserved D residue (D33 in CmlS) for hydrogen bonding with the cofactor. All four enzymes contain the conserved GGGxxG motif, but PHBH lacks the second conserved GWxWxIP motif that is unique to FDHs. The major overall structural difference between CmlS and the three homologues is the structure of the C-terminal domain, which likely dictates substrate specificity.

5.4.4 The FAD binding site: a covalently bound FAD cofactor

The entire FAD cofactor is bound in the solvent-exposed groove between the two bottom lobes of CmlS. As in other GR family members, FAD adopts an elongated conformation, with the adenine and isoalloxazine moieties distal from each other. Within the bottom right lobe, contacts to the AMP moiety of the cofactor are provided by side chains of the strictly conserved to moderately conserved residues belonging to β1–α1 and the loop joining them, as well as the loops joining β2–α2, β7–β8, and β11–α5. Within the bottom left lobe, contacts to the flavin mononucleotide moiety of the cofactor are formed by the side chains of the residues from the
loop joining β18–α9, and one residue each from β12, β14, and β16. The strictly conserved E33 residue (Figure 5.1) hydrogen bonds to the ribose 2′-hydroxyl of the adenosine moiety, while the strictly conserved R39 forms a hydrogen bond with the 3′ hydroxyl. R35 and D295 make hydrogen-bond contacts with the adenosine N7 and ribitol O3 hydroxyl, respectively, while R41 and R107 stabilize oxygens of the pyrophosphate moiety. Unlike previously reported halogenase structures, CmlS did not contain an unambiguous density for bound chloride ion. The crystal structure reported here was obtained from crystals grown with trace amounts of chloride arising from the Tris–HCl buffer. Subsequent crystallization attempts at higher concentrations of NaCl (100–400 mM) did not produce crystals, while structures obtained from crystals grown at lower NaCl concentrations (50 mM) did not contain the chloride ion. Soaking of the original crystals in the cryoprotectant solution containing additional NaCl resulted in crystal cracking and poor diffraction pattern quality.

An unprecedented feature of the interaction of CmlS with FAD is the covalent linkage between D277 and the 8α carbon of the flavin ring. Unambiguous electron density connecting Oδ1 of D277 and the 8α carbon was observed independently in selenomethionine (SeMet)-labeled and native CmlS structures (Figure 5.3; data not shown). In the final model, the D277 Oδ2–8α bond length is 1.45 Å, and the flavin 8–8α–D277 Oδ2 bond angle is 110.8°, consistent with a covalent ester linkage to an sp3-hybridized carbon.

A structural alignment of the FAD binding sites from CmlS and CndH reveals an excellent alignment of the local secondary structures β15, β16, and α6 (Figure 5.4a), with the exception of the kink formed at the middle of β16 containing D277 of CmlS, the site of FAD attachment, and W278. The side chains of D277 and W278 reside 180° opposite to each other, with W287 jutting from the solvent-exposed surface of CmlS. In CndH, the FAD cofactor is not covalently bound, although the position of the cofactor superimposes very well with the
Figure 5-3 (a) An FAD cofactor covalently attached to D277 of CmlS. The omit $F_o-F_c$ electron density map for FAD and D277 is shown as a blue mesh and is contoured at 3σ level. FAD and D277 were omitted for refinement and electron density map calculations. The $8\alpha$ carbon of the flavin ring is indicated. (b) The $F_o-F_c$ electron density map with the radius of 4.0 Å around the covalent bond between Oδ1 of D277 and C8α of FAD, shown at 3σ level.
Figure 5-4 (a) Structural alignment of the FAD binding regions of CmlS (gray) and CndH (orange). Structures were aligned with DALI Lite (Z=44.0, RMSD=2.1 Å). Residues for CndH are given in parentheses. Broken lines indicate hydrogen bonds. Chloride bound in the CndH structure is shown as a yellow sphere. (b) Sequence alignment of CmlS and FDHs in the vicinity of R234 and D277 (highlighted by black arrows). The number of the first residue in each sequence is shown in parentheses. Shown are the following: CndH (PDB ID 3w1t), *Actinosynnema pretiosum* Asm12 (GenBank accession no. AAM54090), *Streptomyces roseochromogenes* Chlohal (GenBank accession no. AAN65237), *C. crocatus* CmdE (GenBank accession no. CAJ46693), *Nostoc sp.* CrpH (GenBank accession no. ABM21576.1), *Streptomyces sp.* halo2544 (GenBank accession no. ABF82430), *Actinomadura madurae* MpdC3 (GenBank accession no. ABY66006), *Streptomyces aculeolatus* NapH2 (GenBank accession no. ABS50460), *Streptomyces tendae* UbiH (GenBank accession no. CAM34371), *L. aerocolonigenes* halogenase RebH (PDB ID 2e4g), PrnA (PDB ID 2ar8), and *S. rugosporus* PyrH (PDB ID 2wet).
The equivalent residues D285 and Y286 in CndH assume an opposite orientation brought about by a 180° flip of the immediate polypeptide backbone comprising the kink. D285 is stabilized on the surface of CndH by R238, while Y286 presents an aromatic face above the flavin ring. The high degree of conservation of D277 in the FDH family as part of a D(W/Y)SY motif is striking (Figure 5.4b) and suggests that this covalent interaction may be utilized by other FDHs at some point during their reaction cycle. The polar character of R234 on the surface of CmlS is likewise highly conserved in FDH sequences (Figure 5.4b). Interestingly, the only FDHs that lack the D(W/Y)SY motif are the tryptophan halogenases. Taken together, the CmlS and CndH structures suggest that this conserved structural element can assume two orientations. If FAD were to remain covalently attached to D277, this would allow the cofactor to exit the binding cleft but remain bound at the surface of CmlS.

5.4.5 Electrospray ionization mass spectrometry analysis of CmlS

The covalent interaction between CmlS and FAD also exists in solution. The electrospray ionization (ESI) mass spectrometry spectrum for CmlS (Figure 5.5), obtained with the enzyme in 1:1 acetonitrile/water with 1% v/v formic acid, shows a broad series of ions with high charge states (+55 to +80), indicative of an unfolded protein in gas phase. Two major ions, which correspond to molecular masses of 66,318 ± 1 Da and 66,496 ±1 Da when averaged, are observed for each charge state. The first mass agrees very well with the molecular mass of CmlS missing the N-terminal methionine (expect 66,314 Da), which was likely generated by methionine aminopeptidase activity during expression in *E. coli*. The second mass, which is 178 Da heavier, corresponds to this form of CmlS covalently linked to D-gluconate, possibly through one of the histidines of the N-terminal hexa-histidine tag, which is a commonly observed modification of proteins overexpressed in *E. coli* due to the accumulation of D-gluconolactone. Upon closer inspection, a small but significant ion peak is observed for all charge states and is particularly
Figure 5-5 ESI mass spectrum of CmlS. Charge states are shown, with the ions corresponding to CmlS covalently bound to FAD indicated with red arrows. Inset: An expansion of the +66 charge state. Three ions are observed: one for apo CmlS missing the N-terminal M (expect 66,314 Da or m/z=1005.75), one for CmlS covalently attached to D-gluconate (expect 66,492 Da or m/z=1008.45), and one for CmlS covalently modified with FAD (expect 67,097 Da or m/z=1017.61).
abundant for the +66 and +65 charge states (see Figure 5.5 inset). The molecular mass of this species (averaged over +66 to +56 charge states) is 67,096 ±50 Da, in agreement with the predicted molecular mass of CmlS missing the N-terminal Met and covalently attached to FAD (66,318 + 785 – 2H = 67,101 Da). A corresponding ion of the gluconylated CmlS linked to FAD is not visible, as this ion appears to be overwhelmed by the adjacent charge state. The low abundance of the FAD-CmlS ion likely reflects a commonly observed decrease in ionization efficiency that accompanies phosphorylated proteins or the fragmentation of D277-FAD ester linkage during the ESI process. Adjustment of the orifice potential of the ESI instrument did not, unfortunately, lead to an increase in the abundance of this ion.

5.4.6 Spectroscopic characterization and denaturation of wild type CmlS and D277N mutant

Additional evidence for a covalent FAD interaction was obtained by denaturing CmlS in 6 M guanidine hydrochloride (GdHCl; pH 7.5) and centrifuging the sample through an ultrafiltration device with a 10-kDa molecular mass cutoff. The UV–visible spectrum of CmlS in 6 M GdHCl, shown in Figure 5.6a, exhibits absorption maxima at 362 nm and 450 nm, which are virtually identical with CmlS in non-denaturing buffer (data not shown). The CmlS/FAD ratio was determined to be 2:1 on a molar basis, calculated with the extinction coefficients for CmlS and FAD at 280 nm and for FAD at 450 nm (see Materials and methods). This sample was subsequently diluted 20-fold in 6 M GdHCl then concentrated to the original volume. Inspection of the spectrum of the sample indicated that flavin absorbance had decreased by only 2-fold relative to the original sample, and the CmlS/FAD ratio had increased slightly to 3:1, indicating that a fraction of FAD was non-covalently bound and was washed through the ultrafiltration membrane. The sample was diluted a second time by 20-fold in GdHCl and reconcentrated. The spectrum this time recorded a modest ~20% decrease in flavin absorbance, and the CmlS/FAD ratio likewise increased modestly to 3.6:1. These results are consistent with a large fraction of
Figure 5-6 Absorbance spectra of wild-type CmlS and the D277N mutant. (a) The spectrum of wild-type CmlS (22 µM) denatured in 6 M GdHCl (pH 7.5) is shown as a continuous black line. The broken black line represents the spectrum after the 20-fold dilution of this sample in 6 M GdHCl, followed by concentration with an ultrafiltration device (10 kDa molecular weight cutoff) to its original volume. The dotted line represents the sample spectrum after this process was repeated a second time. (b) The continuous black line corresponds to the spectrum of CmlS D277N (4.6 µM) denatured in 6 M GdHCl. The dotted black line represents the spectrum after the sample was diluted and reconcentrated as described above. The spectrum of free FAD (11 µM) in 6 M GdHCl is shown as a continuous red line in both plots.
FAD covalently bound to unfolded CmlS, which is retained by the membrane of the ultrafiltration device. The D277N mutant of CmlS co-purified with essentially the same amount of FAD as the wild-type enzyme (CmlS/FAD ratio, 1.84), suggesting that the non-covalent interaction with the cofactor is substantial (Figure 5.6b). However, unlike wild-type CmlS, denaturation and dilution of the D277N mutant in 6 M GdHCl, followed by reconcentration, efficiently removed the flavin, as expected for a noncovalently bound cofactor. As shown in Figure 5.6a, a hypsochromic shift is observed for the near-UV band of FAD bound to CmlS ($\lambda_{\text{max}}$=362 nm initially, $\lambda_{\text{max}}$=353 after washing) relative to free FAD ($\lambda_{\text{max}}$=375 nm). This shift is commonly observed when the 8α methyl group of the flavin ring is substituted. In contrast, the near-UV band observed for FAD noncovalently bound to the D277N mutant has a $\lambda_{\text{max}}$ of 375 nm (Figure 5.6b). A similar value is also observed for FAD noncovalently bound to the tryptophan halogenase RebH.

### 5.4.7 Interaction of the C-terminus with the active site

The final 21 residues of CmlS comprise a random coil (Figure 5.2a). Following β25 and a short turn, the C-terminus runs along α13 directly to the proposed halogenation active site and forms another turn centered on G563 and G564 (Figure 5.7a). Surprisingly, access of a substrate to the halogenation active site is blocked by the final few residues of the C-terminus (Figure 5.7b). Although the position of the C-terminus seems unusual, the significant number of tertiary interactions that the C-terminus makes suggests that this is a structurally and functionally important part of the enzyme. On the descent to the active site, the C-terminus makes hydrophobic contacts with α13 and α16 through V554 and V556. Hydrogen bonds are also formed between the carbonyl of Q557 and the indole NH of W442, as well as between the backbone carbonyl of F555 and R449. F562 precedes the final turn and physically blocks access to the active site, anchored in position by a cationic π-interaction with H561 (3.1 Å away) and a
Figure 5-7 The halogenation active site of CmlS. (a) The electrostatic surface model of CmlS, with the C-terminus shown in green stick form. The C-terminal tail blocks access to the proposed substrate binding site. (b) Surface representation of the enclosed CmlS halogenation site blocked by F562. The approximate position of the chloramine intermediate formed on K71 is indicated as ‘Cl+’. (c) Cutaway view of the enclosed halogenation active site. The perspective is similar to that in (a), but with the C-terminus removed for clarity. (d) Active-site residues of the halogenation active site of CmlS (gray) superimposed with the corresponding residues of CndH (orange). CndH numbering is given in parentheses.
hydrophobic contact with C\textsuperscript{6} of F87 (Figure 5.7a). Clearly at some point during catalysis, the blocking segment of the C-terminus must be removed to allow a substrate to enter the active site.

### 5.4.8 The halogenation active site

The CmlS halogenation active site is notable for its lining of hydrophobic residues (Figure 5.7c), which are highly conserved with CndH (Figure 5.7d). As shown in the sequence alignment (Figure 5.1), CmlS lacks the catalytically crucial E346 and adjacent stabilizing histidine residues found in the tryptophan halogenases (E346, H101, and H395 in PrnA\textsuperscript{128}). Instead, F87, F304, and F357 are found at these positions. One notable polar region of the putative substrate-binding pocket includes the hydroxyl group of Y350 (Figure 5.7c), which is occupied by a phenylalanine residue in CndH (Figure 5.7d). Also located near this polar pocket is H309, which could potentially act as a general base; however, in the observed structure, this residue is pushed away from the active site by G564, hence the blocking segment in the C-terminus of CmlS would need to be removed from the active site for H309 to perform this role (Figure 5.7b). As a FDH family member, CmlS shares the strictly conserved K residue (K71), which is proposed to form a chloramine intermediate that serves as the substrate chlorinating agent (Figure 5.7d). K71 hydrogen bonds with S45 (Figure 5.8a), a residue that is shared with CndH but is absent in PrnA (Figure 5.1). Intriguingly, the tunnel that guides the HOCl generated at the flavin ring to K71 is clearly marked by a chain of water molecules (Figure 5.8b). S305 marks the approximate midpoint of this tunnel and engages two of the water molecules through hydrogen bonds. Despite the strict conservation of this residue in FDHs, mutation to alanine had little effect on PrnA activity\textsuperscript{129}. Aside from the aforementioned hydrophilic residues, the tunnel is lined almost exclusively with hydrophobic residues. The only other polar residues are Y206, which hydrogen bonds with S45, and S208, which does not appear to participate in any hydrophilic contacts but is strictly conserved in FDHs.
Figure 5-8 The tunnel connecting the halogenation active site and the FAD binding site. (a) Key residues lining the proposed HOCl tunnel leading from the flavin ring to K71. Water molecules are shown as blue and green spheres. The water molecule shown in green suggests the position of the peroxide of the FAD(C4α)–OOH intermediate. (b) Surface representation of the tunnel. White arrows indicate the proposed path of the substrate to the active site and the access of E44 to bulk solvent.
5.4.9 A potential general acid catalyst

The water molecules observed in the tunnel, along with S305, form a seamless hydrogen-bond network beginning at E44 and running to K71 (Figure 5.8b). One of the water molecules, through a hydrogen bond with the flavin ring nitrogen, is positioned approximately where the peroxide of the FAD(C4α)--OOH intermediate would be expected (Figure 5.8a, green sphere). E44 is also engaged by hydrogen bonds to Y191 and S227 (Figure 5.8a). A nearly identical arrangement for the equivalent residue (E48) is observed in CndH<sup>60</sup>. Indeed, E44 is strictly conserved in FDHs (Figure 5.1). In the CmlS structure, E44 is in an optimal position to act as the ultimate donor and acceptor of protons from the active-site tunnel, delivered in relay fashion by the observed water molecules. The proximity of E44 to the exterior of CmlS (Figure 5.8b) would ultimately facilitate proton transfer with bulk solvent. Long-range proton transfer to and from the active site of PHBH has also been observed, albeit through a pathway different from that proposed here<sup>149</sup>.

5.5 Discussion

As shown in Figs. 5.1 and 5.2, the C-termini of FDHs have little sequence or structural homology after approximately 380 residues. The structure of CmlS provides further evidence that the FDHs follow a pattern of maintaining a conserved halogenation active site, based upon the flavin monooxygenase domain, while tailoring the C-terminal region to direct substrates into the active site to achieve the desired specificity. The structure of CndH<sup>60</sup> pointed to the existence of two structurally distinct families of FDHs that appear to correlate with the type of substrate that is halogenated: a free small molecule (variant A) or a substrate bound to an acyl carrier protein as a thioester (variant B). The former includes the tryptophan halogenases that have well-structured C-termini and completely envelop their substrates upon binding. Residues near the C-terminus are
used to bind tryptophan in specific orientations according to the desired regioselectivity, as illustrated by the contrasting selectivities of PrnA and PyrH. Based on the CndH structure, variant B enzymes have less structured C-termini and considerably more open halogenation active sites, which would be able to accommodate the steric bulk of a substrate bound to an acyl carrier protein via the phosphopantetheine linker. For CmlS, the C-terminus is well resolved, and it appears that the substrate must enter a tunnel before proceeding to the halogenation active site (barring any major conformational changes). In this classification scheme, CmlS would belong to the variant A family and, accordingly, will likely act on a free small molecule. It is noteworthy that assignment of CmlS as a variant A halogenase would not have been possible based on sequence alone, as CmlS has a greater sequence homology to the variant B halogenase CndH (27% identity) than to the variant A tryptophan halogenase PrnA and RebH (~13% identity). This is primarily due to the wide divergence in sequences found in the C-termini of FDHs. Although the biosynthesis of chloramphenicol involves aminoacyl intermediates bound to the acyl carrier domain of CmlP, the sterically restricted access to the active site of CmlS is strongly suggestive that halogenation does not occur on one of these intermediates. This implies that CmlS acts on a simple acyl group, or derivative thereof, which is subsequently transferred to the chloramphenicol precursor on CmlP.

A major question with CmlS is the form of the acyl group that undergoes halogenation. The operon for chloramphenicol biosynthesis encodes an enzyme CmlK that has sequence homology to acyl-CoA synthetases, which, along with CmlS, is essential for the installation of the dichloroacetyl moiety on chloramphenicol. It is therefore possible that CmlS halogenates the free acyl group directly (e.g., acetate, acetoacetate, or malonate) or as the corresponding CoA thioester produced by CmlK. This latter acyl donor is proposed to be utilized by CmlG to acylate a chloramphenicol precursor bound to CmlP. If an enolate mechanism is hypothesized for CmlS (Figure 5.9a), a major catalytic hurdle will be proton abstraction. The $pK_a$ of acetate is 33.5.
Figure 5-9 Proposed mechanism for halogenation by CmlS. (a) Potential reaction of acetoacetyl-CoA with CmlS, with E44 providing general acid catalysis to neutralize the nitranion-leaving group ($pK_a \approx 30$) of the chloramine intermediate. Y350 may stabilize the enolate intermediate. (b) An alternative role for E44 acting as a general acid catalyst for the generation of HOCl.
and that of its thioester is 21\textsuperscript{151}, —making these challenging substrates for CmlS, which lacks an obvious general base residue. Likewise, direct chlorination of an acetyl group on a chloramphenicol precursor is unlikely (in addition to the steric restraints noted above), as this would involve a poor carbon acid, acetamide (pK\textsubscript{a,CH} = 28.4)\textsuperscript{150}. More thermodynamically forgiving substrates would be 1,3-dicarbonyl substrates such as malonate (pK\textsubscript{a,CH} = 13.5 for the diester) or acetoacetate (pK\textsubscript{a,CH} = 8.5 for the thioester). Indeed, the heme-dependent haloperoxidases have been shown to successively halogenate 3-ketoacids, with concomitant decarboxylation, to form α-haloketones and even haloform\textsuperscript{152,153}. Likewise, a manganese-dependent haloperoxidase was shown to brominate malonic acid\textsuperscript{154}. Analogous to CmlS, the active site near the ferryl center of *Caldariomyces fumago* chloroperoxidase where HOCl is produced is relatively hydrophobic\textsuperscript{155}. In these cases, the production of hypohalous acid in the active site is sufficient to halogenate 1,3-dicarbonyl substrates. In the case of CmlS, dichlorination of the corresponding CoA thioesters, followed by a retro-Claisen reaction (for acetoacetyl-CoA) or decarboxylation (for malonyl-CoA), would afford dichloroacetyl-CoA (Figure 5.9a).

The structure of the CmlS active site indicates that halogenation could be facilitated by Y350 by stabilizing negative charge on an enolate intermediate (Figure 5.9a). General base catalysis by H309 to form the enolate may also be possible, but this would require a conformational change that moves this residue closer to the active site. However, sufficiently acidic carbon acids such as acetoacetyl-CoA (pK\textsubscript{a} = 8.5) may obviate the need for a general base residue. The conserved residue E44 on the surface of CmlS appears appropriately positioned to initiate proton donation into the tunnel that typifies FDHs. There are two roles in which a general acid catalyst could serve a conserved catalytic function in the FDH reaction cycle. First, E44 could deliver a proton to the chloramine intermediate formed on K71 using multiple water molecules in the tunnel as a relay (Figure 5.9a). The pK\textsubscript{a} of the conjugate acid of a chloramine is 114
approximately $0^{156}$, indicating that the chloramines intermediate will highly favor its neutral form. However, reaction of the chloramine intermediate with a substrate will likely require acid catalysis to avoid the formation of a highly unfavorable nitranion-leaving group ($pK_a^{LG} \approx 30$). It is well known that chloramines require either specific or general acid catalysis to react with nucleophiles, the type of catalysis dictated by the strength of the nucleophile$^{156,157}$. A proton relay, initiated by E44, could deliver the required proton during the chlorination transition state. It is also possible that E44 could act remotely as a general base to assist the generation of the enolate for subsequent chlorination. However, this role would be unique to CmlS (and possibly unnecessary for activated dicarbonyl substrates), as the FDHs acting on aromatic substrates, including tryptophan halogenases, are highly unlikely to require general base catalysis to restore aromaticity from the carbocation intermediate (with the preceding destruction of aromaticity to form the carbocation likely being the chemically rate-determining step). As a second role, E44 could protonate the proximal oxygen of the FAD(C4α)–OOH intermediate, thereby improving group-leaving ability, as chloride attacks the distal peroxo-oxygen to form HOCl (Figure 5.9b). Such general acid activation of a peroxide intermediate has been proposed for haloperoxidases$^{155,158}$ and horseradish peroxidase$^{159}$. Mutation of this residue in PyrH (E46) to aspartate or glutamine reduced $k_{cat}$ by 60-fold$^{62}$, indicating a substantial contribution to catalysis. This would be consistent with general acid catalysis acting in the rate-determining step, which is substrate chlorination in the case of the tryptophan halogenase RebH$^{61}$. However, it is important to note that the role of E44 is further complicated by its location on a highly mobile loop in the tryptophan halogenases PrnA, RebH, and PyrH, the conformation of which depends on the presence of a substrate (or product) in the halogenation active site, as well as FAD in the flavin binding site$^{128,62,137,160}$. Hence, a strong case can be also be made for the role of this residue in FAD binding dynamics$^{137}$.  

115
The covalent attachment of FAD to D277 of CmlS represents a new type of the posttranslational modification of a protein. Flavin-dependent enzymes with covalent attachments to 8α of the flavin ring via histidine (both N1 and N3 connections), cysteine, and tyrosine side chains are known (reviewed by Heuts et al\textsuperscript{161}). However, no example of an ester linkage to E or D residues has been reported. The mechanism of covalent attachment in these enzymes is believed to occur autocatalytically (and reversibly) with the iminoquinone methide tautomer of the oxidized flavin ring\textsuperscript{161}. In the case of bacterial p-cresol methylhydroxylase, a conformational change is required to initiate covalent attachment to FAD\textsuperscript{162}. Considering that a number of conformational intermediates have been detected in the tryptophan halogenase RebH en route to forming the FAD(C4α)–OOH intermediate\textsuperscript{61}, similar dynamics in CmlS and CndH may explain why these are crystallographically observed to have covalently and noncovalently bound FAD cofactors. Since D277 of CmlS is conserved in FDHs, with tryptophan halogenases being the notable exception, this covalent interaction is also likely to be a conserved feature of the FDH family. It is well known that electron-withdrawing substituents at 8α raise the redox potential of the flavin ring, which is also observed for enzymes with covalently bound flavins\textsuperscript{161}. The attachment of the carboxyl of D277 to 8α would have the effects of making FADH\textsubscript{2} a weaker reducing agent and of making the FAD (C4α)–OOH intermediate a more potent oxidant towards halides. However, perturbation of the flavin redox potential is clearly not an essential element of FDH catalysis, as reduction of O\textsubscript{2} by FADH\textsubscript{2} and oxidation of halide ion by the resulting peroxide intermediate are common to all FDHs, including the tryptophan halogenases (PrnA, RebH, PyrH, etc.), which have a noncovalently bound FAD. A more likely explanation is the advantage of retaining FAD near the enzyme, wherein FAD on the surface of the halogenase could be reduced by a flavin reductase then immediately sequestered into the cofactor binding cleft before autooxidation by solution O\textsubscript{2}. It has been shown through stopped-flow kinetic analysis of RebH that FADH\textsubscript{2} must bind to the enzyme before reducing O\textsubscript{2} to form the FAD (C4α)–OOH.
intermediate\textsuperscript{61}, and that autooxidation of free FADH\textsubscript{2} prior to binding dramatically reduces the reaction yield of 7-chlorotryptophan formed by this enzyme (which is improved under low-O\textsubscript{2} conditions)\textsuperscript{61,57}. The ‘in’ conformation observed for D277 in CmlS and the ‘out’ conformation observed for D285 in CndH (Figure 5.4a) suggest that FAD could readily exit the cofactor binding site, yet remain tethered to the enzyme. A second advantage of a covalently linked FAD may be enhancement of the structural stability of the enzyme. Covalent flavin attachments in monoamine oxidase A, chitooligosaccharide oxidase, and cholesterol oxidase type II have been shown to enhance the soluble expression of these enzymes, as well as their resistance to unfolding or aggregation\textsuperscript{161}. This would confer a selective advantage to FDHs as well.

5.6 Conclusions

The FDHs represent a stunning example of the evolutionary adaptation of a preexisting catalytic scaffold—that of flavin-dependent monooxygenases—to perform oxidative halogenation. With the structure of CmlS, there are now representative FDH structures for three substrate classes: alkyl, phenyl (CndH), and indole (PrnA, RebH, and PyrH). CmlS also further illustrates how the sequence and structural diversity observed in the C-termini of FDHs reflect the diversity of substrates that are halogenated by these enzymes. The bulky C-terminal lobe of CmlS suggests that this is a ‘variant A’ halogenase. Accordingly, the preferred substrate for CmlS will likely be a free small molecule, such as a simple acyl group or the corresponding CoA thioester. In the absence of a dramatic conformational change, the C-terminal lobe of CmlS will prevent a substrate tethered to an acyl carrier protein from reaching the active site. Nevertheless, a conformational change of some sort is required to remove the C-terminal ‘plug’ that completely blocks access to the halogenation active site. Understanding the mechanism of substrate access to the active site will be of considerable interest in future studies. The halogenation active site of
CmlS is very nonpolar and is essentially identical with CndH, a phenol-group specific halogenase, with the exception of Y350, which could be used to stabilize enolate intermediates during the halogenation reaction. By analogy to known FDH and haloperoxidase reactions, we propose that the reactive chloramine intermediate formed by CmlS directs two halogenation events, possibly on a 1,3-dicarbonyl substrate. General acid catalysis is also likely to be important for the reactivity of the flavin peroxide and chloramine intermediates, and we propose that the strictly conserved E44 located at the beginning of the HOCl tunnel fulfills this role. Finally, the observation of a covalent link between FAD and CmlS raises a number of new questions concerning flavin binding dynamics and reactivity in the FDH family. Efforts to reconstitute the halogenation activity of CmlS in vitro are currently underway to address these issues.
Chapter 6
Discussion, summary, and conclusions

6.1 Importance of studying microbial competition mechanisms

Microorganisms constitute the largest proportion of biomass on the planet, and have adapted to surviving in most exotic and inhospitable environments, ranging from permafrost in the Arctic tundra, to thermal vents deep in the oceans\(^3\). Each biological niche has its own survival benefits and challenges, but all share a common theme of nutrient limitation at some time. Microorganisms have evolved a stunning array of mechanisms to give them a competitive advantage over other species and increase their chances of survival. Strategies for nutrient competition range from concerted antibiotic release to kill off competing species, to evolving complex enzymatic pathways that scavenge nutrients from sources unexploitable by other organisms\(^4\).

The chloramphenicol biosynthesis pathway presents an example of a naturally produced antibiotic whose power is harnessed by humans to treat serious infections like typhoid fever, meningitis, and rickettsial infections\(^5\). Natural antibiotics also provide a rich source of new drug development, as they present scaffolds unimagined by the human mind. Synthetic production of the antibiotics is challenging because of the complexity of the compounds and the difficulties associated with directing regio- or stereoselectivity of chemical reactions. Nature has evolved complex pathways of antibiotic-producing enzymes that direct chemical tailoring of their substrates with desired specificity and efficiency. Therefore, understanding the mechanisms of enzymatic halogenations is of great scientific and pharmaceutical interest.

The Pn utilization operon is another example of a survival mechanism that is activated during phosphate limitation\(^2\). C-P lyase is capable of cleaving off phosphate groups from very
stable organophosphonates – a source of phosphorous that is available only to certain species of bacteria, giving them a survival advantage. Not only do Pn occur in nature, but they are also a product of human agricultural and industrial activity. Due to the stability of the C-P bond these compounds accumulate in soil and water, raising environmental concerns and pressing need for the development of remediation strategies. Our understanding of the C-P bond cleaving mechanism and phosphate utilization by bacteria is essential for developing microbiological approaches for pollutant removal.

6.2 Insights into the mechanism of the C-P lyase pathway

Deciphering the mechanisms of complex biological pathways usually involves identifying gene loci responsible for a specific activity through random transposon insertions. Once identified, individual genes are usually disrupted though insertions, and the resulting phenotypes are assessed for function of interest and/or viability. This mutational analysis often allows delineating of which genes are essential for producing a specific activity, and which are accessory or regulatory. The pathways are further studied by growing cells harboring mutations with radioactively labeled precursors. Accumulation of reaction intermediates prior to the step involving the mutated gene indicates the involvement of the corresponding enzyme in the reaction. Often identification of intermediates is possible through NMR and mass-spectrometry, and this allows establishing the sequence of events leading to a specific activity. Deducing function in this manner is not straightforward, however, since insertion mutations may result in a translation of a functional truncated protein, or inadvertent disruption of transcription of downstream genes (called a “polar effect”), thus yielding inaccurate or un-interpretable results. Therefore, mutagenesis should employ various truncations for each gene, and transcription of the
downstream genes must be confirmed to exclude the possibility of identifying the product of disrupted genes as necessary for function.

In the case of C-P lyase pathway, insertional mutagenesis studies were able to establish the gene requirement for C-P bond cleavage as well as cell growth. It was determined that the cleavage of C-P bonds requires seven core enzymes encoded by $\text{phnGHIJKL}$. Expression of these seven core genes alone is not sufficient to support cellular growth on Pn, as measured by cell growth assays on solid plates or in liquid media, supplemented with Pn as a sole source of phosphorous$^{35}$. Utilization of Pn requires the presence of functional $\text{phnCDE}$ genes that encode a membrane transporter. In addition, the $\text{PhnP}$ gene is absolutely required for cell growth on Pn, while disruption of the $\text{phnN}$ gene results in poor growth on Pn. These results prompted a hypothesis that the cleavage of C-P bond might occur on the periplasmic side of the inner membrane of $E. coli$ cells. The hydrocarbon moiety of Pn would be released in the form of a corresponding alkane, while the phosphorous product must be transported into the cytoplasm by the transporter, where it would be funneled into a metabolic pathway by PhnP and PhnN. This data is in agreement with the observation that Pn cannot serve as the carbon source for $E. coli$ growth$^{26}$.

The next step in disentangling the function of the pathway is translation of the genes of interest and comparison of the deduced amino acid sequences with existing sequences in the GenBank. Identification of conserved sequence motifs or general sequence similarity often allows assignment of the putative protein functions. Sequence-based homologies may be insufficient, misleading or too general for assignment of specific function. The hurdle of correct physiological assignment is encountered because superfamily members adopt the same fold and constellation of active site residues for catalysis of chemically related reactions that can be part of physiologically diverse functions.
This exact difficulty was encountered with a sequence-homology base functional assignment of PhnP, which was identified as a member of the metallo-β-lactamase superfamily on the basis of a conserved HxHxDH motif. The highest sequence homology was with phosphodiesterases, and tRNAses in particular, however pre-tRNA processing activity did not appear to be physiologically relevant for C-P lyase pathway. Structure determination of PhnP revealed not only the overall close fold homology to tRNAses, but also perfect conservation of the active site residue identity, positioning, as well as coordination of two metal ions, suggesting that the catalytic mechanism may also be very similar. The main structural difference was a presence of another metal coordination site in PhnP, as well as lack of a long arm used by tRNases to clamp their substrate. Subsequently, PhnP was shown to be hydrolytically inactive against short stretches of unstructured RNA. Screening PhnP with a library of phosphodiesterase substrates provided a number of hits, with highest activity detected toward the regiospecific ring opening of 2’, 3’-cyclic nucleotides. Research revealed that both production of such nucleotides and pho regulon activation are involved in the stringent response pathway. It seemed physiologically plausible that once the cellular demand for phosphorous has been met through the activity of the C-P lyase pathway, PhnP would be required to degrade ppG 2’,3’p, an alarmone derivative, that have been observed to bind to SpoT and adenylsuccinate synthase in an inhibitory fashion\textsuperscript{124}. This would allow SpoT to degrade alarmones and shut down the stringent pathway response, as well as alleviate cell growth inhibition through restoration of AMP synthesis by adenylysuccinate synthase. However, our repeated unsuccessful attempts to observe ligand density upon co-crystalization of PhnP or its mutants with 2’,3’cyclic nucleotides alone or in combination with orthovanadate seemed suspicious, raising a possibility that they are not a physiological substrate of PhnP.

Analysis of a number of functional bacterial phn operons (as implied by the presence of core catalytic enzymes for C-P bond cleavage) revealed that the processing of cyclic phosphates...
appears to be a conserved function in this pathway. The *phnN* gene was almost always followed either by *phnP*, or an *rcsF* gene encoding a cyclic phosphodiesterase from 2H family\textsuperscript{121}. Not only is this phosphodiesterase capable of hydrolyzing 2',3’-cyclic nucleotides, but it also works on 1’,2’-cyclic phosphates as part of tRNA splicing reactions. Analogously, the physiological role of PhnP may also include catalysis of 1’, 2’-cyclic phosphates. In a few cases where *phnP* or *rcsF* genes were absent, cyclic phosphodiesterase activity may have been encoded by a promiscuous hydrolase from other phosphodiesterase families.

Mutational analysis and structural information may suggest a function that can often be tested *in vitro* through the use of a generic or non-physiologically relevant substrate, but deducing a true biological function is a much more challenging endeavour, requiring a physiologically-relevant biochemical assay. When designing an assay to test a function of an enzyme from a complex pathway, many other factors must be taken into account, such as the requirement for an intact membrane or particular cellular localization, requirement for more than one enzyme or assembly of an enzymatic complex onto a scaffold, and the requirement for external cofactors, redox agents or carrier proteins.

In the case of PhnP, deducing its physiological function would be aided by determining the reactions performed by the neighbouring Phn proteins, but biochemical characterization of the C-P lyase pathway is seriously hindered by the lack of a cell-free assay. Several lines of evidence point to the fact that the C-P lyase pathway may require intermediates of other intact pathways for utilization of phosphate moiety. Accumulation of α-1-(ethylphosphono)ribose (EtPnR) in the cell media was detected in the cryptic *E. coli* mutant grown on a mixture of Pi and \[^{32}P\]-ethylphosphonate (EtPn)\textsuperscript{25}. The same study reported that mutants incapable of C-P bond cleavage failed to accumulate ribosylated EtPn, suggesting that ribosylation occurs as part of the C-P bond cleavage process. PhnN phosphorylates ribose-1,5-bisphosphate (R1,5P) to produce 5-
phospho-D-ribosyl α-1-diphosphate (PRPP), which is a precursor for NAD biosynthesis, as well as a purines, pyrimidines and aromatic amino acids histidine and tryptophan\textsuperscript{34}. In addition, PhnC, PhnK and PhnL have sequence homology to the nucleotide-binding domains of ABC transporters\textsuperscript{30}. Taken together these observations suggest that nucleotides, nucleotide derivatives or other ribose-containing moieties may act as acceptors of Pn and this transfer is essential for C-P bond cleavage to take place. If the process happens in the periplasm, the ribose would most likely be dephosphorylated. This is anticipated because compounds are usually dephosphorylated in the periplasm prior to transport inside the cell. The uptake of phosphorylated molecules, like glycerol phosphate, is less common and requires the use of specific transporters\textsuperscript{163}. Also, dephosphorylated nucleotide occurrence in the periplasm has been reported, but the mechanism for such export is not understood.

Recent work by our collaborator, Dr. Hove-Jensen, provided another piece of the puzzle that may place phosphodiesterase activity of PhnP into the context of a C-P lyase pathway. \textit{E. coli} cells used in this experiment harbored a \textit{pstS} mutation (a phosphate transporter permease subunit) that enabled constitutive expression of the C-P lyase pathway regardless of the amount of Pi present. Radiolabeled Pi was added to the minimal media, which was subsequently supplemented with organophosphonates; this enabled growth of otherwise non-viable \textit{phn} mutants, as well as allowed analysis of accumulated metabolic intermediate through thin layer chromatography (TLC). The TLC analysis of the culture media revealed accumulation of several intermediates when \textit{phnH}, \textit{phnP} or \textit{phnN} genes, but not \textit{phnO} gene, were disrupted. Lack of intermediate accumulation in the \textit{phnO} mutant is consistent with the role of PhnO as an accessory protein. Interestingly, \textit{phnH} and \textit{phnN} mutants had the same pattern of radiolabeled intermediates, while \textit{phnP} mutant possessed one additional spot\textsuperscript{164}. This result suggests that in absence of PhnN, the C-P lyase intermediates may be processed through another auxiliary pathway, which would still allow for phosphate entrance into the metabolic pool, albeit slower.
than the WT, and this is reflected by poor growth of *phnN* mutants on Pn. In contrast, no other enzyme or pathway seems to be able to substitute for PhnP, strengthening the case for a critical role of PhnP for cellular growth on Pn.

Further analysis of this additional intermediate accumulated by the *phnP* mutant revealed that this compound “S” is converted to a product “P” when purified PhnP was added to the media. The reaction is specific for PhnP, as no other Phn enzyme could convert compound “S” into “P”. It is important to note that conversion of compound “S” to compound “P” proceeds without addition of external cofactor or substrate. This is indicative of a reaction involving rearrangement of the substrate, such as hydrolysis. The separation of compounds on the polyethyleneimine TLC medium, an anion exchanger, occurs primarily due to charge differences. It can be concluded that compound “P” must have at least one more negative charge than compound “S”, which displays slower migration on the TLC plate. This observation is consistent with the cyclic phosphate ring opening activity of PhnP with several 2’, 3’ cyclic nucleotides, as reported in Chapter 3. However, compound “S” is different from the 2’, 3’ cyclic nucleotides, as it migrates much slower on a TLC plate, suggesting that the true physiological substrate of PhnP may have more negative charges.

Guided by the previous finding of ribosylated Pn intermediates and by the fact that PhnP and PhnN usually occur together in the operons, Dr. Hove-Jensen hypothesized that PhnP phosphodiesterase activity may provide a substrate for the PhnN reaction. The formation of cyclic phosphates from PRPP has been known since 1958. *In vitro* such a reaction spontaneously occurs at alkaline pH in the presence of Ba\(^{2+}\) ions and the product is 5-phosphoribose-1,2-cyclic phosphate (5PR1,2cP). When PhnP was added to 5PR1,2cP, the time course of ribose-1,5-bisphosphate formation was observed by \(^{31}\)P NMR (Hove-Jensen, unpublished results). Mirroring the observations of the synthetic substrate assay, the peak
corresponding to a 1’,2’ cyclic phosphate was also observed in the $^{31}$P NMR spectra of the culture medium of phnP mutant. The most prominent peak in the spectrum was that corresponding to the phosphate of the MePn, and another small peak with a larger chemical shift was also observed, likely corresponding to ribosylated Pn. Interestingly, no peak for 5’ phosphate was observed in the culture medium, which coincides with the expected dephosphorylation of the ribose moiety in the periplasm. When PhnP was added to this culture medium, the 1’,2’ cyclic phosphate peak decreased in intensity, while a new peak corresponding to the 1’phosphate has appeared. The identity of this new peak was confirmed by spiking the reaction mixture with ribose-1-phosphate (R1P), which cause this peak to increase in intensity.

The results from this series of experiments suggest that the physiological substrate of PhnP is 5-phosphoribose-1,2-cyclic phosphate.

The formation of such a cyclic phosphate in vivo is likely if we assume that C-P bond cleavage occurs on ribosylated Pn. Such a reaction may be more energetically favourable than on alkylphsophonate alone, or such a molecule may be better accommodated in the active site of the enzyme or enzyme complex. Methyl radical departure will result in the formation of electrophilic meta-phosphate on C$_1$. The nucleophilic 2’ hydroxide will promptly react with metaphosphate to form a 1’,2’ cyclic phosphate. This intramolecular reaction would be highly favoured over the metaphosphate attack by water due to the high effective molarity caused by the physical proximity of these two groups. Also, the enzymatic active site where C-P bond cleavage might occur would exclude the bulk solvent, favouring the intramolecular reaction.

Our proposed mechanism of Pn utilization is shown in Figure 6.1. It begins with a nucleotide (either common or rare) or its derivative forming an ester bond with a Pn. The core enzymes of C-P lyase would cleave off the alkyl moiety, resulting in the release of a corresponding alkane and concomitant formation of a ribose cyclic phosphate. Cells deficient in
Figure 6-1 A proposed model for Pn utilization. The first step of the process is hypothetical, while the reactions needed to carry out C-P bond cleavage have not been determined in detail thus far. Reactions carried out by PhnP and PhnN enzymes have been demonstrated \textit{in vitro}, The process allowing R1,5P to enter metabolic pool is not known.
allow for NAD biosynthesis through another pathway, while the phosphorous from PhnN substrate might enter the cell’s metabolic pool through another pathway. This could be a reason why poor growth is observed for phnN mutants, while phnP mutants fail to grow on Pn. It remains to be conclusively demonstrated that this 5PR1,2cP compound corresponds to the compound “S” observed on the TLC plates of culture media from phnP mutants, however the evidence presented here makes a strong case for our hypothesis.

6.3 Insights into the chloramphenicol biosynthesis pathway

In the case of the chloramphenicol biosynthesis pathway, mutagenesis and analysis of radiolabeled intermediates allowed assignment of the events leading from the shikimate pathway intermediate to the formation of p-aminophenylalanine (PAPA)\textsuperscript{72}, which is considered to be a substrate of the first committed step towards formation of Cm. The sequence of events leading to dichlorination of the acetyl moiety on the Cm precursor is unknown, with one of the obstacles being ambiguity of the putative functional assignment of the remaining genes. Sequence analysis of the operon encoding Cm biosynthesis revealed that CmlS is the only enzyme with potential halogenation function\textsuperscript{73}.

Structure solution of CmlS did confirm that it is an FAD-binding protein with close structural homology to halogenases PrnA\textsuperscript{128} and CndH\textsuperscript{60}. The major structural differences appeared in the C-terminus, and are thought to reflect substrate specificity. Alignment of the CmlS active site with those of PrnA and CndH revealed that both CmlS and CndH lack the critical catalytic base residue present in PrnA, raising questions about the reaction mechanism. In the case of CndH, which acts on a substrate bound to a protein carrier, it is hypothesized that the catalytic base would be supplied by either a carrier protein or by the disordered C-terminus, which would become more structured upon the docking of the carrier protein. The bulky C-
terminal domain of CmlS makes carrier docking questionable, while the C-terminus blocks the entrance to the active site and will have to undergo a conformational change to bring its only potential general base residue H561 to face the active site. Absence of density for the chloride ion, observed in structures of other halogenases, is another hurdle in elucidation of the catalytic mechanism of CmlS. This is likely a crystallographic issue, since the protein was crystallized in PEG and salt addition adversely affected crystal formation. Still, the possibility of non-functional protein cannot be excluded.

Gene knockout studies in *S. venezuelae* have revealed that in addition to CmlS, the CmlK protein is also necessary for the formation of dichloroacetyl group. CmlK has sequence homology to acyl Co-A synthases and would potentially activate the halogenation substrate or product for transfer onto the Cm precursor. Mutants lacking both cmlK and cmlS were shown to incorporate a propionyl group in place of the dichloroacetyl group, yielding cornycenecin II instead of chloramphenicol. It was also determined that a number of later enzymatic tailoring steps of the Cm precursor are carried out on the carrier protein CmlP, but it is not known whether chlorination (or chlorinated group transfer) occurs on the CmlP-bound intermediate or after this intermediate is released from the carrier.

Three scenarios are possible for the sequence of events (Figure 6.2). In the first case, CmlS carries out two halogenation events on a free substrate to introduce two chlorines. This hypothesis was tested on a number of plausible substrates, like acetate, chloroacetate (a potential reaction intermediate after the first halogenation step), malonate and acetoacetate. As discussed in detail in Chapter 5, halogenation of free carbon acids would be mechanistically challenging due to the relatively high pKa of such compounds. A proton abstraction in the first step of the reaction would be a major catalytic hurdle, since CmlS lacks an obvious general base in the active site. Substrates containing a 1,3-dicarbonyl moiety would be somewhat more amenable to such a
Figure 6-2 Possible sequence of events leading to formation of dichloroacetyl group of Cm (Figure generously provided by Dr. David Zechel).
reaction due to lower pKa. Halogenation of such small molecules would be preferable from the structural perspective, since the active site is located in a deep pocket that is obstructed by the bulky C-terminal domain. After the halogenation reaction CmlK would convert a halogenated substrate to a corresponding acyl CoA derivatives; CmlH would then transfer the dichloroacetyl moiety onto a CmlP-bound Cm precursor.

In the second scenario adenylation of acetate or similar molecules by CmlK might be required to create an appropriate substrate for CmlS. Acyl CoA thioesters of free carboxylic acids would have an advantage over the free acids because the presence of coenzyme A would lower the pKa’s of the corresponding groups, thus making them easier chlorination substrates. This hypothesis was tested by using Co-A derivatives of the substrates mentioned in the first scenario, again with no detected reaction. In the absence of any conformational change, chlorination of such substrates might be less favoured from the structural point of view due to a potential steric clash with the bulky C-terminal domain of CmlS. As in the first scenario, CmlH would subsequently catalyze the transfer of the dichloroacetyl moiety onto a CmlP-bound Cm precursor.

Finally, the third scenario involves halogenation of the acetyl group or similar group on the CmlP-bound precursor. This hypothesis is challenging to test, as it would require expression of CmlP, synthesis of several potential reaction intermediates, and their covalent attachment to the carrier protein. The crystal structure of CmlS argues against this possibility, as the steric clash between the bulky C-terminal domain and the peptidyl carrier protein will hinder the substrate’s access to the active site. However, a conformational change in solution cannot be excluded. In this case the C-terminal domain would swivel, like a lid, allowing for approach of CmlP and providing more room for a substrate tethered to CmlP by a long phosphopantetheine arm. This possibility can be tested by mutating two small polar uncharged residues to two
cysteine residues in the conserved FAD binding domain around helix α9, and around the area of the C-terminal domain contact with the FAD domain in the strands β21 or β22. Reduced cysteines will allow for a wild-type-like movement of the C-terminal lid and, if there is any, the protein will appear as having more extended conformation. Oxidation of cysteins will result in a disulfide bridge formation that would tether the lid in place. Small angle X-ray scattering experiments can then be used to determine if conformational flexibility exists in the hinge area connecting the C-terminal domain to the core of the protein by comparing the shapes of the two constructs. Of course, these shape differences may be too subtle to detect, or may occur only in the presence of the interacting partner, like a carrier protein.

Despite numerous trials, no chlorination was detected by $^1$H-NMR, $^{13}$C-NMR and ESI-MS for any of the substrates tested (Ryan Latimer, personal communication). Since the true substrate is yet to be identified, we hypothesized that co-crystallization with the potential product of the reaction might provide some clues to the mode of binding. Out of several potential products tested, co-crystallization with dichloroacetate yielded diffraction quality crystals, however we were unable to observe any ligand density in the active site of CmlS at 2.1 Å resolution.

We then questioned if the C-terminal tail blocking the entrance to the active site in the crystal structure was representative of its state in solution, or was simply a crystallographic artifact. As discussed in Chapter 5, an extensive number of tertiary interactions of this tail with the rest of the protein suggested that this is a structurally, and possibly functionally, important part of the enzyme. In this case, the tail would have to be displaced to allow substrate entrance, and for the halogenation reaction to occur. The role of the C-terminal tail was tested by creating truncation mutants of the last 15 and 8 residues resolved in the crystal structure. The first mutant was insoluble, reinforcing the suggested role of the C-terminal tail in protein stabilization. The
preliminary activity trials with the latter mutant were not successful (Ryan Latimer, personal communication), leaving the possibility that the correct substrate is needed. Additionally, the tail may be involved in the reaction by providing reaction intermediate stabilization or performing some other function.

Another possibility that may account for the lack of activity is the requirement for FAD reduction. Flavin-dependent halogenases utilize a flavin cofactor for transfer of electrons from NADH to molecular oxygen (reviewed by Blasiak\textsuperscript{53}). Each reaction cycle begins with regeneration of oxidized FAD into FADH\textsubscript{2}, which bacteria achieve using NADH-dependent reductases. In such systems FAD must freely diffuse between halogenase and reductase to complete the cycle. Unfortunately, freely diffusing FADH\textsubscript{2} can be spontaneously oxidized by oxygen from the solvent, therefore reducing the efficiency of the cycle. CmlS revealed an unprecedented covalent modification of FAD cofactor by a covalent bond with aspartate residue, which, curiously, appears to be conserved in a number of halogenases that act on the acyl-carrier bound substrates, but not in FADHs acting on free small molecules. Based on the solvent-exposed conformation of the equivalent residue in the close structural homologue CndH, we hypothesized that the reduction of covalently attached FAD in CmlS might occur by a polypeptide backbone twist that would flip the D277-FAD into the solvent, where FAD would get reduced by a NADH-dependent reductase. The reduced FAD would then promptly get sequestered back into CmlS, decreasing the chances of non-specific oxidation in the solvent. Elimination of free diffusion step between halogenase and reductase would increase the overall efficiency of the system.

Activity assays carried out with \textit{P. fluorescens} halogenase PrnA, a close structural homologue of CmlS, demonstrated that FAD reduction could be achieved using the reductases SsuE from \textit{E. coli} or Frp from \textit{T. thermophilis}\textsuperscript{126}. Similarly, CmlS activity assays (Ryan Latimer,
personal communication) were carried out in presence of a Fre flavin reductases from *E. coli*, and it was assumed that Fre would be able to carry out many cycles of FAD reduction. Free FAD was also added to the reaction, since purified CmlS contained only ~50% FAD, and not all of it was covalently bound, as shown by denaturation studies described in Chapter 5. At this point it is not known if externally added FAD becomes covalently attached to CmlS (presumably through an autocatalytic mechanism), whether this reaction is reversible, or what is the time scale for either process. If additional FAD remained non-covalently bound to CmlS within the halogenation reaction experiment time scale (6 hours), the free diffusion of that proportion of FAD and reduction by Fre in solution should not be a problem. If, however, a predominant population of CmlS develops a covalent attachment with added FAD quickly, the hurdle of reduction becomes more relevant. It is possible that due to a covalent linkage of the FAD, a non-specific Fre reductase is not able to approach CmlS closely enough to carry out the reduction.

The precedent for protein-protein interaction during FAD reduction has been shown for three pairs of oxygenases and their partner reductases. For example, Lee and Zhao reported that contact between PrnD oxygenase and PrnF reductase was required for efficient FAD reduction\(^{166}\). This interaction was not required for an oxygenation reaction to take place, and the efficiency of the reaction when the protein partner is separated by semi-permeable membrane was only ~5% of that observed under normal conditions. These results show that the reaction rate is limited by the spontaneous oxidation of FAD during diffusion.

The possibility of CmlS requirement for a specific flavin reductase is complicated by the fact that the Cm biosynthesis cluster does not encode such an enzyme. Similarly, biosynthesis of antibiotic pyrrolnitrin in *P. fluorescens* requires only four genes, *prnABCD*, none of which encode a reductase\(^ {167}\). *E. coli* cultures supplied with a plasmid carrying *prnABCD* genes are able to produce pyrrolnitrin, indicating that a non-specific *E. coli* reductase can substitute for a native one and reduce FAD used by halogenases PrnA and PrnC. This is not surprising, as in this case
FAD is bound to PrnA non-covalently and is able to diffuse freely. Curiously, the prn gene cluster contains a flavin reductase PrnF. Its activity is not necessary for PrnA and PrnC halogenases, but is required to enhance the arylamine N-oxygenation carried out by PrnD, as discussed above.

### 6.4 Conclusion

We have demonstrated that insights from structural investigation of enzymes of unknown or putative function can guide biochemical characterization and placement of the enzymatic activity in the physiological context. In particular, the structures of PhnP, a cyclic phosphodiesterase from the C-P lyase pathway, and CmlS, an FAD-dependent halogenase from the Cm biosynthesis pathway are presented.

In the case of PhnP we were able to demonstrate promiscuous phosphodiesterase activity towards a number of PDEse substrates, however, the inability to observe these ligands in crystal structures suggested that these compounds may not be true biological substrates. Preliminary crystallization trials with the newly identified biological substrate suggest that this time we may be more successful in trapping a relevant complex. Cocrystallization with a non-physiological substrate in combination with mutagenesis studies allowed assignment of the function of conserved residues and provided the basis of binding cooperativity. The power of crystallography in combination with ICP-MS and biochemical activity assays allowed assignment of metals to the unique metal sites of PhnP, while in vivo studies combined with NMR allowed identification of the potential biological substrates. Amalgamation of our results with existing knowledge and most current research has allowed us to propose a plausible mechanism for Pn utilization. This model represents a crucial step in deciphering the mechanism of Pn utilization, which has resisted characterization for decades.
In the case of CmlS, the structural solution confirmed its putative role assignment as a halogenase and also allowed identification of a novel covalent modification of the FAD cofactor. However, lack of substrate and chloride ion in the structure highlight the difficulties associated with trapping these complexes crystallographically. In addition, the difficulties encountered with determining the biological substrate of CmlS or even showing a halogenation reaction likely reflect that CmlS requires a specific form of an intermediate or interaction partners that have not yet been possible to obtain in vitro.

Despite these obstacles, the studies presented here have made a significant contribution to our understanding of Pn utilization by bacteria. Given the environmental concerns raised by accumulation of toxic Pn as byproducts of human activity, our understanding of the C-P lyase mechanism is critical for designing bioremediation programs for toxic chemical removal from the environment. In addition, we have laid the foundation for further mechanistic analysis of halogen introduction into chemically inert alkyl groups. Understanding such mechanisms would be beneficial for the areas of new drug design and synthesis – an area of research that is crucial to human health given the rising resistance of infection bacteria to existing antibiotics.
Appendix A
Additional data

Preface:
This Appendix contains the data that was not shown in Chapters 2 and 4 due to the manuscript style format of the chapters.
1. **Purification of PhnP**

PhnP was purified following the procedure described in Chapter 2, Section 2.3.1. The crude lysate and elution fractions were analyzed by 15% SDS-PAGE and visualized by Coomassie Blue G-250 staining.
Figure 1 Purification of PhnP. (A) 15% SDS-PAGE of the elution fractions from the Ni–NTA agarose column; L = crude lysate supernatant from E. coli; M = BioRad Precision Plus protein marker. (B) Size exclusion chromatogram of PhnP. Inset: 15% SDS-PAGE of the indicated fractions from Superdex 200 column. (The figure is courtesy of Shumei He).
2. **Purification of CmlS**

CmlS was purified following the procedure described in Chapter 4, Section 4.3.1. The crude lysate, flow through, insoluble cell pellet and elution fractions were analyzed by 12% SDS-PAGE and visualized by Coomassie Blue G-250 staining.
Figure 2 Purification of CmlS. 12% SDS-PAGE of the elution fractions from the Ni–NTA agarose column; M = Fermentas protein marker; L = crude lysate supernatant from *E. coli*; FT = lysate after loading on the Ni column; P = insoluble fraction. (The figure is courtesy of Ryan Latimer)
References


H. Bujard, unpublished work


148


