Structural and functional characterization of the aminoacetone utilization microcompartment from *Mycobacterium smegmatis* MC2 155

by

Evan Mallette

A Thesis

presented to

The University of Guelph

In partial fulfilment of requirements

for the degree of

Doctor of Philosophy

in

Molecular and Cellular Biology

Guelph, Ontario, Canada

© Evan Mallette, May, 2019
ABSTRACT

Structural and functional characterization of the aminoacetone utilization microcompartment from *Mycobacterium smegmatis* MC2 155

Evan Mallette  
University of Guelph, 2019  
Advisor: Dr. Matthew Kimber

Bacterial microcompartments are proteinaceous complexes made by bacteria, which metabolize volatile or cytotoxic chemicals, by sequestering a series of reactions within a selectively permeable shell. This thesis focused on characterizing the catalytic functions and thus the metabolism of an aminoacetone utilization microcompartment (AAUM) found in *Rhodococcus* and *Mycobacterium* species, using the prototypical operon found in *Mycobacterium smegmatis* MC2 155 as a subject for investigation.

Of the four enzymes associated with the AAUM, catalytic functions and structures were determined for two of the enzymes from *M. smegmatis* and for a homolog of a third enzyme. The first enzyme characterized was a stereospecific alcohol dehydrogenase catalyzing the reduction of 1-amino-2-propanone forming S-(+)-1-amino-2-propanol. The second enzyme was characterized as a 1-amino-2-propanol O-kinase with a preference for the S-isomer, able to also phosphorylate the R-isomer and amino-alcohols of varying lengths. An ortholog of the AAUM associated class-III aminotransferase from *Mesorhizobium loti* was characterized as a phosphopropanolamine phospholyase, forming ammonia, inorganic phosphate, and propionaldehyde. The remaining enzyme from the AAUM is annotated as a coenzyme A acylating aldehyde dehydrogenase, proposed to acylate coenzyme A with propionaldehyde produced by the phospholyase enzyme resulting in propionyl-CoA production. From the
determined enzyme functions, I propose a metabolic pathway for the AAUM, converting aminoacetone to propionyl-CoA for use in central metabolism.

The structures of the four shell proteins from the AAUM were determined by X-ray crystallography; proposed to form an icosahedral shell, with hexagonal oligomers forming the facets capped at the apices by pentagonal oligomers. The hexagonal hexameric bacterial microcompartment shell protein (BMC-H) formed a negatively charged pore, proposed to function in transport of a positively charged molecule across the shell. Both hexagonal trimeric bacterial microcompartment shell proteins (BMC-Ts) formed stacked trimeric rings with loops forming a tight interface at the center of each ring, proposed to undergo conformational shifts to open a large diameter pore for intermittent large molecule exchange. The pentagonal pentameric bacterial microcompartment shell protein (BMC-P) formed a roughly pyramidal oligomer proposed to fulfill an apex capping function. Using current modeling techniques, we were able to propose an architecture for the microcompartment shell.
ACKNOWLEDGEMENTS

During my thesis research I have had the opportunity to work with many individuals, without whom my research would have been substantially more difficult. I would like to first thank my parents Gerry and Joanne Mallette for their continual support on all matters scientific or otherwise, you have allowed me to follow my every aspiration and my gratitude is immeasurable. I would like to also thank my sister, Michelle Mallette, and my brother-in-law, Tim Boothroyd, as well as my ever curious niece, Ellie, and their darling of a pup, Molly, for always supporting my endeavors and welcoming me to your home for a glimpse of respite in times of stress. My extended family have always shown interest in my research giving me opportunity to go on long winded rants about whatever I was researching at the time and for that I commend all of you. The continual support I receive on a daily basis from my lovely girlfriend, Brenda Christensen, and for enduring my exasperation, especially during the process of preparing my thesis manuscript, I thank you wholeheartedly and hope that I can provide an equal level of support through all of your endeavors to come. I’d like to also thank my friend Ryan Tole, who has been like a brother to me, you certainly have influenced who I am today.

I cannot think of a way substantial enough to thank my supervisor, Dr. Matthew Kimber, but I offer my gratitude for everything you have allowed me to do over the years of my thesis research. The continual support and willingness to allow me to investigate my theories, despite their sometime obscure direction, has allowed me to become a confident researcher and you have supplied me with a tremendous amount of knowledge that could only be acquired through years of experience in molecular biology. The often casual conversation of Sci-Fi or comical scientific anecdotes of the nerdy type made me as well as everyone else in the lab feel as though we were a close-knit group exploring the curiosities of biochemistry. The techniques in crystallography and sometimes elaborate theories will stay with me for a lifetime, and I aspire to pass as much of the information I have learned on to others in hopes that they too will be inspired to investigate biology on a molecular scale.

I would like to thank my advisory committee members, Dr. Stephen Seah, Dr. Lucy Mutharia, and Dr. Cezar Khursigara, for supporting me and contributing to my project over the years. I would also like to thank Dr. Chris Whitfield, for allowing me to work on collaboration projects with his lab and supporting me in my transition from graduate studies into post-doctoral research.

Although I would like to name everyone I have had the privilege to interact with individually from the MCB department, over the years there have been so many that I would likely unintentionally leave too many people out. Instead I would like to thank a selection of individuals who have been memorable and influential to me over the years. First I would like to thank all of the staff who have always provided a welcoming patient experience while figuring out how things are done around the Science Complex, including Leanne Krick and Karen Ingram from chem-stores, David Leishman in wash-up, as well as Susan Casey and the entire purchasing department staff. I would like to also thank those members of the administrative services for helping me navigate all the bureaucratic interstices, including but not limited to Laleh Hatefi, Bertilla Moroni, Vanessa Breton, and Cate Mennega. As well I would like to thank all the staff who maintain the facilities who I do not feel receive the appreciation they deserve for the work we take for granted on a daily basis.
Now for the far too numerous associate students that I have worked alongside. From day one and seemingly unshakably still Charles Wroblewski and Dave Martynowicz, despite all the shenanigans, I would not be myself without having known the two of you and I thank you for that. Graduate students and undergraduate students alike from the Kimber and Graether labs (past and present) the daily conversations and sometimes confusing antics have kept me nearly sane throughout my thesis so I’d like to thank you all, but I would name a few of you seeing as if I didn’t I don’t think you’d let me live it down so here we go; Matt Clarke a fellow whisky lover and now a full grown adult, cheers and best of luck; Tom Keeling, probably the realest guy I’ve met and will very likely be running a major biologicals company sooner than later; Kanu Singh just a great guy to hang out with and talk science or play euchre with although you will likely lose to him; Patrick Ryan, he might have two first names but that’s ok since he was twice the friend to anyone he hung out with; Taylor Forrester, it has only been a couple years that I’ve know you but the daily antics make sure that we have fun in the lab. As for the Merrill Lab I would like to acknowledge everyone but I’ll have to thank a selection of you by name; Amanda Poole, Bronwyn Lyons, Stephanie Carlin, Kayla Heney, Madison Turner, and Oliver Tremblay, we have all shared way too much information amongst ourselves and spent seemingly endless time together, and for that I may as well consider you family and I appreciate all of you.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>x</td>
</tr>
<tr>
<td>LIST OF EQUATIONS</td>
<td>xi</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>xii</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Bacterial Microcompartments</td>
<td>1</td>
</tr>
<tr>
<td>1.1.1 Overview</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Microcompartment Shell</td>
<td>2</td>
</tr>
<tr>
<td>1.2.1 Superstructure Architecture</td>
<td>2</td>
</tr>
<tr>
<td>1.2.2 BMC-H</td>
<td>2</td>
</tr>
<tr>
<td>1.2.3 BMC-T</td>
<td>3</td>
</tr>
<tr>
<td>1.2.4 BMC-P</td>
<td>4</td>
</tr>
<tr>
<td>1.3 Characterized Microcompartment Metabolisms</td>
<td>7</td>
</tr>
<tr>
<td>1.3.1 Carboxysome</td>
<td>7</td>
</tr>
<tr>
<td>1.3.2 Ethanolamine and Propanediol Utilization Microcompartments</td>
<td>7</td>
</tr>
<tr>
<td>1.3.3 Partially Characterized Glycyl-radical Microcompartments</td>
<td>8</td>
</tr>
<tr>
<td>1.4 Rhodococcus/Mycobacterium Microcompartment</td>
<td>10</td>
</tr>
<tr>
<td>1.4.1 Operon Complement</td>
<td>10</td>
</tr>
<tr>
<td>1.4.2 Mycobacterium smegmatis MC2 155 RMM</td>
<td>10</td>
</tr>
<tr>
<td>1.4.3 Shell Proteins</td>
<td>12</td>
</tr>
<tr>
<td>1.4.4 Short-chain Alcohol Dehydrogenase/Reductase (SDR)</td>
<td>12</td>
</tr>
<tr>
<td>1.4.5 Aminoglycoside Phosphotransferase</td>
<td>13</td>
</tr>
<tr>
<td>1.4.6 Aminotransferase Class-III</td>
<td>15</td>
</tr>
<tr>
<td>1.4.7 Acylating Aldehyde Dehydrogenase</td>
<td>17</td>
</tr>
<tr>
<td>1.4.8 Encapsulation Peptides and the Protein Containing a Domain of Unknown Function</td>
<td>18</td>
</tr>
<tr>
<td>1.5 Non-microcompartment associated RMM like operon</td>
<td>19</td>
</tr>
<tr>
<td>1.6 Thesis objectives and justification of research</td>
<td>21</td>
</tr>
<tr>
<td>2 A complete structural inventory of the Mycobacterial microcompartment shell proteins constrains models of global architecture and transport</td>
<td>24</td>
</tr>
<tr>
<td>2.1 Abstract</td>
<td>25</td>
</tr>
<tr>
<td>2.2 Introduction</td>
<td>25</td>
</tr>
<tr>
<td>2.3 Results</td>
<td>30</td>
</tr>
<tr>
<td>2.3.1 The structure of MSM0273</td>
<td>30</td>
</tr>
<tr>
<td>2.3.2 The structure of MSM0272</td>
<td>35</td>
</tr>
<tr>
<td>2.3.3 The structures of MSM0271 and MSM0275</td>
<td>37</td>
</tr>
</tbody>
</table>
5 Structural And Functional Characterization of a Homolog of the *Mycobacterium* Microcompartment O-Phosphopropanolamine Phospholyase from *Mesorhizobium loti*  
5.1 Introduction .......................................................................................................................... 129
5.2 Materials and methods ........................................................................................................... 132
  5.2.1 *Mycobacterium smegmatis* MC\(^2\) 155 genomic DNA purification .................................. 132
  5.2.2 *Mesorhizobium loti* MAFF 303099 genomic DNA purification ........................................ 133
  5.2.3 Molecular cloning ............................................................................................................ 133
  5.2.4 Protein expression ............................................................................................................ 134
  5.2.5 Selenomethionine protein labeling .................................................................................... 134
  5.2.6 Protein purification .......................................................................................................... 135
  5.2.6 Crystallography ............................................................................................................... 135
  5.2.7 Phosphoethanolamine phospholyase activity assay ......................................................... 136
  5.2.8 Phosphopropanolamine phospholyase activity assay .................................................... 137
5.3 Results ................................................................................................................................... 137
  5.3.1 Protein purification .......................................................................................................... 137
  5.3.2 Enzyme activity ................................................................................................................ 138
  5.3.3 Structural overview ......................................................................................................... 138
  5.3.4 PLP binding site ............................................................................................................... 144
  5.3.5 Substrate binding site ....................................................................................................... 144
5.4 Discussion .............................................................................................................................. 145
6 Conclusions .............................................................................................................................. 148
  6.1 Structural characterization of the aminoacetone utilization microcompartment shell proteins ................................................................. 148
  6.2 S-(-)-1-amino-2-propanol alcohol dehydrogenase ............................................................... 151
  6.3 1-amino-2-propanol kinase ................................................................................................ 152
  6.4 O-Phosphopropanolamine phospholyase ........................................................................... 153
  6.5 Aminoacetone utilization microcompartment metabolism ................................................ 155
  6.6 Non-microcompartment aminoacetone utilization metabolism ....................................... 159

REFERENCES ........................................................................................................................... 161
LIST OF TABLES

Table 1.1 Genes of the Mycobacterium smegmatis MC2 155 RMM operon .......................... 11
Table 1.2 Genes of Mycobacterium smegmatis MC2 155 non-microcompartment aminoacetone utilization operon .................................................................................................................. 20
Table 2.1 X-ray data collection and refinement statistics ........................................................................ 62
Table 3.1 Cloning and Site Directed Mutagenesis PCR Primer Table .............................................. 69
Table 3.2 Steady-State Kinetic Parameters for Mycobacterium smegmatis MC2 155 APDH variants\textsuperscript{a} ........................................................................................................................................... 78
Table 3.3 X-ray data collection and refinement statistics ........................................................................ 84
Table 4.1 Steady-state kinetic parameters for Mycobacterium smegmatis MC2 155 APK variants\textsuperscript{a} ........................................................................................................................................... 106
Table 4.2 Cloning and mutagenesis primer sequences .......................................................................... 123
Table 4.3 X-ray data collection and refinement statistics ........................................................................ 127
# 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>Microcompartment shell proteins</td>
<td>4</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>Diagrammatic representation of the metabolisms of anabolic microcompartments (carboxysomes) and catabolic microcompartments</td>
<td>9</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>Architecture of bacterial microcompartments</td>
<td>28</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>Overall structures of RMM shell proteins</td>
<td>31</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>Details of potential transport routes through RMM shell proteins</td>
<td>33</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>Crystal packing and potential interactions mediated by RMM shell proteins</td>
<td>39</td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>Modeling the RMM vertex by docking</td>
<td>47</td>
</tr>
<tr>
<td>Figure 2.6</td>
<td>Overall model for the organization of the RMM shell</td>
<td>58</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Michaelis-Menten nonlinear regression fitting of APDH&lt;sub&gt;MSM0269&lt;/sub&gt;</td>
<td>75</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>Nonlinear regression of APDH&lt;sub&gt;MSM0269&lt;/sub&gt; and APDH&lt;sub&gt;MSM0779&lt;/sub&gt; kinetics data</td>
<td>76</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>Structure of APDH</td>
<td>81</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>Details of interactions in the NADP+ binding site</td>
<td>83</td>
</tr>
<tr>
<td>Figure 3.5</td>
<td>Bioinformatics analysis of APDH orthologs</td>
<td>91</td>
</tr>
<tr>
<td>Figure 3.6</td>
<td>Enzyme kinetics</td>
<td>105</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>Structure of APK&lt;sub&gt;MSM0270&lt;/sub&gt;</td>
<td>108</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>Bioinformatics analysis of APK homologs</td>
<td>115</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>Proposed organization of, and pathway mediated by, the aminoacetone utilization microcompartment.</td>
<td>120</td>
</tr>
<tr>
<td>Figure 5.1</td>
<td>Structural overview of the <em>Mesorhizobium loti</em> phosphopropanolamine phospholyase.</td>
<td>140</td>
</tr>
<tr>
<td>Figure 5.2</td>
<td>Sequence alignment of aminoglycoside aminotransferase class-III like phospholyses.</td>
<td>142</td>
</tr>
<tr>
<td>Figure 6.1</td>
<td>Proposed metabolisms of the aminoacetone utilization operons</td>
<td>158</td>
</tr>
</tbody>
</table>
LIST OF EQUATIONS

Equation 1 ........................................................................................................................................... 137
LIST OF ABBREVIATIONS

AAUM - aminoacetone utilization microcompartment
ack - acetate kinases
ADH or AldDH - aldehyde dehydrogenase
ADP - atomic displacement parameters
Aminoacetone - 1-amino-2-propanone
Aminopropanol - 1-amino-2-propanol
APDH - aminopropanol dehydrogenase
APH - aminoglycoside phosphotransferase
APK - S-1-amino-2-propanol kinase
ATP - adenosine triphosphate
BLAST - basic local alignment search tool
BMC - bacterial microcompartment
BMC-FP - fused circularly-permuted BMC domain
BMC-H - bacterial microcompartment domain hexamer
BMC-P - bacterial microcompartment domain pentamer
BMC-T - bacterial microcompartment domain trimer
CA - carbonic anhydrase
CoA - coenzyme A
EP - encapsulation peptide
EUT - ethanolamine utilization
GRE - glycyl-radical enzyme
GRM - glycyl-radical microcompartment
GTP - guanosine triphosphate
HTH - helix-turn-helix
IC - inter-chain
ID - inter-domain
IMAC - immobilized metal affinity chromatography
IPTG - isopropyl β-D-1-thiogalactopyranoside
LLP - N’pyridoxal-lysine-5’-monophosphate
NAD(P)(H) - nicotninamide adenine dinucleotide (phosphate)(reduced)
NTP - nucleotide triphosphate
PDB - protein data bank (protein structure archive)
PDU - propanediol utilization
PEtN - O-phosphorylethanolamine
PHP - proline-histidine-proline
PK-like - protein kinase-like domain superfamily
P-loop - phosphate binding loop
PLP - Pyridoxal 5’phosphate
PMP - pyridoxamine monophosphate
PPrN - O-phosphorylpropanolamine
pta - phosphate transacetylases
r.m.s.d. - root mean square deviation
RMM - Rhodococcus/Mycobacterium Microcompartment
Rubisco - ribulose-1,5-bisphosphate carboxylase/oxygenase
SAD - single wavelength anomalous diffraction
SDR - short-chain alcohol dehydrogenase/reductase
SSN - sequence similarity network
1 Introduction

1.1 Bacterial Microcompartments

1.1.1 Overview

Bacterial metabolism of volatile or cytotoxic chemicals in the cytoplasm presents challenges due to membrane leakage of volatile chemicals or high reactivity of cytotoxic chemicals with biological chemicals. Bacterial microcompartments provide an enclosure with restricted diffusion, limiting dispersion of intermediate chemicals produced within. In these systems a series of enzymes catalyze steps of a metabolic pathway that produces, then consumes, either a volatile or cytotoxic intermediate within the confines of a proteinacious shell. Microcompartments can be broadly divided as being either anabolic or catabolic\(^1\). \(\alpha\)- and \(\beta\)-carboxysomes are currently the only known anabolic class of microcompartments, catalyzing the fixation of carbon dioxide with ribulose-1, 5-bisphosphate, forming 2 molecules of 3-phosphoglycerate\(^1\). Catabolic microcompartments include diverse metabolisms involving the production of an aldehyde intermediate which, through a series of subsequent reactions, produces an acyl-phosphate and alcohol product.

Microcompartments allow bacteria to catalyze problematic reactions under controlled conditions. Carboxysomes allow cyanobacteria to fix carbon dioxide as a sole carbon source under reduced concentrations of atmospheric carbon dioxide\(^2\). Bacteria, including commensal and opportunistic pathogenic species, express microcompartments able to catabolize chemicals found in the human gut environments (e.g. ethanolamine, propanediol, choline, or fucose/rhamnose)\(^3\). These bacteria can have one or more metabolite utilization microcompartments contributing to survival, fitness, and pathogenesis\(^4,5,6\). Some species of soil
and aquatic bacteria have microcompartments proposed to be involved in carbohydrate degradation or aminoketone utilization, that are hypothesized to play a role in carbon recycling in their respective environments\textsuperscript{3,7}. The breadth of chemicals used by bacterial microcompartment metabolism has only recently become apparent; understanding the biochemistry of these systems gives insight to the diversity of chemicals available to support bacterial survival.

1.2 Microcompartment Shell

1.2.1 Superstructure Architecture

Microcompartments were first visualized in electron micrographs of carboxysome expressing \textit{Thiobacillus neapolitanus}, where they appeared as large polyhedral inclusions\textsuperscript{8,9}. Further investigation of carboxysomes and other bacterial microcompartments depicted the structures as having icosahedral geometry (20 facets and 12 vertices), with diameters ranging from 55 nm to 600 nm\textsuperscript{10}. The shell facets are composed of hexagonal subunits formed by hexagonal oligomers of proteins containing a conserved pfam00936 domain. Vertices are capped by a pentagonal oligomer of a protein containing a conserved pfam03319 domain. Hexagonal subunits are formed by two structural variants having either a single domain and forming a hexamer (BMC-H, \textit{bacterial microcompartment hexameric protein}) or a double domain and forming a trimer (BMC-T, \textit{bacterial microcompartment trimeric protein}). Pentagonal oligomers are formed from a pentamer of wedge-shaped single domain protomers (BMC-P, \textit{bacterial microcompartment pentameric protein}).

1.2.2 BMC-H

The most abundant protein found in the shell structure is the BMC-H single pfam00936 domain protomer. The pfam00936 domain is composed of an \(\alpha/\beta\) fold with a four strand anti-parallel \(\beta\)-sheet with two \(\alpha\)-helices on one side and a third on the opposing side with a roughly
triangular profile (Figure 1.1). Assembled as a hexagonal shell unit, six protomers form a ring oligomer around a six-fold rotational axis of symmetry. BMC-H oligomers have distinct concave and convex faces, converging at the rotational axis of symmetry where a pore is formed. The pore opening is lined by a loop connecting β2 to β3 with a conserved G-X-G sequence\textsuperscript{11}, where X is the pore lining residue that is commonly a small hydrophilic residue such as serine or glycine. The pores formed by these shell proteins have an average diameter of 5 Å and are positively charged due to the perimeter of the opening being lined by the pore lining residue backbone amino group. The charge and pore opening diameter are proposed to function in controlling substrate/product transport, although this function has not been demonstrated experimentally\textsuperscript{12}. A conserved triad of residues (lysine, asparagine, and arginine) form lateral edge interactions with other BMC-H hexamers\textsuperscript{13}, capable of assembling two dimensional sheets in crystals analogous to the facets found in the microcompartment shell\textsuperscript{11}.

1.2.3 BMC-T

The other hexagonal facet oligomer found in microcompartment shells is formed by BMC-T proteins. The BMC-T class of shell proteins is composed of permuted pfam00936 domains that fold into consecutive α/β fold domains analogous to two fused BMC-H proteins (Figure 1.1). BMC-T protomers form trimeric rings with three-fold rotational symmetry with concave and convex facets analogous to BMC-H hexamers. In contrast to BMC-H oligomers, a pore is not found at the center of each oligomer; instead loops from each BMC domain form a tight interface at the rotational axis of symmetry. BMC-T trimers have, however, been observed in an alternate conformation with the occluding loops in a retracted conformation forming a large pore like structure (~14 Å diameter)\textsuperscript{14,15,16}. Trimeric rings commonly form a dimer of trimers where the concave facets of each ring pack on one another, forming a central cavity within the hexamer. Together, pore conformational shifts and the cavity between the trimers are proposed
to operate as an airlock mechanism for large molecule transport across the microcompartment shell\textsuperscript{16,17}. The BMC-T class of shell protein is the most varied shell component with some BMC-T proteins being formed by simple fusion of BMC domains\textsuperscript{18}, some are involved in metal binding for redox chemistry\textsuperscript{19}, while some microcompartments lack BMC-T proteins all together in their shell structure\textsuperscript{7}.

\subsection*{1.2.4 BMC-P}

As the microcompartment shell has an icosahedral geometry, five facets formed by hexagonal subunits meet at the apices to form a pentagonal “defect”. To ensure the restricted permeability of the shell, a BMC-P oligomer fills the void at each apex\textsuperscript{20,21}. A minimum of one BMC-P homolog is necessary to fulfill the requirements of a microcompartment shell\textsuperscript{22}, however, multiple paralogs have been found in some microcompartment operons\textsuperscript{3}. BMC-P protomers contain a pfam03319 domain consisting of an anti-parallel 5 stranded β-barrel with a short α-helix at one end (Figure 1.1). Five BMC-P protomers oligomerize around a five-fold rotational axis with β-barrels pointing toward the rotational axis of symmetry with the short α-helices meeting at the center. BMC-P pentamers have a roughly pentagonal circumference and a conical geometry suitable for interaction with edges of each facet\textsuperscript{23}. A pore is formed at the center of the pentamer with a narrow restriction (<5 Å diameter) lined by small hydrophilic residues, often occluded by small molecules in structural studies, leading to the consensus that these oligomers have a strictly structural function\textsuperscript{21,23,24}.

\textbf{Figure 1.1 Microcompartment shell proteins}

Canonical microcompartment shell protein protomeric and oligomeric structures. From top to bottom the BMC-H single pfam00936 domain shell protein (PDB 5L38), the BMC-T permuted double pfam00935 domain shell protein (PDB 3FCH), and the BMC-P single pfam03319 domain shell protein (PDB 5L37). Protomers are colored blue to red from the N to C termini.
with one protomer in each oligomer colored and oriented as in the protomer frame. Rotational symmetry of the oligomers is depicted by the black shapes (hexagon 6-fold; triangle 3-fold; pentagon 5-fold).
1.3 Characterized Microcompartment Metabolisms

1.3.1 Carboxysome
The first identified microcompartment was the carboxysome \(^{25}\) (Figure 1.2). Carboxysome metabolism is catalyzed by two enzymes: carbonic anhydrase (CA) and ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco). Bicarbonate from the cytoplasm is dehydrated to carbon dioxide by CA within the carboxysome, which is then used to carboxylate ribulose-1, 5-bisphosphate at the 2’ carbon forming two molecules of 3-phosphoglycerate by co-localized Rubisco. The carboxysome shell acts as a diffusion barrier, concentrating the volatile carbon dioxide in close proximity to inefficient Rubisco to enhance turnover rate and reducing side reactions with molecular oxygen \(^{26}\). Carboxysomes are subdivided into two types based on the Rubisco form associated with the microcompartment being either Proteobacterial like Rubisco, form 1A, (α-carboxysome) or plant like Rubisco, form 1B, (β-carboxysome) \(^{27}\). Despite catalyzing an identical metabolism, α and β carboxysomes have profound evolutionary differences including operon composition, shell composition, internal organization, assembly mechanism, and type of carbonic anhydrase enzymes.

1.3.2 Ethanolamine and Propanediol Utilization Microcompartments
Of the identified catabolic microcompartments, the propanediol and ethanolamine utilization microcompartments (PDU and EUT, respectively) have been the most thoroughly investigated. Consistent across currently characterized catabolic microcompartments is a core set of four enzymes including: a “signature” aldehyde producing enzyme, an alcohol dehydrogenase, an acylating aldehyde dehydrogenase, and a phosphate transacylase \(^{28}\) (Figure 1.2). The signature enzymes of PDU (propanediol dehydratase) \(^{29}\) and EUT (ethanolamine ammonia lyase) \(^{30,31}\) form cytotoxic aldehydes propanal and acetaldehyde, respectively, necessitating compartmentalization of the metabolic pathway within a microcompartment shell \(^{31}\). The aldehyde produced by the
signature enzyme is either oxidized by an aldehyde dehydrogenase or reduced by alcohol dehydrogenase within the microcompartment, recycling the redox cofactor NAD(P)(H)\(^{32}\). Acyl-coenzyme A produced by the aldehyde dehydrogenase undergoes transacylation by a phosphate transacylase producing the product acyl-phosphate and regenerating coenzyme A\(^{33}\). Additional enzymes outside of the core enzymes are commonly found associated with microcompartment operons including enzymes for cobalamin synthesis (the cofactor required by both PDU and EUT signature enzymes), carboxylate kinase (converting acyl-phosphate products to an acid product and ATP), and ammonia lyase reactivating factor (for EUT ethanolamine ammonia lyase)\(^{31}\). These microcompartments have been identified alone or together in species of enteric bacteria, where propanediol and ethanolamine are readily available and may contribute to the pathogenesis of opportunistic species\(^{4,5,6}\).

1.3.3 Partially Characterized Glycyl-radical Microcompartments

A poorly-studied class of microcompartments that involve glycyl-radical enzymes (GRE), termed glycyl-radical microcompartments (GRM), have been the subject of recent attention. Subclasses of these microcompartments include metabolisms for choline utilization, propanediol utilization, and fucose/rhamnose utilization\(^{34}\). The choline utilization GRE catalyzes a trimethylamine lyase\(^{35}\) activity forming trimethylamine and acetaldehyde, whereas both propanediol and fucose/rhamnose GREs catalyze propanediol dehydratase activity producing propanaldehyde\(^{36}\). In addition to the “signature” GREs and remaining catabolic microcompartment core enzymes, a GRE activating protein is required for catalysis, and in the case of fucose/rhamnose utilization, a fuculose aldolase and propanediol producing lactate reductase, are included in the metabolism\(^{34,36}\). Similar to PDU and EUT microcompartments, GRM microcompartment operons are found in enteric bacteria and use substrates available in the gut as a carbon and energy source for proliferation\(^{3}\).
Figure 1.2 Diagrammatic representation of the metabolisms of anabolic microcompartments (carboxysomes) and catabolic microcompartments.

The left panel shows canonical anabolic microcompartments metabolism, converting carbonic acid from the cytoplasm into volatile carbon dioxide within the microcompartment shell where it is fixed by rubisco with ribulose 1,5-bisphosphate to form two molecules of 3-phosphoglycerate. In the right panel the common scheme for catabolic microcompartment metabolism is shown where an initial substrate is converted to an intermediate aldehyde by a “signature” enzyme. The intermediate aldehyde is either reduced to an alcohol by-product or oxidized to an acyl-CoA intermediate. The acyl-CoA intermediate is transferred to inorganic phosphate, releasing as an acyl-phosphate, which is then used to make ATP and an acid product by an associated acyl kinase. Redox cofactors are recycled between the alcohol and aldehyde dehydrogenase, and CoA is recycled between the aldehyde dehydrogenase and transacylase enzymes maintaining an internal pool of cofactors required for metabolism.
1.4 *Rhodococcus/Mycobacterium* Microcompartment

1.4.1 Operon Complement

The RMM (*Rhodococcus/Mycobacterium* Microcompartment) operons were first identified by Axen *et al.*³, who identified a class of catabolic microcompartment with conserved shell components and a non-canonical enzyme complement. Four shell proteins, four enzymes, a transcription regulator, an amino-acid like transporter, and a protein with a domain of unknown function make up the core complement of the RMM operon. The four shell proteins include two BMC-Ts, a BMC-H, and a BMC-P that meet the composition requirements to form a complete microcompartment shell. The four core enzymes include a short-chain alcohol dehydrogenase/reductase (SDR), an aminoglycoside phosphotransferase (APH), a class-III aminotransferase, and a coenzyme A acylating aldehyde dehydrogenase. Prior to this study, little investigation of the RMM enzymes or shell proteins had been described.

1.4.2 *Mycobacterium smegmatis* MC2 155 RMM

Mycobacterial species identified with an RMM operon belong to both non-pathogenic and opportunistic pathogen classes of Mycobacteria. Included in the non-pathogenic species of Mycobacteria containing an RMM operon is the common lab model species *Mycobacterium smegmatis* MC2 155. The RMM operon in this species contains all of the core operon genes and represents the minimal gene composition for an RMM, making the *M. smegmatis* RMM an ideal candidate for investigation of the microcompartment’s metabolism (Table 1.1).
<table>
<thead>
<tr>
<th>Gene name</th>
<th>NCBI Gene ID</th>
<th>Annotation (pfam#)</th>
<th>Proposed function</th>
</tr>
</thead>
<tbody>
<tr>
<td>msmeq_0268</td>
<td>4533722</td>
<td>GntR-like bacterial transcription factor (00392)</td>
<td>Transcription enhancer</td>
</tr>
<tr>
<td>msmeq_0269</td>
<td>4531547</td>
<td>Short-chain dehydrogenase/reductase (00106)</td>
<td>S-(+)-1-amino-2-propanol dehydrogenase</td>
</tr>
<tr>
<td>msmeq_0270</td>
<td>4536001</td>
<td>Aminoglycoside phosphotransferase (01636)</td>
<td>1-amino-2-propanol kinase</td>
</tr>
<tr>
<td>msmeq_0271</td>
<td>4532263</td>
<td>Double domain bacterial microcompartment shell protein (00936)</td>
<td>Double domain hexagonal shell protein</td>
</tr>
<tr>
<td>msmeq_0272</td>
<td>4532935</td>
<td>Single domain bacterial microcompartment shell protein (00936)</td>
<td>Single domain hexagonal shell protein</td>
</tr>
<tr>
<td>msmeq_0273</td>
<td>4535004</td>
<td>Vertex bacterial microcompartment shell protein (03319)</td>
<td>Vertex capping shell protein</td>
</tr>
<tr>
<td>msmeq_0274</td>
<td>4531449</td>
<td>Uncharacterized protein</td>
<td>Organizing protein</td>
</tr>
<tr>
<td>msmeq_0275</td>
<td>4537562</td>
<td>Double domain bacterial microcompartment shell protein (00936)</td>
<td>Double domain hexagonal shell protein</td>
</tr>
<tr>
<td>msmeq_0276</td>
<td>4533133</td>
<td>Acylating aldehyde dehydrogenase (00171)</td>
<td>Coenzyme A acylating propanal dehydrogenase</td>
</tr>
<tr>
<td>msmeq_0277</td>
<td>4537174</td>
<td>Aminotransferase class III (00202)</td>
<td>O-phosphopropanolamine phospholase</td>
</tr>
<tr>
<td>msmeq_0279</td>
<td>4532110</td>
<td>Amino acid permease (13520)</td>
<td>Aminopropanone permease</td>
</tr>
</tbody>
</table>
1.4.3 Shell Proteins

The *M. smegmatis* RMM operon includes representatives of all 3 classes of the shell forming proteins. This includes a homolog of each BMC-H and BMC-P proteins and two homologs of BMC-T proteins. Both the BMC-H (MSM0272) and the BMC-P (MSM0273) sequences are most closely related to homologs from the ethanolamine utilization microcompartment (EutM and EutN, respectively). MSM0272 shares conservation of key sequence features of the BMC-H type shell proteins including the GXG motif, of the pore lining loop, and edge interaction residues\(^{11,13}\). MSM0273 likewise contains key BMC-P sequence features, including a small hydrophilic residue at an identical position as the pore lining residue in EutN and similar edge interaction motifs\(^{21,37,38}\). Together, the conserved features of these proteins with characterized homologs suggest that they fulfill similar roles in the RMM microcompartment shell, acting as a substrate/product pore (MSM0272) and a shell vertex cap (MSM0273). The BMC-T proteins in the *M. smegmatis* RMM (MSM0271 and MSM275) are of the fused circularly-permuted BMC domains (BMC-FP) type, most similar to homologs from α and β carboxysomes (CsoS1D and CcmP, respectively). Consistent with the carboxysome homologs, loops contributing to occlusion of the central pore opening are of equivalent length and share conservation of a defined set of interacting residues (arginine, glutamine, and glutamate)\(^{15,16}\). The RMM BMC-T proteins likely fulfill a similar function as the BMC-T proteins in carboxysomes, contributing to the shell facet structure and potentially as large molecule transporters.

1.4.4 Short-chain Alcohol Dehydrogenase/Reductase (SDR)

The short-chain alcohol dehydrogenase/reductase (SDR) family of enzymes are structural conserved, while having weak sequence conservation\(^{39}\). The protomeric tertiary structure includes a dinucleotide binding Rossmann-fold domain and a variable C-terminal domain, with a
catalytic tetrad of residues (serine, tyrosine, lysine, and asparagine) that form a proton transfer chain with NAD(P)H and a substrate\textsuperscript{40}. This family is mostly comprised of enzymes that catalyze the reversible oxidation of secondary alcohols to ketones, although examples of members catalyzing dehydratase, dehalogenase, isomerase, decarboxylase, or epimerase reactions have been identified\textsuperscript{39}. Classical SDRs represent the bulk of the family, catalyzing oxidoreduction of small molecules including a range of alcohols, polyols, and steroids using NAD(P)(H) as a cofactor. Consistent within the classical SDRs is a short C-terminal substrate specifying domain and cofactor binding motif “TGxxx[AG]xG” (where x is any amino acid)\textsuperscript{41}. Cofactor preference is selected for by the presence of a basic residue in the cofactor binding motif forming hydrogen bonds with the phosphate of NADP(H)\textsuperscript{42}.

The \textit{M. smegmatis} RMM SDR (MSM0269) is a classical SDR that contains the conserved catalytic tetrad residues and a lysine containing cofactor binding motif suggesting a preference for NADP(H) over NAD(H). High sequence conservation (64% sequence identity over 100% of the sequence) with the previously characterized S-1-amino-2-propanol alcohol dehydrogenase from \textit{Rhodococcus erythropolis} \textsuperscript{43}, suggests that this enzyme catalyzes the same reaction with similar specificity. Inclusion of an alcohol dehydrogenase in microcompartment metabolism is common, however previously characterized alcohol dehydrogenases function in cofactor regeneration acting on aldehydes instead of a ketone, producing a primary alcohol side product\textsuperscript{44}.

1.4.5 Aminoglycoside Phosphotransferase

The phosphotransferase enzyme encoded in the \textit{M. smegmatis} RMM (MSM0270) operon belongs to the aminoglycoside phosphotransferase family (APH) of kinases, part of the protein kinase-like (PK-like) superfamily. Conserved across the PK-like superfamily is a tertiary structure of an N-terminal nucleotide triphosphate (NTP) binding α/β-fold and a mostly α-helical
substrate specifying C-terminal domain$^{45}$. A cleft is formed at the interface of the two domains that is lined by a conserved Brenner motif “HxDhxxxN” (where x is any residue and h is a bulky hydrophobic residue)$^{46}$. The Brenner motif contributes residues for magnesium coordination with the donor NTP and the catalytic base. Additional NTP coordination is contributed by a flexible phosphate binding loop (P-loop)$^{47}$, which undergo a structural conformation change when bound to a NTP. Binding of substrate to the active site incurs a rotation of the N-terminal domain that contributes additional substrate binding residues from the P-loop and brings the substrates together for catalysis$^{48}$. Members of the APH family have a preference for adenosine triphosphate (ATP) as the phosphate donor; however, guanosine triphosphate (GTP) preferring enzymes have been characterized$^{49}$. Nucleotide selectivity in GTP specific enzymes is due to a steric clash of adenosine with a tyrosine residue in the nucleotide residue binding pocket$^{49}$.

Low sequence conservation in the APH family of enzymes results in few sequences aligning with MSM0270 when searched in the BLAST database. The APH family includes kinases acting on alcohols of varying sizes including antibiotic aminoglycosides such as spectinomycin, kanamycin, or gentamicin, and small alcohols such as homoserine or choline. An alcohol kinase is uncommon in microcompartment metabolism. However, there is commonly a carboxylic acid kinase (acetate kinase) associated with catabolic microcompartments phosphorylating ADP using an acyl-phosphate, produced by the microcompartment metabolism, forming ATP$^{3,32}$. Although an ATP consuming enzyme in a catabolic microcompartment system is unusual, inclusion of an ATP requiring enzyme has some precedence; carboxysomes include a rubisco ATPase activase rescuing rubisco from inhibitory side reactions$^{28}$. Absolute conservation of the APH in the RMM microcompartment indicates that its function is integral to the metabolism catalyzed by the microcompartment. A kinase identified by a group investigating
aminoacetone degradation in a species of *Pseudomonas* catalyzed the phosphorylation of aminopropanol produced by an aminoacetone reductase, followed by a dephosphorylation and deamination by a phospholyase enzyme\(^5^0\). The co-occurrence of an APH with an aminoacetone reductase and aminotransferase (with putative phospholyase activity as outlined below) in the RMM suggests that the RMM may catalyze a similar set of reactions as observed in the *Pseudomonas* system.

### 1.4.6 Aminotransferase Class-III

Pyridoxal 5’phosphate (PLP) dependent aminotransferase enzymes have been divided into five classes based on sequence similarity. Class III aminotransferases include enzymes that catalyze amino transfer from ornithine, gamma-amino butyric acid, and omega amino acids with common amino acceptor keto-acids such as alpha-ketoglutarate, oxaloacetate, or pyruvate\(^5^1\). Class III aminotransferases belong to the fold type 1 PLP dependent enzymes, conserving a two domain tertiary architecture with a homodimeric quaternary structure\(^5^2\). Each protomer contains a central cofactor binding domain (7-8 stranded α/β-fold) and a substrate binding domain (4 stranded β-sheet with α-helices on one face). The active site is formed at the interface of the dimer, with both cofactor domains sandwiching PLP bound to a conserved lysine as an “internal” aldimine. The class III aminotransferases catalyze amine transfer through a two-step process: first the donor substrate transfers an amine to PLP forming pyridoxamine monophosphate (PMP) and a carbonyl product, second the amine from PMP is transferred to an acceptor keto-acid reforming PLP and an amino-acid product\(^5^3\). Alternating substrate specificity for the two half reactions is controlled by a glutamate-arginine switch, changing between the longer amine donor substrate and the keto-acid amino acceptor substrate, through pairing of the carboxylate group of the acceptor with arginine\(^5^4\).
A subclass of the class III aminotransferase enzymes catalyzes a phospholyase reaction instead of transamination. The subclass of enzymes shares a tertiary fold type 1 structure, with a modified substrate binding domain compared to other class III aminotransferase enzymes. The phospholyase activity is proposed to occur through a reaction mechanism starting with a similar first half reaction, wherein the substrate forms an external aldimine with PLP. The second half reaction releases the substrate phosphate as inorganic phosphate, followed by reformation of the internal aldimine and release of a product alkenylamine product\textsuperscript{55,56}. Spontaneous hydrolysis of the alkenylamine releases a molecule of ammonia and an aldehyde. Examples of these enzymes have been characterized in eukaryotic and prokaryotic species, catalyzing the dephosphorylation and subsequent deamination of \textit{O}-phosphorylethanolamine (PEtN). Recent structural characterization of a PEtN phospholyase from \textit{Arthrobacter aurescens} TC1 (A1RDF1) with PEtN bound as an external aldimine, demonstrated the mechanics by which the phospholyase activity may be catalyzed\textsuperscript{56}. The phosphate group of PEtN was coordinated with an electropositive phosphate binding pocket proposed to facilitate phosphate release from the phosphoryl-substrate with protonation by the catalytic lysine (otherwise bound to PLP as an internal aldimine). Additional changes in the active site have resulted in the loss of the arginine-glutamate switch in identified PEtN phospholyases as a result of extensive mutation and deletions in the peptides forming the involved structures.

The aminotransferase enzyme from the \textit{M. smegmatis} RMM (MSM0277) has been annotated as a member of the class III aminotransferase family. Annotation as a class III aminotransferase is based on conservation of sequence features, however, sequence similarity with the closest characterized homolog, a putrescine aminotransferase (Ygig from \textit{Escherichia coli}; UniProt code P42588), is weak (34\% sequence identity over 86\% of the sequence). Similar
to other members of the aminotransferase class III enzymes, the PLP binding pocket residues are conserved in MSM0277, however, residues determined to be involved in forming the substrate binding pocket of Ygjg are not conserved\textsuperscript{57}. Sequence comparison of MSM0277 with characterized PEtN phospholyase A1RDF1 does not indicate that phosphate binding residues of the PEtN phospholyase active site are conserved by MSM0277\textsuperscript{56}. MSM0277 does, however, contain numerous basic residues in the vicinity of the substrate binding pocket. Together, with the absence of the aminotransferase arginine-glutamate switch and an alternative phosphate binding pocket to the previously characterized phosphate binding pocket, convergent evolution of phospholyase activity for the aminotransferase enzyme is plausible in the RMM.

1.4.7 Acylating Aldehyde Dehydrogenase

The aldehyde dehydrogenase family of enzymes catalyze the reversible oxidation of aldehydes using the redox cofactor NAD(P). Conserved across this family of enzymes is a central nucleotide binding domain composed of a Rossmann fold, with a variable N-terminal catalytic domain and a C-terminal oligomerization domain. Catalysis occurs through a mechanism where deprotonation of the conserved catalytic cysteine sulfhydryl, by a conserved glutamate or histidine, enables the resulting thiolate to attack the substrate carbonyl forming an intermediate thioester with NAD(P) reduction\textsuperscript{58}. Release of NAD(P)H and hydrolysis of the thioester, by a glutamate coordinated water\textsuperscript{59}, releases a carboxylic acid product. Alternatively, in coenzyme A acylating aldehyde dehydrogenases, the water coordinating glutamate is absent and an alternate glutamate\textsuperscript{58} or histidine\textsuperscript{36} deprotonates the CoA sulfhydryl. The CoA sulfhydryl attacks the thioester intermediate, releasing an acyl-CoA product. As these enzymes catalyze reversible oxidoreductase activities, cofactor specificity often indicates reduction or oxidation propensities for the enzyme. NAD(H) specific enzymes commonly catalyze aldehyde oxidation, whereas NADP(H) specifying enzymes catalyze reduction of carboxylic acid or acyl-CoA
substrates\textsuperscript{50}. Preference for NAD(H) is dictated by a conserved proline-histidine-proline (PHP) motif that results in a steric clash with the 2’ phosphate of adenosine in NADP(H)\textsuperscript{58,61}.

The aldehyde dehydrogenase found in the RMM operon (MSM0276) is annotated as a coenzyme A acylating aldehyde dehydrogenase by sequence similarity. Compared to structurally and catalytically characterized homologs including the CoA propionylating aldehyde dehydrogenase from the \textit{Clostridium phytofermentans} fucose utilization microcompartment\textsuperscript{58}, MSM0276 contains all of the key motifs and catalytic residues required for coenzyme A acylation. This includes the catalytic cysteine, sulphydryl deprotonating histidine, CoA deprotonating glutamate, and PHP NAD(H) specifying motif. The water coordinating glutamate conserved by hydrating aldehyde dehydrogenases is replaced by a small hydrophobic residue. Together, the conservation of key catalytic and substrate binding residues of the RMM aldehyde dehydrogenase with characterized aldehyde dehydrogenases suggests that MSM0276 catalyzes an oxidation of an aldehyde with coenzyme A acylation.

\subsection*{1.4.8 Encapsulation Peptides and the Protein Containing a Domain of Unknown Function}

Assembly of microcompartments is proposed to operate through an aggregation mechanism of organization using short 20 residue amphipathic helices termed encapsulation peptides (EPs)\textsuperscript{2}. Identified EPs have weak sequence conservation, but they show conservation in a motif of alternating pairs of hydrophobic and hydrophilic residues\textsuperscript{62,63}. EPs are found in all identified catabolic microcompartments, proposed to operate as aggregating centers between EPs and in microcompartment targeting through interaction with shell proteins\textsuperscript{64,65}. EP helices are often found at the termini of microcompartment enzymes connected to the globular domain of the enzyme by unstructured linker peptides of variable length and composition. Some enzymes, including the signature enzymes of the PDU (PduCDE)\textsuperscript{29} and EUT (EutBC)\textsuperscript{31} microcompartments, form heterocomplexes where only one protomer encodes an EP with
microcompartment targeting activity. In these systems a “piggybacking” mechanism is proposed to ensure encapsulation of all required enzymes into the microcompartment\textsuperscript{62}. Non-canonical encapsulation mechanisms are proposed in other systems such as the GRM3 class of microcompartment, using encapsulation peptide like sequences within the signature enzyme for enzyme packaging\textsuperscript{36}. In the RMM operon putative EPs are found only in the aldehyde dehydrogenase (MSM0276) and the protein containing a domain of unknown function (MSM0274). Low sequence similarity of MSM0274 to any characterized protein and exclusive association with RMM operons suggests a potential role in microcompartment enzyme targeting similar to the EutBC ethanolamine ammonia lyase complex. The remaining enzymes in the RMM may be incorporated into the microcompartment through heterocomplexation with EP carrying proteins or a discrete encapsulation mechanism may be involved similar to the GRM3 system.

1.5 Non-microcompartment associated RMM like operon

In \textit{M. smegmatis}, a second operon contains homologues of the short-chain alcohol dehydrogenase (SDR), aminoglycoside phosphotransferase (APH), and aminotransferase class-III enzymes from the RMM, as well as enzymes annotated as an acetate kinase and an acetyltransferase. The SDR and APH enzymes share high sequence identity with their RMM homologs, particularly with regards to their substrate binding domains, suggesting that they catalyze similar reactions. The aminotransferase class-III enzyme, however, contains all key catalytic residues of the phosphohydrolase subclass of enzymes. This includes all PLP binding pocket residues and substrate-phosphoryl binding residues characterized in the phosphoethanolamine phosphohydrolase from \textit{A. aurescens}\textsuperscript{56}. The remaining enzymes in the operon share high sequence similarity with characterized acetate kinases (ack) and phosphate transacetylases (pta) and likely catalyze similar reactions (carboxylate phosphorylation and phosphate-coenzyme A
transacylation). Similar to the RMM operon, the non-microcompartment RMM like operon includes genes for a transcription regulator and an amino acid permease.

**Table 1.2 Genes of *Mycobacterium smegmatis* MC2 155 non-microcompartment aminoacetone utilization operon**

<table>
<thead>
<tr>
<th>Gene name</th>
<th>NCBI Gene ID</th>
<th>Annotation (pfam#)</th>
<th>Proposed function</th>
</tr>
</thead>
<tbody>
<tr>
<td>msmeg_0778</td>
<td>4535574</td>
<td>GntR-like bacterial transcription factor (00392)</td>
<td>Transcription enhancer</td>
</tr>
<tr>
<td>msmeg_0779</td>
<td>4534081</td>
<td>Short-chain dehydrogenase/reductase (00106)</td>
<td>S-(+)-1-amino-2-propanol dehydrogenase</td>
</tr>
<tr>
<td>msmeg_0780</td>
<td>4532870</td>
<td>Aminoglycoside phosphotransferase (01636)</td>
<td>1-amino-2-propanol kinase</td>
</tr>
<tr>
<td>msmeg_0781</td>
<td>4534735</td>
<td>Amino acid permease 2 (13520)</td>
<td>Aminopropanone permease</td>
</tr>
<tr>
<td>msmeg_0782</td>
<td>4535062</td>
<td>Aminotransferase class-III (00202)</td>
<td>O-phosphopropionolamine phospholyase</td>
</tr>
<tr>
<td>msmeg_0783</td>
<td>4533640</td>
<td>Phosphate acetyl transferase (01515)</td>
<td>Phosphate propionyl transferase</td>
</tr>
<tr>
<td>msmeg_0784</td>
<td>4531693</td>
<td>Acetate kinase (00871)</td>
<td>Propionate kinase</td>
</tr>
</tbody>
</table>
1.6 Thesis objectives and justification of research
The microcompartment operon found in *M. smegmatis* is a model system with all the core components of the *Rhodococcus/Mycobacterium* microcompartment (RMM). Variations of microcompartment metabolism had been the focus of much investigation especially with regards to carboxysomes, ethanolamine utilization, and propanediol utilization microcompartments. However, little was known about the RMM metabolism. The minimal composition of the microcompartment shell and limited number of enzymes compared to other catabolic microcompartments suggested that this microcompartment could be viewed as a prototypical catabolic microcompartment. The research presented here focused on determining the catalytic functions of the enzymes from the RMM in *M. smegmatis*, as well as determining the structures of the shell forming proteins to gain insight into transport mechanisms involved in the microcompartment metabolism. Determination of the metabolism of the RMM increases our understanding of the diversity of both the substrates that catabolic microcompartments degrade, and the strategies they use to do so. By characterizing orthologous enzymes from the RMM and non-RMM aminoacetone utilization pathways, insights into the metabolic advantages that compartmentalization offers their hosts could be proposed.

In chapter 2, the structural characterization of all the shell proteins associated with the RMM operon from *M. smegmatis* is presented. Three of the shell proteins retained structural features identified in homologs from carboxysome microcompartments; however, the BMC-H protein (proposed to operate in substrate transport) represented the only characterized example of a negatively charged pore forming BMC-H protein. In chapter 3, the structural and catalytic characterization of the S-1-amino-2-propanol alcohol dehydrogenase from the *M. smegmatis* RMM is presented. The enzyme catalyzed reversible stereospecific oxidation of S-1-amino-2-propanol with a low-micromolar *K_M* for either 1-amino-2-propanol or aminoacetone and strict
specificity for NADP(H). Apo and ternary complex, with bound NADP\(^+\) and aminoacetone, structures were presented for the S-1-amino-2-propanol dehydrogenase. Substrate specificity was apparent from the determined structures, with a substrate binding pocket forming a small hydrophobic pocket and a negatively charged pocket suitable for binding a 3-carbon amino-alcohol. In chapter 4, the structural and functional characterization of a 1-amino-2-propanol \(O\)-kinase is presented. Micromolar \(K_m\) constants were determined for the \(O\)-phosphorylation of amino-alcohols with 3-4 carbons with preference for S isomers. The structure of the aminopropanol kinase resembled other members of the aminoglycoside phosphotransferase family of enzymes. The substrate binding pocket was smaller than those of previously characterized aminoglycoside phosphotransferases; with a tight hydrophobic pocket at one end and acidic residues at the opposing end. A model for substrate binding is proposed wherein the hydrophobic pocket binds the methyl group of aminopropanol and the amino group is bound by the acidic residues. In chapter 5, the structural and functional characterization of a \(O\)-phosphopropanolamine phospholyase from \textit{Mesorhizobium loti} is presented. Structural and function investigation of this close homolog to the aminotransferase enzyme from the \textit{M. smegmatis} RMM was pursued as a soluble alternative to the insoluble RMM enzyme. The enzyme showed structural similarity to other members of the aminotransferase class-III enzymes, including the PLP cofactor binding domain. The substrate binding domain, however, did not resemble other members of the family; with much of the substrate binding pocket being occupied with basic residues proposed to form a phosphate binding pocket. Aldehyde production was detected with \(O\)-phosphorylated aminoalcohols analogous to the catalysis observed for phosphoethanolamine phospholyase enzymes. From the determined function of this enzyme, phosphopropanolamine phospholyase activity is proposed for the \textit{M. smegmatis} RMM homolog.
opposed to the aminotransferase annotation. In chapter 6, a metabolic pathway for the *Rhodoococcus/Mycobacteria* microcompartment is proposed, with a proposed renaming of this class of microcompartment to aminoacetone utilization microcompartment or AAUM.
2 A complete structural inventory of the *Mycobacterial* microcompartment shell proteins constrains models of global architecture and transport.

This research was originally published in the Journal of Biological Chemistry. Mallette, E. and Kimber, M. S. A complete structural inventory of the *Mycobacterial* microcompartment shell proteins constrains models of global architecture and transport. *J. Biol. Chem.* 2017; 292:1197-1210

Statement of Contribution

The primary author, Evan Mallette, designed and conducted experiments, analyzed the data, and contributed to writing the manuscript and visualization. Matthew S. Kimber conducted modeling work, acquired funding, and contributed to writing of the manuscript and visualization.

Acknowledgements

We thank J. Alexiou and M. Connor who initiated studies on these proteins and P. Grochulski and S. Labiuk at Canadian Light Source for assistance in collecting diffraction data.
2.1 Abstract

Bacterial microcompartments are bacterial analogs of eukaryotic organelles in that they spatially segregate aspects of cellular metabolism, but do so by building not a lipid membrane but a thin polyhedral protein shell. While multiple shell protein structures are known for several microcompartment types, additional uncharacterized components complicate systematic investigations of shell architecture. We report here the structures of all four proteins proposed to form the shell of an uncharacterized microcompartment designated the *Rhodococcus* and *Mycobacterium* Microcompartment (RMM), which, along with crystal interactions and docking studies, suggests possible models for the particle’s vertex and edge organization. MSM0272 is a typical hexameric β-sandwich shell protein thought to form the bulk of the facet. MSM0273 is a pentameric β-barrel shell protein that likely plugs the vertex of the particle. MSM0271 is an unusual double-ringed bacterial microcompartment shell protein whose rings are organized in an offset position relative to all known related proteins. MSM0275 is related to MSM0271, but self-organizes as linear strips that may line the facet edge; here, the presence of a novel extendable loop may help ameliorate poor packing geometry of the rigid main particle at the angled edges. In contrast to previously characterized homologs, both of these proteins show closed pores at both ends. This suggests a model where key interactions at the vertex and edges are mediated at the inner layer of the shell by MSM0271 (encircling MSM0273) and MSM0275, while the facet is built from MSM0272 hexamers tiling in the outer layer of the shell.

2.2 Introduction

Bacteria differ fundamentally from eukaryotes in that they lack the ability to form membrane-bound organelles in order to segregate aspects of their metabolism. Some bacteria are, however, able to bypass this limitation by encapsulating enzymes catalyzing critical steps of a problematic metabolic pathway in a protein shell; the resulting 100 nm scale bodies are known as
bacterial microcompartments (BMCs). Recent surveys suggest 23 phyla of bacteria contain operons harbouring candidate genes for construction of these microcompartments\(^3\). Ecologically, the most important BMCs are the anabolic carboxysomes, which facilitate carbon fixation in cyanobacteria and chemoautotrophs\(^{67,68}\); a variety of BMCs in heterotrophic bacteria also support the catabolism of a variety of small organic compounds\(^69\). Well-characterized examples of microcompartments include the two varieties of carboxysomes (\(\alpha\) and \(\beta\)) that fix carbon dioxide using Rubisco, propanediol utilization (\(pdu\)) and ethanolamine utilization (\(eut\)) microcompartments. Available evidence for less well-characterized microcompartments suggests roles in the catabolism of ethanol, choline, 6-deoxy-hexoses and other small molecules\(^{35,70}\), while bioinformatics analysis suggest 37 distinct microcompartment gene patterns, some of which are highly likely to encode novel pathways\(^3\). The emergent pattern is that microcompartments typically catalyze successive steps in pathways where the intermediates are toxic and/or volatile\(^28\). Encapsulation of the enzymes within a large (~100 nm diameter) selectively permeable polyhedral proteinacious shell therefore serves to bring sequentially acting active sites into spatial proximity, while the selectivity and limited porosity of the shell helps impede the escape of problematic intermediates\(^{12,71}\).

The organization of the microcompartment shell has been of considerable research interest; assembly appears to echo patterns previously identified in the organization of icosahedral viruses, but the constituent proteins are not homologous. The facets of the shell are built from members of a single protein family (pfam00936) with a core structure comprised of a four stranded anti-parallel \(\beta\)-sheet flanked by three \(\alpha\)-helices in a thioredoxin-like topology; six copies of this domain then form a basic hexagonal tile that can be packed into extended sheets (Figure 2.1)\(^{13,72,73,74}\). At least one (and generally several) such protein has been identified in
every candidate microcompartment identified to date, and, in systems where the shell composition has been quantified, these bacterial microcompartment - hexameric shell proteins (BMC-H) are the most abundant component of the shell. More recently, a second, deeply divergent branch of this family was identified in Cso1D in α-carboxysomes\textsuperscript{15}. This protein has only very weak sequence similarity to BMC-H proteins, and, indeed, the topology of the basic structural domain is permuted, with two permuted domains linked in a head-to-tail fusion. These proteins are here termed BMC-FP (BMC fused, permuted) proteins (note that we use BMC-FP to denote only fused, permuted, double-ring proteins, such as CcmP and Cso1D). Only three copies of the protein are required to form a pseudohexagonal ring, but in both known examples, two such rings oligomerize face to face forming a double-ring that has been suggested to form a locally double-layered shell\textsuperscript{15,16}. BMC-FPs are only associated with a subset of microcompartments, and have only been previously characterized in carboxysomes. Other variants on the pfam00936 family are also found, including permuted hexameric proteins (e.g. EutS and PduU)\textsuperscript{75,76} and double-domain non-permuted proteins (PduB, EutL and CcmO)\textsuperscript{75,77,78}, though these are generally associated with only a subset of microcompartment types, and likely play more specialized functional roles.
Figure 2.1 Architecture of bacterial microcompartments.

The bulk facet is formed by tiling of BMC-H/FP proteins. The organization around the vertex, and especially the edge, are much more poorly understood.
In addition to BMC-H proteins, the other protein family that seems absolutely required for forming a functional microcompartment shell is pfam 03319. These small proteins are built as a five-stranded anti-parallel β-barrel that oligomerize with four other chains to form a truncated pentameric pyramid. Available evidence suggests that voids left in the five-fold symmetric vertices of the shell are capped by these bacterial microcompartment - pentameric shell (BMC-P) proteins. Interestingly, many microcompartments contain several paralogs of these BMC-P proteins (up to seven), suggesting that they may have roles beyond acting as simple plugs.

In this work, we investigate a member of a family of functionally uncharacterized microcompartments, previously identified in members of the Actinomycetales genus; in the absence of an experimentally confirmed function, we adopt the terminology of Axen et al., terming these RMM (Rhodococcus and Mycobacterium Microcompartment) microcompartments. We have focused on Mycobacterium smegmatis MC for this work; M. smegmatis is a close saprophytic relative of M. tuberculosis, and being non-pathogenic and fast growing, a variety of tools (including transformation, protein expression and genome-wide knockouts) have been developed in M. smegmatis to facilitate use of this organism as a model. The operon associated with the RMM in M. smegmatis contains open reading frames for four shell protein homologs and four enzymes, as well as a possible additional structural protein, a transcription regulator, and an amino acid permease-like transporter. The only enzyme from an RMM operon that has been enzymatically characterized is a short chain alcohol dehydrogenase (SDR) that has been shown to possess L-1-amino-2-propanol dehydrogenase activity. Encoded enzymes that have not been catalytically characterized but from sequence homology the operon includes a class III amino-transferase, an aldehyde-alcohol dehydrogenase, and a protein with
distant homology to aminoglycoside phosphotransferases. This operon can be induced in *Rhodococcus* by the addition of 1-amino-2-propanol\(^{80}\). Together, this suggests that the initial substrate is either 1-amino-2-propanol, and/or a small amino alcohol or closely related metabolite. The products which would ultimately enter general metabolism remain unclear. The shell proteins comprise a BMC-H protein, a BMC-P protein, and two BMC-FP proteins. Intriguingly, this is the smallest number of shell proteins associated with any known microcompartment (tied with the α-carboxysome in *Prochlorococcus* MED4, which has two BMC-P proteins); this suggests that that the RMM may be a usefully tractable system for characterizing and modeling microcompartment shell assembly. We report here for the first time x-ray structures for all shell components of a single microcompartment (all obtained from a single organism), revealing intriguing new insights into how these shell proteins might interact and function.

2.3 Results

2.3.1 The structure of MSM0273

The pentameric shell protein (BMC-P) MSM0273 crystallized in space group P3\(_1\)21 and diffracted to a resolution of 1.6 Å. The asymmetric unit contained five protomers arranged into a pentamer. Of the 87 native amino acids, one chain is missing only the last residue, while the last 9 – 10 residues are disordered in the other four. Each protomer consists of a single pfam03319 domain, organized around a five-stranded antiparallel β-barrel (1, -2, 3, 5, -4 topology) (Figure 2.2A). The axis of this barrel lies within the plane of the pentamer, radiating out from the central pore. The long β1 - β2 loop (β1a) extends out towards an adjacent protomer, which in turn pairs with β5 and an extended region of the C-terminus. In the one chain where ordered, the extended C-terminus forms a β-strand along the outside edge of β1 extending outward from the β-barrel. A short six residue α-helix is inserted into the β4 - β5 loop; the N-terminal end of this helix lines
the central pore of the pentamer.

Searching with DALI reveals that EutN (PDB # 2Z9H), GrpN (PDB # 4I7A), CcmL (PDB # 2QW7) and CsoS4 (HnCsoS4, PDB # 2RCF) are the closest structural homologs, with Z-scores in the range of 12 – 13\textsuperscript{81}. Sequence identity to EutN is closest at 35%. These structures all generally resemble one another, with MSM0273 having a shorter β2 - β3 loop and shorter C-terminus than most homologs. Similar to homologous structures, a pore is formed at the 5-fold axis, and is lined by the N-terminal end of α1; the hydroxyl group and amide nitrogen of Ser56 lines an electropositive, hydrophilic pore approximately 5 Å in diameter (Figure 2.3A-C).

**Figure 2.2 Overall structures of RMM shell proteins.**

A) MSM0273, the BMC-P. B) MSM0272, the BMC-H. C) MSM0271 D) MSM0275, the BMC-FP proteins. The top sub-panels show the protomeric structure (colored blue to red N- to C-terminus) with secondary structure elements labeled. The second sub-panel shows the oligomer (pentamer in A, otherwise a hexamer), shown looking down the axis of maximal symmetry, and in the same orientation as the protomer in the top panel. Additional chains are shown in white, except for the BMF-FP proteins, where one protomer in the opposite ring is shown in beige to show the relative orientations of the two rings. Small triangles indicate the position the three-fold symmetry axis would be found should the top (black) or bottom (white) trimer be tiled in an arrangement analogous to classical BMC-H proteins. The third panel shows the oligomer from a viewpoint orthogonal to that in the second panel.
Figure 2.3 Details of potential transport routes through RMM shell proteins.

A) Electrostatic surface of MSM0273. B) Organization of the MSM0273 pore, seen from the concave side. Similar to other BMC-P proteins, MSM0273 is restricted to a narrow ~3.5 Å pore with 5-fold symmetry lined by the side chain of Ser56. C) Cross section of the MSM0273 pore. Note that the pore is noticeably electropositive. D) Electrostatic surface of MSM0272; note that the central pore area is highly electronegative. E) Details of the more open MSM0272 pore. F) The same pore as in E, but shown in the plane of the pore. G) Details of the pore of MSM0272 in the more closed ring, where Asp40 from one protomer inserts into the pore, partially occluding it. H) MSM0275 electrostatic surface, sliced along the three-fold symmetry axis. Note that the hexamer is closed at both the upper and lower surface. However, there are narrow pores open to the side in protomers where the β1-α1 loop is in an extended conformation. I) Details of the interactions at the three-fold axis of MSM0275. The interactions between Gln70 and Arg69 possibly add additional stability to this interface. J) Electrostatic surface for MSM0271, sliced along the three-fold symmetry axis. Note that both top and bottom surfaces are sealed, and there are no lateral passages. K) Details of the interaction at the three-fold axis in MSM0271.
2.3.2 The structure of MSM0272

MSM0272 is a BMC-H protein that is 68 % identical to *Clostridium* EutM and 59 % identical to *Salmonella* PduA. The structure of MSM0272 structure was determined at a resolution of 2.2 Å, with an R-free of 23.3 %. Crystals were in the space group P43 with two complete hexamers in the asymmetric unit. At minimum, all protomers have residues 4-88 (of 93) ordered, while chain A (the most complete chain with the lowest atomic displacement parameters) is missing only residue 1, which is likely removed by methionine amino peptidase. MSM0272, as expected for a BMC-H protein, is built as a hexamer of single pfam00936 domains, with a four-stranded antiparallel -sheet (with a 2,-3,1,-4 topology) that has two -helices packed onto one face (Figure 2.2B). The C-terminus forms a third helix-like structure which packs onto the equivalent sheet face on an adjacent protomer. Six protomers are organized into a hexameric ring that has a regular hexagonal outline, with near straight edges and distinct convex and concave (where α3 is located) faces.

While α3 is well defined with up to three hydrogen bonds in a 310 pattern in one protomer, in others only these residues form a single hydrogen bond and more resemble two adjacent turns. The structural flexibility of this region, especially the disorder of residues 89-93 in many protomers is interesting, as this region in the close homolog PduA is proposed to mediate interactions with terminal helical peptides from encapsulated residues82,64. In PduA, this helix is required to recruit PduP to the microcompartment, with His81, Val84 and Leu88 (here His81, Leu84 and Phe88) being the critical residues. The relative mobility of these residues, with Phe88 disordered in 7 of the 12 protomers, suggests that this helix may be able to reorganize to maximize interactions with the targeting peptide.

In common with other BMC-H proteins, a small pore is formed where all six protomers meet at the six-fold rotational axis and is lined by a β-hairpin motif (between β2 and β3). In
common with most BMC-H proteins, this turn has a sequence ΦGZGX, where Φ is a small hydrophobic residue, X is any residue, and Z in the residue that immediately lines the pore (note that the full diversity of BMC-H proteins includes more radical variants, including significant insertions and deletions into this loop\(^3\); however, these have not been structurally characterized, and an understanding of the characteristics of the pore they form is lacking). This pore lining residue is important as its ability to interact with adjacent copies of the motif helps define pore diameter, conformation and flexibility, while its properties define the interactions any ligands will be able to form as they pass through the pore. Most commonly, this pore lining residue is small and polar, generally serine (in CcmK1, CcmK2 and CcmK4 from the β-carboxysome) or glycine (CsoS1A and CsoS1C from the α-carboxysome).

In MSM0272, the pore-lining residue is an aspartate, Asp40, which by default carries a negative charge, resulting in a pore environment with crowded carboxylate groups that should strongly repel one another. In the structure, the side chains of Asp40 protrude on the convex face of the protein, forming a ring of carboxylates that face the solvent (Figure 2.3 D-G). The carboxylate generally makes hydrogen bonds only to water molecules, though the nearby placement of the Lys12 ζ-amine from the same or an adjacent protomer (generally around 3.5 Å, though not hydrogen bonded) helps stabilize the excess negative charge while adopting several alternative conformations; in some orthologs, Arg12 is found to substitute. Two additional nearby acidic groups – Glu10 and Glu70 – ensure that the convex face of the hexamer is highly acidic. Pore diameter is 5 Å (i.e. the diameter of the largest object that could readily pass through) at the Asp40 amide nitrogen (which is relatively fixed), and a slightly narrower 4.5 Å at the level of the carboxylate group, which is more mobile.

Interestingly, the second independent hexamer in the crystal displays an asymmetric
conformation. Here Asp40 of one chain is inserted into the pore, making a hydrogen bond to the amide backbone of an adjacent Asp40; this conformation constricts the pore to 3.0 Å diameter (Figure 2.3G).

2.3.3 The structures of MSM0271 and MSM0275

MSM0271 crystallized in space group P6$_3$22 and diffracted to a resolution of 2.7 Å, and was refined to an $R_{\text{free}}$ of 22.4 %. The asymmetric unit contains a trimer of protomers arranged in a ring, with each copy of the chain being largely identical (r.m.s.d. values of 0.2 Å or less). A second trimer related by crystallographic symmetry completes the hexameric double ring structure. Residues 10 to the C-terminus (218) are ordered in all three protomers; however, residues 31 – 36 in the $\beta_1$ – $\alpha_1$ are disordered in all protomers (with 29 and 30 also disordered in one chain).

MSM0275 crystallized in space group C222$_1$ and diffracted to a resolution of 2.1 Å. The asymmetric unit contains six protomers arranged as a dimer of trimers. Of the 202 amino acids in the native protein 189 (chain E) to 201 (chain A) amino acids could be traced. Up to 7 additional amino acids derived from the hexa-histidine tag were ordered in chains A and B, forming a short, antiparallel four $\alpha$-helical bundle with tags from adjacent hexamers; this motif mediates the main packing interactions in the $c$-direction between layers of molecules arranged in the $a/b$ plane (not depicted).

MSM0271 and MSM0275 are tandem domain shell proteins. The protomer is built from a pair of fused, permuted pfam00936 domains, each with a four-stranded antiparallel $\beta$-sheet (1, -2, 4, -3 and 5, -6, 8, -7 topology for the N- and C-terminal repeats respectively). Two $\alpha$-helices $\alpha_2$ and $\alpha_3$ ($\alpha_5$ and $\alpha_6$ in the C-terminal domain) pack on one face of the $\beta$-sheet, while $\alpha_1$ ($\alpha_4$) is part of a long loop-helix-loop motif inserted between $\beta_1$ and $\beta_2$ ($\beta_4$ and $\beta_5$) packs on the
opposite face and mediates interaction between adjacent rings.

While the protomeric organizations of MSM0271 and MSM0275 closely resemble one another (see Figure 2.2C, D; the chains can be superimposed with an r.m.s.d. of 0.69 Å), MSM0275 and MSM0271 hexamers show emergent differences. MSM0275 rings associate in a manner similar to CsoS1D and CcmP, where interactions are mediated largely by α1 packing on its C-terminal domain equivalent helix, α4. The two-fold symmetry axis run through the diagonals of the hexagon. MSM0271, uniquely, has its domains offset by 60°, so α1 packs onto α1 and α4 onto α4; the internal symmetry axis therefore runs through the centers of the facets of the hexagon (Figure 2.2C, D). This unique arrangement requires two distinct sets of interactions to hold the rings together.

It is worth noting that the two rings in MSM0271 and MSM0275 are not properly oriented to form a double-layered sheet with canonical BMC-H-like interactions. This is because, as previously noted for CcmP16 the two trimeric rings are not appropriately aligned, but rather are offset by about 10°. One way to envision this is to map the locations of the three-fold axes at the corners of the (pseudo)hexamers; were these hexamers to partake in canonical facet packing (as seen in PduA; Figure 2.4A). These axes are offset in the two layers (triangles in Figure 2.2C, D), meaning that BMC-FP proteins cannot pack into facet-like sheets with both rings simultaneously.

The concave-to-concave interactions between rings in MSM0271 and MSM0275 create a substantive enclosed volume of space. Similar to what has previously been seen with the CsoS1D and CcmP structures15,16, access to this space is blocked by the meeting of symmetry related copies of Arg81 (Arg69 in MSM0275) meeting at the three-fold axis with Arg81 hydrogen bonding with Glu80 (Glu68) from an adjacent protomer (Figure 2.3 I, K). In
MSM0271 and MSM0275 this interaction seems especially strong, as it is stabilized by an additional inter-chain hydrogen bond contributed by Gln82 (Gln70). Arg81 also packs on Gly171 and Ala172 of the topologically equivalent loop from the C-terminal repeat. In CsoS1D and CcmP, this conformation is present in only one ring, while the second ring adopts an open conformation that reveals a substantial (~13 Å wide) pore. In MSM0271 and MSM0275 the rings on both protomers are closed, blocking access to the central cavity from outside. Atomic displacement parameters for these residues are similar to the rest of the structure, giving little indication of a tendency to opening by order-to-disorder transitions.

**Figure 2.4 Crystal packing and potential interactions mediated by RMM shell proteins.**

All are shown from the concave side of the respective (pseudo)hexamer, unless indicated otherwise. A) Crystal packing in PduA from *Salmonella enterica* Typhimurium (PDB # 3ngk), a hexameric shell protein, as a well-established model for the canonical interactions mediating facet tiling. Individual hexamers interact along their edge, forming a seal. Key residues include Lys26, Asn29, and Arg79 which reaches across the interface to interact with to the C-terminal end of α1. B) The corresponding facet of MSM0272. MSM0272 is 59 % identical to PduA, and conserves all of the important interface residues. C) The ID interaction interface of MSM0271. Note that key residues are not conserved, including Arg79 and Lys26 (here Arg171). D) The IC interaction interface of MSM0275. Here again, Arg67 is substituted for the usual Lys26. E) Crystal packing in MSM0275 in the *a/b* plane; chains A-F are colored green, red, magenta, yellow, pink white, respectively. Note that interactions between the A and B protomers follows the canonical tight facet interaction shown in Figure 2.4A. This results in a zig-zag pattern of rings linked in a continuous line (bright vs. faded). The interactions between the magenta protomers involves the β1-β2 loop adopting an extended conformation, and packing between
molecules (Figure 2.4F). F) Details of the conformation change in the β1-β2 loop. This region shows large changes in conformation between protomers, with individual residues moving up to 15 Å. The loop adopts a fully compacted conformation (cyan) where close contacts exist with adjacent molecules, and an extended conformation (orange) where there is available space between molecules. The blue conformation appears to be an intermediate conformation where most of the loop is disordered. Note that the structural transition also involves substantive rearrangement of α1. G) Interactions mediated by the tight (ID – ID) interface of MSM0275. Lys158 and Arg12 are equivalent to Lys26 and Arg79 in PduA, and mediate similar interactions. The interface is gapless, and the interactions appear at least as strong as in PduA. H) Interactions mediated by the second pseudohexamer at the same interface as in Figure 2.2G. Note that these interactions are very different than the canonical facet interactions. While Lys56 appears in the equivalent position, the geometry of the two rings of MSM0275 (and by extension, all structurally characterized BMC-FP proteins) means making the tight interactions in Figure 2.4C precludes these lysine residues from interacting similarly. Instead they are displaced laterally, so the two rings only interact across two thirds of their length. This arrangement precludes a tight seal forming between these rings, though additional stabilization is afforded by the interface formed by the opposite rings. Near identical copies the interactions shown in C and D recur at a second interface in the crystal, offset by 120°. I) Alternative, long-range ID – ID interaction in MSM0275. This interaction positions the domains much further apart. The gap between protomers is instead filled by the β1 – α1 loop, which extends into this space and interacts with the opposite ring. Note that this interface is skewed, with the interfaces angled about 10° apart. This panel shows both pseudohexameric rings, with the ID-ID interface shown from the convex side of the ring.
2.3.4 Structural plasticity in the MSM0275 $\beta_1 - \alpha_1$ loop.

All six copies of MSM0275 in the asymmetric unit can be superimposed with an r.m.s.d. 0.29 to 0.55 Å. The only region where the protomeric structures diverge significantly is in the $\alpha_1$ helix and the loops connecting this motif to $\beta_1 - \beta_2$ in the N-terminal domain (Figure 2.4F). Residues Trp24-Asn33 are disordered in two of the six protomers. In one protomer, chain F, all residues are ordered and very closely match the conformation of the equivalent residues in CsoS1D and CcmP, in a conformation that we will refer to as the retracted conformation. Two additional chains (A and B) closely match this conformation, but residues 29-32 are displaced up to 2 Å from the site of chain F. In chain D, the C-terminal end of the helix has become slightly unwound, with Trp24 sitting in the position usually occupied by Ile25, with residues 25 – 32 being disordered. The largest change is seen in chains C and E. Note that in these chains, density is weak with high atomic displacement parameters; for chain F the central residues of this region are disordered, while for chain E, residues 26-33 are partially stabilized by interactions with an adjacent molecule. In these chains, helix $\alpha_1$ partly unwinds and shifts 3.8 Å. This results in Thr23 occupying the approximate pocket occupied by Pro30 in the other chains. This rearrangement leaves residues 26-33 protruding from the surface of the hexamer, with Ala28 of chain E shifted 21 Å from the position it occupies in chain B. This conformation also leaves a narrow channel open from the central pore cavity to the surface of the facet. We refer to this loop as the extensible loop, and conformations similar to that in chain E as the extended conformation.

In the MSM0271 structure, all chains are similar. Residues 29-35 (equivalent to residues 20-24 in MSM0275) are disordered in all three chains of the MSM0271 structure, but the positioning of these residues implies that the disordered residues are located in the central cavity, rather than exposed on the surface.
2.3.5 Interaction faces and packing interactions

BMC-H proteins preferentially adopt a specific edge-to-edge interaction geometry; the interface in PduA, for example, shows the key conserved residue Lys26 packing antiparallel to its symmetry related mate, while Arg79 reaches across the interface and makes a hydrogen bond with the C-terminal end of α1 (Figure 2.4A); this interface straddles a two-fold symmetry axis, so these interactions are repeated on the other side of the interface. Lys26, Asn29 and Arg79, have all been shown to be required to form a stable, functional shell\textsuperscript{13}. In crystal structures of several BMC-H proteins, this interface is repeated at each face, resulting in a continuous, gapless sheet of molecules that is believed to serve as a close model for the packing in the microcompartment facets. We term this the “canonical” interaction geometry.

The MSM0272 crystal form obtained packs individual hexamers in non-sheet like orientations; however, MSM0272 is a close homolog of PduA, and all key residues of the canonical interface are conserved in MSM0272 (with identical numbering) (Figure 2.4B). Modeling indicates that MSM0272 should be capable of packing in very similar sheet-like arrangements, with essentially identical packing interactions.

Exploring the possible interactions of MSM0275 and MSM0271 in the microcompartment shell requires awareness of two complicating factors. Firstly, because of the fused nature of the dimers, there are two distinct potential canonical interfaces in each protein ring: one face is characterized by the meeting of two domains of one chain (inter-domain or ID facet), and the other by the meeting of adjacent chains (inter-chain or IC facet). In both proteins, these facets differ in the amino acids presented at the interface. Secondly, because these proteins form rigid back-to-back double-ringed hexamers, interactions in one ring forces a specific interaction geometry in the second ring. Interactions therefore likely require optimizing both interfaces simultaneously.
MSM0271 fails to form any meaningful lateral interactions in the crystals. Moreover, the potential interaction interface shows some substantive differences from characterized BMC-H and BMC-FP proteins. Lys26, otherwise near absolutely conserved, is replaced by arginine, Arg67/171. The arginine that usually interacts with the adjacent ring is replaced by an aspartate, Asp21/127 (Figure 2.4C, D). Despite the lack of the usual interacting motifs, the facet remains flat, and there is a strong similarity between the potential interaction surfaces formed at the ID and IC faces.

MSM0275 hexamers pack in the crystal as layers of flat sheets of molecules. The hexamer in the asymmetric unit abuts six others related by two-fold rotational and screw symmetry operations, generating four crystallographically distinct interfaces. Two of these interfaces (where chain A packs on A and chain B on B) mediate a tight interface that places the center-to-center distance between hexamers at 67.6 Å, very close to that seen in crystal structures that are the basis of the canonical single sheet facet model - PduA (3ngk; 67.2 Å) and CcmK2 (3dnc, 67.4 Å). Equivalent side chains also mediate very similar interactions (Figure 2.4G); symmetry related copies of Lys158 are in van der Waals contact, while making hydrogen bonds to Asp11 in the same chain; Arg12 reaches across the interface and interacts with the C-terminal end of helix α5. In addition, Gln187 hydrogen bonds to Asp11 carbonyl oxygen, while Thr80 hydrogen bonds with Asp184, providing additional stabilizing interactions. The packing and specific interactions involved closely match the canonical interaction; we will refer to this as the canonical MSM0275 interaction. Note that at this interface, the extendable β1 – α1 loop is fully ordered in the retracted conformation and indeed, extending it out into the space between hexamers would result in clashes. This tight interaction simultaneously forces a specific geometry on the second trimeric ring of these hexamers, which interact through the IC interface
(Figure 2.4H). These rings interact more weakly, with the only hydrogen bond occurring between symmetry related copies of Ser114, and some van der Waals contacts between Ala183 and the carbonyl oxygen atoms of Lys56 and His57. The IC facet lacks the equivalent of Asp11 and Arg12 (the equivalent residues are Ser113 and Ser114), or Gln187 (Ala85). Despite the lack of strong contacts, this surface leaves no large gaps along its interface. However, the subunits at the interface are effectively displaced 13 Å from their more usual packing positions, meaning any attempt to tile a surface with these interactions will leave large gaps at the three-fold axes. These interactions therefore appear to further stabilize the ID – ID interface, but do not themselves form a well-sealed IC - IC interface.

Unlike what has been observed for BMC-H proteins, the crystal does not display canonical packing at each face of the hexamer, so uniform two-dimensional tiling is absent. Rather, the placement of these canonical interfaces results in strips of molecules being tightly connected in extended lines, with the 120° offset of connections resulting in a zig-zag pattern (Figure 2.4E). Note that CcmP form 1 crystals pack in a very similar pattern of zig-zagged linear strips, though with an interface that is displaced from the canonical packing interaction; electron microscopy suggests that this protein prefers this organization even when unconstrained in a crystal lattice. It should be noted that these canonical ID-ID MSM0275 interactions cannot readily be extended into a model for packing MSM0275 into a continuous facet, as the ID facet edge is found alternatively in the top and bottom trimers of the hexamer as one proceeds around the ring. Forming the canonical interface alternatively through the upper and lower rings results in severe clashes, as the off-centered placement of the neighboring IC – IC interface causes it to protrude into the space required for the ID-ID interaction.

A third independent interface also features ID-ID packing, between chain C and chain E
This packing interface is looser, reflecting a significantly longer center – to center distance between hexamers (72.4 Å vs 67.6 Å), and the facets are angled about 10° apart in plane, so packing along one face is closer than the other (possibly because the constraints of crystal packing preclude optimal interactions). This interface is characterized by interactions mediated by the α1 – β1 loop which is in the fully extended conformation from both chains, though only fully ordered in chain E, where interactions are closer (Figure 2.4I). Leu29 from this loop forms van der Waals interactions with Pro15C, Ile25C and Pro134F, while Arg31E hydrogen bonds with Asp161C. Additional interactions are formed by Glu79E hydrogen bonding with Gln187C, and the guanidinium groups of Arg12E/C being in van der Waals contact. Residues 29 – 32 of chain C are disordered, but were they to adopt a conformation similar to those of chain E, a sealed interface would seem at least possible. In the opposing ring, chains A and F interact through van der Waals interactions over a reasonably wide interface, but chains B and D (the second half of this IC interface) are too far apart to interact.

The final independent interface in the crystal is again different. This interface pairs an ID with an IC interface, with centers of hexamers 69.2 Å apart. The resulting interface is weakly stabilized by van der Waals packing, but appears loosely packed and porous.

2.3.6 Modeling the vertex by docking

Tanaka et al., in the original study investigating the organization of the vertex by computational modeling, suggest that the BMC-P have an appropriate size and shape to plug a pentameric defect where five facets of BMC-H proteins meet. However, this pioneering study was limited by trying only one of the six candidate hexameric shell proteins for modeling, and the use of rigid body docking. In addition, the authors modeled the pentameric defect between CcmK2 and CcmL by simply bending the canonical interface to the required 141° angle. However, the acute angle between BMC-H/FP proteins at the vertex introduces the possibility
that they pack substantively differently, as the vertex is not an extension of the interactions in the facet, but rather of those at the edge. We instead systematically investigated the local organization of the RMM vertex by docking (with side chain flexibility) each of the three candidate BMC-H/FP proteins in a geometry appropriate for forming the juxta-vertex ring, followed by docking of the BMC-P into the resulting pentameric defect (Figure 2.5A). Docking results were compared on the basis of their interface score (I_sc; the difference in energy between the complex and its component structures: typically an I_sc of -5 to -10 is considered “good”) as well as total buried surface area.

**Figure 2.5 Modeling the RMM vertex by docking.**

A) Overall logic of the docking. The formation of the vertex requires forming two distinct protein-protein interfaces; these are investigated sequentially using Rosetta dock. First (1) the interface between BMC-H (MSM0272) or BMC-FP (MSM0271 or MSM0275) was investigated. Two copies of a given protein (blue and cyan) were positioned with the appropriate geometry for a vertex, and allowed to rotate and translate along their rotational symmetry axis. The interface was then repacked and evaluated using Rosetta dock. For the BMC-H protein this was done separately with the concave side facing outward (as shown) or inward. For the BMC-FP proteins, the ID and IC faces were evaluated separately. The highest scoring pose for each protein was then fixed and used to dock MSM0273 (2), the BMC-P, again systematically investigating rotation and translations around the five-fold symmetry axis. B) Table summarizing results of docking of each BMC-H/FP shell protein to itself, then to MSM0273. N.D. indicates that the docking was not done, due to poor docking results in the first stage. C) I_sc vs rmsd (versus top scoring pose) for docking of MSM0271 on MSM0271. The presence of a funnel, where all of the lowest energy solutions closely resemble the best solution, is indicative of a successful docking
run. D) $I_{sc}$ vs rmsd for docking MSM0273 (concave outward) into the previously docked MSM0271 ID-ID complex. E) Final model of the complex. MSM0271 is in cyan, MSM0273 in orange. The second ring of each interacting MSM0271 hexamer is shown as cartoon. F) Details of the MSM0271-MSM0271 interface. Inset on right shows the same interaction looking along the interface. Note that the interface occurs over a wide surface. G) The same MSM0271-MSM0271 interface, shown looking down the interface from the side. H) Details of the MSM0271-MSM0273 interface.
As Tanaka et al. point out\textsuperscript{23}, single ringed BMC-H proteins can plausibly meet at a five-fold axis with either their concave or convex face on the inner edge of the vertex. For BMC-FP proteins, however, we observe that the presence of a second bulky ring on the concave face implies that these proteins can only interact at the vertex with the convex face inward (i.e. localized to the lumenal side of the compartment), so that the second rings diverge from one another rather than collide. We also assume that if BMC-FP proteins line the vertex, they are likely do so consistently with a given face (ID or IC) forming BMC-FP self-interactions, and the other face forming BMC-FP to BMC-P interactions; any other organization would force a complex mixture of interfaces, with an asymmetric pentameric defect housing the symmetric MSM0273. In order to minimize assumptions about vertex organization, we started by modeling the self-interactions of the ring of BMC-H/FP molecules immediately adjacent to the vertex. This necessitates modeling six different interactions (MSM0272 with the convex and concave side inward, and MSM0271 and MSM0275 each interacting through their distinct IC and ID faces). The highest scoring model for each protein was then used to model interactions with MSM0273, with MSM0273 in either possible orientation (base in or base out).

BMC-H/FP self-docking yielded solutions with roughly comparable interaction scores for all three proteins; however, none of these interactions generated the canonical-facet-bent model originally proposed for lining the vertex. MSM0272 gave significantly better packing with the convex side facing inward than outward ($I_{sc}$ of -6.99 vs -5.24; see Figure 2.5B). Interestingly, the optimal (convex inward) interaction geometry occurs where the edges are offset laterally 5 Å from their positions in the canonical facet interaction; indeed, no good packing arrangement was found for any BMC protein with the distorted canonical-like interactions originally assumed by Tanaka \textit{et al.}\textsuperscript{23}. This MSM0272 interaction results in a tight seal along the edge, with no
discernible gaps. The interactions extend the interacting surface to include interactions between side chains at the C-terminal end of \( \alpha 2 \). Docking MSM0273 onto this optimal complex, however, does not yield any convincing solutions; the highest observed \( I_{sc} \) is -18.05, but examination of the model shows that the protruding helices \( \alpha 1 \) and \( \alpha 2 \) mediating most of the contacts, and the clefts left between these helices leave large gaps (~10 Å) between molecules (note that five separate interfaces are being summed in these models, resulting in much higher \( I_{sc} \) scores). In the optimal concave inward packing, \( \alpha 2 \) from MSM0272 forms a prominent ridge that precludes a tight fit with MSM0273, with docking giving poor packing (\( I_{sc} \)-12.38), resulting in even larger (~14 Å) gaps between molecules. Similarly poor packing was observed for MSM0273 in either orientation docking onto the optimal convex-outward MSM0272 complex. While MSM0272 can form plausible self-interactions at the vertex, these position the hexameric rings in a way that makes productive docking of MSM0273 impossible.

Modeling MSM0275 self-interactions around the vertex suggests that this protein interacts significantly more strongly through the IC face than the ID face (\( I_{sc} \)-5.21 vs -3.31). The optimal packing is displaced about 5 Å laterally from the canonical interface, but in the opposite direction that that seen in MSM0272; this then results in a very different interface. The model has hydrogen bonds between Lys56 and Ser114, and between Asp92 and Arg193, with a small pore (~3 Å) at the center of the interface. Docking MSM0273 into the IC complex however yields sub-optimal packing, with \( I_{sc} \) of -15.11, and significant gaps in the interface where interacting MSM0275 rings meet the MSM0273 pentamer.

In contrast with MSM0275, MSM0271 packs with noticeably better geometry along its ID face than its IC face (\( I_{sc} \)-6.09 vs -5.22). Optimized packing at the IC face pulled the molecules into configuration with geometry incompatible with a juxta-vertex ring; this solution
was therefore not considered further. The interaction along the ID face positions the rings in an orientation analogous to that seen as optimal for MSM0275 (Figure 2.5C, F, G). Similar to that interaction, Arg171 hydrogen bonds with Glu20 at the facet-like interface, while Arg116 hydrogen bonds with Glu112, residues provided by the longer helix α6 and the β4-β5 loop (joining domains). Leu93 also packs against Arg171 and Val173, contributing hydrophobic stabilization. Effectively, the ID face in MSM0271 is padded out by these features, extending the interacting area when the proteins are oriented at a 141° angle. The buried interface is considerably larger (1802 Å²) than that observed for any of the other self-interactions modeled. Docking MSM0273 into this model results in a considerably better fit than any of the other models, with an I_sc of -21.38. This arrangement allows favourable electrostatic interactions between Glu95 (MSM0271) and Arg80 (MSM0273) and between Arg98 and Asp78, as well as various hydrogen bonds and favourable hydrophobic interactions (Figure 2.5D, E, H). This configuration has no large gaps in the interface, allowing a good seal. Docking therefore suggests the preferred geometry for organizing the RMM vertex is with MSM0271 self-interacting through its ID face, with MSM0273 docking into the resulting pentameric hole with its base facing outwards.

2.4 Discussion

The four proteins whose structures are described above are presumed to together build the shell of the RMM microcompartment. In particular, the organization of these building blocks must somehow account for all architecturally distinct environments in the polyhedral shell (at minimum bulk facet, edge, vertex and juxta-vertex) and provide passage for all metabolites that need to enter or leave the microcompartment.

2.4.1 Substrates, transport and selectivity

Bacterial microcompartment shells need to allow passage of all substrates and products
required in the reactions of the encapsulated enzymes. However, the need to transport stoichiometric quantities of bulky co-factors is now understood to be bypassed by pairing reactions which generate and consume each co-factor within the shell\textsuperscript{32} (for example, the \text{NAD(P)H} required for the alcohol dehydrogenase and aldehyde dehydrogenase encoded in the RMM operon is likely to be handled in this way). The presence of an amino-transferase homolog, as well as work indicating that the alcohol dehydrogenase acts on 1-amino-2-propanol strongly indicates that the initial substrate for the RMM is an amine. MSM0272 has a highly electronegative central pore lined with aspartate residues (Figure 2.3D-G). This motif seems well suited to interact with an amine substrate, and likely provides the substrate ingress channel. The nature of the product(s) and route of egress is more difficult to discern. The presence of the aldehyde dehydrogenase implies that one of the products is likely anionic (e.g. a carboxylic acid or a phosphate ester), making it a problematic substrate for the highly anionic MSM0272 pore; an amine acceptor and its product (e.g. α-ketoglutarate/glutamate) may also require passage. Possibly the more electropositive central pore in MSM0273 might provide a channel, though there are likely only 12 copies of this complex in the shell (one at each vertex of an icosahedral particle) limiting flux, and there is little that differentiates this pore from all related BMC-P proteins, meaning that there is unlikely to be much specificity associated with this pore. The fact that the number of distinct central pores is likely smaller than the number of substrates passaged is interesting as it implies that either these pores are not the only available route, or the shell has limited substrate specificity. An alternative model is that specific metabolite channels form where packing of specific faces of specific shell proteins leaves a gap. This model has the advantage that the pores can, like the substrates, be asymmetric, allowing a greater degree of specificity; using double-shelled BMC-FP proteins in this manner has the advantage that the
second pseudo-hexamer can provide an additional set of buttressing interactions (as seen for the MSM0275 canonical interaction) that compensates from the loss of interaction energy where a pore is formed. This idea of specific pores emerging from the interaction between appropriate pairs of shell proteins is interesting, but difficult to address experimentally.

2.4.2 Transport roles of BMC-FP proteins

Previous work on BMC-FP proteins has lead to the proposal that these act as gated channels, where weak binding of a metabolite within the pore opens the exterior gate\textsuperscript{16}. It is not clear whether this would apply to these MSM0271 and MSM0275, especially as stochastic gate opening events would be required to first allow entry of gate triggering metabolites from the microcompartment lumen. We propose instead that the gates at either end may open randomly through thermal fluctuations, allowing any small enough molecule (~1 kDa) to enter or leave from either the cytoplasmic or lumenal face. These permuted double-ring shell proteins would then function as sub-stoichiometric, non-specific transporters for larger metabolites, most importantly co-enzymes. Recent results have demonstrated that bacterial microcompartments maintain co-enzyme pools separate from the general metabolic pools, where e.g. NADH consumed in one reaction will be regenerated in another\textsuperscript{32}. However, regeneration is likely not completely efficient over time. Any escape from the microcompartment of the metabolic intermediates between the two opposed reactions will lead to a gradual accumulation of the co-factor in one state or the other. In addition, NAD(P)H can be oxidized non-enzymatically by oxidizing agents, and can also be damaged in side reactions\textsuperscript{83}; without some mechanism to replenish these cofactor pools, microcompartments would gradually lose efficacy as they age. Slow, sub-stoichiometric exchange of co-factors with the larger cytoplasmic pool would ensure that these co-factors retain access to repair mechanisms, and can be replenished if depleted. The cationic nature of the space between gates may help recruit co-enzymes, which are typically
anionic due to their phosphate-rich nucleotide groups. In carboxysomes, this mechanism may be required to allow encapsulated Rubisco activase access to the ATP it requires to function\textsuperscript{84,85}.

### 2.4.3 Organization of the RMM shell

The RMM microcompartment investigated here is the first BMC for which the structures of all candidate shell proteins are known. These structures together strongly constrain the possible organization of the shell, given the requirement that a largely sealed capsule needs to be generated from interactions between these four proteins. In keeping with what has previously been discovered with other microcompartments, MSM0272 is a straightforward hexameric BMC closely related to PduA that has canonical edge interaction motifs that should allow it to readily tile in flat sheets to form the shell facets\textsuperscript{19}. MSM0273 seems a typical BMC-P protein; the role of these proteins in plugging pentameric defects in the icosahedral shells that are created where five facets meet is generally accepted in the field\textsuperscript{21,23,38}. The role of the two BMC-FP proteins is more difficult to ascertain, as there are no published models that clearly suggest how they contribute to shell organization. We argue that, at least in the RMM microcompartments, these proteins play a critical role in forming the edges and vertex, a conclusion that has interesting geometric implications.

Incorporating BMC-FP protein(s) into a BMC shell introduces an important inner / outer layer distinction into descriptions of the shell. In particular, assuming that BMC-H/FP proteins can only interact strongly along their edges when arrayed in parallel (i.e. with their convex face facing the same direction), then the outer layer has its concave side, and the inner layer its convex side facing the lumen. Because the inter-ring interaction geometry of BMC-FP proteins seems to preclude them from readily organizing in continuous sheets, and the BMC-H protein should occupy only a single layer, the bulk of the facets probably show a mostly single-layered, BMC-H protein dominated, organization, possibly with isolated BMC-FP proteins embedded
(though this organization may not be true of β-carboxysomes, where abundant CcmK2 can form double-ringed dodecamers with geometry appropriate for extended tiling\(^{86}\)). One critical constraint on shell organization is that, in common with many BMC-H proteins, MSM0272 is likely required to interact with helix-forming peptides fused onto the termini of encapsulated enzymes. Since this occurs through the α3 helix on the BMC-H concave face\(^{64}\), MSM0272 proteins should strongly prefer to be localized in the outer shell layer. However, modeling of possible interactions between BMC-H/FP proteins and the BMC-P at the vertex seems to strongly favour models with concave-outward interactions in the inner layer. Docking between MSM0272 rings does not produce favorable packing in vertex models where these hexamers are assumed to be localized to the outer layer; placing MSM0272 at the inner layer of the vertex also seems unlikely, given that it does not seem to form favourable packing interactions with MSM0273, and interactions with the encapsulated enzyme targeting peptides will in any case tend to pre-orient it with its concave side inward. Docking studies suggest that the vertex is built from MSM0271 interacting with MSM0273 through its inner ring, with ID faces mediating MSM0271 self-interactions, and the IC faces mediating interactions with MSM0273. However, this model is only possible if the edges of the particle also meet in the inner layer of the shell.

To date, there has been relatively little discussion in the literature of the organization of the microcompartment edge. The edge of the facet truncates the regular pattern of horizontal tiling, and therefore leaves a serrated edge that must somehow be welded to a similar edge. It is worth noting that the particle vertex is formed by the terminal members of this edge pattern. The only detailed model for the organization of the edge of a microcompartment proposed to date is for the \textit{Eut} microcompartment, where EutS forms a bent wedge shaped hexamer that could join two facets at the appropriate angle\(^{75}\); however it should be noted that both the \textit{Clostridium} EutS
homolog, and the closely related protein PduU form conventional symmetric hexameric rings, raising questions as to the wider applicability of this model\textsuperscript{76,87}. The RMM seems to lack a similar protein, meaning that one or more of the other shell proteins is required to stabilize the edge through a different mechanism. In MSM0275, tight, canonical interfaces are found at alternating facets, organizing the molecules into linear strips in a zig-zagged fashion (Figure 2.4E)\textsuperscript{35}. The close lattice match between MSM0275 and PduA (a close homolog of MSM0272) suggests that these strips could be seamlessly added to the edge of a facet made of MSM0272 hexamers. Two such edging strips would then meet at a 141° angle to form an edge. While the interaction geometry likely does not allow tight interactions between the bodies of the rings in all positions, the extensible $\beta_1-\alpha_1$ loop, unrestricted by tight packing, could extend and organize so as to maximize interactions with the abutting strip, sealing the gap. The emergence of this motif from the equator of the hexamer is useful in this regard as these motifs would be exposed from both rings where molecules are bent back at the edge. Parallel lines of MSM0275 molecules that meet in a hinge-like arrangement and are sealed by this adaptable loop would therefore provide a mechanism for mediating sealing the edges where two facets meet (see Figure 2.6).

In conclusion, we present for the first time the structures of a complete set of microcompartment shell proteins. These strongly constrain the possible architecture and transport strategies of the RMM microcompartment. The BMC-FP proteins in particular present several novel features that may point to an important role in shell architecture, including the ability to form well-sealed edges in the icosahedral particle.
Figure 2.6 Overall model for the organization of the RMM shell.

MSM0273 (orange), the BMC-P protein occupies the vertices, surrounded by a ring of the BMC-FP protein MSM0271 (blue) at each vertex. Interactions at the vertex occur through only the inner ring of MSM0271. The other BMC-FP protein MSM0275 (green) is organized in long strips that line the edge of the facets. The bulk of the facet is made of MSM0272 (yellow) the BMC-H protein. Two lines of MSM0275 proteins meet at the edge. Here imperfect geometric match between the MSM0275 strips is compensated for by packing the extendable loops between the adjacent rings.
2.5 Experimental procedures

2.5.1 DNA methods

*Mycobacterium smegmatis* mc² 155 was grown in LB, and after heat killing and centrifugation, genomic DNA was purified following the Invitrogen mini-prep kit procedure. For cloning, genes were amplified from genomic DNA using PfuX7 (a generous gift from Dr. D. Christendat) and ligated into vectors so as to leave either an N-terminal (pET28a) or a C-terminal (pET22b) hexa-histidine tag. MSM0271 (UniProt # A0QP48_MYCS20) was amplified using primers GGTTCGCATATGGTCGCACCGGAAACCGAGAGGATCCGTACC and GGTTCGCTCGAGTCAGTGTTCCTTCGTCCATCGCGCTCAGC and ligated into the NdeI and XhoI sites of pET28a. MSM0272 (UniProt # A0QP49_MYCS2) was amplified using primers GAGCTACATATGTCCAGCAACGCAATCGGATTGATC and GATGCTCTCGAGTCAGTGTTCCTTCGTCCATCGCGCTCAGC and ligated into the NdeI and XhoI sites of pET22b. MSM0273 (UniProt # A0QP50_MYCS) was amplified using primers ACGTATCTCGAGCTCGGCGGGTTGCTATCTGATCTG and TAGCAACATATGTGAGAGCGACCGAGAATTGC and ligated into the NdeI and XhoI sites of pET22b. MSM0275 (UniProt # A0QP52_MYCS) was amplified using primers TCGGGTCATATGGCAGAATTACGTTCTCCTTCATCTTCATCG and ATTTCAAAGCTTCCACCCTGCAGGACCGGCAGG and ligated into the NdeI and HindIII sites of pET22b. Plasmids were propagated in *E. coli* DH5α cells and sequenced at the University of Guelph Agriculture and Food Laboratory Services facility. Plasmids containing the correct sequences were used to transform *E. coli* BL21 cells for over-expression.

2.5.2 Protein Expression and Purification

Large volume cultures (1 L 2YT media with appropriate antibiotic - kanamycin 30 μg/mL for pET28a or ampicillin 50 μg/mL for pET22b) were inoculated with small volume overnight
cultures and incubated at 37°C with shaking until an optical density of 0.8 at 600 nm was reached. Cultures were induced for protein expression by addition of 0.4 mM isopropyl β-D-1-thiogalactopyranoside, followed by incubation for 18 - 20 hours at 16 °C with shaking. Cultures were pelleted by centrifugation at 4 °C for 20 min at 4400 x g. Cell pellets were resuspended in 35 mL lysis buffer (20 mM Tris pH 8, 500 mM NaCl; lysis buffer for MSM0272 contained added 10% glycerol) prior to freezing at -20 °C.

Pellets were thawed at 4°C prior to cell disruption by sonication. The insoluble fraction of cell lysate was pelleted by centrifugation (at 4°C for 30 min at 48,000 x g) and supernatant was filtered through a 0.45 μm filter prior to loading on a 1 mL Ni-NTA agarose column for immobilized metal affinity chromatography (IMAC) purification. Bound proteins were washed with 10 mL of the respective lysis buffer followed by 10 mL of wash buffer (lysis buffer plus 40 mM imidazole). Proteins were eluted with 10 mL of elution buffer (lysis buffer plus 500 mM imidazole).

2.5.3 Crystallization
For crystallization, MSM0271, MSM0273 and MSM0275 were dialyzed into crystallization buffer (20 mM Tris pH 8.0, 150 mM NaCl) while MSM0272 was kept in elution buffer. Proteins were concentrated by centrifugal concentration and the final concentration was measured using ninhydrin. MSM0271 (10 mg/mL) crystallized as hexagonal prisms by sitting drop vapour diffusion at room temperature with a protein solution to crystallization solution (0.8 M Na/K tartrate, 0.1 M Tris pH 8.5, 0.5% (w/v) PEG 2000MME) ratio of 2:1 or 1:1. MSM0272 (7 mg/mL) crystallized as tetragonal prisms by sitting drop vapour diffusion at room temperature with a protein solution to crystallization solution (0.1 M Tris pH 7.5, 11% ethylene glycol, 10% PEG 8000) ratio of 2:1 or 1:1. MSM0273 (10 mg/mL) crystallized as hexagonal prisms by sitting drop vapour diffusion at room temperature with a protein solution to crystallization solution (90
mM sodium acetate pH 4.5, 1.8 M NaCl, 10 mM TCEP) ratio of 2:1 or 1:1. MSM0275 (20 mg/mL) crystallized as needles by sitting drop vapour diffusion at room temperature with a protein solution to crystallization solution (0.5 M Na citrate pH 5.6, 20% PEG3350) ratio of 2:1 or 1:1. Crystals were cryoprotected with Paratone® N (MSM0271, MSM0272 and MSM0275) or a 3:5 vol:vol glycerol/mother liquor mixture (MSM0273) prior to flash freezing in liquid nitrogen.

2.5.4 Data collection and structure determination

All datasets were collected at the Canadian Light Source beamline 08ID at 100 K using a wavelength of 0.97949 Å. Data were processed using XDS, and scaled using XSCALE. All structures were phased using molecular replacement by Phaser in Phenix and initially rebuilt using Autosol (Phenix). Structures were iteratively manually rebuilt in Coot, followed by refinement in Phenix. MSM0273 was in space group P3121 and diffracted to 1.6 Å; a pentameric polyalanine model of CcmL (PDB # 2QW7) was used for molecular replacement searches. MSM0272 crystallized in spacegroup P43 and diffracted to 2.2 Å, a EutM hexamer (PDB # 4AXJ) was used as a search model for molecular replacement. MSM0275 was in the space group C2221, diffracted to 2.1 Å, and was solved using a hexamer of CcmP (PDB # 4HT5) as a molecular replacement search model. MSM0271 was in the P6322 space group and diffracted to 2.7 Å, a trimer of MSM0275 was used as a search model for molecular replacement. Ramachandran statistics for percent outliers/favoured were 0/99.5, 0.2/97.2, 0/97.8 and 0.3/96.6 for MSM0273, MSM0272, MSM0271 and MSM0275, respectively. Structure statistics are reported in Table 2.1. All structure figures were prepared in Pymol.
Table 2.1 X-ray data collection and refinement statistics

<table>
<thead>
<tr>
<th></th>
<th>MSM0273</th>
<th>MSM0272</th>
<th>MSM0271</th>
<th>MSM0275</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCSB I.D.</td>
<td>5L37</td>
<td>5L38</td>
<td>5SUH</td>
<td>5L39</td>
</tr>
<tr>
<td><strong>Data collection</strong>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P3\textsubscript{1}21</td>
<td>P4\textsubscript{3}</td>
<td>P6\textsubscript{3}22</td>
<td>C222\textsubscript{1}</td>
</tr>
<tr>
<td><strong>Cell dimensions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$, $b$, $c$ (Å)</td>
<td>66.97, 66.97, 140.09</td>
<td>93.60, 93.60, 192.54</td>
<td>139.89, 139.89, 150.12</td>
<td>132.14, 143.48, 116.99</td>
</tr>
<tr>
<td>$\alpha$, $\beta$, $\gamma$ (°)</td>
<td>90, 90, 120</td>
<td>90, 90, 90</td>
<td>90, 90, 120</td>
<td>90, 90, 90</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50-1.6 (1.64-1.6)</td>
<td>50-2.2 (2.26-2.2)</td>
<td>50-2.7 (2.77-2.70)</td>
<td>50-2.1 (2.15-2.10)</td>
</tr>
<tr>
<td>$R_{\text{sym}}$</td>
<td>0.064 (0.875)</td>
<td>0.063 (0.745)</td>
<td>0.133 (0.979)</td>
<td>0.109 (0.897)</td>
</tr>
<tr>
<td>$I/\sigma I$</td>
<td>17.7 (2.3)</td>
<td>18.4 (2.5)</td>
<td>19.2 (2.1)</td>
<td>12.0 (2.4)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>0.992 (0.984)</td>
<td>1.00 (1.00)</td>
<td>0.999 (1.00)</td>
<td>1.00 (1.00)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>7.3 (7.1)</td>
<td>5.1 (5.1)</td>
<td>9.6 (11.1)</td>
<td>7.5 (7.5)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>45.8 - 1.6</td>
<td>48 – 2.2</td>
<td>47.1 – 2.7</td>
<td>48.6 – 2.1</td>
</tr>
<tr>
<td>No. reflections</td>
<td>52789</td>
<td>83642</td>
<td>24385</td>
<td>64967</td>
</tr>
<tr>
<td>$R_{\text{work}} / R_{\text{free}}$</td>
<td>0.154/0.181</td>
<td>0.20/0.233</td>
<td>0.173/0.224</td>
<td>0.182/0.224</td>
</tr>
<tr>
<td>No. atoms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>2890</td>
<td>4228</td>
<td>4685</td>
<td>8826</td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>-</td>
<td>5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Water</td>
<td>370</td>
<td>925</td>
<td>26</td>
<td>334</td>
</tr>
<tr>
<td><strong>B-factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>24.3</td>
<td>32.5</td>
<td>75.8</td>
<td>45.8</td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>-</td>
<td>34.4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Water</td>
<td>37.1</td>
<td>43.0</td>
<td>57.2</td>
<td>46.7</td>
</tr>
<tr>
<td><strong>R.m.s. deviations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.017</td>
<td>0.004</td>
<td>0.010</td>
<td>0.005</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.525</td>
<td>0.684</td>
<td>1.278</td>
<td>0.872</td>
</tr>
</tbody>
</table>

*All datasets were collected from a single crystal. Values in parentheses are for the highest-resolution shell.
2.5.5 Docking

Docking was performed using Rosetta_Dock with flexible side chains in Rosetta 3.6\textsuperscript{92}. Interacting molecules were pre-positioned using a customized script, then minimized using local docking refinement. For BMC-H/FP self-docking, poses were iteratively rotated in 0.5° increments, and translated along the rotation axis in 0.5 Å increments, with a total of ~1000 docking poses evaluated per candidate complex. Docking with MSM0273 in both a concave-in and concave-out conformation was conducted into a pentameric ring of previously docked BMC-H/FP proteins, utilizing rotations and translations around the central symmetry axis. Because of the computational expense of docking the large number of atoms involved and larger angular and translational ranges, sampling here was coarser (1°, 1 Å); ~600 poses were tested per candidate interaction. Buried surface area was assessed using the CCP4 program areaimol\textsuperscript{93}. 
3 Structure and kinetics of the S-(+)/1-amino-2-propanol dehydrogenase from the RMM microcompartment of *Mycobacterium smegmatis*

This chapter has been published in the journal *Biochemistry* under the same title (Mallette and Kimber, 2018).94.

Statement of Contribution

The primary author, Evan Mallette, designed and conducted experiments, analyzed the data, and contributed to writing the manuscript and visualization. Matthew S. Kimber conducted bioinformatics analysis, acquired funding, and contributed to writing of the manuscript and visualization.

Acknowledgements

We thank P. Grochulski and S. Labiuk at Canadian Light Source for assistance in collecting diffraction data.
3.1 Abstract
S-(+)-1-amino-2-propanol dehydrogenase (APDH) is a short chain dehydrogenase/reductase associated with the incompletely characterized Rhodococcus and Mycobacterium bacterial microcompartment (RMM). We enzymatically characterized the APDH from M. smegmatis, and showed it is highly selective, with a low-micromolar $K_m$ for S-(+)-1-amino-2-propanol, and specificity for NADP(H). A paralogous enzyme from a non-microcompartment associated operon in the same organism was also shown to have similar activity. We determined the structure of APDH in both apo form (at 1.7 Å) and as a ternary enzyme complex with NADP$^+$ and aminoacetone (at 1.9 Å). Recognition of aminoacetone was mediated by strong hydrogen bonds to the amino group by Thr145 and by Glu251 from the C-terminus of an adjacent protomer. The substrate binding site entirely encloses the substrate, with close contacts between the aminoacetone methyl group and Phe95, Trp154, and Leu195. Kinetic characterization of several of these residues confirm their importance in enzyme functioning. Bioinformatics analysis of APDH homologs implies that many non-microcompartment APDH orthologs partake in an aminoacetone degradation pathway that proceeds via an aminopropanol O-phosphate phospholyase. RMM microcompartments may mediate a similar pathway, though possibly with differences in the details of the pathway that necessitates encapsulation behind a shell.

3.2 Introduction
1-amino-2-propanone (aminoacetone) is something of an “accidental metabolite” in that while it is commonly found in cells, its presence is generally not due to any purposeful enzymatic reaction or dedicated transporter. Instead, aminoacetone is evolved through the unwanted spontaneous decarboxylation of 2-amino-3-ketobutyrate, an intermediate in the widespread glycine/threonine interconversion pathway. Humans dispose of aminoacetone by
deaminating it to methylglyoxalate, which is subsequently oxidized to pyruvate; both aminoacetone and methylglyoxyl have toxic cellular effects, something that may be exploitable in cancer cells that utilize glycine catabolism to thrive in ischaemic conditions\textsuperscript{95,96}. Work in the sixties and seventies, predominantly by Turner and colleagues, established that some bacteria can utilize aminoacetone as a carbon and nitrogen source\textsuperscript{97,98,99}. Working with enzyme enrichments from aminoacetone-grown bacterial cultures, the first enzyme in this pathway was established as a NADP\textsuperscript{+} dependent S-(+)-1-amino-2-propanol dehydrogenase (APDH; E.C. 1.1.1.75), which catalyzes the reversible reduction of aminoacetone (aminopropanone) to S-(+)-1-amino-2-propanol (aminopropanol). Characterization of natively purified enzyme from \textit{Rhodococcus} showed activity with a variety of short chain amines, though aminopropanol was the best substrate, and no activity was found with ethanolamine\textsuperscript{43}. However, the focus of this research was on the utility of this enzyme as a means to catalyze the stereospecific manufacture of ephedrine\textsuperscript{100}, and the basic biochemistry of APDH remains poorly characterized.

The eventual sequencing of APDH from \textit{Rhodococcus erythryopolis} MAK154\textsuperscript{43,80} showed this protein to be a member of the highly abundant short chain dehydrogenase/reductase (SDR) family. SDR enzymes have a Rossmann fold, and catalyze mostly redox reactions, though some variants act instead as dehydratases, dehalogenases, decarboxylases, carbohydrate epimerases or steroid isomerases\textsuperscript{39}. The SDR reaction mechanism depends upon a tyrosine residue that, in coordination with a conserved lysine through a ribose hydroxyl group, acts as the general acid, while an additional conserved serine residue acts to stabilize and polarize the carbonyl oxygen\textsuperscript{39}; all of these key amino acids appear to be conserved in APDH sequences. Under modern SDR classification, APDH is considered to be a classical SDR, of the bacteria-only family SDR207C\textsuperscript{101}. 
APDH in *Rhodococcus erythryopolis* MAK154 was noted to be located within an operon that encodes a largely uncharacterized bacterial microcompartment. Microcompartments are cytosolic bodies with an enzyme core that produces, and then consumes, a metabolite (generally an aldehyde) that is volatile and/or cytotoxic. Escape of this problematic metabolite is limited by encapsulating the key enzymes in a protein shell, built from several structural proteins from two distinct families. These shells are generally polyhedral or icosahedral, and are thought to share some architectural principles, but not evolutionary origins, with large T-number icosahedral viruses. Because microcompartment shells are generally though to preclude facile passage of large metabolites, cofactors needed by encapsulated enzymes are generally found to be recycled within the microcompartment. The microcompartment in the operon encoding APDH is now termed the RMM microcompartment, named for the fact that it is commonly found in the actinobacterial groups *Rhodococcus* and *Mycobacterium*. While extensive work characterizing the ethanolamine, propanediol and α- and β-carboxysome microcompartments has been published, many microcompartments, including the RMM microcompartment, remain little understood. There are several direct and indirect indications that the RMM microcompartment catabolizes an aminoalcohol or amino ketone, including: 1) aminopropanol as well as longer chain aminoalcohols (aminobutanol and aminohexanol) both induce expression of the operon and serve as good substrates for the encoded APDH; 2) the operon contains a transporter related to amino-acid permeases; and 3) recent work on the organization of the shell of the *Mycobacterium smegmatis* RMM shows that the main pore forming protein has a uniquely acidic pore consistent with passage of an amino containing metabolite. Fuller understanding of this pathway will require detailed characterization of the associated enzymes. We report here the detailed structural and functional characterization of the APDH from the RMM.
microcompartment of *Mycobacterium smegmatis* MC\(^2\) 155, and analyze the functional implications for orthologs found in different genetic contexts, and the larger SDR207C family of enzymes.

3.3 Materials and methods

3.3.1 Cloning and Site-Directed Mutagenesis

*Mycobacterium smegmatis* MC\(^2\) 155 was grown in lysogeny broth; after heat killing and centrifugation, genomic DNA was purified following the Invitrogen mini-prep kit procedure. For cloning, APDH\(_{MSM0269}\) (YP_884684) and APDH\(_{MSM0779}\) (YP_885184) were amplified using PfuX7 polymerase (a generous gift from Dr. Dinesh Christendat, University of Toronto) from genomic DNA (Table 3.1) and ligated into the multiple cloning site of pET28a so as to include a N-terminal hexa-histidine tag. All primers were purchased from Invitrogen. Mutations were introduced by site-directed mutagenesis (Table 3.1) using PfuX7 polymerase to generate variant enzyme constructs. Plasmids were transformed into chemically competent *Escherichia coli* DH5\(\alpha\) cells, and colonies were sequenced at the University of Guelph Agricultural and Food Laboratory Services facility. Plasmids containing correct sequences were used to transform chemically competent *E. coli* BL21(DE3) cells for overexpression.
<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSMEG0269 FWD</td>
<td>5’-GAGCTACATATGTTACCTCTCTTTGAGGGCCGATCG-3’</td>
</tr>
<tr>
<td>MSMEG0269 REV</td>
<td>5’-GATGCTGATCTATACACATTCTCGCGATCGGAGGTGC-3’</td>
</tr>
<tr>
<td>MSMEG0779 FWD</td>
<td>5’-CGTACATATGTTCGACCTGCAGGGCCGCTCAGTGGTG-3’</td>
</tr>
<tr>
<td>MSMEG0779 REV</td>
<td>5’-GCTACTCGAGCTACGTCGATCGCGATCGCGATCAGGACT-3’</td>
</tr>
<tr>
<td>W154A FWD</td>
<td>5’-TACCGGCGGCTGACTACCGGGCCAGCAAGGCG-3’</td>
</tr>
<tr>
<td>W154A REV</td>
<td>5’-GATGCGAGCGCGCGATCGGATCGGACC-3’</td>
</tr>
<tr>
<td>T145A FWD</td>
<td>5’-TCGTCGATCGGGTGTCATCAGGGGATATCCGCTCAGGA-3’</td>
</tr>
<tr>
<td>T145A REV</td>
<td>5’-CGAATCAGGCGGCTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC-3’</td>
</tr>
<tr>
<td>E251A FWD</td>
<td>5’-GAGGTGATCGGGCAGTTGACGTGCGAGGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACG-3’</td>
</tr>
</tbody>
</table>
3.3.2 Protein expression and purification

Large volume cultures (1 liter of 2xYT media with 30 μg/ml kanamycin for selection) were inoculated with overnight cultures and incubated at 37 °C with shaking until an optical density of 0.8 at 600 nm was reached. Cultures were induced for protein expression with 1 mM isopropyl β-D-1-thiogalactopyranoside, followed by incubation for 18-20 h at 16 °C with shaking. Cultures were pelleted by centrifugation at 4,400 x g. Cell pellets were resuspended in 35 mL of lysis buffer (20 mM Tris pH 8.0, 500 mM NaCl) prior to freezing at -20 °C.

Pellets were thawed at 4 °C prior to cell disruption by sonication. The insoluble fraction of the cell lysate was pelleted by centrifugation (at 4 °C for 30 min at 48,000 x g), and supernatant was filtered through a 0.45 μm filter prior to loading onto a 1 mL HisTrap FF Crude IMAC column for immobilized affinity chromatography purification. Protein bound to the column was washed with 10 mL each of lysis buffer, lysis buffer plus 20 mM imidazole, and lysis buffer plus 40 mM imidazole. Protein was eluted with 10 mL of lysis buffer plus 500 mM imidazole. Protein eluate was exchanged into reaction buffer (20 mM Tris pH 8.0, 150 mM NaCl) by buffer exchange on a HiPrep 26/10 desalting column. Protein was concentrated by centrifugal concentration, and the concentration was determined by spectrophotometric absorbance at 280 nm on a NanoDrop. Protein concentration was determined for W154A variant by the ninhydrin method\(^2\).

3.3.3 Steady state kinetics assays

Candidate substrates were screened for activity in the direction of NADP\(^+\) reduction at room temperature (~22°C) in 500 μL reactions containing 20 mM Tris pH 8.0, 150 mM NaCl, 500 μM NADP, 50 μg/mL enzyme with 1 mM alcohol substrate (2-propanol, ethanolamine, R-(−)-1-amino-2-propanol, S-(+)-1-amino-2-propanol, and (R/S)-1-amino-2-butanol). The increase in
absorbance at 340 nm was measured for 1 min following addition of substrate on a Cary Bio 300 spectrophotometer.

Kinetic constants were determined at 25°C in 20 mM Tris pH 8.0, 150 mM NaCl, by varying the concentration of one substrate while keeping the other substrate at a constant concentration (100 μM NADP(H) and S-(-)-1-amino-2-propanol or 50 μM 1-amino-2-propanone) with appropriate enzyme concentrations. Reaction progress was monitored by change in absorbance at 340 nm due to change in concentration of NADPH on an Applied Photophysics SX20 stopped-flow apparatus equipped with a 2 mm path length cell. Reactions consisted symmetric mixing of enzyme solution with substrate solution followed by data collection for 20 seconds with a total of 400 absorbance readings per reaction. Initial rates were calculated by averaging the initial linear slope for 7-8 reactions per concentration measured by change in NADPH concentration per concentration of enzyme per second (ε_{340} 0.00622 Lμmol⁻¹cm⁻¹). Kinetic parameters \(k_{cat}, K_M, \frac{k_{cat}}{K_M}\), and \(K_I\) were calculated by fitting to initial rate data using non-linear regression in GraphPad Prism 6. The Michaelis-Menten equation was used to model the data, with substrate inhibition for aminoacetone.

Kinetic parameters were determined for the substituted phenylethylamine substrate 2-aminoacetophenone in the direction of NADPH oxidation at room temperature (~22°C) in 400 μL reactions containing 20 mM Tris pH 8, 150 mM NaCl, 200 μM NADPH, with an appropriate amount of enzyme with varying concentrations of 2-aminoacetophenone. The decrease of absorbance at 340 nm was measured over 1 minute following addition of substrate on a Cary Bio 300 spectrophotometer. Kinetic parameters were calculated by fitting the data by non-linear regression in GraphPad Prism 6 fit to the Michaelis-Menten equation with substrate inhibition.
3.3.4 Variant activity

Enzyme variants were assayed for activity in the direction of NADP reduction at room temperature (~22 °C) in 500 μL reactions containing 20 mM Tris pH 8, 150 mM NaCl, 500 μM NADP, 50 μg/mL enzyme with 1 mM S-(+)-1-amino-2-propanol. The increase in absorbance at 340 nm was measured for 1 minute following addition of substrate for 4 replicates each on a Cary Bio 300 spectrophotometer.

Kinetic parameters for W154A reduction of 1-amino-2-propanone were determined at room temperature (~22°C) in 400 μL reactions containing 20 mM Tris pH 8, 150 mM NaCl, 200 μM NADPH, with an appropriate amount of enzyme by measuring the decrease in absorbance at 340 nm on a Cary Bio 300 spectrophotometer and fitting the data by non-linear regression in GraphPad Prism 6 fit to the Michaelis-Menten equation.

3.3.5 Crystallization

APDH<sub>MSM0269</sub> without substrates crystallized by sitting drop vapour diffusion at room temperature with a protein solution (20 mg/mL) to crystallization solution (1 M tri-sodium citrate pH 6.3) ratio of 1.5:1 or 1:1. The prismatic crystals were cryoprotected with Paratone N prior to flash-freezing in liquid nitrogen. APDH<sub>MSM0269</sub> was co-crystallized with substrates by sitting drop vapor diffusion at room temperature with a protein solution (10 mg/mL protein, 5 mM S-(+)-1-amino-2-propanol, 5 mM NADP<sup>+</sup>) to crystallization solution (0.1 M Tris pH 8.5, 0.2 M MgCl<sub>2</sub>, 13% PEG 4000) ratio of 1:1. Crystals were again prismatic and were cryoprotected in a mixture of 25% (v/v) MPD with 75% (v/v) crystallization solution prior to flash-freezing in liquid nitrogen.
3.3.5 Structure determination.
   Data was processed in XDS and scaled in XSCALE\textsuperscript{88}. Structure determination was completed in Phenix\textsuperscript{90}, and model building was performed in Coot\textsuperscript{91}. The structure of apoAPDH was determined at 1.7 Å from the trigonal spacegroup P3\textsubscript{1}21, using molecular replacement with a putative short-chain dehydrogenase/reductase from \textit{Mycobacterium abscessus} (PDB code 3RIH) as a search model. There were two molecules in the asymmetric unit, representing half a tetramer. The final structure has all amino acids for both chains, while one chain also has three residues from the His\textsubscript{6} tag ordered at the N-terminus. Structure figures were prepared in Pymol v1.8.

3.3.6 Bioinformatics
   The sequence of MSM0275 (YP_884684) was used for a BLAST search against the NCBI protein sequence database. This protein is well conserved among RMM microcompartments, with E values < -80, and good discrimination against non-orthologs (the few hits with -80 > E > -60, appear to possibly be microcompartments with RMM-like shell proteins repurposed with new enzymes). Gene neighbourhoods of hits were checked to ensure other RMM proteins were co-localized. The sequence similarity network (SSN) was produced using EFI-EST\textsuperscript{103} searching with APDH\textsubscript{MSM0269} with a cutoff of E\textsuperscript{-49}. Gene neighbourhoods were analyzed using this SSN with an edge cutoff of E\textsuperscript{-100} in EFI-GNT. Data was visualized and analyzed in cytoscape\textsuperscript{104}.

3.3.7 Accession codes
   APDH\textsubscript{MSM0296} is YP_884684 and APDH\textsubscript{MSM0779} is YP_885184. The apo APDH\textsubscript{MSM0296} structure has been deposited at the RCSB with PDB i.d. 6CI9; the holo structure with PDB i.d. 6CI8.
3.4 Results

3.4.1 Substrate specificity and kinetics of APDH$_{\text{MSM0269}}$

APDH from the RMM microcompartment of *Mycobacterium smegmatis* MC$^2$ 155 (APDH$_{\text{MSM0269}}$) was successfully cloned into pET28a and purified using IMAC chromatography. The enzymatic activity of APDH was then measured (monitoring changes in nicotinamide absorbance) for a variety of candidate substrates. Maximal activity was measured for aminoacetone reduction by NADPH, with a $k_{cat}$ of 5.1 s$^{-1}$ and a $K_m$ of 2.9 μM (Figure 3.1, Table 3.1). $K_m$ for the cofactor NADPH in this reaction was 16.4 μM. Aminoacetone also showed what appears to be weak substrate inhibition, with an apparent $K_I$ of 593 μM. Substrate inhibition is a common phenomenon in NAD(P)H dependent dehydrogenases in general, including SDRs$^{105}$. For the reverse reaction, the oxidation of S-(+)-1-amino-2-propanol using NADP$^+$, a $k_{cat}$ of 0.534 s$^{-1}$ and a $K_m$ of 6.8 μM was observed (Figure 3.1, Table 3.2). The $K_m$ for NADP$^+$ in this reaction was 2.0 μM. In contrast, no activity was measurable with R-(−)-1-amino-2-propanol, indicating that the reaction is strictly stereospecific. In addition, no activity could be observed with either 2-propanol or ethanolamine, indicating that both the amino group and methyl group are required for turnover. On the other hand, bulkier analogs of the methyl group appear to be tolerated, albeit with reduced turnover and higher $K_m$ (Figure 3.2, Table 3.2). (R/S)-1-amino-2-butanol oxidation proceeded with a $k_{cat}$ of 0.334 s$^{-1}$ and a $K_m$ of 29.8 μM (presumably enantiomerically pure (S) substrate would exhibit a $K_m$ half of this value). We also tested the commercially interesting substrate 2-aminoacetophenone, which substitutes a bulky phenyl group for aminoacetone’s methyl group. This molecule showed a $k_{cat}$ of 3.30 s$^{-1}$ with a $K_m$ of 34.1 μM, indicating that this bulky substituent can be accommodated with a relatively modest loss in catalytic efficiency.
3.4.2 Kinetics of a non-microcompartment associated APDH, APDH<sub>MSM0779</sub>

Recombinant non-RMM microcompartment associated APDH from Mycobacterium smegmatis MC<sup>2</sup> (APDH<sub>MSM0779</sub>) was successfully expressed and purified. Candidate substrates identified in the above work were similarly tested for activity (Figure 3.2, Table 3.2). APDH<sub>MSM0779</sub> proved to be slightly less active than APDH<sub>MSM0269</sub> with aminoacetone, with a $k_{cat}$ of 1.59 s<sup>-1</sup> and a $K_m$ of 5.1 µM; however, there was no substrate inhibition observed in the range substrate concentration tested. S-(-)-1-amino-2-propanol showed a five-fold lower activity ($k_{cat}$ of 0.103 s<sup>-1</sup>) with APDH<sub>MSM0779</sub> than APDH<sub>MSM0269</sub> and a higher $K_m$ of 20.2 µM.

![Figure 3.1 Michaelis-Menten nonlinear regression fitting of APDH<sub>MSM0269</sub>.](image)

Initial rates of reaction measured as the absolute change in NADPH concentration per second for A) reduction of aminoacetone by NADPH and B) oxidation of S-(-)-1-amino-2-propanol by NADP<sup>+</sup>.
Figure 3.2 Nonlinear regression of APDH$_{MSM0269}$ and APDH$_{MSM0779}$ kinetics data

Nonlinear regression of steady-state kinetic data for: APDH$_{MSM0269}$ with A) NADPH, B) NADP, C) 2-aminoacetophenone, and D) R/S-1-amino-2-butanol; APDH$_{MSM0779}$ with E) 1-amino-2-propanone, and F) S-(+)-1-amino-2-propanol; APDH$_{MSM0269W154A}$ with G) 1-amino-2-propanone. Nonlinear regression fitting of data and graphing was performed in GraphPad Prism6, as outlined in Materials and Methods.
Table 3.2 Steady-State Kinetic Parameters for *Mycobacterium smegmatis* MC2 155 APDH variants<sup>a</sup>

<table>
<thead>
<tr>
<th>APDH&lt;sub&gt;MSM0269&lt;/sub&gt; WT</th>
<th>Substrate</th>
<th>$k_{cat}$ (s&lt;sup&gt;−1&lt;/sup&gt;)</th>
<th>$K_m$ (μM)</th>
<th>$k_{cat}/K_m$ (M&lt;sup&gt;−1&lt;/sup&gt;s&lt;sup&gt;−1&lt;/sup&gt;)</th>
<th>$K_i$ (μM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1-amino-2-propanone</td>
<td>5.1 ± 0.1</td>
<td>2.9 ± 0.3</td>
<td>1.76 x 10&lt;sup&gt;6&lt;/sup&gt;</td>
<td>593 ± 98</td>
</tr>
<tr>
<td></td>
<td>2-aminoacetophenone&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3.30 ± 0.07</td>
<td>34.1 ± 2.1</td>
<td>9.7 x 10&lt;sup&gt;4&lt;/sup&gt;</td>
<td>4230 ± 356</td>
</tr>
<tr>
<td></td>
<td>NADPH</td>
<td>4.92 ± 0.07</td>
<td>16.4 ± 0.8</td>
<td>3.00 x 10&lt;sup&gt;5&lt;/sup&gt;</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>S-(-)-1-amino-2-propanol</td>
<td>0.534 ± 0.009</td>
<td>6.8 ± 0.4</td>
<td>5.2 x 10&lt;sup&gt;4&lt;/sup&gt;</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>S/R-1-amino-2-butanol</td>
<td>0.334 ± 0.008</td>
<td>29.8 ± 2.4</td>
<td>1.1 x 10&lt;sup&gt;4&lt;/sup&gt;</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>NADP&lt;sup&gt;+&lt;/sup&gt;</td>
<td>0.471 ± 0.007</td>
<td>2.0 ± 0.2</td>
<td>2.36 x 10&lt;sup&gt;5&lt;/sup&gt;</td>
<td>n/a</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>APDH&lt;sub&gt;MSM0779&lt;/sub&gt; WT</th>
<th>Substrate</th>
<th>$k_{cat}$ (s&lt;sup&gt;−1&lt;/sup&gt;)</th>
<th>$K_m$ (μM)</th>
<th>$k_{cat}/K_m$ (M&lt;sup&gt;−1&lt;/sup&gt;s&lt;sup&gt;−1&lt;/sup&gt;)</th>
<th>$K_i$ (μM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1-amino-2-propanone</td>
<td>1.59 ± 0.02</td>
<td>5.1 ± 0.2</td>
<td>2.94 x 10&lt;sup&gt;5&lt;/sup&gt;</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>S-(-)-1-amino-2-propanol&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.103 ± 0.003</td>
<td>20.2 ± 2.3</td>
<td>5.15 x 10&lt;sup&gt;3&lt;/sup&gt;</td>
<td>n/a</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>APDH&lt;sub&gt;MSM0269&lt;/sub&gt; W154A</th>
<th>Substrate</th>
<th>$k_{cat}$ (s&lt;sup&gt;−1&lt;/sup&gt;)</th>
<th>$K_m$ (mM)</th>
<th>$k_{cat}/K_m$ (M&lt;sup&gt;−1&lt;/sup&gt;s&lt;sup&gt;−1&lt;/sup&gt;)</th>
<th>$K_i$ (μM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1-amino-2-propanone&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5.6 ± 0.1</td>
<td>1.26 ± 0.06</td>
<td>4.4 x 10&lt;sup&gt;3&lt;/sup&gt;</td>
<td>n/a</td>
</tr>
</tbody>
</table>

<sup>a</sup> Reactions were assayed under conditions including 20 mM Tris-HCl, 150 mM NaCl at pH 8 and 25˚C or<sup>b</sup> room temperature ~22˚C. Errors for kinetic parameters are the standard deviations determined from nonlinear regression fitting of kinetic data as described in Materials and Methods. Nonlinear regression fitting had R<sup>2</sup> values greater than 0.99 with the exception of<sup>c</sup> which had a R<sup>2</sup> value of 0.9876.
3.4.3 Structure of apo APDH

The structure of apo APDH<sub>MSM0269</sub> was determined in spacegroup P3<sub>1</sub>21 at a resolution of 1.7 Å. APDH is a classic short chain dehydrogenase/reductase (SDR) with a single Rossmann-fold domain (Figure 3.3A). The two chains in the asymmetric unit form half of a D<sub>2</sub> symmetric tetramer with an overall square shape (Figure 3.3B). The protomer is organized around a central seven-stranded parallel β-sheet, with three α-helices packed on either side. α4 and α5 pack adjacently along one face of the sheet, with their N-termini protruding well above the β-sheet; these protruding helix ends are stabilized by packing on the C-terminal ends of a symmetry related helix pair. α5, and the loop preceding it, contribute all the key catalytic residues to the active site pocket. α6 and α7 are loosely packed relative to the rest of the structure, with elevated atomic displacement parameters in both chains; these two helices form a helix-turn-helix (HTH) motif that lines one side of the substrate binding pocket. APDH<sub>MSM0269</sub> has an unusual additional feature in α9 - a short α-helix at the end of an extended C-terminal loop that does not interact with the rest of the protomer; instead, this motif extends diagonally across the tetrameric interface to pack between α5 and α7 of a symmetry related molecule. α9 may help stabilize the HTH motif in an open conformation; in many apo SDR structures (which lack α9), the HTH motif is generally too disordered to reliably trace.

3.4.4 Structure of the ternary APDH complex

The structure of APDH<sub>MSM0269</sub> as a ternary complex was determined at a resolution of 1.9 Å in the monoclinic spacegroup P2<sub>1</sub> (see Table 3.3 for structure statistics). This structure has four complete tetramers in the asymmetric unit, where two have relatively low atomic displacement parameters (ADPs; also known as B-factors; chains A-D and E-H), and two have relatively high ADPs (chains I-L and M-P). The HTH motif and α9, for all 16 chains, are located in solvent channels, are unconstrained by crystal packing contacts and display the highest ADPs
in the structure. In some active sites the density for both motifs is fairly well defined, but in others there is clear indications of two or more competing conformations, with Gln200-Met206 in α7 showing weak, ambiguous density; typically this correlates with weaker density for NADP⁺ and aminoacetone, and these sites generally refine with less than complete occupancy for these ligands (with a minimum occupancy of 0.85). The quality of the density does not allow a well-defined alternative conformation to be modeled, but does imply that access of the substrate/product to the active site may be mediated by shifting or partially unwinding the α-helix, with Tyr202 flipping out of the active site. While some active sites suggest a tendency towards disorder, strong clear density for these more mobile motifs and ligands is present in multiple active sites, most notably chains C, D and H, allowing substrate binding to be unambiguously analyzed.

Comparing the ternary complex structure to the apo conformation reveals that the most pronounced displacements are of the HTH motif (α6 and α7), which moves substantially to close the active site, while most of the rest of the structure seems to form a rigid platform (Figure 3.3C). In the ternary complex structure Leu195, Met198, Phe202, and Met206 pack tightly against the substrate, and aromatic residues (contributed by α5 and the extended β4-α4 and β5-α5 loops) most notably Phe95, Tyr151 and Trp154 pack tightly against α6 and α7. The HTH motif rotates as an approximately rigid body, with the fulcrum passing approximately through Asn189 and Leu210, and Gln200 at the tip moving approximately 7 Å inward. One unique aspect of this repacking is that α9 occupies a position in the apo structure that is occluded by α7 in the ternary complex. As a result, this motif becomes partially disordered in most protomers, with a large increase in ADPs or complete disorganization of residues C-terminal to Ser252.
Figure 3.3 Structure of APDH.

A) organization of the protomer. B) Organization of the tetramer. One chain is coloured with a
colour gradient, and the protomer contributing its C-terminus into the catalytic site is in pale
cyan. C) overlay of apo (blue tones) and holo complex (white, with mobile regions highlighted in
beige) of APDH structure. The structures are overall highly similar with r.m.s.d. values of 0.22
Å. Significant differences are apparent in two regions that contribute to the active site – the
helix-turn-helix motif, which shifts significantly to close the active site, and the C-terminal helix,
which is displaced by the incoming HTH motif and becomes largely unstructured. D) details of
the active site. Aminoacetone is shown in green, with electron density contoured around it at 1.5
σ. Residues from the HTH motif are shown in beige. Residues contributed by an adjacent
protomer are shown in cyan. Residues covering the active site are partially transparent.
Figure 3.4 Details of interactions in the NADP+ binding site.

Colouring is identical as in figure 3.3 D.
### Table 3.3 X-ray data collection and refinement statistics

<table>
<thead>
<tr>
<th></th>
<th>APDH apo</th>
<th>APDH holoenzyme</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RCSB I.D.</strong></td>
<td>6C19</td>
<td>6C18</td>
</tr>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wavelength</td>
<td>0.97949</td>
<td>0.979</td>
</tr>
<tr>
<td>Space group</td>
<td>P3121</td>
<td>P2₁</td>
</tr>
<tr>
<td>Cell dimensions, a, b, c (Å)</td>
<td>104.24 104.24 80.86</td>
<td>111.29, 163.57, 128.00</td>
</tr>
<tr>
<td>Cell dimensions, α, β, γ (°)</td>
<td>90, 90, 120</td>
<td>90, 95.713, 90</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50 - 1.7 (1.74 – 1.7)</td>
<td>50 – 1.9 (1.95 - 1.9)</td>
</tr>
<tr>
<td>R&lt;sub&gt;sym&lt;/sub&gt;</td>
<td>0.054 (0.68)</td>
<td>0.088 (0.612)</td>
</tr>
<tr>
<td>I / σI</td>
<td>19.9 (3.04)</td>
<td>10.04 (2.36)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>100 (100)</td>
<td>99.9 (99.9)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>6.2 (6.2)</td>
<td>3.8 (3.8)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50 - 1.7</td>
<td>50 – 1.9</td>
</tr>
<tr>
<td>No. reflections</td>
<td>56008</td>
<td>356596</td>
</tr>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt; / R&lt;sub&gt;free&lt;/sub&gt; (%)</td>
<td>14.4 / 17.0</td>
<td>20.4 / 23.3</td>
</tr>
<tr>
<td>No. atoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>3782</td>
<td>29710</td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>2 (Cl⁻)</td>
<td>873</td>
</tr>
<tr>
<td>Water</td>
<td>604</td>
<td>2461</td>
</tr>
<tr>
<td>B-factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>22.2</td>
<td>40.9</td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>24.6</td>
<td>34.5</td>
</tr>
<tr>
<td>Water</td>
<td>32.6</td>
<td>42.9</td>
</tr>
<tr>
<td>R.m.s. deviations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.017</td>
<td>0.007</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.33</td>
<td>0.882</td>
</tr>
<tr>
<td>Ramachandran plot</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favored (%)</td>
<td>96.9</td>
<td>96.0</td>
</tr>
<tr>
<td>Allowed (%)</td>
<td>100</td>
<td>99.9</td>
</tr>
</tbody>
</table>

*Each dataset was collected from a single crystal. Values in parentheses are for the highest-resolution shell.*
3.4.5 NADP$^+$ cofactor binding

Examination of the active site showed clear electron density that was best interpreted as NADP$^+$ (the nicotinamide ring is flat, without the characteristic pucker of NADPH) and aminoacetone bound. This then is a non-productive complex, with both substrates oxidized. Electron density for both ligands is clear and unambiguous in all active sites. NADP$^+$ binds with a conformation typical for SDRs, with interactions to the pyrophosphate moiety being mediated by T$_{13}$GGSKGIG in the $\beta$1-$\alpha$1 loop, which corresponds to the “TGxxx[AG]xG” cofactor-binding motif characteristic of classic SDRs (see Figure 3.4). The O2 phosphate of NADP$^+$ sits in an empty pocket formed by Gly38, with Arg39 forming three hydrogen bonds and, along with Lys17, providing positive charge; the presence of these two basic residues serve as well-studied markers for NADP(H) specificity in SDRs$^{106}$. In one chain per tetramer, the O2 phosphate coordinates a magnesium ion, where the other five ligands are water molecules. While the geometry of this complex is very similar in all four instances, the coordinating waters are stabilized by a crystal packing interaction (most notably one of the water ligands is hydrogen bonded to the carbonyl oxygen of Gly84), meaning this interaction is likely not relevant in vivo. The nicotinamide amide nitrogen hydrogen bonds with the $\beta$-phosphate group. Lys161 (the catalytic lysine) hydrogen bonds with both the O2 and O3 atoms of the nicotinamide ribose. Multiple interactions are made with the HTH motif, with hydrogen bonds by Ile190 (N and O), Thr192 (O$\gamma$) and Gly194 (N) being made to the nicotinamide amide, $\beta$- and $\alpha$-phosphate respectively. These interactions likely help position this motif correctly for substrate binding.

3.4.6 Aminoacetone binding

Aminoacetone is entirely buried in the active site, with no contact with any water molecules and no exposure to the surface (Figure 3.3D). The shape of the pocket conforms extremely closely to the shape of the ligand, helping explain the observed specificity. The
carbonyl carbon stacks on C4 of the nicotinamide ring with approximately 3.5 Å separating the groups; in a productive complex with NADPH, this distance may be reduced by the nicotinamide ring pucker. As seen in all SDR-ternary complexes, the carbonyl oxygen atom forms hydrogen bonds with the catalytic tyrosine and serine, here Tyr157 and Ser143. These two residues typically serve to polarize the carbonyl bond, and act as a general acid, with a proton being relayed from Lys161 through the ribose O2.

The aminoacetone amino group sits in a hydrophilic pocket, and makes strong hydrogen bonds to Thr145, the backbone oxygen of Gly188, and the carboxylate group of Glu251 of a second protomer. This glutamylate is the only negatively charged residue in this pocket, and serves to stabilize the positive charge on the amino group. We tested these key residues by determining the relative activity of a series of single point mutations. The T145A mutation completely abolished turnover, while E251A retained 4% of wild type activity. Glu251’s role is reminiscent of glucose 1-dehydrogenase, where the terminal Gly261 carboxylate (positioned somewhat differently) hydrogen bonds with O4 and O6 of the substrate\(^{107}\). In these hexose dehydrogenases, the C-terminal tail contributes to sugar binding and closes off the active site. In hexose dehydrogenases, however, the helix-turn-helix motif does not shift, but rather forms a well-defined binding pocket that is present even in the apo structure.

The aminoacetone methyl group sits in a non-polar pocket formed by Phe95, Trp154, Tyr157, and Leu195. Here the methyl group is in van der Waals contact with nearby residues; longer substrate side chains (such as the butyryl group) can likely only be accommodated by repositioning one or more side chains. We propose that conservation of these residues serves as a determinant for aminoacetone specificity. Consistent with this idea, the W154A variant had a \(k_{\text{cat}}\) of 5.6 s\(^{-1}\) but a \(K_m\) of 1.26 mM (Figure 3.2), indicating severely compromised substrate binding.
Unlike the wildtype enzyme, this variant shows no appreciable substrate inhibition in the substrate concentration range tested.

It is worth noting that the only substrate interactions mediated by the HTH motif are a single H-bond (Gly188) and a small amount of non-polar interaction (by Leu195). The polar residues interacting with the amino group are, however, enmeshed in an extended series of hydrogen bonds that connect to the nicotinamide ring, Tyr202, Thr149, Gln247. This network couples the two adjacent active sites, and may serve to couple NADP(H) binding, substrate binding, and closure of the α6 - α7 lid domain.

3.4.7 Similarity with related structures
SDRs are known to be one of the largest enzyme families; searching the pdb using apoAPDH co-ordinates in DALI\textsuperscript{81} finds over a thousand hits with Z-scores greater than 14. The closest hits, 3PK0 and 3RIH\textsuperscript{108}, have Z-scores of over 40, and sequence identities around 60\%. This work has shown that 3PK0 (APDH\textsubscript{MSM0779}) is an ortholog and analysis (see below) suggests that 3RIH is also likely a \textit{Mycobacterial} APDH ortholog. Multiple functionally characterized proteins are also found with Z-scores > 35, r.m.s.d. values between 1.5 and 2.0 Å; these proteins all have sequence identities between 30 and 39\%, indicating that they are not especially close homologs. These structures include dihydroanticapsin 7-dehydrogenase (5ITV), glucose 1-dehydrogenase (3AY7), 1-(4-hydroxyphenyl)-ethanol dehydrogenase (4URF), human retinal short-chain dehydrogenase/reductase 3 (1YDE), human 17 β-hydroxysteroid dehydrogenase type 14 (5HS6), alcohol dehydrogenases from \textit{Lactobacillus kefir} (4RF4) and \textit{L. brevis} (1ZK4), tropinone reductase-II (1IPF), and noroxomaritidine/norcraugsodine reductase (5FF9). Note that none of these enzymes have substrates that closely resemble aminoacetone – none have primary amino groups alpha to the keto/hydroxyl and none prefer substrates as small.
3.4.8 Phylogenetic distribution, genetic context and variability of SDR207C

SDR207C family members appear to be most commonly found in Actinobacteria, Bacilli, Fusobacteria and Proteobacteria (α, β, and γ), though a few isolated examples are found in Verrumicrobia and eukaryotic algae (e.g. *Emiliania huxleyi*). SDR207C encompasses a fair amount of sequence diversity, and most are not associated with microcompartment gene clusters. Interestingly, BLAST searches using the characteristic shell protein MSM0275 in NCBI (followed by inspection of the genomic neighbourhood) identified RMM gene clusters not only in Actinobacteria (*Mycobacterium* and *Rhodococcus* as previously identified, as well as *Nocardioides*, *Gordonia*, and *Streptomyces*), but also in a handful of α-proteobacterial strains (including *Fodinicurvata fenggangensis* [WP_026988258], *Afifella pfennigii* [WP_026379786], *Bauldia litoralis* [SDB54821], *Roseovarius* sp. HI0049 [KZY36693]) as well as one β-proteobacterium, *Verminephrobacter eiseniae* [WP_041949776]). APDH genes associated with RMM microcompartments share sequence identities of 55% or more, with the proteobacterial homologs being relative outliers but conserving all key APDH functional residues (Figure 3.5A). Together this would seem to imply that RMM microcompartments are significantly more phylogenetically widespread than previously identified.

A search against Uniprot using MSM0269 using BLAST with an E\(^{-49}\) cutoff captures all identifiable APDH-like SDRs (defined here as those with conservation of both amino group coordinating residues – Thr145 and Glu251) while excluding other SDR families. A sequence similarity network was generated from these proteins using EFI-SSN, and visualized in Cytoscape\(^{103,104}\). Using a stringent edge cutoff of E\(^{-100}\), these sequences can be resolved into several large clusters which are united by phylogeny, conserved variations in critical functional residues, and/or co-localization within similar gene clusters (Figure 3.5B). In order of cluster size, cluster I (red) contains sequences from α- and γ-Proteobacteria, and Bacilli. Cluster II
(green in Figure 3.5B), contains Firmicute, α- and γ-Proteobacterial proteins in which the GR motif that binds 3´ phosphate of NADP(H) is replaced by a conserved DI motif (implying likely NAD(H) specificity and a reaction biased towards aminopropanol oxidation). Cluster III contains all RMM-associated APDH orthologs (note that Uniprot seems to lack the Proteobacterial strains that have RMM operons, so these are all Actinobacteria). Cluster IV is also populated exclusively by Actinobacterial proteins (including MSMEG_0779), but these sequences are all found outside of RMM gene clusters. In all four of these clusters, the residues that bind aminoacetone in the APDH_{MSM0269} structure appear to be conserved; these proteins are likely all to be APDH orthologs. Two groups appear likely to act on different substrates: cluster VIII is a group of Rhizobiales (α-Proteobacteria) sequences which shows a conserved Trp154 to His variant, and utilize NAD(H) as a co-factor. Cluster X are from Pseudomonas strains (γ-Proteobacteria) and have W154 replaced with Leu, while Asn189 and Gln247 (which co-ordinate an active site hydrogen bond network) are replaced with small hydrophobic residues; in addition, there is a two amino acid deletion in the HTH motif. Together this implies a considerable reorganization of the substrate binding pocket. This pattern of changes extends to several smaller clusters that are united at slightly less stringent E values (highlighted in pink in Figure 3.5B) which contain sequences from Actinobacteria, Firmicutes or other γ-Proteobacteria.

This SSN was further analyzed with ESI-GNT, which identifies proteins (as classified by pfam) commonly associated with each cluster of proteins within an SSN network (Figure 3.5C). Consistent with earlier analysis, sequences from cluster III (which includes APDH_{MSM0269}) are found to be closely associated with seven PFAM families. These proteins include a regulator (pfam00392), an amino acid permease (pfam13520), two microcompartment shell proteins (pfam00936 is the BMC-H; PF03319 is the BMC-P), and three enzymes: an aminotransferase

class III – pfam00202; a phosphotransferase – pfam01636; and an aldehyde dehydrogenase – pfam00171). Of interest is the fact that two of these families - the aminotransferase and the phosphotransferase - are also associated with the majority of APDH homologs in clusters I and IV; in addition, cluster I proteins also have representatives of the permease and regulator proteins. This indicates that the different clusters may be operating in similar pathways which nevertheless differ in their details.
Figure 3.5 Bioinformatics analysis of APDH orthologs.

A) Multiple sequence alignment of APDH homologs. Reference numbers for sequences: MSMEG_0269 is YP_884684; MSMEG_0779 is YP_885184; Rhodo_eryth is A1IG83; Roseovarius is KZY36693.1; Pseudomonas is A6V011; Agrobacterium is B9K0Q7; Staph is G5JJ69. Secondary structure for apo APDH is marked above the alignment. Green circles below the alignment denote key catalytic residues, pale yellow are residues that bind the co-factor, dark blue are residues that bind the O3 phosphate and specify NADP or NAD as the cofactor; orange circles residues that contribute to the methyl binding pocket, light blue are residues that bind the substrate amino group; pale yellow are residues that contribute indirectly to the organization of the aminoacetone binding site.

B) Sequence similarity network for APDH homologs. Sequence nodes are connected when the pairwise BLAST E value exceeds $E^{-100}$. Group 1 (red) contains $\alpha$- and $\gamma$- Proteobacteria and Bacilli; group 2 (green) $\beta$- and $\gamma$-Proteobacteria, and Staphylococcus strains, group 3 (blue) contains RMM-associated proteins from Actinobacteria (with non-RMM associated sequences designated in cyan), while group 4 (yellow) is again Actinobacteria, but these proteins are not associated with RMM gene clusters. Among the smaller clusters, those with Leu154 or His154 are denoted in pink or orange, respectively. C) Gene Neighbourhood Network for clusters I, III and IV from B. Red, blue and yellow hexagons represent APDH homologs from each cluster. Other shapes represent genes found within 10 open reading frames of APDH in more than 50 % of the genomes represented by the cluster. Grey lines linking APDH an APDH of a given cluster to a protein indicate a pattern of co-association of these genes within a genome. APH – aminoglycoside phosphotransferase; ADH – aldehyde dehydrogenase; AT3 – aminotransferase; Reg – regulatory protein; BMC-H, BMC-H, BMC-FP1 and BMC-FP2 are the hexagonal,
pentameric, and two fused permuted bacterial microcompartment shell proteins, respectively. Other clusters in B do not share these gene associations. D) MSM0269 is organized in an operon with structural components of the bacterial microcompartment. Colours and gene indicators as in C. E) MSM0779 is not associated with the RMM shell proteins, but is associated with more distant homologs of two of the enzymes, as well as with the permease.
3.5 Discussion

Our kinetic characterization of APDH\textsubscript{MSM0269} found that the enzyme will oxidize aminopropanol but completely lacks activity against ethanolamine – a candidate substrate that will be routinely present in cells. In studies of the \textit{Rhodococcus} APDH enzyme, similar levels of activity were reported with 1-amino-2-butanol, and appreciable activity was found with 2-amino cyclohexanol. However, the \textit{Rhodococcus} enzyme assays used 50 mM substrate\textsuperscript{80} - five orders of magnitude higher than the $K_m$ – conditions where substrates with high and low micromolar affinities would not be readily distinguishable. The low micromolar $K_m$ APDH\textsubscript{MSM0269} displays for aminoacetone is surprisingly low, especially given the small size of the substrate. This specificity may be required to avoid reactions with off-target substrates such as ethanolamine, or may reflect \textit{in vivo} substrate concentrations being very low. Analysis of the structure indicates that this high affinity is attained by utilizing every possible interaction to the fullest extent. In particular, Thr145 and Glu251 make strong directed hydrogen bonds to the amino group, and appear absolutely conserved over this protein family, while Phe95, Trp154 and Leu195 envelop the methyl group, maximizing van der Waals interactions. In summary, both structural analysis and enzyme kinetics strongly argue that APDH is evolutionarily optimized to reduce aminoacetone.

Bioinformatics analysis of APDH homologs reveals that this SDR207C protein family likely encompasses some diversity in both the substrate specificity, and metabolic pathway in which the enzyme is embedded. Most identified homologs outside of RMM gene clusters appear to be NADP(H)-dependent APDH orthologs. Interestingly, \textit{M. smegmatis} MC\textsuperscript{2} 155 contains one such non-RMM gene (APDH\textsubscript{MSM0779}) in addition to APDH\textsubscript{MSM0269} within the RMM operon (see Figure 3.5E). Sequence comparisons do not suggest any obvious conserved differences between the RMM associated and non-RMM associated versions of the APDH proteins, and the structure
of this protein (3PK0) is very similar to the apo structure of \( \text{APDH}_{\text{MSM0269}} \). Cluster II sequences (Figure 3.5B) conserve the hallmarks of aminopropanol specificity, but replace \( \text{Gly39/Arg40} \) motif with \( \text{Asp/Ile} \); this \( \text{Gly39} \) to \( \text{Asp} \) and loss of \( \text{Arg40} \) combination has been shown to result in a drastic switch from \( \text{NADP(H)} \) to \( \text{NAD(H)} \) specificity in the SDR, alcohol dehydrogenase\(^{106}\). These enzymes are also not associated with the characteristic enzyme cluster discussed below. This suggests that these proteins are APDH homologs that \textit{in vivo} oxidize L-aminopropanol, rather than reduce aminoacetone. More radical alterations are seen in homologs such as \textit{Pseudomonas aeruginosa} PA07 (PSPA7_1010), where \( \text{Trp154} \) is substituted with \( \text{Leu} \), the catalytic site sequestering HTH motif is two amino acids shorter, and two polar residues that help orient the key C-terminal glutamate are substituted with small non-polar residues. Together, these divergences likely reflect a significantly altered substrate specificity. The SDR207C family therefore likely act on a range of substrates, united by the presence of a primary amino group immediately adjacent to the reactive hydroxyl group.

Over forty years ago, Turner and colleagues characterized enzymatic activities in protein enrichments from \textit{Pseudomonas sp.} cultures grown on aminoacetone as the sole carbon and nitrogen source\(^{97}\). This work lead to the discovery of four enzymatic activities that together could function as a pathway that converts aminoacetone to propionate: aminoacetone dehydrogenase, aminopropanol kinase, aminopropanol \( O \)-phosphate phospholyase, and propionaldehyde dehydrogenase. Complicating interpretation of their results is the fact no nucleotide or protein sequence has subsequently been identified associated with the kinase and phospholyase activities. While no genomic sequence exists for the strain investigated in their study, our analysis indicates that RMM microcompartments are absent from all known \( \gamma \)-proteobacterial genomes, while \( \gamma \)-proteobacteria, including \textit{Pseudomonas} strains, do routinely contain APDH
homologs outside of RMM operons (cluster I in Figure 3.5C) that associate with an aminotransferase and a phosphotransferase homolog. Phosphoethanolamine phospholyase was recently structurally characterized, and the key determinant of phospholyase activity was found to be the presence of a basic pocket that accommodates the phosphate group of the substrate\(^{56}\). The \textit{M. smegmatis} aminotransferase from the non-RMM gene cluster (MSM0782) is 42\% identical to this enzyme, with all of the catalytic residues and basic phosphate-binding residues conserved; this protein is then a candidate aminopropanol \(O\)-phosphate phospholyase. The phosphotransferase protein also found in this gene cluster lacks any well-characterized close homologs, but a role as a aminopropanol kinase is not unreasonable given that other family members typically phosphorylate aminoglycosides\(^{109}\). Taken together, it seems plausible that these non-RMM gene clusters with APDH, pfam00202 and pfam01636 genes encode orthologs of the aminoacetone reductase, aminopropanol kinase and aminopropanol \(O\)-phosphate phospholyase enzymes purified, but not identified, by Faulkner. These gene clusters (which include a permease and regulator in cluster I) would then represent the core of an aminoacetone catabolism pathway, with other elements (such as the propionaldehyde dehydrogenase) encoded elsewhere on the genome.

The RMM-associated APDH homologs are closely related to their non-RMM counterparts. However, the other proteins in the pathway, while in the same pfam family, have weaker sequence identity. Comparing non-RMM associated clusters with the RMM associated counterparts in \textit{M. smegmatis}, the permease proteins share about 64\% identity, the phosphotransferases share approximately 45\% identity, and the aminotransferases 27\% identity. The RMM-associated aminotransferases in particular do not conserve most of the basic residues found in the ethanolamine phospholyase catalytic pocket\(^{56}\), meaning if they are phospholyases,
they have evolved convergently and may exhibit substantive mechanistic differences. An aminopropanol kinase encapsulated in a microcompartment would also require access to ATP: microcompartments typically recycle cofactors such as NADP(H) and CoA\(^{32}\), but there are no candidate enzymes known to be associated with the RMM that might regenerate ATP. Microcompartments generally do not seem to transport stoichiometric amounts of nucleotide cofactors, though the two unusual BMC-FP proteins in the shell might possibly be capable of such transport, making the RMM an exception to this general rule\(^{66}\). An additional puzzle is, if the RMM microcompartment simply recapitulates a pathway routinely utilized in the absence of a microcompartment, what selective advantage does building the microcompartment offer? Organisms such as *M. smegmatis*, where both versions of the pathway seem to be present, make this question especially perplexing. Possibly the pathways are chemically identical, and encapsulation within an RMM microcompartment is only advantageous under specific circumstances – such as where enough intermediates accumulate for the toxicity costs or efficiency loss to outweigh the resource investment required to build the microcompartment. Alternatively, there may be differences between pathways. There are operons in some organisms (in a subset of *Mycobacterial* strains (e.g. NAZ190054) as well as *Cryptosporangium aurantiacum*; light blue in Figure 3.5B; also *Rhizobium* and *Mesorhizobium* strains that assort to cluster I) where the APDH, aminotransferase and phosphotransferase genes are closely related to those in the *M. smegmatis* RMM operon, but neither the aldehyde dehydrogenase nor the shell proteins are present. This suggests that encapsulation only becomes required when the RMM-associated aldehyde dehydrogenase is used to process the final aldehyde product. One possibility is that this aldehyde dehydrogenase produces a high-energy product (propionate phosphate or propionyl-CoA rather than propionate) at the cost of a greater accumulation of the volatile and/or
toxic propionaldehyde intermediate; microcompartment encapsulation is then required to cope with propionaldehyde accumulation. In summary, there is at the very least a plausible argument to be made on the basis of available biochemical and bioinformatics evidence that the RMM microcompartment catabolizes aminoacetone using the pathway originally identified by Turner. If so, this microcompartment requires either novel transport capabilities or an unusual pattern of enzyme encapsulation, as the order in the enzymes operate requires that either at least one co-factor is imported across the shell rather than recycled, or that only the phospholyase is encapsulated, with all other enzyme residing outside the shell.
4 Structural and kinetic characterization of (S)-1-amino-2-propanol kinase from the aminoacetone utilization microcompartment of *Mycobacterium smegmatis*

This research was originally published in the Journal of Biological Chemistry. Mallette, E. and Kimber, M. S. Structural and kinetic characterization of (S)-1-amino-2-propanol kinase from the aminoacetone utilization microcompartment of *Mycobacterium smegmatis*. *J. Biol. Chem.* 2018; 293:19909-19918

Statement of Contribution

The primary author, Evan Mallette, designed and conducted experiments, analyzed the data, and contributed to writing the manuscript and visualization. Matthew S. Kimber conducted bioinformatics analysis, acquired funding, and contributed to writing of the manuscript and visualization.

Acknowledgements

We thank P. Grochulski and S. Labiuk at Canadian Light Source for assistance in collecting diffraction data.
4.1 Abstract

Bacterial microcompartments encapsulate enzymatic pathways that generate small, volatile, aldehyde intermediates. The *Rhodococcus* and *Mycobacterium* microcompartment (RMM) operon from *Mycobacterium smegmatis* encodes four enzymes, including S-1-amino-2-propanol dehydrogenase and a likely propionaldehyde dehydrogenase. We show here that a third enzyme (and its non-microcompartment-associated paralog) is a moderately specific (S)-1-amino-2-propanol kinase. We determined the structure of apo-aminopropanol kinase at 1.35 Å, revealing that it has structural similarity to hexosamine kinases, choline kinases, and aminoglycoside phosphotransferases. We modeled substrate binding, and tested our model by characterizing key enzyme variants. Bioinformatics analysis established that this enzyme is widespread in Actinobacteria, Proteobacteria and Firmicutes, and is very commonly associated with a candidate phospho-lyase. In Rhizobia, aminopropanol kinase is generally associated with aromatic degradation pathways. In the RMM (and the parallel pathway that includes the second paralog), aminopropanol kinase likely degrades aminoacetone through a propanolamine-phosphate phospho-lyase dependent pathway. These enzymatic activities were originally described in *Pseudomonas*, but the proteins responsible have not been previously identified. Bacterial microcompartments typically co-encapsulate enzymes which can regenerate required co-factors, but the RMM enzymes require four biochemically distinct co-factors with no overlap. This suggests that either the RMM shell can uniquely transport multiple co-factors in stoichiometric quantities, or that all enzymes except the phospho-lyase reside outside of the shell. In summary, aminopropanol kinase is a novel enzyme found in diverse bacteria and multiple metabolic pathways; its presence in the RMM implies that this microcompartment degrades aminoacetone, using a pathway that appears to violate some established precepts as to how microcompartments function.
4.2 Introduction

Bacterial microcompartments are specialized cellular inclusions built by diverse bacteria to mitigate the challenges presented by the toxic or volatile intermediates produced by certain metabolic pathways. Most microcompartments are catabolic, and contain a set of enzymes that produce, and then consume, aldehyde intermediates. Well-studied examples include the propanediol and ethanolamine utilization microcompartments, but more recently ethanol, choline, rhamnose/fucose degrading microcompartments have also been proposed, and many more remain uncharacterized. The exceptions to this general rule are the environmentally important α- and β-carboxysomes, which encapsulate rubisco and carbonic anhydrase and mediate anabolic carbon fixation in cyanobacteria and chemoautotrophs. Microcompartments are defined by the presence of a thin icosahedral shell built from proteins from two distinct families (pfam00936 & pfam03319); this shell encapsulates key enzymes and is likely selectively permeable, allowing initial substrates and products to pass freely, while retaining enzymes, co-factors and metabolic intermediates.

Recent surveys of microbial genomes have uncovered a deep diversity in microcompartment families, many which are poorly- or wholly un-characterized. One such operon has been identified predominantly in species of the Actinobacteria phylum, and has been provisionally designated the RMM (Rhodococcus and Mycobacterium Microcompartment). While formally uncharacterized, Urano and colleagues previously showed that this operon is induced 100-fold by the addition of aminopropanol in the culture media. The RMM encoding operon contains a regulatory protein, an amino acid permease, four structural proteins that are close homologs to those that form the only structurally characterized shell, along with enzyme sequences with clear homology to alcohol dehydrogenases, aminoglycoside kinases, aminotransferases, and aldehyde dehydrogenases respectively. Building on earlier work, we
recently published a detailed structural and functional characterization of the alcohol dehydrogenase from *Mycobacterium smegmatis* MC² 155, identifying it as an NADP-dependent S-1-amino-2-propanol (aminopropanol) dehydrogenase (APDH). However, the other enzymes in this pathway remain uncharacterized; specifically, understanding of the function of this microcompartment is hindered by the fact that no functions have been ascribed in the literature to any close homologs (i.e. with sequence identity > 25%) of either the aminotransferase- or kinase-like enzymes.

The alcohol kinase homolog from the *Mycobacterium smegmatis* MC² 155 microcompartment operon (msmeg_0270) is annotated as a member of the aminoglycoside phosphotransferase family (APH, pfam01636), a group which is, in turn, part of the protein kinase-like domain superfamily (PK-like, SSF56112). Members of the PK-like family of enzymes have low sequence conservation but exhibit conserved tertiary structure including an N-terminal α/β-fold nucleotide triphosphate binding domain and a mostly α-helical C-terminal substrate-binding domain. Enzymes of this family phosphorylate diverse substrates, ranging from proteins to small alcohols, using ATP (or, in some cases, GTP) as the donor. Kinases are not usually found in microcompartment pathways, likely because microcompartments generally lack the capacity to import stoichiometric quantities of cofactors, instead recycling a fixed cofactor pool through complementary enzymes. Among microcompartments, the only identified phosphotransfer reaction is a phosphotransacylation reaction used to recycle CoA in the propanediol microcompartment, a reaction that occurs outside of the shell. Many organisms (mostly α-Proteobacteria and Actinobacteria) also encode a homologous operon that has no shell proteins; interestingly, many Mycobacterial strains encode both (RMM and non-RMM) variants of this operon. Early work by Turner and colleagues have shown that an enzyme with S-1-
Amino-2-propanol kinase (APK) activity can be purified from many of the same organisms that express APDH\(^9\). Given the strict specificity of APDH, we therefore tested both the RMM-associated (msmeg_0270, henceforth APK\(_{\text{MSM0270}}\)) and the non-RMM associated (msmeg_0780, henceforth APK\(_{\text{MSM0780}}\)) APH homologs from *Mycobacterium smegmatis* MC\(^2\) 155, showing that both can act as aminopropanol kinase. We also report the structure, kinetics and a model for substrate recognition for APK\(_{\text{MSM0270}}\).

4.3 Results

4.3.1 Enzyme kinetics.

Kinetic parameters for APK were determined using a coupled assay. Briefly, ADP released by APK functioned as a limiting reagent for phosphoenolpyruvate conversion to pyruvate by pyruvate kinase; released pyruvate was reduced by lactate dehydrogenase at the expense of oxidizing NADH, and the resulting decrease in absorbance of NADH was monitored spectrophotometrically at 340 nm (Table 4.1). APK\(_{\text{MSM0270}}\) exhibits Michaelis-Menten kinetics against all substrates tested (Figure 4.1). The best substrate proved to be S-aminopropanol, with a \(k_{cat}\) of 8.7 s\(^{-1}\) and a \(K_M\) of 0.373 mM. APK\(_{\text{MSM0270}}\) is enantioselective, but not absolutely so, with R-aminopropanol being phosphorylated with a ten-fold lower efficiency. (R/S)-1-aminoo-2-butanol and ethanolamine were also both substrates, with \(k_{cat}\) values similar to S-aminopropanol, though with a \(K_M\) roughly 3-fold and 600-fold higher, respectively. This suggests that APK may readily tolerate (slightly) larger substrates, but strongly discriminates against ethanolamine, which will commonly be present in cells for lipid biosynthesis, and whose diversion into degradation might be counterproductive.

*Mycobacterium smegmatis* also encodes a second APK homolog (APK\(_{\text{MSM0780}},\) YP_885185) from a non-microcompartment gene cluster, with sequence identity of 44% with APK\(_{\text{MSM0270}}\). The full sequence (as annotated by UniProt) proved insoluble, so the protein was
re-cloned without the first 23 annotated amino acids, as alignment with APK\textsubscript{MSM0270} suggested that the start methionine may be misannotated. APK\textsubscript{MSM0780} was able to catalyze the phosphorylation of S-(+)-1-amino-2-propanol and is approximately 3-times more efficient than its microcompartment-associated paralog due to a significantly lower $K_M$ ($k_{cat}$ of 4.35 s\(^{-1}\) and $K_M$ of 0.069 mM). This difference is intriguing, and possibly hints that the non-RMM pathway may preferentially operate when less substrate is available.
Figure 4.1 Enzyme kinetics.

APK_{MSM0270} initial reaction rates were measured by change in NADH concentration due to ADP production coupled with pyruvate kinase/lactate dehydrogenase, and fit to the Michaelis-Menten equation by non-linear regression. A) Kinetics data for S-(+)-1-amino-2-propanol and B) kinetics data for ATP.
Table 4.1 Steady-state kinetic parameters for *Mycobacterium smegmatis* MC2 155 APK variants

<table>
<thead>
<tr>
<th>Substrate</th>
<th>k\textit{cat} (s\textsuperscript{-1})</th>
<th>K\textit{M} (mM)</th>
<th>k\textit{cat}/K\textit{M} (M\textsuperscript{-1} s\textsuperscript{-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(S)-1-amino-2-propanol</td>
<td>8.7 ± 0.05</td>
<td>0.373 ± 0.006</td>
<td>2.3 x 10\textsuperscript{4}</td>
</tr>
<tr>
<td>(R)-1-amino-2-propanol</td>
<td>3.2 ± 0.2</td>
<td>1.54 ± 0.15</td>
<td>2.1 x 10\textsuperscript{3}</td>
</tr>
<tr>
<td>(R/S)-1-amino-2-butanol</td>
<td>8.8 ± 0.2</td>
<td>1.11 ± 0.066</td>
<td>7.9 x 10\textsuperscript{3}</td>
</tr>
<tr>
<td>Ethanolamine\textsuperscript{b}</td>
<td>8.1 ± 0.5</td>
<td>193 ± 13</td>
<td>4.2 x 10\textsuperscript{1}</td>
</tr>
<tr>
<td>ATP</td>
<td>8.0 ± 0.2</td>
<td>0.27 ± 0.02</td>
<td>3.0 x 10\textsuperscript{4}</td>
</tr>
</tbody>
</table>

**APK\textsubscript{MSM0270} WT**

<table>
<thead>
<tr>
<th>Variant</th>
<th>k\textit{cat} (s\textsuperscript{-1})</th>
<th>K\textit{M} (mM)</th>
<th>k\textit{cat}/K\textit{M} (M\textsuperscript{-1} s\textsuperscript{-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>E33A\textsuperscript{b}</td>
<td>0.068 ± 0.003</td>
<td>26 ± 2</td>
<td>2.6</td>
</tr>
<tr>
<td>R290A\textsuperscript{b}</td>
<td>0.44 ± 0.07</td>
<td>72 ± 15</td>
<td>6.1</td>
</tr>
<tr>
<td>D235A</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>R213A</td>
<td>0.85 ± 0.01</td>
<td>2 ± 0.1</td>
<td>4.3 x 10\textsuperscript{2}</td>
</tr>
</tbody>
</table>

**APK\textsubscript{MSM0270} variants\textsuperscript{c}**

<table>
<thead>
<tr>
<th>Substrate</th>
<th>k\textit{cat} (s\textsuperscript{-1})</th>
<th>K\textit{M} (mM)</th>
<th>k\textit{cat}/K\textit{M} (M\textsuperscript{-1} s\textsuperscript{-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-(+)-1-amino-2-propanol</td>
<td>4.35 ± 0.03</td>
<td>0.069 ± 0.003</td>
<td>6.3 x 10\textsuperscript{4}</td>
</tr>
<tr>
<td>ATP</td>
<td>2.79 ± 0.02</td>
<td>0.074 ± 0.004</td>
<td>3.8 x 10\textsuperscript{4}</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Reactions were assayed in 20 mM Tris pH 8 and 150 mM NaCl at 25°C. Errors for kinetic parameters are the standard deviation determined from nonlinear regression fitting of kinetic data as described in materials and methods. Nonlinear regression fittings had R\textsuperscript{2} values greater than 0.99. \textsuperscript{b} Kinetic parameters are calculated by extrapolation from collected data due to inability to reach substrate saturation under assay conditions. \textsuperscript{c} Variant kinetic parameters are for S-(+)-1-amino-2-propanol with saturating concentrations of ATP.
4.3.2 Structure of APK_{MSM0270}

We determined the structure of APK_{MSM0270} using selenomethionine SAD in spacegroup C2, and then refined the structure against an isomorphous native dataset at 1.35 Å. A single protomer was found in the asymmetric unit, with most of the structure, excluding a short loop (97 – 101), being ordered. As predicted by the distant sequence homology, the fold of APK broadly resembles other atypical kinases of the aminoglycoside phosphotransferase family (APH; pfam01636), with an N-terminal α/β-fold nucleotide binding domain and a C-terminal, mostly α-helical substrate binding domain (Figure 4.2A). The N-terminal domain is built as a five-stranded anti-parallel β-sheet (topology 1,-2,3,-5,4) flanked by two α-helices. The C-terminal domain is built from seven helices (α3-α9), with four short β-strands (β6-β9) in two two-stranded antiparallel sheets, contributed by the long α5-α6 loop that packs into the cleft between the domains.

Searching the entire PDB for structural homologs using DALI retrieves 329 chains with Z-scores > 10. The closest functionally characterized homologs are bacterial N-acetylhexosamine kinases (PDB # 4OCQ), with Z = 22.2, an r.m.s.d. of 3.4 Å, and 14 % sequence identity (Figure 4.2B). High r.m.s.d. values are typical for this enzyme family, and reflect inter-domain flexibility as well as considerable variability in the secondary structure complement of the C-terminal domain. Bacterial choline kinase (PDB # 4R78; Z = 20.1) and spectinomycin phosphotransferase (PDB # 3I0O, Z = 18.9) are similarly closely related. No retrieved homologs had r.m.s.d. values < 3.0 Å, or sequence identities > 19 %. APK is therefore a distant homolog of hexosamine kinases, choline kinases and aminoglycoside phosphotransferase families, without being especially close to any individual enzyme group.
**Figure 4.2 Structure of APK\textsubscript{MSM0270}**

A) structure of the APK\textsubscript{MSM0270} protomer. Secondary structure elements are labelled. B) superposition of APK\textsubscript{MSM0270} with representative structural homologs. APK\textsubscript{MSM0270} is shown in dark blue. N-acetylhexosamine -1-phosphate kinase (PDB # 4OCQ) is shown in pale yellow. Choline kinase (PDB # 4R78) in pale orange. Spectinomycin phosphotransferase (PDB # 3I0O) is shown in pale cyan. C) details of the catalytic site of APK\textsubscript{MSM0270}. D) Model of the ternary complex of APK\textsubscript{MSM0270} with ATP and propanolamine. E) Multiple sequence alignment of APK with structural homologs. Secondary structure is coloured as in panel A. APK\textsubscript{MSM0270} is aligned with APK\textsubscript{MSM0780} (greyed out residues are present in UniProt but are likely a misannotation), *S. pneumoniae* bacterial choline kinase (BCHK, UniProt # Q8DPI4), *H. sapiens* choline kinase B (HCHKB, UniProt # Q9Y259), *B. longum* N-acetylhexosamine 1-kinase (NAHK, UniProt # E8MF12), and *E. casseliflavus* aminoglycoside phosphotransferase (2’)-Iva. The secondary structure and residue numbering of APK\textsubscript{MSM0270} are aligned with the multiple sequence alignment. Colored circles below the multiple sequence alignment indicate residues of the hydrophobic methyl binding pocket (grey), substrate α-amino binding (pink), conserved Brenner motif (green), and the phosphate-product stabilizing residue (light-blue).
4.3.3 Active site modeling

APK’s active site architecture resembles that of other APH family members, with three key conserved motifs which together form the ATP binding site: the hinge loop (which connects the N- and C-terminal domains F109-F127, yellow in Figure 4.2)\textsuperscript{46}, the Brenner motif (H209xD211xxxxN216; dark red in Figure 4.2), and the P-loop (Leu29-Asn34, cyan in Figure 4.2). ATP binding in APH enzymes is generally accompanied by major structural rearrangements including a 16° rotation of the N-terminal domain toward the C-terminal domain, and repositioning of the P-loop (as seen in comparing PDB # 4OCK to PDB # 4OCP)\textsuperscript{113}. Comparisons to reported APH ATP-analog co-structures (including \textit{B. longum} N-acetylhexosamine kinase (PDB # 4WH3); \textit{S. pneumonia} choline kinase (PDB # 4R78); and \textit{E. faecalis} 3',5''-aminoglycoside phosphotransferase type IIIa (PDB # 1J7U)) indicates that in APK, the adenine ring likely binds with its non-polar face stacked between Ile47, Leu218, Ile231, and with hydrogen bonds with the backbone groups contributed by the hinge loop. Conserved residues of the Brenner motif Asp211 and Asn216 along with Asp232 superpose closely with those used to co-ordinate the two magnesium ions that bridge the α, β- and γ-phosphates. Additional hydrogen bonds to ATP are generally contributed by backbone amide groups in the conserved phosphate binding loop or P-loop. Arg213 in APK is less usual, and appears to be analogous to Lys210 in N-acetylhexosamine kinase lysine which forms additional contacts with the γ-phosphate.

While APK’s substrate is very different from that of all well-studied APH family members, modeling of aminopropanol is simplified by the fact that this is a small, chemically simple molecule with few degrees of freedom. The position of the attacking hydroxyl group is also largely fixed by the position of the γ-phosphate, and the key catalytic residue Asp211 (from the Brenner motif) which activates this hydroxyl as a nucleophile (Figure 4.2C). Arg290 is an
additional charged residue conserved in the APK catalytic pocket. This residue is homologous to Arg284 in the active site of spectinomycin phosphotransferase (PDB # 3I0O), a residue which forms a hydrogen bond with the catalytic aspartate (Asp212) in the substrate complex, but whose role has not been specifically investigated by mutagenesis. In analogy with spectinomycin phosphotransferase, we suggest that this residue helps position the catalytic Asp211 in the reaction. Asp235 in APK is situated in an identical position to Asp220 in aminoglycoside-2''-phosphotransferase-IVa (PDB # 3SG8), suggesting that the propanolamine amine forms a salt bridge to the carboxylate of this residue (as does Asp220 in 3SG8 to tobramycin’s 3’N, a group analogously adjacent to the phosphorylation site 2’OH); this residue may also help co-ordinate magnesium as seen in an unpublished aminoglycoside phosphotransferase structure (PDB # 4N57). An additional interaction with this amine group is possible with a second highly conserved acidic residue, Glu33, which is positioned close by in the apo structure (6 Å between the carboxylate groups), and is in a region of the structure that is typically rearranged upon substrate binding.

To test the roles of these possible catalytic residues, we generated and investigated the site directed enzyme variations E33A, R213A, D235A and R290A. All of these variants expressed well, were soluble, and showed co-operative unfolding in a fluorimetric thermal denaturation assay, confirming that they fold. E33A and D235A had a T_m very similar to the wild type enzyme (53 °C), while the R213A and R290A had a T_m 7 and 10°C lower, possibly indicating that these residues help stabilize the enzyme by neutralizing excess negative charge in the active site. All variants except D235A showed additional stabilization by 6 – 10 °C upon binding of Mg^{2+} and ATP; this suggests that Asp235 may be positioned to bridge the substrate amine and a magnesium ion. As suggested by our model, all of these variants resulted in
significant losses in enzyme activity: the R213A variant has two orders of magnitude lower activity, the E33A and R290A variants have activity reduced by four orders of magnitude, and the D235A does not detectably turn over (Table 4.1).

A set of non-polar residues (Trp167, Phe251, Leu294, and Trp297) contributed by the helical domain form a hydrophobic pocket is found immediately adjacent to this cluster of charged residues. Leu294 and Trp297 are found in successive turns of helix α8, while Phe251 is contributed by an extension of the α7 – α8 loop. Trp167 is notable as this position is normally not positioned in the active site: in other APH enzymes, this residue forms part of the long helix α4, while in APK this helix is broken into two short segments, with the center bulging out as a loop that positions Trp167 adjacent to Trp297 (orange in Figure 4.2). Leu294 and Trp297 are the structural analogs of Trp251 and Trp254, which in the bacterial choline kinase LicA (PDB # 4r7b), interact with the methyl modifications of the betaine group of choline. We suggest that, in APK, this hydrophobic pocket is optimized to envelop methyl group of the substrate. The importance of interactions with the methyl group is underscored by the very poor affinity (extrapolated $K_m$ of 190 mM) exhibited by ethanolamine.

4.3.4 Sequence and gene association analysis of APK homologs. We used BLAST to search the UNIREF90 database (i.e. a subset of the full sequence databases with no pairs of sequences having > 90 % sequence identity) on the UniProt website with the APKMSM0270 sequence; sequences were accepted as being likely APK orthologs on the basis of conservation of the amino acids Glu33, Trp167, Arg213, Phe251, Arg290, Leu294 and Trp297 (with three of these being experimentally verified as being functionally important residues). We found that a large group of proteins (>600) conserve all of these amino acids, in some cases with as little 34 % overall sequence identity (e.g. UniProt # A0A1T5CUT6 from Sphingomonas; E-value 2e-38, 34% sequence identity). Relaxing these criteria slightly, there is
also a larger family of proteins that conserve most, but not all of these motifs (e.g. UniProt # A0A1H7ZUK6 from *Gemmobacter aquatilis* has an E value of 1e-20, and 31% sequence identity and conserves all of the identified residues except Phe251 (Gln) and Leu294 (Val). These proteins, while possibly not authentic APK orthologs, are likely to be significantly more similar to APK than any other functionally characterized enzyme. APK-like proteins were most abundant in Proteobacteria, Firmicutes and Actinobacteria, with significant numbers also in Chloroflexi, Cyanobacteria, and Bacteroidetes, and a few scattered representatives also found in another dozen bacterial phyla, and two archaeal phyla. APK is thus a widespread and common enzyme, with high prevalence in three important bacterial phyla.

We clustered these sequences (selected using the more permissive 1e-20 cutoff) in EFI-EST using e-50 as an edge cutoff and displaying the results in cytoscape (Figure 4.3A). This classification places most of the Actinobacterial and Proteobacterial sequences in a single large cluster, along with a subset of the Firmicute sequences and other sequences. This cluster is bilobed, and separates into two large groups (with some smaller clusters) at stringencies < e-60. We examined the association of genes within these clusters using EFI-GNT. In cluster I, a very tight association is found between APK and an aminotransferase III gene. This gene is found within 10 genes in 72 % of 1271 organisms; inspection of individual organisms shows that that both proteins are in a single operon, and are generally separated by no more than one ORF. Together this implies a tight functional association. Other proteins strongly co-occurring in this cluster include a short chain alcohol dehydrogenase (36%), an aldehyde dehydrogenase (27%), an amino-acid permease (46%), and various proteins associated with the degradation of catechol-like molecules, such as isochorismatase (11%). Looking at how genes cluster in individual genomes reveals two distinct patterns underpinning the association between the APK and the
aminotransferase. In the left lobe of cluster I, with Actinobacterial strains dominant, APK, the
aminotransferase and the permease associate with an APDH homolog to form an operon, often
with additional proteins. A subset (about 10% of cluster I) of these operons also encode BMC
shell proteins and an aldehyde dehydrogenase to form an RMM (Figure 4.3B). The
microcompartment operon is found primarily in Actinobacteria (though we also noted examples
here in Alphaproteobacteria and Bacilli e.g. *Bacillus azotoformans* UniProt # MEV2011). As
presented below, we strongly suspect that both of these operons encode a pathway capable of
incorporating aminoacetone into central metabolism as propionyl-CoA.
Figure 4.3 Bioinformatics analysis of APK homologs.

A) Sequence similarity network of APK homologs. Sequences are connected by edges if $E < e^{-50}$. Sequences are coloured by phyla, as noted in the key. Phyla marked (c) are candidate phyla. APK homologs highlighted in B are indicated. B) Gene organization of representative examples of APK homologs. Gene labels: BMC – bacterial microcompartment shell proteins; AT3-aminotransferase III; trans. - transporter; 4HPH: 4-hydroxyphenylacetate 3-hydroxylase; FD: flavin reductase; HHMI: 5-carboxymethyl-2-hydroxymuconatedelta-isomerase; CHMD: 5-carboxy-2-hydroxymuconate semi aldehyde dehydrogenase; DHPD: 3,4-dihydroxyphenylacetate 2,3-dioxygenase; FAAH: fumarylacetoacetate hydrolase; 2KPH: 2-keto-4-pentenoate hydratase; OHDH: 2-oxo-hept-3-ene-1,7-dioate hydratase; CTTD: Catechol 2,3 dioxygenase
For APK orthologs in the right lobe of Cluster I (mostly Rhizobia and Actinobacteria), APK is associated with the aminotransferase III in an operon that is adjacent to (though often divergently transcribed from) an operon encoding homologs of aromatic degradation pathway enzymes\textsuperscript{116}. These operons generally contain a 3,4-dihydroxyphenylacetate 2,3-dioxygenase and a 5-carboxy-2-hydroxymuconate semi aldehyde dehydrogenase homolog; these two enzymes open the aromatic ring by oxygenation and then oxidize the resulting semi-aldehyde. Additional enzymes associated with preparatory steps (e.g. hydroxylating the ring) or further process the resulting partially saturated hydrocarbon products into common metabolic intermediates (including isomerases, dehydratases, reductases, aldolases and hydrolases) are also found associated. We propose that these pathways degrade a variety of phenyl derivatives with differing ring substituents; the aminotransferase and APK may be required to incorporate a 2-amino or 2-keto (these pathways commonly have an additional aminotransferase associated) product back into central metabolism.

Outside of the two main lobes of the largest cluster, gene associations are weaker, possibly implying a role in a variety of minor pathways. Acetyltransferases, amidohydrolases, glyoxalases, and HAD-like hydrolases are among the enzymes more commonly associated with APK homologs, but no individual enzyme seems associated in more than 10 – 20 % of organisms. Hindering analysis is the fact that many of the associated proteins are poorly characterized. However, there are hints that APK plays a role in synthetic pathways. For example, APK is associated with an exopolysaccharide operon in \textit{Streptomyces} sp. FBKL.4005, implying that aminopropanol may be incorporated into polysaccharides and other complex macromolecules.
4.4 Discussion

S-1-Amino-2-propanol O-kinase, as characterized here, represents a new group of atypical kinases within the APH group, related to hexosamine kinases, choline kinase, and aminoglycoside phosphotransferases. The conservation of key signature motifs within the substrate binding site in a very diverse set of bacterial genomes argues that this enzyme family is widespread and important. The fairly weak substrate specificity APK\textsubscript{MSM0270} displays, coupled with the occurrence of amino acid variants in the active site argue that family members likely are involved in phosphorylating a variety of substrates, though likely mostly 1-amino-2-ols. In general, these enzymes appear to mostly occur in catabolic pathways, where after various preparatory steps, an aminoalcohol is converted to an aldehyde by the sequential actions of a kinase and a phospho-lyase (see below). Another very common occurrence of this enzyme pair in Rhizobobacteria, where they co-occur with protocatechuate 2,3-like oxidizing catabolic pathways, is certainly worthy of further investigation. Some rarer homologs of APK are also associated with what appear to be biosynthetic pathways, for example ones for lipid or polysaccharide biosynthesis. Of note is that aminopropanol is commonly found as a modification on cobalmin (though cobalmin is generally synthesized without APK involvement), and there is no reason to suspect that bacteria have not found other uses for 1-amino 2-propanol-O-phosphate.

Our original interest in this enzyme arose from its association with the RMM microcompartment; this microcompartment is induced in \textit{Rhodococcus jostii} by the addition of aminopropanol in the media\textsuperscript{43}, suggesting a possible role in aminoacetone (which commonly arises from an uncatalyzed reaction in threonine/glycine pathway) catabolism. Our previous work established that the short chain dehydrogenase is a highly specific (S)-1-amino-2-propanol dehydrogenase\textsuperscript{94}, while among the other three enzymes present in the operon, the aldehyde
dehydrogenase has clear homology to CoA-dependent propionaldehyde dehydrogenase. The outstanding question remained how an APH-like enzyme and a class III aminotransferase could convert aminopropanol (or aminoacetone) to propionaldehyde. The demonstration that MSM0270 is an aminopropanol kinase allows the problem to be narrowed down to one of converting 1-amino-2-propanol phosphate to propionaldehyde, a reaction catalyzed by phospho-lyases. The only well characterized example is ethanolamine-phosphate phospho-lyase (EC # 4.2.3.2.)\textsuperscript{56,55}, a homolog of class III aminotransferases. Strongly supporting this idea, propanolamine-phosphate phospho-lyase activity was found associated with APDH, APK and propionaldehyde dehydrogenase activities in enriched \textit{Pseudomonas} or \textit{Erwinia} extracts by Turner and colleagues in the 70's\textsuperscript{97,112}. We therefore propose that the RMM microcompartment utilizes aminoacetone using the pathway shown in Figure 4.4, and that this microcompartment should therefore be renamed the aminoacetone utilization microcompartment, or \textit{aaum}. Of note, a second APK paralog (MSM0780) in \textit{M. smegmatis} is found in an operon which parallels the enzymes of the \textit{aaum} operon, but lacks the shell proteins; this version of the aminoacetone degradation pathway is phylogenetically more widespread than the \textit{aaum}. This is, to our knowledge, the first instance where a catabolic pathway occurs in both a microcompartment-dependent and microcompartment independent variant; in addition, the common co-occurrence of these two pathways in a single organism possibly suggests that these two pathway variants are likely optimized for distinct environmental conditions.
Figure 4.4 Proposed organization of, and pathway mediated by, the aminoacetone utilization microcompartment.

Aminoacetone is imported into the cell via a aminopropanol permease (AAP). It is then reduced to S-aminopropanol by aminopropanol dehydrogenase (APDH), which is in turn phosphorylated by aminopropanol kinase (APK). Both of these reactions occur in the cytosol. Aminopropanol-phosphate then enters the microcompartment through the shell, and is transformed into propionaldehyde by (S)-1-amino-2-propanol-phosphate phospho-lyase (APPL). Propionaldehyde escapes through the shell, and is captured by propionaldehyde dehydrogenase (AldDH) which adheres to the exterior face of the microcompartment via its encapsulation peptide.
Bacterial microcompartments are thought to be organized around a signature enzyme, which generates an aldehyde, and an aldehyde dehydrogenase that consumes it. These are proposed to be always encapsulated together within a shell whose pores are likely evolved to minimize aldehyde escape\textsuperscript{3}. These pores generally do not passage large co-factors in stoichiometric quantities, so additional enzymes that regenerate co-factors – termed the enzyme core – need to be co-encapsulated\textsuperscript{3,32}. In short, every encapsulated enzyme that requires a co-factor must be paired with a second enzyme which regenerates that co-factor. In the \textit{aaum}, there are three enzymes that together require four different co-enzymes or co-substrates: APDH requires NADPH\textsuperscript{94}, APK requires ATP, and the aldehyde dehydrogenase has all of the sequence hallmarks of a CoA and NAD+ dependent aldehyde dehydrogenase (including the PHP motif which prevents NADP+ binding, and His387 which acts as a general base in the acylation reaction\textsuperscript{58}). Note that in addition to the aldehyde dehydrogenase having co-factor requirements incompatible with APDH, \textit{aaum} operons do not contain a phosphotransacylase gene, so the enzyme which in other microcompartments regenerates free CoA by transferring propionate to a phosphate group is missing\textsuperscript{33}. This strongly suggests that the \textit{aaum} must either passage stoichiometric quantities of at least two co-factors (NAD+ and CoA) or localize all enzymes except the phospho-lyase (which regenerates its own co-factor) outside the shell. We propose that only the phospho-lyase is encapsulated, while the propionaldehyde dehydrogenase is cytosolic, but anchored on the exterior wall of the shell. The C-terminus of this aldehyde dehydrogenase contains a conserved encapsulation peptide\textsuperscript{29,82} of the type known to bind the concave surface of BMC-H proteins (such as MSM0272)\textsuperscript{64}; however, the concave surface of a very closely homologous BMC-H was recently shown to localize to the cytoplasmic face of the shell, which presumably then targets the attached proteins to this face\textsuperscript{10}. This suggests an
alternative model for microcompartment function, where only the “signature enzyme” that generates the aldehyde is encapsulated, while problematic cofactor recycling requirements are avoided by localizing the aldehyde dehydrogenase to the immediate exterior surface of the shell, where it can maximize capture of escaping aldehyde. This arrangement, and the existence of a parallel aminoacetone degradation pathway that does not depend on building a microcompartment, suggests a possible model for a “primitive”, less specialized microcompartment where encapsulation is non-essential but acts to more efficiently channel aldehydes to the aldehyde dehydrogenase by co-localizing these two key enzymes within a small volume in the cytosol.

4.5 Experimental procedures

4.5.1 Molecular biology

*Mycobacterium smegmatis* mc² 155 was grown in 5 mL LB (50 μg/mL carbenicillin, 10 μg/mL cycloheximide) over 5 days at 37°C with shaking; genomic DNA was isolated by miniprep (Invitrogen PureLink). Genes encoding *msmeg_0270* (UniProt # A0A0D6G1D9) and *msmeg_0780* (UniProt # A0QQJ7) were amplified by PCR using PfuX7 polymerase (a generous gift from Dr. Dinesh Christendat, University of Toronto) for cloning into plasmids pET28a to include an N-terminal hexahistidine affinity tag (Table 4.2). Plasmids and amplified inserts were prepared by restriction digest, and ligated together after DNA purification. Ligated constructs were then transformed into DH5α cells for propagation. Plasmids suspected to contain the gene of interest were sequenced at the Advanced Analysis Center (AAC) University of Guelph. Site directed mutagenesis was performed using PCR with overlapping primers to generate the desired nucleotide substitutions and verified by sequencing (Table 4.2). BL21(DE3) or B834(DE3) (methionine-auxotrophic genotype for selenomethionine labeling) cells were transformed with plasmid containing the gene of interest for overexpression.
<table>
<thead>
<tr>
<th>Primer name</th>
<th>Cloning primers (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>APK270pet28aFWD</td>
<td>ATACAGCTCGAGTCAGGTCAGGCGAAGGTCCGTCCGAGGG</td>
</tr>
<tr>
<td>APK270pet28aREV</td>
<td>TTAGATCATATGGGCGTGATCAACGGACGATGTCACCATC</td>
</tr>
<tr>
<td>APK780Δ1-23pet28aFWD</td>
<td>CGAAGCTACGATGCGGGGACTTCAGCTGACCACCAGAAGG</td>
</tr>
<tr>
<td>APK780Δ1-23pet28aREV</td>
<td>GCTAAAGCTTCTATGCTGTCGGGGGTTGAGGGGTGACGACAAA</td>
</tr>
<tr>
<td>E33AFWD</td>
<td>AACCTGTCCGCGAAGCGGACGCACC</td>
</tr>
<tr>
<td>E33AREV</td>
<td>GGTCGCGTTCGCGGACAGGTT</td>
</tr>
<tr>
<td>D235AFWD</td>
<td>GATTTGCAGTCCTGCGGATT</td>
</tr>
<tr>
<td>D235AREV</td>
<td>GAATCCGCAGCGCTCGAAATC</td>
</tr>
<tr>
<td>R213AFWD</td>
<td>GCCGACCTGGCGGGACCAA</td>
</tr>
<tr>
<td>R213AREV</td>
<td>GTTGGCCAACGCAGGTCCG</td>
</tr>
<tr>
<td>R290AFWD</td>
<td>GTGGTCTGAGGCTCTGG</td>
</tr>
<tr>
<td>R290AREV</td>
<td>CAAGAGCTGGCCAGGACAC</td>
</tr>
</tbody>
</table>
4.5.2 Protein expression and purification

BL21 cells were grown overnight in 5 mL LB at 37°C and used to inoculate 1 L of 2YT media. Once an optical density of 0.8 at 600 nm was reached, cultures were induced for expression with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubated for 18-20 hours at 16°C. Cells were pelleted by centrifugation at 4,400 x g and resuspended in 35 mL of lysis buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl) prior to freezing at -20 °C for storage. Selenomethionine labeled protein was expressed in B834 cells by an analogous protocol, except cells were pelleted once the culture reached an optical density of 0.6 at 600 nm and resuspended in selenomethionine media (prepared according to guidelines from the EMBL Heidelberg Protein Expression and Purification Core Facility https://www.embl.de/pepcore/pepcore_services/protein_expression/ecoli/seleno/) prior to induction with IPTG.

Harvested cells were incubated on ice and disrupted by sonication with a Sonicator XL2020 at a power level of 7 with 10 sec on / 20 sec off pulses for 10 minutes of processing time. The insoluble fraction of the lysate was separated by centrifugation at 48,000 x g for 30 minutes at 4°C. Lysate supernatant was loaded onto a 2 mL Ni-IMAC column, and unbound proteins were eluted with 20 mL lysis buffer. Partially purified protein was eluted with 10 mL lysis buffer with added 500 mM imidazole. The buffer was exchanged by desalting the protein on a HiPrep 26/10 column equilibrated with 10 mM Tris pH 8. Pure protein was isolated by anion exchange on a HiTrap Mono Q column with a gradient of 0 to 1 M NaCl (with 10 mM Tris-HCl pH 8.0). Aliquots of pure target protein, as indicated by SDS-PAGE, were pooled and exchanged into 20 mM Tris-HCl pH 8.0, 150 mM NaCl by desalting before being concentrated by centrifugal concentration and quantified spectrophotometrically by absorbance at 280 nm.
4.5.3 Protein crystallization

N-terminally hexahistidine tagged protein was crystallized using serial seeding in a sitting drop configuration. One microliter of protein solution (30 mg/mL protein in 20 mM Tris-HCl pH 8.0, 150 mM NaCl) was mixed with 1 μL of precipitating solution (1.1 M sodium malonate pH 7.0, 0.1 M HEPES pH 7.0, 1 % (v/v) Jeffamine E.D. 2001 pH 7.0), and was then equilibrated against 50 μL precipitating solution for 24 hours. At this time, 0.2 μL of 1/1000 diluted crushed crystals were added in precipitating solution. Crystals grew as large, single monoclinic prisms and were cryoprotected using a 2 M sodium malonate soak followed by immersion in paratone N oil prior to freezing in liquid nitrogen.

4.5.4 Data collection and structure determination

Datasets were collected at the Canadian Light Source beam-line 08ID at 100 K, and were processed using XDS and scaled using XSCALE\textsuperscript{88}. Data collected at the anomalous peak from a selenomethionine derivatized crystal diffracted to 1.6 Å and was phased by single wavelength anomalous diffraction using AutoSol in Phenix\textsuperscript{90}. All four encoded methionine residues (after post-translational methionine trimming) were found as anomalous peaks, producing a phase solution with a 0.276 figure of merit. Subsequent automated model building by Phenix Autobuild was able to trace most of the structure. After manual rebuilding in Coot\textsuperscript{91} and automated refinement by Phenix Refine, this model was used to phase an isomorphous 1.35 Å native dataset by molecular replacement using Phaser in Phenix\textsuperscript{89}. Manual rebuilding in Coot and automated refinement by Phenix Refine. In the final structure, residues 7-334 are traced in the map, except residues 99 – 101 which are too disordered to reliably trace. The structure contains additional density in the nucleotide binding site that appears to be benzoate (source unknown) and catalytic site density that was tentatively identified as a terminal group of Jeffamine ED 2001. Data
collection and structure refinement statistics are presented in Table 4.3. Structure figures were prepared in Pymol v2.0.
Table 4.3 X-ray data collection and refinement statistics

<table>
<thead>
<tr>
<th></th>
<th>APK&lt;sub&gt;270&lt;/sub&gt; selmet</th>
<th>APK&lt;sub&gt;270&lt;/sub&gt; native</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCSB I.D.</td>
<td>6EF6</td>
<td></td>
</tr>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wavelength</td>
<td>0.97921</td>
<td>0.97949</td>
</tr>
<tr>
<td>Space group</td>
<td>C2</td>
<td>C2</td>
</tr>
<tr>
<td>Cell dimensions (a, b, c) (Å)</td>
<td>72.18, 70.19, 60.83</td>
<td>72.01, 70.05, 60.72</td>
</tr>
<tr>
<td>Cell dimensions (α, β, γ) (°)</td>
<td>90.00, 95.36, 90.000</td>
<td>90.00, 95.12, 90.00</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50 – 1.60 (1.64 – 1.60)</td>
<td>50 – 1.35 (1.39 – 1.35)</td>
</tr>
<tr>
<td>(R_{sym})</td>
<td>0.061 (0.55)</td>
<td>0.058 (0.52)</td>
</tr>
<tr>
<td>(I/σI)</td>
<td>16.7 (2.57)</td>
<td>11.0 (2.04)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.9 (99.9)</td>
<td>99.6 (98.4)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>3.8 (3.7)</td>
<td>3.3 (2.8)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50 – 1.35 (1.39 – 1.35)</td>
<td></td>
</tr>
<tr>
<td>No. reflections</td>
<td>65635</td>
<td></td>
</tr>
<tr>
<td>(R_{work} / R_{free}) (%)</td>
<td>14.9 / 16.8</td>
<td></td>
</tr>
<tr>
<td>No. atoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>2638</td>
<td></td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>231</td>
<td></td>
</tr>
<tr>
<td>(B)-factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>28.5</td>
<td></td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>42.9</td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>34.5</td>
<td></td>
</tr>
<tr>
<td>R.m.s. deviations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.16</td>
<td></td>
</tr>
<tr>
<td>Ramachandran plot</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favored (%)</td>
<td>96.3</td>
<td></td>
</tr>
<tr>
<td>Allowed (%)</td>
<td>99.7</td>
<td></td>
</tr>
</tbody>
</table>

*The dataset was collected from a single crystal. Values in parentheses are for the highest-resolution shell.
4.5.5 Enzyme kinetics
Enzyme kinetic constants were determined by monitoring the ADP released by APK using a coupled pyruvate kinase/lactate dehydrogenase enzyme pair. Experiments were conducted on a MultiSkan FC plate reader at 340 nm at 25°C. Data was collected for 8 replicates of 300 μL volumes containing 200 μM NADH, 2.5 mM phosphoenolpyruvate, 50 mM KCl, 10 mM MgCl₂, 2 μL/mL pyruvate kinase/lactate dehydrogenase solution (P0294 Sigma), and an APK concentration that was adjusted to optimize signal (2-5 μg/mL wild-type enzyme or 20-100 μg/mL variant enzyme) with one substrate (alcohol substrate or ATP) held at 1 mM and the other varied over a range of concentrations. Kinetic constants were calculated by non-linear regression in SigmaPlot fitting the data to the Michaelis-Menten equation.

4.5.6 Thermal denaturation
Structural stability of protein variants were tested by the thermofluor method using an Applied Biosystems StepOnePlus thermocycler real-time PCR system. 50 μL reaction aliquots contained with 0.2 mg/mL protein (or 0.2 mg/mL protein plus 2 mM MgCl₂ and 2 mM ATP) in 20 mM Tris pH 8.0, 150 mM NaCl and 1X SYPRO orange dye. Solution temperatures were increased in 1°C increments and fluorescence was monitored at 570 nm. Each condition was evaluated in triplicate and melting temperatures are reported as the average temperature to denature half of the protein in the reaction.

4.5.7 Bioinformatics analysis
The sequence of APK_MSMS0270 was searched using ExPASy BLAST against the UniProt90 database. Sequence similarity networks were generated using EFI-EST¹⁰³, and a gene neighbourhood network was generated using EFI-GNT. The resulting data relationships were visualized using cytoscape¹⁰⁴.
5 Structural And Functional Characterization of a Homolog of the

*Mycobacterium* Microcompartment *O*-Phosphopropanolamine

Phospholyase from *Mesorhizobium loti*.

This chapter reports experiments contributing to an ongoing project.

Acknowledgements

We thank P. Grochulski and S. Labiuk at Canadian Light Source for assistance in collecting diffraction data.
5.1 Introduction

The *Rhodococcus* and *Mycobacterial* Microcompartment (RMM) is a newly identified class of catabolic microcompartment. Included in the RMM are four shell proteins and four enzymes, including previously characterized 1-amino-2-propanol alcohol dehydrogenase and 1-amino-2-propanol kinase. A third enzyme in the RMM is classified as a class-III aminotransferase enzyme, and is proposed to be the aldehyde producing “signature” enzyme of this class of microcompartment.

The class-III aminotransferase enzymes catalyze transfer of ammonia from an amine donor to an acceptor keto-acid, forming an amino acid and a carbonyl product. Included in this family are the ornithine, gamma-aminobutyric acid, and putrescine aminotransferases. These enzymes catalyze aminotransferase reactions through a two-step reaction using a prosthetic group pyridoxal-5-phosphate (PLP). The first reaction transfers the substrate amine to PLP forming pyridoxamine phosphate (PMP); the second reaction transfers the amine to a keto-acid amine acceptor such as α-ketoglutarate or pyruvate forming glutamate or alanine, respectively.

A subclass of the class-III aminotransferase enzyme catalyzes the dephosphorylation of *O*-phosphoryl aminoalcohols. Currently the only characterized example of this subclass is the *O*-phosphorylethanolamine (PEtN) phospholyase. PEtN phospholyases are proposed to function though a similar first reaction between the substrate amine and PLP forming a PMP-PEtN intermediate. This intermediate is protonated releasing inorganic phosphate, then an alkenylamine product that undergoes hydration releasing ammonia and an aldehyde. PEtN phospholyases are found in bacteria and in eukaryotes including humans, contain a conserved active site with a basic phosphate-binding pocket formed by lysine and arginine residues.

*O*-phosphorylpropanolamine (PPrN) is an uncommon chemical; with the only characterized biological synthesis sources originating from the previously characterized
aminopropanol kinase from the *Mycobacterium smegmatis* RMM, and aminoacetone degradation pathways in *Erwinia* and *Pseudomonas* species\textsuperscript{97,112}. The identified aminoacetone degradation pathways included an aminopropanol dehydrogenase, an aminopropanol kinase, and a phospholyase enzyme, which catalyzed the dephosphorylation and deamination of PPrN producing propanal. The common catalytic activities of the characterized *M. smegmatis* RMM enzymes and aminoacetone degradation pathways led to the proposal that the class-III aminotransferase enzyme from the *M. smegmatis* RMM may be a second example of the phospholyase subclass of enzymes, dephosphorylating PPrN\textsuperscript{110}.

The aminotransferase class-III enzyme from the *M. smegmatis* RMM (msmeg\_0277; MSM0277) is most closely related to the putrescine aminotransferase from *Escherichia coli* (YgjG; 34\% sequence identity). Active site residues involved in PLP binding are conserved in these proteins, however the substrate binding pocket of MSM0277 does not share conservation of key substrate binding and catalytic residues with YgjG\textsuperscript{57}. The substrate binding pocket of MSM0277 (as determined from alignment with YgjG) has substituted putrescine binding residues with basic residues, and the characteristic glutamate-arginine switch and surrounding residues for keto-acid binding are not conserved \textsuperscript{54,57}. This altered basic substrate binding pocket suggests a possible convergent evolution of PLP-dependent phospholyase activity in the RMM analogous to that of the PEtN phospholyase.

We have identified a second aminoacetone degradation pathway in *M. smegmatis* with enzymes catalyzing aminopropanol dehydrogenase and aminopropanol kinase reactions\textsuperscript{94,110}. Included in this alternate pathway is an aminotransferase class-III enzyme (msmeg\_0782; MSM0782). The aminotransferase enzyme from this pathway, however, shares conservation of all key substrate pocket and catalytic residues with the PEtN phospholyase identified from
Arthrobacter aurescens TC1 (A1RDF1; 42% sequence identity). Analogous to the RMM pathway, MSM0782 conceivably catalyzes a PLP-dependent phospholyase activity with PPrN.

We investigated the aminotransferase class-III enzymes from the RMM and the alternative aminoacetone degradation pathways to determine their catalytic activities. Recombinant constructs of the putative phospholyase, MSM0277, resulted in only insoluble protein. As a means to characterize structural and functional aspects of this enzyme, an ortholog not associated with a RMM from Mesorhizobium loti (63% sequence identity with MSM0277) was explored as a substitute. The aminotransferase enzyme (UniProt mlr6714; henceforth ML0277) was analyzed using x-ray crystallography and catalytic assays. The PEtN phospholyase homolog from the alternate aminoacetone degradation pathway was similarly assayed for phospholyase activity. Both aminotransferase enzymes catalyzed reactions producing aldehyde from O-phosphoryl amino alcohols supporting the proposed PPrN phospholyase functions in their respective aminoacetone degradation pathway. The active site of ML0277 was reminiscent of the PEtN phospholyase active site, with a basic substrate binding pocket formed by a pair of conserved histidine residues and a more weakly conserved tryptophan.

5.2 Materials and methods

5.2.1 Mycobacterium smegmatis MC² 155 genomic DNA purification

Mycobacterium smegmatis MC² 155 (a generous gift from Dr. Lucy Mutharia) was cultured in lysogeny broth with added 10 µg/mL cyclohexamide, 50 µg/mL carbenecillin, and 0.05% (v/v) Tween 80 at 37°C with shaking for 72 hours. Cells were collected by centrifugation and genomic DNA was extracted by column purification with the Invitrogen Purelink miniprep kit.
5.2.2 Mesorhizobium loti MAFF 303099 genomic DNA purification

*Mesorhizobium loti* MAFF 303099 (a generous gift from Dr. Krzysztof Szczyglowski) was cultured on YT agar for 5 days at 30°C. A single large colony was collected and pelleted in sterile water, genomic DNA was extracted by column purification with the Invitrogen Purelink miniprep kit.

5.2.3 Molecular cloning

Genes (msmeg_0277, msmeg_0782, or UniProt mlr6714; *Mesorhizobium loti* MAFF 303099 locus BAB52954.1) were amplified from genomic DNA purified from in house cultured source bacteria. PCR amplification of genes was carried out by initial denaturation of template DNA and primers (msmeg_0277_pET28a: 5’-AGGGATCATATGTACGACTACGGCACGTTCGTTTCG-3’ and 5’-GATACTAAGCTTTTACCCGTTCTCCGTCCGTTGCAGTGC-3’, msmeg_0277_pET22b: 5’-AGGGATCATATGTACGACTACGGCACGTTCGTTTCG-3’ and 5’-GATACTAAGCTTTTACCCGTTCTCCGTCCGTTGCAGTGC-3’, msmeg_0782_pET28a: 5’-GCTAGCTAGCAGGTTTTCCAACATCATGGATTCCAAC-3’ and 5’-GCTACTCGAGCTAGGGCAGATGTGCCAGCGCCCCCGCAAA-3’, mlr6714_pET28a: 5’-GATCCATATGATGTACGACTACGGCACGTTCGTTTCG-3’ and 5’-GATCAAGCTTTCAACTCGCTTGCGCTGCGCTGATG-3’) at 96°C for 5 minutes followed by 25 cycles of 96°C (30 sec), 50-60°C (30 sec), 72°C (2 min/kb) and final extension at 72°C for 15 min. PCR amplicons were purified by a PCR clean up kit to remove contaminating chemicals and evaluated for purity and molecular weight by agarose gel electrophoresis, then quantitated spectrophotometrically. Insert DNA and vector DNA (pET 22b or pET 28a) were digested by restriction endonucleases and vector DNA dephosphorylated by Antarctic phosphatase before ligation by T4 DNA ligase. Ligated plasmids were transformed into *E. coli* DH5α cells and
propagated with selective antibiotics; resulting colonies were screened for successful ligation by agarose gel separation of restriction endonuclease plasmid digestion followed by verification of sequences by dye terminator sequencing. Plasmids carrying verified gene inserts were used to transform *E. coli* BL21 (DE3) cells for protein expression.

### 5.2.4 Protein expression

Protein over expression was carried out in *E. coli* BL21 (DE3) cells with plasmids containing the pBR322 origin of replication under expression regulation by the LacI repressor. A single colony, or a scraping from 25% (v/v) glycerol cryoprotected frozen (-80°C) aliquot of cells derived from single colony transformations, were used to inoculate 5 mL small volume cultures of LB medium with antibiotic selection. After overnight growth (16-18 hours) of small volume cultures at 37°C with shaking, 1 L of 2YT media with antibiotic selection were inoculated with the small volume culture and grown at 37°C, with shaking, until an optical density at 600 nm (OD$_{600nm}$) of 0.8 was reached. Induction of transcription and protein expression was initiated by addition of 1 mM IPTG to the media followed by overnight (18-20 hours) incubation at 16°C with shaking. Following overnight expression, cells were pelleted by centrifugation at 4,400 g for 30 minutes at 4°C and resuspended in lysis buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 500 μM pyridoxal-5-phosphate with added 10% (v/v) glycerol for MSM0277 purification) prior to storage at -20°C.

### 5.2.5 Selenomethionine protein labeling

Labeling of ML0277 protein with selenomethionine was carried out in the methionine auxotrophic *E. coli* BL21 (DE3) derivative B834 cells. To remove alternative sources of methionine from the growth media, cells were grown until an OD$_{600nm}$ of 0.6 was reached then cells were collected by centrifugation at 4,400 g for 30 minutes. The cell pellet was resuspended in selenomethionine media (prepared according to guidelines from the EMBL Heidelberg Protein
Expression and Purification Core Facility [https://www.embl.de/pepcore/pepcore_services/protein_expression/ecoli/seleno/] and incubated at 37°C with shaking until an OD$_{600nm}$ of 0.8 was reached at which point protein expression was induced with 1 mM IPTG. The prior expression and subsequent purification steps were carried out with an identical method as unlabeled protein expression and purification.

5.2.6 Protein purification

Resuspended pellets were thawed at 4°C prior to lysis, and cooled by immersion in an ice bath during lysis. Lysis of cells was carried out by sonication with a total of 10 min processing time with 10 second pluses and 20 seconds of cooling at power level 7 on a Misonex sonicator 2020XL. The insoluble fraction of the lysate was separated by centrifugation at 48,000 g for 30 min at 4°C. Supernatants were applied to agarose Ni-IMAC resin, followed by elution of non-specifically bound proteins with lysis buffer containing 50 mM imidazole. Target proteins were eluted from the IMAC resin with lysis buffer containing 500 mM imidazole. Pure protein aliquots were combined based on SDS-PAGE evaluation of protein composition and exchanged into minimal buffer (20 mM Tris-HCl pH 8.0 and 150 mM NaCl) by desalting with a Hi-Prep desalting column equilibrated with minimal buffer. The protein eluate was concentrated by centrifugation in a centrifugal concentrator. Quantification of protein concentrations was determined by spectrophotometric absorbance at 280 nm (ML0277: molecular weight 49,551 Da and ε: 64860 M$^{-1}$cm$^{-1}$; MSM0782: molecular weight: 49,595 Da and ε: 33,350 M$^{-1}$cm$^{-1}$).

5.2.6 Crystallography

Initial sparse matrix screening of ML0277 at 5 mg/mL (including 2 mM PLP) produced hexagonal prisms after incubation at room temperature for 7 days in 0.1 M HEPES pH 7.5, 25% (w/v) PEG 3350 in a 1:1 ratio. Selenomethionine labeled protein crystals were grown by an
identical method to that used for the unlabeled protein. Protein crystals were cryoprotected by immersion in Paratone N prior to freezing in liquid nitrogen.

Datasets were collected at the Canadian Light Source beam-line 08ID at 100 K, and were processed using XDS and scaled using XSCALE\textsuperscript{88}. X-ray diffraction data for selenomethionine labeled ML277 crystals was collected at the K absorption edge for selenium multi-wavelength anomalous dispersion phasing. Phases were determined in Phenix\textsuperscript{90} AutoSol using anomalous diffraction site detection, with experimental phasing, and molecular replacement. The initial model determined by AutoSol was processed by Phenix Autobuild model building and refinement to expand the initial model. A final model was produced by iterative manual building and refinement in Coot\textsuperscript{91} and automated refinement in Phenix Refine. Data collected from the selenomethionine labeled protein crystal was initially processed in the P6\textsubscript{5}22 space group. Inability to determine reliable phases in this space group led to re-examination of alternative allowable space groups and the reprocessing of the data in the P6\textsubscript{5} space group with merohedral twinning.

5.2.7 Phosphoethanolamine phospholyase activity assay

Activity for phosphopropanolamine phospholyase enzymes were determined by coupling aldehyde production with reduction by alcohol dehydrogenase. Reaction coordinates were recorded as a change in absorbance due to oxidation of NADH measured on a MultiSkan FC plate reader at 340 nm at 25°C. Data was collected for 8 replicates of 300 µL volume reactions containing 50 µM pyridoxal-5-phosphate, 100 µM NADH, 25 U/mL \textit{Saccharomyces cerevisiae} alcohol dehydrogenase (A7011 Sigma-Aldrich), and 50 µg/mL of enzyme in minimal buffer with 10 mM O-phosphoethanolamine (P0503 Sigma-Aldrich). Initial rates of reaction were converted from change in absorbance per second to change in NADH concentration per concentration of
enzyme per second \((\varepsilon_{340} 0.00622 \text{ L}\mu\text{mol}^{-1}\text{cm}^{-1})\), enzyme molecular weights of 49,550.70 g/mol for ML0277 or 49,595.94 g/mol for MSM0782, and pathlength \(l\) of 0.84 cm with equation 1).

5.2.8 Phosphopropanolamine phospholyase activity assay

The proposed target substrate, phosphopropanolamine, is not available for purchase and required in vitro synthesis from available chemicals. Using the aminopropanol kinase MSM0270 identified in chapter 4, phosphopropanolamine was synthesized in situ in a two-fold concentration pre-reaction containing 50 \(\mu\text{g/mL}\) MSM0270, 1 mM magnesium chloride, 50 mM ATP, and 4 mM S-(-)-1-amino-2-propanol in minimal buffer. Substrate synthesis reactions were mixed 1:1 with enzyme solution containing 25 \(\mu\text{g/mL}\) ML0277 or 2 \(\mu\text{g/mL}\) MSM0782 with a final concentration of 50 \(\mu\text{M}\) pyridoxal-5-phosphate, 25 U/mL S. cerevisiae alcohol dehydrogenase, and 100 \(\mu\text{M}\) NADH in a final reaction volume of 300 \(\mu\text{L}\). Reaction coordinates were recorded as a change in absorbance due to oxidation of NADH measured on a MultiSkan FC plate reader at 340 nm at 25°C with 8 replicate reactions for each enzyme. Initial rates of reaction were converted from change in absorbance per second to change in NADH concentration per concentration of enzyme per second \((\varepsilon_{340} 0.00622 \text{ L}\mu\text{mol}\cdot\text{cm}, \text{enzyme molecular weights of } 49,550 \text{ g/mol for ML277N or } 49,695 \text{ g/mol for MSM0782, and pathlength } l \text{ of } 0.84 \text{ cm with equation 1}).

Equation 1

\[
V_o(\text{sec}^{-1}) = \frac{(V_o(A_{340}/t))/((\varepsilon_{340}(L/\mu\text{mol} \times \text{cm}) \times l(\text{cm}))/([E](\text{mg/mL})/M(\text{g/mol}))) \times 1000000(\mu\text{mol})}
\]

5.3 Results

5.3.1 Protein purification

The microcompartment associated phospholyase MSM0277 experienced issues with solubility regardless of affinity tag location, growth conditions, and purification conditions. Issues with solubility may be due to the innate tendency of the enzyme to aggregate in order to
facilitate encapsulation within a microcompartment. Therefore a non-microcompartment associated ortholog of MSM0277 from *Mesorhizobium loti* (ML0277) was explored as an alternative subject for structural and activity investigations. After successful cloning of the gene for ML0277 from genomic DNA into the expression vector pET28a, purification of the N-terminally hexahistidine tagged protein produced soluble protein when stored in minimal buffer. The non-microcompartment RMM-like operon aminotransferase gene (msmeg_0782) amplified from *M. smegmatis* genomic DNA was successfully cloned into pET28a. Soluble MSM0782 recombinant protein was purified and stored in minimal buffer for enzymology experiments.

### 5.3.2 Enzyme activity

PLP dependent phospholyase activity, as proposed for the phosphoethanolamine phospholyase, catalyzes a dephosphorylation of an α-amino β-phosphoryl substrate with spontaneous deamination resulting in aldehyde production. Catalytic activity was evaluated by coupling aldehyde production with NADH oxidation by *S. cerevisiae* alcohol dehydrogenase, measuring the change in absorbance at 340 nm due to change in NADH concentration. Phospholyase activity was observed with phosphoethanolamine and phosphopropanolamine with either MSM0277 or MSM0782. MSM0277 had a turnover rate of 0.016 s⁻¹ with 10 mM phosphoethanolamine or 0.32 s⁻¹ with 2 mM phosphopropanolamine. MSM0782 had a turnover rate of 0.15 s⁻¹ with 10 mM phosphoethanolamine or 4.9 s⁻¹ with 2 mM phosphopropanolamine.

### 5.3.3 Structural overview

ML0277 crystallized in the P6₅ space group with a dimer in the asymmetric unit. The structure was refined to 2.2 Å with a $R_{\text{work}}$ of 0.22 and a $R_{\text{free}}$ of 0.24. Data collected for this crystal proved to be merohedrally twinned, making solving the structure challenging. The tertiary structure retains an aminotransferase class-III type fold, consisting of a substrate binding domain and a cofactor binding domain. The substrate binding domain is formed by two
antiparallel β-sheets and two α-helices from the N-terminus (residues 1-72), which pack along
the C-terminal three stranded antiparallel β-sheet and a three helices α/β fold (residues 314-427).
The cofactor binding domain composed of a mostly parallel six stranded β-sheet and six α-
helical α/β-fold (residues 73-313) (Figure 5.1 A). The active sites of the dimer are found at the
interface of the substrate binding domain and cofactor binding domain of one protomer, with
contribution of active site residues from the adjacent protomer (Figure 5.1). Nearly the entire
polypeptide chain (2Y-Q422), with the exception of the initial methionine and the terminal 5
residues, could be traced by electron density. An addition volume of electron density could be
found in the active sites of both protomers with characteristic features suitable to ascribe the
electron density as lysine (K263) covalently bound pyridoxal-5’-monophosphate (LLP;
N’pyridoxal-lysine-5’-monophosphate).
**Figure 5.1 Structural overview of the *Mesorhizobium loti* phosphopropalanolamine phospholyase.**

A) Secondary structure labeled protomer model with the substrate binding domain coloured in blue and cofactor binding domain in gold; bound N’pyridoxal-lysine-5’-monophosphate is shown in magenta. B) Dimeric organization with one protomer coloured as in A. C) Active site with conserved residues proposed to be involved in phosphate binding (pink), PLP binding residues (green), and the catalytic internal aldimine forming lysine (blue) with the internal aldimine bound PLP (magenta). D) Superimposed active site residues (including phosphate binding residues, labeled in yellow) of the phosphoethanolamine phospholyase A1RDF1 (PDB 5G4J) with PLP as the phosphoethanolamine (green) bound external aldimine state showing hydrogen bonding of the phosphate group with the active site (red dotted lines).
Figure 5.2 Sequence alignment of aminoglycoside aminotransferase class-III like phospholyases.

Sequence alignment of the phosphopropionolamine phospholyase from *Mesorhizobium loti* (ML0277, UniProt m1r6714) with the *Mycobacterium smegmatis* AAUM phosphopropionolamine phospholyase (MSMEG0277, UniProt A0QP54), *Rhodococcus erythropolis* AAUM phosphopropionolamine phospholyase (Rhoer_5062, UniProt C3JQC8), *Arthrobacter aurescens* phosphoethanolamine phospholyase (A1RDf1, UniProt A1RDF1), and *Mycobacterium smegmatis* non-microcompartment phosphopropionolamine phospholyase (MSMEG0782, UniProt A0QQJ9). Coloured dots indicate conserved residues proposed to be involved in phosphate binding (pink or yellow), PLP binding residues (green), and the catalytic internal aldimine forming lysine (blue).
5.3.4 PLP binding site

Analogous to characterized members of the aminotransferase class-III type enzymes, ML0277 contains all canonical PLP binding site features (Figure 5.2). Besides the Schiff base covalent bond between the 4’-aldehyde group of PLP and lysine (K263), aspartate (D234) hydrogen bonds with the PLP pyridine ring nitrogen, the PLP hydroxyl forms a hydrogen bond with glutamine 237, and the pyridine ring is sandwiched between two hydrophobic residues (V236 and the ring of Y148) (Figure 5.1). The phosphate of PLP is found in a canonical phosphate binding pocket, where it forms hydrogen bonds with Y148, T295, and T296 and the backbone amides of A122, G121, and T295.

5.3.5 Substrate binding site

The noticeably unique feature of ML0277 compared to previously characterized enzymes of the aminotransferase class-III type enzymes is the substrate binding site. Compared to the closest characterized structural and sequence related homolog, *Escherichia coli* putrescine aminotransferase YgjG (PDB code 4UOX, 32% sequence identity), the entire substrate binding site has been altered for the specialized activity of ML0277. Characteristics of the substrate binding site resembles that of the functionally similar enzyme, phosphoethanolamine phospholyase from *Arthrobacter aurescens* TC1 (A1RDF1; PDB 5G4J) (Figure 5.2D). The glutamate and arginine switch (E243:R426, 4UOX numbering) used by aminotransferase enzymes to specify amino acceptor α-carboxylic acid, is absent from the active site in both A1RDF1 and ML0277. In the active site of the putrescine aminotransferase YgjG, the α-amino group and aliphatic characteristics of putrescine are mirrored by the hydrophobicity of the substrate binding pocket culminating in an electronegative pocket, specifying the terminal amine. The PLP dependent phospholyase enzymes instead have distinct electropositive substrate binding pocket for interacting with the substrate phosphate group. Despite differences
in the proposed phosphate binding pockets, their function in polarizing phosphate facilitating its release as inorganic phosphate, appears to be similar. A1RDF1 tetrahedrally coordinates with the substrate phosphate with the basic residues lysine (K412), and two arginines (R90, R414) with additional hydrogen bonding by Y61 and Q254 (Figure 5.1D). A basic pocket is formed in the active site of ML0277 by the residues H60 and H91, with amine groups oriented for hydrogen bonding with phosphate. Additional hydrogen bonding may be contributed by adjacent N62 and W92. W92 is observed in two distinct rotational conformations. In one conformation, the indole ring amine is pointing toward the phosphate binding pocket in a suitable position for hydrogen bonding with substrate phosphate; in the other conformation the indole ring is pointing away from the phosphate binding pocket.

5.4 Discussion

In the RMM operon of *M. smegmatis*, the only potential aldehyde producing “signature” enzyme is the aminotransferase class-III enzyme MSM0277. We attempted to purify a recombinant construct of the enzyme under optimal conditions, however, only insoluble protein was produced. Insolubility issues observed with the microcompartment enzyme may be due to aggregation of the enzyme involved in microcompartment assembly. Alternatively an unidentified binding partner from the microcompartment may be required to stabilize the enzyme structure or associate with surface features to prevent aggregation, similar to the ethanolamine ammonia lyase EutBC heterocomplex. To determine the function of the microcompartment phospholyase, we investigated the closest available homolog not associated with a microcompartment operon from *Mesorhizobium loti* (ML0277; 64% sequence identity over 97% of the sequence with MSM0277). Aldehyde production was observed with ML0277 and either phosphopropionaldehyde or phosphopropionaldehyde without the addition of an amine acceptor keto-acid. Phospholyase activity was significantly lower compared to the phospholyase
characterized from the *M. smegmatis* non-microcompartment aminoacetone utilization operon (MSM0782), as indicated from preliminary catalytic investigations. The discrepancy in activity is likely due to differences in the phosphate binding pockets; MSM0782 possesses the canonical PLP phospholyase pattern of basic phosphate binding residues (Arg-Arg-Lys), whereas ML0277 has an alternative less electropositive phosphate binding pocket (His-His-Trp). Substitution of W92 (ML0277) for a third more electropositive histidine in the phosphate binding pocket, similar to homologs from Actinomycetales species including those in the AAUM, may enhance the catalytic efficiency of the phospholyase activity.

We propose the phosphopropanolamine phospholyases identified here operate by an analogous mechanism to that proposed for the phosphoethanolamine phospholyase enzymes. Similar to the first half reaction of PLP dependent aminotransferases, the prosthetic group PLP is transferred from an internal aldimine Schiff base to an external aldimine Schiff base with the substrate. With proton abstraction by the substrate phosphate from the catalytic lysine, inorganic phosphate is released followed by release of the alkenylamine product with transfer of PLP back to the catalytic lysyl amine. Spontaneous hydrolysis of the alkenylamine produces ammonia and an aldehyde. To facilitate dephosphorylation, polarization of the substrate phosphate group by the electropositive phosphate binding pocket drives the reaction forward. Two variations of the phosphate binding pocket appear to have arisen from convergent evolution. The first phosphate binding pocket variation is exemplified by the previously characterized PLP dependent phospholyase A1RDF1, consisting of three conserved basic residues lysine/arginine/arginine and a tyrosine. This type of phosphate binding pocket is conserved by phospholyases found in Actinomycetales including *Bifidobacteria* and *Mycobacteria* species and is not associated with microcompartment forming proteins. The alternative phosphate binding pocket variation is
composed of basic residues derived from distant residues compared to the conserved residues of the phosphoethanolamine phospholyases. This variation of phosphate binding pocket is present in the structure of ML0277 reported here. Residues H60, N62, H91, and W92 are proposed to form the alternative phosphate binding pocket and are conserved in homologs found in Actinomycetales, including all *Mycobacteria* and *Rhodococci* species associated with microcompartment operons, and in Firmicutes and α-Proteobacteria species (not associated with microcompartments). Conservation of W92 is only observed in homologs from α-Proteobacteria species, otherwise replaced by histidine in homologs from Actinomycetales or Firmicutes.
6 Conclusions

6.1 Structural characterization of the aminoacetone utilization microcompartment shell proteins

In chapter 2, the structural characterization of all the microcompartment shell proteins from the AAUM from *M. smegmatis* was presented. At the time of publication it represented the first microcompartment for which all of the shell proteins had been experimentally characterized. The shell proteins share conservation with classical microcompartment shell protein architectures with a single BMC domain BMC-H protein (MSM0272), two fused-permuted BMC domain BMC-T proteins (MSM0271 and MSM0275), and a single pfam003319 domain BMC-P protein (MSM0273). The MSM0272 BMC-H protomer resembled the common BMC domain forming a hexagonal hexamer with a pore found at the central axis of symmetry. The pore formed by MSM0272, proposed to function in substrate transport in homologous systems, was lined by negatively charged aspartate residues. This negatively charged pore was therefore proposed to facilitate transport of positively charged substrates, such as aminoacetone, across the shell for metabolism by the AAUM pathway. Alternatively this pore may be involved in transport of propionaldehyde derived from the dephosphorylation and deamination of O-phosphorylpropanolamine by the microcompartment enclosed phospholyase as suggested in chapter 3. Uptake of only the aldehyde intermediate has previously been proposed for the ethanolamine utilization microcompartment, where the ethanolamine lyase signature enzyme is located just outside the substrate transporting pore and the aldehyde produced by the amino lyase is taken in by the microcompartment and rapidly converted to acetyl-phosphate and ethanol. The C-terminal helix of MSM0272 was partially flexible and prone to disorder from our structural experiments. This plasticity, however, has been proposed to be involved in microcompartment targeting as encapsulation peptides readily form complexes with these
helices. The two BMC-T proteins from the RMM share conservation with classical fused-permuted BMC-T architecture, with loops connecting the third and fourth β strands of each BMC domain occluding the pore structure at the three-fold rotational axis. A conformational shift of the occluding loops to an open conformation is plausible for either BMC-T oligomer as observed in either CcmP or Csos1D BMC-T proteins of carboxysomes. Opening of either trimeric ring pore in the hexameric oligomer would provide access to the internal positively charged cavity formed between trimers. Intermittent opening and closing would provide a means of transport between the lumen and exterior of the microcompartment potentially acting to transport negatively charged phosphorylated products or nucleotide-pyrophosphate cofactors (such as ATP, NADPH, or coenzyme A). An alternate alignment of trimer-trimer oligomerization was observed in the structure of MSM0271 indicating a possible alternative oligomeric state for BMC-T hexamers, where trimers are aligned opposed to having one ring offset by 60 degrees, as observed in all other known examples. The MSM0273 BMC-P pentamer formed a small diameter pore; however unlike structurally characterized homologs the pore was not occluded by a small molecule; instead a network of hydrogen bonded waters occupied the pore. BMC-P oligomers are proposed to solely act as structural proteins; however conservation of pore lining residues and observation of multiple homologs in single microcompartment operons suggests they may play a role in transport that is yet to be elucidated.

In section 2 we proposed a model for the AAUM shell complex from edge interactions predicted by docking and energy reduction computation in Rosetta. The model we proposed was since disproved by a publication reporting a combined Cryo-EM and x-ray crystallography determined structure of the Haliagium ochraceum microcompartment shell. In the H. ochraceum recombinant microcompartment assembly experiment, coexpression of all the shell
proteins with the exception of a BMC-P protein and purification of assembled uncapped shells by sucrose gradient allowed for packaging of desired contents followed by capping of the shell apices with recombinant BMC-P pentamers. In the *H. ochraceum* shell, edges were formed by two BMC-H hexamers with a single BMC-T hexamer at the center of the facet, and apices were capped by a pentamer of BMC-P proteins. Our model for the microcompartment failed to predict the characterized shell structure due to plasticity of edge interactions not previously observed in x-ray crystallographic studies, as well as assumptions made based on encapsulation sidedness of BMC-H oligomers.

Conservation of structural features including fused-permuted type BMC-T oligomers, edge interaction residues of BMC-H oligomers, and similar edge charge distribution/dimensions of BMC-T and BMC-P oligomers suggests that the AAUM would form a very similar shell architecture as observed for the *H. ochraceum* microcompartment shell. Synthesis of AAUM shells from expression of the four shell proteins (MSM0271, MSM0272, MSM0273, and MSM0275) in a recombinant operon in *E. coli*, followed by sucrose gradient purification with imaging of the structures by TEM would indicate whether the AAUM shell proteins assemble a shell structure. Subsequent Cryo-EM imaging and x-ray crystallography would provide a structure for comparison of shell ultrastructure with the structure determined for the *H. ochraceum* microcompartment. Aside from the architectural aspect of the shell, a recombinant system would allow for investigation of enzyme targeting properties and shell permeability properties. Enzyme localization and chemical permeability in the AAUM could be tested by coexpression of each enzyme with microcompartment shells. Enzyme localization could be analyzed by coexpression of individual or multiple AAUM enzymes with full shell protein complement followed by shell purification. Proteins associated with the microcompartment
shells would copurify with the shells. Inclusion of a protease cleavage site in encapsulation linker loop regions of proteins would allow for release of enzymes by digestion with the specific protease if they are solely associated with the exterior of the shell. Only luminally located proteins will copurify with the shell proteins after protease treatment. Chemical permeability properties of the shell could be analyzed by coexpression of each enzyme individually or with multiple enzymes from the AAUM with the microcompartment shell with substrates/cofactors supplied in the presence or absence of capping by BMC-P pentamers. Catalytic turnover and product release comparisons between capped and uncapped microcompartment shells would indicate what chemicals in the AAUM metabolism are shell-permeable.

6.2 S-(-)-1-amino-2-propanol alcohol dehydrogenase

In chapter 3, I reported the structure of the RMM 1-amino-2-propanol alcohol dehydrogenase (MSM0269) with and without substrates aminoacetone and NADP⁺. The structure agreed with the annotation as a member of the SDR family of enzymes, with a conserved Rossmann fold nucleotide binding domain and a short C-terminal substrate binding domain. Substrate binding induced a conformational change of the C-terminal loop, with closure of the C-terminal domain forming a highly occluded substrate binding pocket. The substrate binding pocket is hydrophobic with dimensions suitable for binding of a methyl group, with a negatively charged pocket at the opposing end. The resulting substrate binding pocket was of appropriate dimensions for the propane backbone of aminoacetone, with the amino group positioning the carbonyl group for proton transfer from the C4 carbon of the nicotinamide ring of NADPH. Specificity for NADP(H) over NAD(H) was conferred by the conserved ribose binding motif lacking an acidic residue for ribose hydroxyl coordination, instead the surrounding motifs form a basic phosphate binding pocket.
Catalytic activity with a low micromolar $K_M$ was detected with aminoacetone and aminopropanol, agreeing with previously determined catalytic activity of the *R. erythropolis* AAUM aminopropanol dehydrogenase ortholog. $\alpha$-amino $\beta$-hydroxyl/ carbonyl substrates with longer alkyl chain backbones were accepted by this enzyme albeit with higher $K_M$ values. Absolute stereospecificity for the S isomer over the R isomer appeared to be a function of the positioning of the 2’ carbon with the C4 of the nicotinamide ring and $\beta$-hydroxyl with the catalytic tetrad tyrosine hydroxyl through positioning of the $\alpha$-amine.

The catalytic activity of the non-microcompartment 1-amino-2-propanol alcohol dehydrogenase from *M. smegmatis* (MSM0779) determined in chapter 3 indicates that this enzyme is a functional equivalent of the AAUM SDR. The structure of this enzyme was previously determined by a structural genomics group, and shows a very similar overall structure with conservation in most of the substrate binding pocket residues. Homologs identified from sequence similarity in chapter 3 from $\alpha/\beta$ Proteobacteria and Bacilli that share conservation of key active site residues could be analyzed for the proposed aminopropanol dehydrogenase activity.

### 6.3 1-amino-2-propanol kinase

In chapter 4, I presented the structural and catalytic characterization of an aminopropanol kinase from the *M. smegmatis* AAUM (MSM0270). I was able to determine only the structure of the substrate-free enzyme, despite efforts to determine a ternary structure. The proposed substrate binding pocket forms a tight hydrophobic pocket with a pair of acidic residues at the opposing end. The hydrophobic pocket was proposed to form hydrophobic interactions with the substrate methyl group, while the acidic residues were proposed to orient the target hydroxyl for phosphorylation through coordination with the substrate amine. Compared to the active site of
choline kinase enzymes, which catalyze phosphorylation of a similar small hydrophobic substrate\textsuperscript{48}, MSM0270 formed a smaller hydrophobic pocket though incorporation of aromatic residues from flexible loops distant from the active site. Replacement of either acidic residue proposed to function in amine binding with alanine resulted in inactivation or minimal activity. A ternary structure with aminopropanol and catalytically inactive analogues of the nucleotide triphosphate (adenylyl-imidodiphosphate or ADP and aluminum tri-fluoride) would validate proposed substrate binding mechanisms. MSM0270 catalyzed the phosphorylation of amino-alcohols with 2 to 4 backbone alkyl chains with a preference for the 3-carbon, S isomer over the R isomer. Aminopropanol kinase activity was also reported for the orthologous enzyme (MSM0780) from the non-microcompartment aminoacetone degradation operon. A lower $K_M$ value for aminopropanol kinase activity of MSM0780 is proposed to be a result of a more hydrophobic methyl binding pocket due to substitution of a phenylalanine (F251) to tryptophan. Variants of the proposed methyl binding hydrophobic pocket residues (W167, F251, L294, and W297) could be explored to determine the effect each residue has on substrate binding and catalytic efficiency. Structural and catalytic characterization of proposed orthologs based on sequence analysis identified in chapter 4 would allow for a comparison of substrate binding and catalytic mechanisms for this newly characterized member of the APH family of enzymes.

6.4 \textit{O-Phosphopropanolamine phospholyase}

The structure of the \textit{Mesorhizobium loti} \textit{O-}phosphopropanolamine phospholyase (ML0277) is presented in chapter 5. The cofactor binding domain conserved structural features and cofactor binding residues with other members of the aminotransferase class-III enzymes\textsuperscript{52}. PLP was found bound to the catalytic lysine (K263) as an “internal” aldimine similar PLP observed in aminotransferase class-III homologues. Substrate binding residues conserved by aminotransferase class-III enzymes, including the amine acceptor switch and substrate binding
pocket, were not conserved by ML0277. The active site did bear resemblance to a subclass of the aminotransferase class-III enzymes, which use a strong electropositive phosphate binding pocket to catalyze a phospholyase reaction with spontaneous deamination of O-phosphoryl amino-alcohol substrates. The proposed phosphate binding pocket in ML0277 is formed by two histidine residues and a tryptophan residue (replaced by a third histidine in Actinomycetales homologues) as opposed to two arginines and a lysine in characterized human and bacterial phosphoethanolamine phospholyases. Further structural characterization should be explored to generate a higher resolution model of the enzyme, and a substrate bound structure to provide direct evidence of the substrate binding in this enzyme.

In chapter 6 the catalytic activity investigation is presented for a phosphopropanolamine phospholyase from *Mesorhizobium loti* with the readily available phosphoethanolamine and *in situ* synthesized phosphopropanolamine. ML0277 catalyzed the proposed reaction, producing aldehyde products with either substrate as determined by coupling the reactions with an alcohol dehydrogenase to detect aldehyde production. Phospholyase activity of the non-microcompartment associated aminoacetone degradation operon aminotransferase (MSM0782) was verified by the same method with either substrate. Full kinetic parameters should be determined for both phospholyase enzymes with phosphoethanolamine and phosphopropanolamine substrates. Alanine variants of proposed substrate binding site residues (H60, H91, or W92) in ML0277 should be investigated for their contributions to the catalytic mechanism. Of particular interest would be a comparison of activity of a tryptophan (W92) to histidine variant in ML0277 relative to the wild type enzyme, representing the proposed phosphate binding pocket of the AAUM phospholyase homolog. Differences in turnover rate of
the enzymes should be investigated for effects of pH and product inhibition, as homologous enzymes appear to experience phosphate inhibition\textsuperscript{55}.

The \textit{M. loti} phospholyase was explored as an alternative to the \textit{M. smegmatis} AAUM (MSM0277) for structural and catalytic investigation of the phospholyase. These homologs share 64\% sequence identity, including residues proposed to be involved in substrate binding and catalysis. From the structural and catalytic properties characterized from the \textit{M. loti} enzyme, I propose that the annotated class-III aminotransferase from the AAUM similarly catalyzes a phospholyase activity. We hypothesize that the insolubility of MSM0277 is a functional feature that allows this protein to aggregate into compact foci in order to be encapsulated within the shell. Solubilization of MSM0277 through residue substitution, addition of solubility protein tags, chemical solubilization, or chemical denaturation and refolding could therefore be explored. Similar to other catabolic microcompartment systems, a binding partner may be required to solubilize the enzyme. Coexpression of MSM0277 with the remaining proteins from the RMM should be tested for any effect on solubility. Soluble protein should be tested for phospholyase activity to confirm the proposed function. Variation of surface exposed residues conserved in AAUM homologs could be used to discern interactions mediating self-association. Recent attempts of solubilizing the AAUM putative phospholyase enzyme by an undergraduate research project student resulted in a partially soluble protein upon replacement of a surface exposed residue. Replacement of this residue is proposed to disrupt a conserved electronegative surface conserved in AAUM orthologs, potentially disrupting aggregation interactions involved in microcompartment assembly.

6.5 Aminoacetone utilization microcompartment metabolism

From the enzyme investigations presented in chapters 3-5, I characterized an aminopropanol dehydrogenase, aminopropanol kinase, and an homolog of the proposed
phosphopropanolamine phospholyase from the AAUM in *M. smegmatis*. The reactions these enzymes catalyze agree with the aminoacetone utilization pathway proposed by Jones and Turner\(^{50}\). The remaining enzyme in the AAUM operon not characterized here is the acylating aldehyde dehydrogenase (MSM0276). This enzyme was tested for aldehyde dehydrogenase activity (not presented in this work), however no activity was detected. Orthologs from AAUM systems should be explored for catalytic activity to verify the proposed propionaldehyde acylating aldehyde dehydrogenase activity. Lack of observable activity may be due to inactivation of the enzyme by oxidation of the conserved catalytic cysteine\(^{61}\). Verification of this modification should be investigated by mass spectrometry analysis of trypsin digested protein. Despite the inability to verify the MSM0276 dehydrogenase activity, a metabolic pathway can be proposed for the AAUM operon from determined functions of the remaining enzymes. We propose the AAUM family of microcompartments reduce aminoacetone to aminopropanol, which is subsequently phosphorylated to *O*-phosphopropanolamine, dephosphorylated and deaminated to propanaldehyde, then oxidized to propionyl-CoA (Figure 6.1). We therefore proposed that the *Rhodococcus/Mycobacteria* family of microcompartments be renamed the aminoacetone utilization microcompartment (AAUM)\(^{110}\).

The current microcompartment assembly mechanism for microcompartments (except \(\alpha\)-carboxysomes) is based on an aggregation model; the enzymes are proposed to form a dense network connected through encapsulation peptides, recruiting shell proteins and pinching off holocomplexes as complete shells form\(^1\). The AAUM operon contains only two putative encapsulation peptides on the aldehyde dehydrogenase and the protein containing a domain of unknown function. Two potential models for the AAUM organization can be proposed based on current research: in one model the AAUM contains only the phosphopropanolamine
phospholyase with remaining enzymes in the bulk cytoplasm or closely associated with the microcompartment shell. An alternative model requires that the enzymes containing encapsulation peptides form heterocomplexes with the enzymes lacking encapsulation peptides allowing all required enzymes to reside within the microcompartment shell. Access to the required cofactors would be addressed by different methods in either model; in the first model cofactors are needed only in the bulk cytoplasm, in the second model rapid exchange of cofactors (NAD(P)H/NAD(P), ATP/ADP, CoA/propionyl-CoA) would be required to occur through large pores formed by BMC-T hexamers. Either model restricts diffusion of the cytotoxic aldehyde intermediate by encapsulation of the phosphopropanolamine phospholyase. The second model would require exchange of large cofactors across the shell contrary to the proposed selective diffusion shell function. Exchange of cofactor sized molecules has, however, been observed in recombinant microcompartment systems supporting the second model\textsuperscript{22}. 
Figure 6.1 Proposed metabolisms of the aminoacetone utilization operons.

Schematic representation of the *Mycobacterium smegmatis* MC² 155 aminoacetone utilization microcompartment and non-microcompartment aminoacetone utilization operons (A). Proposed aminoacetone utilization metabolic pathway for the microcompartment (B) and non-microcompartment (C) operons.
6.6 Non-microcompartment aminoacetone utilization metabolism

Chapters 3-5 present the characterized catalytic activities of the first three reactions proposed to be involved in a non-microcompartment associated aminoacetone degradation operon from *M. smegmatis*. Similar to the proposed AAUM metabolism, the characterized enzymes catalyzed aminopropanol dehydrogenase, aminopropanol kinase, and phosphopropionolamine phospholyase activities. The remaining enzymes from the operon, phosphate acetyl transferase (MSM0783) and propionate kinase (MSM0784), were assayed for their proposed activities and demonstrated catalytic activity with their proposed substrates (not presented here). Analogous to the AAUM metabolism, propanal is proposed to be formed by the consecutive reactions of the alcohol dehydrogenase, kinase, and phospholyase. In the non-microcompartment aminoacetone degradation pathway propanal is proposed to be oxidized by an endogenous aldehyde dehydrogenase for phosphorylation and transacylation producing a final propionyl-CoA product (Figure 6.1). Following the preliminary catalytic assays for MSM0783 and MSM0784, full kinetic parameters should be determined for their respective acetyl kinase and phosphotransacylase activities.

Presence of two aminoacetone utilization operons in *M. smegmatis* introduces a question as to the evolutionary advantage for retention of two pathways that seemingly use the same substrate to produce the same product. As proposed in chapter 3, a metabolic capacity for aldehyde production and consumption may be the underlying justification. Relative gene expression assayed by quantitative PCR using the signature phosphopropionolamine phospholyase of each operon as an indicator of operon expression should be evaluated with growth of *M. smegmatis* under conditions of varying aminoacetone concentration. This would determine if there are differences in expression of each operon based on substrate availability. Initial induction experiments with the AAUM operon from *R. erythropolis* indicated that
inclusion of aminopropanol in growth media alone induced expression of AAUM genes\textsuperscript{80}. The transcription regulators found in both the AAUM and non-microcompartment aminoacetone degradation operons are from the HutC subfamily of GntR family of regulators\textsuperscript{120}, sharing 43\% sequence similarity. Similar to other members of the HutC regulators the DNA binding domain and dimerization interface are conserved in both regulators, however the ligand binding pocket forming region of the sequences are varying in composition suggesting differences in ligand binding properties. Dissociation constants for aminoacetone or aminopropanol could be determined with either regulator, using an oligonucleotide carrying a consensus HutC operator sequence as a bait target in surface plasmon resonance.
REFERENCES


(47) Saraste, M., Sibbald, P. R., and Wittinghofer, A. (1990) The P-loop- a common motif in ATP- and GTP-


