REVEALING THE MOLECULAR DETAILS OF AZOLINE FORMATION IN RIBOSOMAL NATURAL PRODUCT BIOSYNTHESIS

BY

KYLE LAWRENCE DUNBAR

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemistry in the Graduate College of the University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Doctoral Committee

Assistant Professor Douglas A. Mitchell, Chair
Professor Wilfred A. van der Donk
Professor Satish K. Nair
Professor John A. Gerlt
ABSTRACT

The thiazole/oxazole-modified microcins (TOMMs) are a recently grouped class of ribosomally synthesized and posttranslationally modified peptides. Encoded by many bacteria and archaea, these natural products occupy a large chemical and functional space and are linked by the presence of the eponymous azole moieties. Previous work has demonstrated that an evolutionarily conserved, heterotrimeric enzyme complex (TOMM synthetase) is responsible for azole biogenesis. The TOMM synthetase transforms select serine, threonine and cysteine residues in the core peptide into azole heterocycles via a two-step cyclodehydration-dehydrogenation reaction. The installation of the azol(in)e heterocycles endows the peptide with rigidity and, in all characterized TOMM natural products, is required for biological activity. Despite being reconstituted nearly two decades ago, many questions remain regarding the role of each protein in the TOMM synthetase and the mechanism of cyclodehydration. Previous attempts to elucidate the molecular details of azole formation were impeded by the poor stability/solubility of reconstituted TOMM synthetases. In this thesis I report the characterization of the biosynthetic machinery from two novel TOMM cluster from *Bacillus* sp. Al Hakam that do not suffer from these issues. Using these robust synthetase complexes, I demonstrate that the cyclodehydration reaction proceeds via a phosphorylated hemi-orthoamide intermediate (chapter 2) and provide definitive roles from each of the proteins in the cyclodehydratase complex (chapter 4 and 5). My findings identify an unprecedented role for ATP and provide the first function to the previously uncharacterized YcaO superfamily of proteins. Furthermore, in order to complete my mechanistic investigations of the TOMM synthetase, I developed a novel strategy for generating peptides with site-specific carbonyl oxygen isotope labels (chapter 3). I also demonstrate that this strategy can be expanded to the site-specific installation of thioamides.
Ar scáth a chéile a mhaireann na daoine.
ACKNOWLEDGEMENTS

I would like to begin by thanking my advisor, Professor Doug Mitchell. You gave me the right amount of guidance to keep me on track but enough free reign to go after the questions that truly excited me. Your enthusiasm for research is contagious. It has been a pleasure working alongside you these last 5 years. I am grateful for my excellent collaborators, Satish Nair, Vinayak Agarwal, and Jonathan Chekan for their patience and persistence. I would also like to thank the Professors that have served on my committee over the course of my graduate career: Professor Wilfred van der Donk, Professor Satish Nair, Professor John Gerlt, Professor Eric Oldfield, Professor Yi Lu, and Professor Huimin Zhao.

During my time at UIUC I have been fortunate to work with fantastic colleagues. I would like to thank every member of the Mitchell lab for contributing to what has become a fantastic work environment. More specifically I would like to extend my gratitude to Joel Melby, Tucker Maxson, Spencer Peck and Tyler “Jones” Takeshita. I will never forget my lost dancing pants, the impassable fence, and all of the good times and rambunctious nights. I would never have survived graduate school were it not for your friendship.

Finally, I would like to thank my mother, my sister Stephanie, my entire extended family, and my “significant other” Evelyn. I cannot express how much your love and support has meant to me. Suffice it to say that none of this would have been possible without you. Thank you...
TABLE OF CONTENTS

LIST OF TABLES AND FIGURES ................................................................. x

CHAPTER 1: INTRODUCTION ....................................................................... 1
  1.1 Ribosomally synthesized posttranslationally modified peptides (RiPPs) .......... 1
  1.2 Thiazole-Oxazole Modified Microcins (TOMMs) ........................................... 5
  1.3 Biosynthesis of TOMMs ............................................................................. 7
    1.3.1 Azole Biosynthesis ............................................................................ 8
    1.3.2 Macrocyclization of TOMM Natural Products ..................................... 9
    1.3.3 Radical-Mediated Transformations in TOMM Biogenesis .................... 13
    1.3.4 Prenylation in Cyanobactin Biosynthesis ........................................ 18
    1.3.5 C-terminal Amidation in Thiopeptides .............................................. 21
    1.3.6 N-terminal Dimethylation in Plantazolicin Biosynthesis .................. 23
  1.4 Summary and Outlook ............................................................................. 24
  1.5 References ............................................................................................. 25

CHAPTER 2: TOMM CYCLODEHYDRATASES UTILIZE ATP FOR CARBONYL
  OXYGEN ACTIVATION ............................................................................. 31
  2.1 Introduction ........................................................................................... 31
  2.2 Minimal Requirements for Balh Cyclodehydratase Activity ....................... 34
  2.3 A Cyclizable Residue Stimulates Rapid ATP Hydrolysis .......................... 37
  2.4 ATP Hydrolysis is Tightly Coupled to Azoline Formation ....................... 39
  2.5 ATP is Hydrolyzed by a Non-Solvent Oxygen Nucleophile .................... 41
  2.6 TOMM Cyclodehydrations Occur via an Activated Carbonyl Intermediate... 43
  2.7 Summary and Outlook ............................................................................ 46
  2.8 Experimental ......................................................................................... 51
    2.8.1 Substitution of BcerB for BalhB ....................................................... 51
    2.8.2 Protein overexpression and purification ........................................... 51
    2.8.3 Purine nucleotide phosphorylase-based phosphate detection assay ...... 52
    2.8.4 Subtractive activity studies ............................................................... 53
    2.8.5 Substrate analog assays ................................................................. 53
    2.8.6 Stoichiometry of ATP utilization .................................................... 54
2.8.7 $^{31}$P-NMR analysis of phosphate isotope incorporation of $^{18}$O-water reactions ............................................................. 54
2.8.8 $^{31}$P-NMR analysis of phosphate isotope incorporation in $[^{18}O_2]$-BalhA1 reactions ................................................................. 55
2.8.9 LCMS detection of ring formation ............................................. 56
2.8.10 Malachite green phosphate detection assay ............................... 57
2.8.11 LCMS detection of $[^{16}O]$-ADP from $^{18}$O-water reactions .......... 57
2.8.13 Bioinformatics ........................................................................ 58

2.9 References .................................................................................... 58

CHAPTER 3: AZOLINE-MEDIATED PEPTIDE BACKBONE LABELING AS A TOOL FOR STUDYING TOMM BIOSYNTHESIS ................................................................. 62
3.1 Introduction .................................................................................... 63
3.2 Oxygen Isotope Incorporation into Peptide Backbones Through Azoline Hydrolysis ........................................................................ 67
3.3 Thiolysis of Oxazoline Heterocycles Allows for Site-Selective Thioamide Incorporation .................................................................................. 70
3.4 Leader Peptide-Independent Peptide Backbone Labeling .................. 74
3.5 Engineering of Dehydrogenase-Dependent Synthetases for Backbone Labeling ... 80
3.6 Azoline Localization Provides Insights into Cyclodehydratase Substrate Processing .................................................................................. 83
3.7 Azoline Stoichiometry of Natural Products Obtained via Backbone Labeling ...... 89
3.8 Summary and Outlook ..................................................................... 93
3.9 Experimental .................................................................................. 94
   3.9.1 General methods ...................................................................... 94
   3.9.2 Protein overexpression and purification ...................................... 94
   3.9.3 Overexpression and purification of Trx-BalhA1 ........................... 94
   3.9.4 BalhA1 $^{18}$O-labeling studies ................................................... 95
   3.9.5 FT-MS/MS localization of $^{18}$O labels ...................................... 96
   3.9.6 Processing of leader peptide-free substrates ............................... 97
3.9.7 Rate of BalhA1 core processing ................................................................. 97
3.9.8 Mcb synthetase reactions ................................................................. 98
3.9.9 Thiolysis of azoline heterocycles .................................................. 98
3.9.10 BalhX synthesis ........................................................................ 99
3.9.11 Dehydrogenase alignments ......................................................... 99
3.9.12 McbC mutagenesis ..................................................................... 99
3.9.13 $^{18}$O labeling of BalhA2 and McbC-K201A/Y202A reactions .......... 99
3.9.14 MS/MS localization thioamides and azol(in)e heterocycles .......... 100
3.9.15 Iodoacetamide labeling of McbA in McbC-K201A and -Y202A reactions .............................................................................. 100
3.9.16 $^{18}$O-labeling of TOMM natural products .................................. 101
3.9.17 BalhA1 core cloning ...................................................................... 101
3.10 References ......................................................................................... 101

CHAPTER 4: BIOCHEMICAL AND STRUCTURAL CHARACTERIZATION OF THE YCAO SUPERFAMILY ................................................................. 105
4.1 Introduction ......................................................................................... 106
4.2 The TOMM D Protein Alone is Sufficient for Azoline Formation .......... 110
4.3 TOMM YcaOs Utilize ATP for Amide Backbone Activation .................. 115
4.4 Non-TOMM YcaO Proteins Also Utilize ATP ........................................ 117
4.5 Structural Characterization of ATP-binding in Ec-YcaO ......................... 119
4.6 The Ec-YcaO ATP-Binding Site is Conserved in TOMM YcaOs ............ 124
4.7 The Ec-YcaO ATP-Binding Site is Vital for BalhD Cyclohydratase Activity ......................................................................................... 130
4.8 The Proline-Rich C-terminus of TOMM YcaOs Affects C Protein Binding and Catalysis .............................................................................. 138
4.9 Summary and Outlook ........................................................................ 151
4.10 Experimental ..................................................................................... 153
4.10.1 General methods ........................................................................ 153
4.10.2 Single-turnover activity studies .................................................... 154
4.10.3 Iodoacetamide labeling ................................................................ 154
4.10.4 Heterocycle localization via FT-MS/MS .............................................. 154
4.10.5 BalhD-only ATP/ring stoichiometry .................................................... 155
4.10.6 Malachite green phosphate detection assay ........................................ 155
4.10.7 $^{31}$P-NMR analysis of phosphate isotope incorporation of $^{18}$O-water reactions ........................................................................................................... 156
4.10.8 Cloning of MBP-YcaO ............................................................................ 156
4.10.9 Preparation of BalhD PxPxP mutants ..................................................... 157
4.10.10 Site-directed mutagenesis ..................................................................... 157
4.10.11 Overexpression and purification of MBP-tagged proteins ..................... 157
4.10.12 Multiple sequence alignments ............................................................... 157
4.10.13 Cytoscape sequence similarity network ............................................... 157
4.10.14 Maximum likelihood phylogenetic analysis .......................................... 157
4.10.15 ATP-binding site conservation ............................................................. 158
4.10.16 Proline-rich C-terminus analysis .......................................................... 158
4.10.17 Ec-YcaO crystallization and structural solution ...................................... 158
4.10.18 Endpoint heterocycle formation assays ............................................... 159
4.10.19 BalhD mutant kinetic studies ............................................................... 159
4.10.20 Size-Exclusion chromatography .......................................................... 159
4.10.21 Mutant BalhD IC$_{50}$ determination .................................................... 160
4.10.22 BalhC “$K_M$” for active BalhD mutants .............................................. 160
4.11 References .................................................................................................. 161

CHAPTER 5: THE ROLE OF TOMM E1-HOMOLOGS IN AZOLINE BIOSYNTHESIS ........................................................................................................... 165
5.1 Introduction ................................................................................................... 165
5.2 TOMM C Proteins Direct YcaO Cyclodehydratase Activity ....................... 168
5.3 TOMM C Proteins are Responsible for Leader Peptide Recognition ............ 170
5.4 TOMM C Proteins Potentiate YcaO Cyclodehydratase Activity ................. 171
5.5 The Role of Auxiliary E1-Homologs in Azoline Biogenesis ......................... 173
5.6 Summary and Outlook .................................................................................. 179
5.7 Experimental ............................................................................................... 182
5.7.1 General methods ................................................................. 182
5.7.2 Balh cluster 2 cloning ............................................................. 182
5.7.3 Protein overexpression and purification ...................................... 182
5.7.4 $^{18}$O labeling of BalhD-treated BalhA1 reactions ..................... 182
5.7.5 BalhC leader peptide recognition ............................................. 183
5.7.6 BalhD-A4LEH6 activity .......................................................... 183
5.7.7 Balh cluster 2 overnight heterocyclization assays ...................... 184
5.7.8 Size-exclusion chromatography .............................................. 184
5.8 References ........................................................................ 184

APPENDIX A: LIST OF PRIMERS ................................................................ 187

APPENDIX B: ADDITIONAL BALH CYCLODEHYDRATASE MUTANTS ......... 192
B.1 Unpublished BalhC Mutants ....................................................... 192
B.2 Unpublished BalhD Mutants ....................................................... 193

APPENDIX C: YCAO PROTEINS FOR CRYSTALLOGRAPHY SCREEN .......... 195
C.1 YcaO proteins screened at UIUC ............................................... 195
C.2 YcaO proteins screened by Steve Almo (Albert Einstein College of Medicine) 196
LIST OF FIGURES AND TABLES

Figure 1.1: RiPP natural product biosynthesis and diversity ........................................................... 4
Figure 1.2: TOMM Natural Product Diversity .................................................................................. 6
Figure 1.3: TOMM Azole Biogenesis .............................................................................................. 8
Figure 1.4: [4+2] cycloaddition in thiopeptide biosynthesis ............................................................ 12
Figure 1.5: The mechanism of MIA formation in nosiheptide biosynthesis ................................. 17
Figure 1.6: Cyanobactin prenylation .............................................................................................. 20
Figure 1.7: C-terminal amidation of thiopeptides ......................................................................... 22
Figure 2.1: TOMM biosynthetic gene clusters and possible mechanisms of ATP utilization
during azole formation ................................................................................................................. 33
Table 2.1: Substrate kinetic parameters for Balh synthetase ........................................................... 35
Figure 2.2: Minimal requirement for cyclodehydratase activity and ATP hydrolysis product
of the Balh synthetase .................................................................................................................... 36
Figure 2.3: BalhA1-NC inhibition and BalhNC-C40 processing ...................................................... 38
Figure 2.4: ATP/azole stoichiometry for substrate processing ......................................................... 40
Figure 2.5: ATP is utilized to directly activate the amide backbone of the substrate ................. 42
Figure 2.6: Isotope composition of ADP produced in [18O]-water ................................................ 42
Figure 2.7: 31P-NMR of [18O5]-BalhA1 reactions ........................................................................... 45
Figure 2.8: TOMM azoline installation is reminiscent of intein-mediated protein splicing .... 48
Figure 2.9: The proposed enzymatic mechanism of PurM and ThiI .............................................. 49
Figure 2.10: Alternative mechanism of azoline formation ............................................................ 49
Figure 2.11: Strategies for azole biosynthesis outside of TOMMs ............................................... 50
Figure 3.1: Azolines in natural products and AMPL overview ..................................................... 66
Figure 3.2: Peptide substrate sequences ....................................................................................... 68
Figure 3.3: Site-specific incorporation of 18O labels into BalhA1 .................................................. 68
Figure 3.4: 18O labeling of BalhA1 is dependent on azoline heterocycles .................................. 69
Figure 3.5: Thiolysis is specific for oxazoline heterocycles .......................................................... 72
Figure 3.6: Potassium hydrosulfide (KHS) treatment of unmodified BalhNC-A40T ................... 73
Figure 3.7: BalhX thiolysis reaction and localization ..................................................................... 73
Figure 3.8: BalhA1_core processing ................................................................. 77
Figure 3.9: Azol(in)e localization on BCD-treated BalhA1_core ................. 78
Figure 3.10: McbA and SagX processing by a non-cognate BCD complex ..... 78
Figure 3.11: Leader peptide-free processing follows previously established rules 79
Figure 3.12: TOMM dehydrogenase alignments ........................................ 81
Figure 3.13: Stability and FMN loading of McbC-K201A and -Y202A ............ 82
Figure 3.14: Azoline formation by the microcin B17 synthetase .................. 82
Figure 3.15: Azoline localization on BalhA2 via 18O labeling ......................... 85
Figure 3.16: Iodoacetamide labeling of McbA following treatment with McbC variants ...... 86
Figure 3.17: Azoline localization on McbA following treatment with McbC-K201A 87
Figure 3.18: Azoline localization on McbA following treatment with McbC-Y201A 88
Figure 3.19: 18O-labeling of plantazolicin ..................................................... 90
Figure 3.20: 18O-labeling of ulithiacyclamide .............................................. 91
Figure 3.21: 18O-labeling of lissoclinamide 4 ............................................... 92
Figure 4.1: YcaO gene clusters and sequence of precursor peptides used in this study ........ 108
Figure 4.2: The biosynthetic gene clusters for trifolitoxin and bottromycin do not contain a recognizable TOMM C protein ........................................................................ 109
Figure 4.3: BalhD is an ATP-dependent cyclodehydratase .............................. 111
Figure 4.4: MS/MS localization of the BalhD installed dehydration product ..................................................... 112
Figure 4.5: BalhC potentiates the BalhD ATPase activity .................................. 113
Figure 4.6: BalhD ATP/ring stoichiometry is dysregulated in the absence of BalhC ........ 114
Figure 4.7: TOMM YcaOs activate carbonyl oxygens with ATP ..................... 116
Figure 4.8: Ec-YcaO hydrolyzes ATP to AMP and pyrophosphate .................. 118
Figure 4.9: Structural homology between Ec-YcaO and TruD and structure of AMP-bound Ec-YcaO ...................................................................................... 121
Figure 4.10: Structure of AMP-bound Ec-YcaO ............................................. 122
Figure 4.11: ATP coordination in Ec-YcaO ..................................................... 123
Figure 4.12: Cytoscape sequence similarity network for the YcaO superfamily .......... 126
Figure 4.13: Conservation of the Ec-YcaO ATP-binding residues in the superfamily .... 127
Figure 4.14: The canonical E1 domain ATP-binding site is not conserved in TOMM C proteins ........................................................................................................ 128
Figure 4.15: The ATP-binding pocket is conserved in characterized TOMM YcaOs ............ 129
Table 4.1: Mutations to the ATP-binding pocket of BalhD decrease cyclodehydratase activity
...................... 132
Figure 4.16: Mapping ATP-binding residues from Ec-YcaO to BalhD ......................... 133
Figure 4.17: Mutations to the BalhD ATP-binding site affect heterocycle formation ........ 134
Figure 4.18: Mutations to the ATP-binding site that decrease cyclodehydratase activity also affect BalhD-only activity .................................................. 135
Figure 4.19: The $K_M$ for ATP does not depend on the concentration of BalhA1 ............ 136
Figure 4.20: Mutations of the Mg$^{2+}$-binding residues prevent ATPase potentiation by BalhC ................................................................................................. 136
Figure 4.21: Effect of the ATP-binding site mutations on BalhC binding ...................... 137
Table 4.2: Mutations to the C-terminus of BalhD disrupt catalysis ............................ 140
Figure 4.22: The C-terminal PxPxP motif conserved in TOMM YcaOs is near the ATP-binding site in TruD .................................................. 141
Figure 4.23: The PxPxP motif of BalhD is critical for cyclodehydratase activity .......... 142
Figure 4.24: Mutations that relocate the C-terminus of BalhD prevent ATPase potentiation by BalhC ................................................................................................. 143
Figure 4.25: C-terminal truncations of BalhD affect BalhC binding ......................... 144
Figure 4.26: Removal of the C-terminal residue of BalhD dysregulates the order of azole formation ................................................................................................. 145
Figure 4.27: Mutagenesis of the C-terminus of BalhD affects cyclodehydratase activity .... 146
Figure 4.28: The sequence of the C-terminus of BalhD is important for binding BalhC ...... 147
Figure 4.29: BalhD P429G displays an altered order of heterocycle formation ............. 148
Figure 4.30: Mutations to the C-terminus of BalhD also affect BalhC-independent catalysis ................................................................................................. 149
Figure 4.31: The conserved TOMM YcaO C-terminus is critical for activity in the microcin B17 cyclodehydratase ................................................................................................. 150
Figure 5.1: Azoline localization on BalhD-treated BalhA1 via $^{18}$O-labeling .................. 169
Figure 5.2: BalhC is responsible for leader peptide recognition .................................. 170
Figure 5.3: A C-terminal His$_6$ tag on BalhD abolishes cyclodehydratase activity .......... 172
Figure 5.4: BalhC modulates ATP binding and hydrolysis by BalhD ......................... 172
Figure 5.5: Overview of the Balh TOMM biosynthetic clusters ........................................ 175
Figure 5.6: Heterocycle formation on Balh2A is Balh2F and Balh2CD dependent ............ 176
Figure 5.7: The cyclized Balh2A product contains 13 thiazolines ................................ 177
Figure 5.8: Balh2F is responsible for leader peptide recognition in Balh cluster 2 .......... 178
Figure 5.9: An updated model for azoline formation by TOMM cyclodehydratases ...... 181
Table A.1: Primers for chapter 3 constructs ................................................................. 187
Table A.2: Primers for chapter 4 constructs ................................................................. 187
Table A.3: Primers for YcaO constructs cloned for crystallography screens .............. 188
Table A.4: Primers for chapter 5 constructs ................................................................. 189
Table A.5: Primers for unpublished cyclodehydratase mutants ................................. 190
Table B.1: Mutant BalhC activity .................................................................................. 192
Table B.2: Mutant BalhD activity .................................................................................. 193
Table B.3: Mutant BalhD BalhA1 and ATP kinetics ..................................................... 194
Table C.1: YcaOs used for crystallography screens ..................................................... 195
Table C.2: YcaOs screened by Steve Almo ................................................................. 196
CHAPTER 1: INTRODUCTION

This chapter was taken in part from Dunbar & Mitchell\textsuperscript{1} and Dunbar Chekan \textit{et al.}\textsuperscript{2} All material is reproduced with permission from the American Chemical Society and Nature Publishing Group, respectively.

I am grateful to Courtney Cox for generating the YcaO Cytoscape network.

1.1 Ribosomally Synthesized Post-Translationally Modified Peptides (RiPPs)

How does one measure the impact that natural products have had on human health and science as a whole? Natural products, or the derivatives thereof, encompass the lion’s share of drugs in therapeutic use.\textsuperscript{3-5} The exquisite complexity of these compounds has served as a source of inspiration for organic chemists for over a century, leading to the development of reaction methodologies too numerous to list.\textsuperscript{5,6} Synthetic organic chemistry as we know it today would likely not exist had the study of natural products been neglected. While the impact on human therapeutics and synthesis has been immense, natural products research also enabled the study of fundamental biological processes. For example, research into the mode of action of antibiotics continues to provide insight into bacterial biochemistry and physiology,\textsuperscript{7-9} while compounds such as phalloidin (filamentous actin stabilizer), rapamycin (mTOR inhibitor), and tunicamycin (\textit{N}-glycosylation inhibitor) have become invaluable tools for probing eukaryotic cell biology.\textsuperscript{10-12} Furthermore, the study of natural products and their biosynthesis has revolutionized the way we think about the chemistry of biological systems. With the vast majority of Nature’s biosynthetic potential remaining to be discovered,\textsuperscript{13,14} the study of natural products will likely continue to inspire future generations of scientists.

A class of natural products receiving significant recent attention is the ribosomally synthesized posttranslationally modified peptides (RiPPs).\textsuperscript{15,16} Comprised of the lanthipeptides, microcins and the thiazole/oxazole-modified microcins (TOMMs) among others, RiPPs occupy a
large chemical, genetic, and functional space, but remain linked by a common mechanism of biosynthesis. In all cases, a ribosomally synthesized precursor peptide undergoes modification by a set of tailoring enzymes usually found in the local genomic region (i.e. within the biosynthetic gene cluster, Figure 1.1). These modifications include, but are not limited to, varied cyclizations, dehydrations, and rearrangements, all of which are responsible for endowing the peptide with a rigidified structure, biological activity, and improved protease stability. The tailoring enzymes encoded in the biosynthetic gene cluster govern the posttranslational modifications received and, in some cases, the structure of the final product can be accurately predicted.\textsuperscript{17-19} In most cases, modification is followed by the proteolytic removal of an $N$-terminal leader peptide (or, less frequently, a $C$-terminal follow peptide) required for recognition by the tailoring enzymes, and the mature compound is exported.\textsuperscript{20} Using this simple strategy, a varied compound library can be constructed from a minimal amount of genetic space. In fact, the investigation of RiPPs has overturned the longstanding paradigm that large genomes are required for an organism to produce architecturally complex natural products.\textsuperscript{21,22}

Adding to the allure of this class of natural products is the gene-encoded nature of the precursor peptide coupled with the often high level of promiscuity of the biosynthetic enzymes, allowing for the facile generation of unnatural compound derivatives.\textsuperscript{15,19,23-25} This approach has garnered a great deal of excitement due to the potential to develop therapeutically relevant derivatives of this underutilized compound class.\textsuperscript{26,27} Although the basis for the high level of promiscuity is not fully understood, the use of an $N$-terminal leader peptide is thought to play a role. By spatially separating peptide recognition from the residues that will be modified the enzymes have the ability to process a substrate multiple times at diverse positions using a single recognition strategy.\textsuperscript{20}
The profound interest in these natural products has fueled extensive studies into biosynthetic mechanisms. Recent successes in the study of RiPP biosynthetic enzymes have greatly expanded our view of Nature’s synthetic organic potential. In addition to the discovery of novel biological chemistry, these studies provide a window through which the evolutionary adaptation of enzymes can be observed. However, there are many tailoring enzymes that have yet to be characterized and many posttranslational modifications whose biosynthesis remains enigmatic. For additional information regarding the advances in the field of non-TOMM, RiPP biosynthesis I direct the reader towards recent reviews.16,28
Figure 1.1 | RiPP natural product biosynthesis and diversity. The general strategy for RiPP biosynthesis is illustrated in the generic gene cluster and schematic below. The functional assignment for each of the open reading frames is displayed. An array of ribosomal natural products generated from this strategy is displayed along with their RiPP subclass and characterized bioactivity or molecular target, when known. Post-translational modifications are colored red. This figure was adapted with permission from Dunbar & Mitchell.¹
1.2 Thiazole-Oxazole Modified Microcins (TOMMs)

Thiostrepton was the first azole/azoline containing RiPP natural product to be structurally characterized.\(^{29}\) Although the structures of other thiazol(in)e and oxazol(in)e-containing natural products were solved over the following 25 years, it was not until the mid-1990s that the first insights into azole biosynthesis were revealed with the study of microcin B17.\(^{30}\) Following this discovery, just two additional biosynthetic gene clusters of this type were identified over the next decade (streptolysin S\(^{31}\) and patellamide A/C\(^{32}\)), and the unification of the TOMM family as a whole was only recently realized.\(^{22}\)

TOMMs are comprised of the thiopeptides, azole-containing cyanobactins, linear azole-containing peptides (LAPs), trifolitoxins, and bottromycins (Fig. 1.1 and 1.2).\(^{16}\) As implied by the name, all members of the TOMM subclass of RiPPs contain Cys-, Ser-, or Thr-derived azol(in)e heterocycles. The installation of these heterocycles onto a precursor peptide endows the molecule with conformational rigidity and, in all cases examined to date, is absolutely required for biological activity.\(^{16,23}\) Consistent with the structural diversity observed in characterized family members, TOMMs display diverse biological functions including, but not limited to, acting as antibacterial and antiviral compounds, antitumor agents, and cytolytic virulence factors.\(^{16,33}\)

Currently well over 400 TOMMs have been structurally and/or functionally characterized;\(^{16}\) however, there are still approximately 700 bioinformatically identifiable gene clusters in GenBank with no discernable structure or function (Fig. 1.2).\(^{2}\) These uncharacterized clusters are found in both Archaea and Bacteria and the characterization of such clusters will undoubtedly expand the structural and functional diversity of this important class of natural products.
**Figure 1.2 | TOMM Natural Product Diversity.** A Cytoscape sequence similarity network of the TOMM D protein (discussed below) is displayed. Each node represents a unique TOMM D protein, and lines are drawn between nodes if the proteins share at least BLAST e-value of at least $e^{-80}$. Nodes are colored if the structure or function of the natural product is known or can be inferred from bioinformatics. Select structures of characterized TOMM natural products are displayed. The biological activity is shown in parenthesis followed by the associated molecular target, if the mechanism of action is known. All post-translational modifications are shown in red, while the blue serine in the microcin B17 structure marks the site of the 9th azole that is found as a minor product during biosynthesis. Ab, denotes antibacterial activity. This figure was adapted with permission from Dunbar & Chekan *et al.*
1.3 Biosynthesis of TOMMs

The first genetic evidence that TOMM natural products were of ribosomal origin was provided with the characterization of the trifolitoxin and microcin B17 biosynthetic clusters in 1993 and 1994, respectively.\textsuperscript{34,35} Shortly after these initial reports, the enzymes responsible for azole formation during microcin B17 biosynthesis were purified and characterized.\textsuperscript{30} The extraordinary studies performed on the microcin B17 synthetase provided the framework for all subsequent work on TOMM biosynthesis. Following these seminal studies, the TOMM biosynthetic clusters for natural products that contain auxiliary post-translational modifications have been discovered.\textsuperscript{23} Recently, many of auxiliary tailoring enzymes responsible for the maturation of diverse TOMMs have been reconstituted and studied.\textsuperscript{1} Although all TOMM tailoring enzymes were originally hypothesized to require an \textit{N}-terminal leader or \textit{C}-terminal follower peptide for substrate recognition,\textsuperscript{1,16,36-39} exceptions to this strategy for peptide substrate binding have emerged.\textsuperscript{40-43} The following sections will describe our current understanding of the biosynthesis of TOMM natural products. While the modifications discussed below provide a flavor of the diversity of biological chemistry involved in TOMM biogenesis, there are many tailoring events that are completely uncharacterized. As such, TOMM biosynthesis is likely to remain a fruitful area of research for years to come.
1.3.1 Azole Biosynthesis

A heterotrimeric protein complex, referred to as the TOMM synthetase, is required to install the characteristic azole heterocycles on a precursor peptide.\textsuperscript{30,32,44} Studies performed on the TOMM enzymatic machinery involved in microcin B17 and azol(in)e-containing cyanobactin biosynthesis have demonstrated that azole heterocycles are installed through two steps: a cyclodehydration to generate an azoline heterocycle and a subsequent flavin mononucleotide (FMN)-dependent dehydrogenation to afford the aromatic azole (Fig. 1.3).\textsuperscript{30,44,45} While the enzymatic complex responsible for these transformations has been partially characterized, the dehydrogenase (B-protein) is the only enzyme with a definitive function.\textsuperscript{46} Dissection of the biochemical functions of the C and D proteins has been stymied by the inability to obtain individual activities for either protein; however, cyclodehydratase activity has been demonstrated for a complex of the two proteins\textsuperscript{32,45} and has been shown to require ATP hydrolysis.\textsuperscript{47,48} Underscoring this linked function, roughly half of known TOMM clusters express C and D as a single polypeptide (CD fusion).\textsuperscript{23,32} Although significant progress has been made towards understanding the rules that govern substrate promiscuity, the mechanism of azoline formation and the role of ATP in the reaction is not currently understood.\textsuperscript{23}

![Diagram of TOMM Azole Biogenesis](image)

**Figure 1.3 | TOMM Azole Biogenesis.** Azole heterocycles in TOMM natural products are installed by the successive action of a cyclodehydratase (C and D proteins) and a flavin mononucleotide (FMN)-dependent dehydrogenase (B protein). The cumulative mass change for each step is shown below the modification.
1.3.2 Macrocyclization of TOMM Natural Products

Multiple subclasses of TOMM natural products are macrocyclized (Fig. 1.1 and Fig. 1.2). Although the mechanisms for macrolactam and macroamidine formation in bottromycin and YM-216391 remain uncharacterized, the mechanism of macrocyclization in cyanobactin and thiopeptide biosynthesis has been partially elucidated.

Cyanobactin biosynthesis begins with the installation of azole(in)e heterocycles by PatD, a TOMM synthetase. Next, a subtilisin-like protease, PatA, removes the leader peptide and separates the heterocyclized patellamide A/C peptides (patellamide precursor peptides often contain multiple core peptides separated by protease cleavage sites), each bearing an AYDG sequence at their C-termini. This tetrapeptide motif serves as a recognition sequence for another subtilisin-like protease, PatG. *In vitro* studies demonstrated that PatG removes the C-terminal AYDG tag and facilitates cyanobactin N-to-C macrocyclization. Given the homology to subtilisin, a mechanism was proposed where the acyl-enzyme intermediate was resolved with the N-terminus, rather than water, to generate the macro lactam.

Recently, the crystal structure of the central PatG macrocyclase domain, PatGmac, was solved. As expected, PatGmac adopts the subtilisin-like fold. However, PatG contains an additional loop positioned over the active site. The PatG-type protease/macro cyclase proteins all contain a 30–50 amino acid insertion between two of the three residues involved in subtilisin catalytic triad, providing a bioinformatic handle for identifying cyanobactin macrocyclases. An analysis of PatGmac co-crystallized with a peptide substrate demonstrated that this insertion was responsible for shielding the active site from water and also binding of the AYDG sequence. Using both site-directed mutagenesis and selective truncations of this insertion sequence, it was determined that any disruption of AYDG recognition or water shielding resulted in a loss of
macrocyclase, but not protease, activity. Combined with the observation that the acyl-enzyme intermediate is remarkably long lived, a mechanism was proposed in which the insertion loop prevents AYDG peptide dissociation after cleavage until the N-terminus displaces it and resolves the covalent intermediate.

The PatG-catalyzed transformation is extraordinary for three reasons: (1) unlike the thioesterase domains in polyketide and non-ribosomal peptide synthetases, PatG macrocyclization occurs on an unactivated substrate without an external energy source, (2) the prime side specificity (recognition is C-terminal to the scissile bond) of PatG facilitates the processing of hypervariable substrates with a minimal sequence tag is in stark contrast to other subtilisin-like proteases and (3) the strong sequence similarity to subtilisin proteases suggests that these proteins have been naturally evolved to catalyze a variety of chemical transformations.\(^{54}\)

While cyanobactins are head-to-tail macrolactams, the central macrocycle found in all thiopeptides is formed through a cycloaddition processes that forms two carbon-carbon linkages.\(^{56}\) This macrocyclization forms the ubiquitous thiopeptide (dehydro)pyridine ring, which is critical for target recognition.\(^{57}\) Recent investigations into thiocillin biosynthesis have demonstrated that the pyridine ring is assembled via the TclM-catalyzed trans-annular heteroannulation of two dehydroalanine residues (Fig. 1.4).\(^{58,59}\) This formal [4+2] cycloaddition has received a great deal of attention, as it has been postulated to proceed through an aza-Diels-Alder mechanism. Although the Diels-Alder is one of the most widely employed reactions in organic synthesis, no naturally occurring catalyst has been definitively proven to carry out a concerted [4+2] cycloaddition.\(^{60}\) Theoretically, \(^{13}\)C primary kinetic isotope effects (KIEs) could be used to determine if the TclM-catalyzed cycloaddition occurs via a stepwise or concerted
To perform this experiment, the two differentially (dual) labeled substrates shown in Figure 1.4 would be required. If the mechanism were concerted, a primary $^{13}$C KIE would be observed with both substrates; however, if cycloaddition occurred via a stepwise mechanism, a KIE would only be observed for one of the dual-labeled substrates. While this experiment may seem relatively straightforward, in practice it will be difficult. TclM has never been reconstituted in vitro, the heterocyclized and dehydrated substrate is not readily accessible, and the expected primary $^{13}$C KIEs require a highly sensitive kinetic assay to measure. A more tractable alternative for the discovery of a bona fide Diels-Alderase may be to examine the enzyme responsible for the [4+2] cycloaddition in spinosin A biosynthesis. While SpnF is not responsible for producing a RiPP, it has been reconstituted in vitro, the reaction allows for the use of $^2$H secondary KIEs (which are typically 30% larger than $^{13}$C primary KIEs), and the substrate is more amenable to chemical synthesis.
Figure 1.4 | [4+2] cycloaddition in thiopeptide biosynthesis. Two potential mechanisms for installing the pyridine moiety of thiocillin I (red) are displayed. Starting from the tautomer form, the orange arrows demonstrate a concerted Diels-Alder mechanism, while the purple arrows follow a stepwise, polar mechanism. The shared step of each mechanism is displayed as a dashed black arrow. After the [4+2] cycloaddition is complete, the elimination of water and the leader sequence (LP) affords the central pyridine ring. The colored circles on the tautomer form of the uncyclized intermediate indicate positions where dual $^{13}$C-labeling could distinguish between stepwise and concerted cyclization mechanisms.
1.3.3 Radical-Mediated Transformations in TOMM Biogenesis

While the presence of the core macrocycle is seen in all thiopetides, select family members also contain a second macrocycle.\(^5^6\) This ring is not formed from the gene-encoded peptide sequence but rather occurs through the posttranslational attachment of a modified tryptophan.\(^6^3^-^6^7\) Currently, two types of this auxiliary macrocycle have been identified. In thiostrepton, the macrocycle is comprised of a L-tryptophan-derived quinaldic acid moiety, whose biosynthesis occurs via a complex and poorly understood pathway and thus will not be discussed further.\(^6^3^,^6^5^,^6^8\) In contrast, nosiheptide contains an indolic acid derivative, which is assembled via the action of two rSAM enzymes, NosL and NosN (Fig. 1.5A).\(^6^4\) Through genetic studies, it was discovered that deletion of nosL abolished nosiheptide production, but that supplementation of 3-methyl-2-indolic acid (MIA) rescued product formation. Moreover, a nosN deletion formed a nosiheptide derivative with MIA attached via a single thioester linkage rather than the expected 3,4-dimethyl indolic acid derivative. These results were used to putatively assign NosL and NosN as the enzymes involved in the rearrangement of L-tryptophan to MIA and the methylation of the unactivated C4 position, respectively.\(^6^4\)

To better understand the transformation of L-tryptophan to MIA, NosL was recently heterologously expressed and investigated in vitro.\(^6^9\) Support for assigning NosL as the sole protein responsible for MIA formation came initially from the detection of MIA in NosL-expressing E. coli. Further characterization demonstrated that NosL reductively cleaved SAM and used the 5'-Ado radical to form MIA. Reactions conducted with [1-\(^1^3\)C]- and [3-\(^1^3\)C]-labeled L-tryptophan showed that both carbons were retained during the transformation, while reactions conducted with [\(^2\)H\(_8\)]-L-tryptophan established that initial hydrogen abstraction does not occur from any of the C-H bonds in tryptophan. Fortuitously, a detailed analysis of the NosL product
profile identified four side products: 3-methylindole, glyoxate, glycine and formaldehyde. This led to the mechanistic proposal presented in Figure 1.5B. Based on the \[^{1}H_8\]-L-tryptophan experiment, the 5’-Ado radical is proposed to abstract the N-H hydrogen to generate a stabilized indole radical. This radical then fragments to form 3-methylene indole and a glycine radical, derivatives of which were identified as shunt products. Subsequent radical attack on the C2 position of the indole ring, and decomposition of the glycine subunit to form formaldehyde and ammonium, affords MIA. A subsequent report on NocL, a NosL homolog in nocathiacin MIA biogenesis, directly detected the glycine radical by EPR, strongly supporting this mechanism.\(^{70}\)

Prior to the study of NosL, no member of the rSAM superfamily had been shown to catalyze both the fragmentation and rearrangement of a substrate.\(^{71}\) While this chemistry is noteworthy, glutamate mutase, a cobalamin-dependent enzyme, catalyzes a similar fragmentation-rearrangement reaction.\(^{72}\) Glutamate mutase uses a cobalamin-generated 5’-Ado radical to perform C-H abstraction. Akin to the NosL mechanism, the resulting radical decomposes into a glycine radical and acrylate. In a final step, these recombine to afford 2-methylaspartate. However, unlike glutamate mutase, NosL catalyzes a fragmentation-recombination reaction with elimination of a portion of the molecule during the transformation.

In addition to the radical-based rearrangements that form the side-rings of thiopeptides, various positions within members of this TOMM subclass are methylated, despite being unactivated and non-nucleophilic.\(^{64,73}\) Recent efforts have identified similar modifications in other RiPP biosynthetic clusters, including the bottromycins and polytheonamides.\(^{50-53,74}\) In certain cases, labeling studies indicated that SAM was the source of these methyl groups;\(^{75,76}\) however, the unactivated nature of the modified position suggested that a radical mechanism would be necessary. Accordingly, recent studies conducted on two rRNA methyltransferases,
RlmN and Cfr, provided the first definitive evidence that rSAM enzymes can perform methylation reactions. These proteins have become the founding members of the radical SAM methyltransferase (RSMT) protein family, of which there are three bioinformatically-identifiable classes. Class A is solely comprised of RlmN and Cfr, contains only a rSAM domain, and is the best characterized class to date. However, as these enzymes are involved in ribonucleotide methylation and not RiPP biosynthesis, they will not be discussed further. Rather, focus will be placed on class B and class C RSMTs because analyses of several RiPP biosynthetic clusters demonstrate that in all cases where the natural product is methylated at an unactivated position, a RSMT of one of these classes is present.

In addition to the canonical rSAM domain, class B RSMTs contain an N-terminal cobalamin-binding domain. Recently, the first class B RSMT to be characterized, TsrM which is involved in thiostrepton biosynthesis, was recently reconstituted in vitro and exhibited a novel strategy for methylation. TsrM catalyzed the SAM and cobalamin-dependent methylation of the C2 position of tryptophan, which has been predicted to be the precursor for the quinaldic acid moiety. Unlike all other rSAM homologs, TsrM did not carry out the reductive cleavage of SAM despite containing a CxxxCxxC ligated Fe-S cluster. Instead, SAM was used as a methyl source for the in situ formation of a methylcobalamin cofactor, which then serves as the methyl source for tryptophan methylation. Although the mechanism of methyltransfer from the methylcobalamin cofactor to tryptophan remains enigmatic, preliminary data suggests that the Fe-S cluster may play a pivotal role as a redox cofactor in the transformation. While this initial study of TsrM is interesting from a mechanistic point of view, it also has larger implications for the classification of the rSAM superfamily. The protein contains the conserved Fe-S cluster ligation motif and is SAM dependent, but does not catalyze the reductive cleavage of the
sulfonium moiety, a previously ubiquitous feature of the enzyme family. This prompts the question as to whether TsrM is a true member of the rSAM superfamily. Moreover, the characterization of additional class B RSMTs will be required to determine if this unusual utilization of SAM by a rSAM domain is a conserved strategy for methylation of unactivated positions.

To date, no member of the class C RSMTs has been reconstituted. These proteins share significant similarity to the rSAM protein coproporphyrinogen III oxidase (HemN), which is responsible for the oxidative decarboxylation of coproporphyrinogen III to form protoporphyrinogen IX in heme biogenesis. Although Class C RSMTs contain both a HemN-like rSAM domain and a domain similar to the C-terminus of HemN, they lack the N-terminal “trip-wire” domain of HemN. Studies on HemN have demonstrated that the rSAM TIM barrel core binds two molecules of SAM and a comparison to class C RSMTs hints that these enzymes may also simultaneously bind two SAM molecules.
**Figure 1.5 | The mechanism of MIA formation in nosiheptide biosynthesis.**

(a) The structure of nosiheptide is displayed with the indolic acid ring highlighted in red. The uncharacterized radical SAM methyltransferase, NosA, and an as of yet unidentified protein methylate and oxidize MIA to afford the second attachment point (blue moiety). (b) The proposed mechanism of indolic acid formation by the radical SAM protein NosL is shown. The structures of the intermediates/products enclosed in the dotted circles were detected in reaction mixtures. The colored asterisks denote $^{13}$C-labeled positions that were used to demonstrate the fate of the carbon backbone during the rearrangement.
1.3.4 Prenylation in Cyanobactin Biosynthesis

Amongst other modifications, select cyanobactins are further decorated by O-prenylation. Deriving from dimethylallyl pyrophosphate (DMAPP), this modification is observed at Ser, Thr and Tyr residues. Ser and Thr prenylation occurs through the C3 position of DMAPP (reverse prenylation; Fig. 1.6A). Conversely Tyr prenylation appears to occur exclusively through the C1 position (forward prenylation; Fig. 1.6A). This modification is notable, as the only known O-prenylated RiPPs are cyanobactins. Based on comparative genomics, TruF and its homologs were tentatively assigned as O-prenyltransferases. However, this assignment was complicated by the fact that TruF-like proteins are not homologous to known prenyltransferases and are essential in the biosynthesis of non-prenylated cyanobactins as well.

In 2011, the first insights into cyanobactin prenylation emerged with the in vitro reconstitution of a prenyltransferase from *Lyngbya aestuarii*, LynF. While the product of the Lyn cluster remains uncharacterized, initial activity screens with various peptides substrates confirmed that LynF was a tryrosine prenyltransferase. Unexpectedly, LynF displayed a strong selectivity for macrocyclized substrates over linear peptides. Moreover, LynF was not able to prenylate substrates with leader peptides regardless of whether they contained azol(in)ene heterocycles. Considering that heterocyclization occurs prior to macrocyclization, this suggested that prenylation is the final step in cyanobactin biosynthesis. Further analysis of LynF’s substrate promiscuity demonstrated that the protein displayed relaxed sequence specificity, but a definitive set of modification rules could not be established. This prompts the question as to how substrate recognition is achieved in vivo given that LynF does not utilize the standard leader peptide recognition strategy used by most RiPP tailoring enzymes. Reaction products were analyzed by 2D-NMR to definitively establish the site of Tyr prenylation. Surprisingly, the spectra showed
that LynF catalyzed forward C-prenylation ortho to the phenolic oxygen, rather than O-prenylation. Similar transformations have been characterized in other biosynthetic pathways, and are proposed to proceed via an electrophilic aromatic substitution mechanism.\textsuperscript{83} To evaluate the mechanism of LynF prenylation, reaction products were analyzed with multiple spectroscopic methods over a reaction time course. The results demonstrated that the substrate initially undergoes reverse $O$-prenylation, but slowly converts to the forward $C$-prenylated phenol, most likely through a Claisen rearrangement (Fig. 1.6B).

Although highly similar reaction schemes have been synthetically employed,\textsuperscript{84,85} LynF represents the first instance of enzymatic $C$-prenylation via a reverse $O$-prenylation/Claisen rearrangement route. Additional experiments concluded that this rearrangement occurs at a slow but measurable rate (0.23 h\textsuperscript{-1}) in the absence of enzyme, demonstrating that reverse $O$-prenylated phenols are prone to spontaneous rearrangement.\textsuperscript{41} This result provides a possible explanation as to why reverse $O$-prenylated phenol rings are not seen in natural products while reverse $O$-prenylated hydroxyl moieties are common. It remains uncertain whether the rearrangement is truly uncatalyzed \textit{in vivo} or if the product of the Lyn cluster is actually $C$-prenylated. Regardless, LynF remains the only RiPP prenyltransferase characterized to date and demonstrates that $C$-prenylation in natural product biosynthesis can be achieved via a Claisen rearrangement, as was predicted over 40 years ago.\textsuperscript{86}
Figure 1.6 | Cyanobactin prenylation. a) The two forms of O-prenylation are displayed along with examples of natural products containing each modification. b) The LynF prenylation mechanism. After reverse O-prenylation of tyrosine, the intermediate undergoes a Claisen rearrangement to afford the C-prenylated product. As the product of the Lyn cluster has not been identified, the predicted product is displayed without stereochemistry. The DMAPP derived prenyl groups are shown in red.
1.3.5 \textit{C-terminal Amidation in Thiopeptides}

The C-termini of many thiopeptides are amidated. This modification, while easy to overlook when considering the structural complexity of these heavily modified RiPPs, has an important role in biological activity.\textsuperscript{42} Recently, the tailoring machinery responsible for C-terminal amidation moiety has been characterized in thiostrepton and nosiheptide biosynthesis.\textsuperscript{42,43} These studies demonstrated that two divergent strategies have been utilized by Nature to perform the amidation in a leader peptide-independent fashion. In nosiheptide biosynthesis, genetic deletions and \textit{in vitro} reactions were used to identify \textit{nosA} as the gene responsible for C-terminal amidation.\textsuperscript{43} The $\Delta$nosA strain produced a nosiheptide analog with an additional C-terminal dehydroalanine moiety, which is derived from the dehydration of the terminal serine of the precursor peptide. Reactions with purified NosA validated this assignment and revealed that pyruvate was a reaction byproduct. Based on the pH dependence of the reaction and the cofactor independent activity of NosA, this reaction is predicted to proceed via the enamide de-alkylation mechanism displayed in Figure 1.7A. Thiostrepton, on the other hand, requires at least three enzymes to perform the same modification.\textsuperscript{42,63,65} An unidentified methyltransferase is hypothesized to generate the C-terminal methylester on the precursor peptide to facilitate the dehydration of the terminal Ser. Subsequently, TsrB (also called TsrU\textsuperscript{65}) catalyzes the saponification of the methylated thiostrepton intermediate. Finally, the carboxylate is modified by an asparagine synthetase-like amidotransferase, TsrC (also called TsrT\textsuperscript{65}), which the forms the amide linkage, presumably though an adenylated intermediate.\textsuperscript{42} The proposed mechanism of this modification is displayed in Figure 1.7B; however, additional studies are necessary to identify the methyltransferase and to establish the mechanism of TsrC.
Figure 1.7 | C-terminal amidation of thiopeptides. The putative mechanism for C-terminal amide formation is displayed for (a) nosiheptide and (b) thiostrepton. The proposed transformation catalyzed by TsrC is boxed. Tables listing the other thiopeptides that utilize each mechanism, along with the responsible enzyme, are shown on the right.
1.3.6 N-terminal Dimethylation in Plantazolicin Biosynthesis

Plantazolicin (Pzn) is a linear TOMM natural product produced by *Bacillus amyloliquefaciens* (for structure see Fig. 1.2) with narrow-spectrum bactericidal activity against *Bacillus anthracis* and, more recently, nematicidal activity. Deletion mutagenesis studies have demonstrated that the dimethylation of the *N*-terminus is critical for antimicrobial activity. N-terminal methylation in RiPPs is very rare. Apart from Pzn, the linaridin antibiotics are the only characterized RiPPs that are *N*-terminally methylated. Akin to the *C*-terminal amidation of thiopeptides and the prenylation of cyanobactins, *N*-terminal methylation is a leader peptide independent process. Reconstitution of the SAM-dependent Pzn methyltransferase, BamL, demonstrated that the enzyme is highly specific for the heterocyclized Pzn precursor peptide. An X-ray crystal structure of BamL and a homolog from *Bacillus pumilus* (BpumL) demonstrated that this remarkable selectivity is due to the presence of a very narrow substrate-binding channel. At the tightest point, the channel is too narrow to accommodate a peptide bearing carbonyl oxygens. Heterocyclization of the Cys, Ser, and Thr side chains, results in the elimination of the carbonyl oxygen and decreases the width of the molecule enough to facilitate binding. Although the channel is deep enough to accommodate the first 4-5 residues of the Pzn core peptide, studies performed with synthetic substrate analogs have demonstrated that the recognition occurs at the *N*-terminal Arg and the first thiazole heterocycle. Based on the crystal structure, the Arg is thought to be coordinated by a conserved Asp in the active site, while the thiazole is believed to be required to allow the peptide to pass through the narrowest part of the channel. As all Pzn core peptides begin with an Arg and Cys, this strategy for substrate recognition by the methyltransferase is likely invariant within this subfamily of TOMMs.
1.4 Summary and Outlook

Recent successes in the study of TOMM, and more generally RiPP, biosynthetic enzymes have greatly expanded our view of Nature’s synthetic organic potential. In addition to the discovery of novel biological chemistry, these studies provide a window through which the evolutionary adaptation of enzymes can be observed. Although marked progress has been made in the last two decades towards understanding the complex post-translational modifications that occur during TOMM biogenesis, there are many aspects of TOMM biosynthesis that remain uncharacterized and many posttranslational modifications whose biosynthesis remains enigmatic. As such, the post-translational modifications discussed above only represent a subset of the characterized TOMM chemical space.

Perhaps the most pronounced gap in our knowledge of TOMM biosynthesis can be found in the mechanism of azoline formation, which is the focus of this thesis. Studies carried out on diverse TOMM biosynthetic pathways have demonstrated that heterocycle formation occurs early in precursor peptide maturation and can be important for downstream processing by auxiliary tailoring enzymes. Despite the central role that azole formation plays in TOMM biogenesis, the finer details of azole formation remain uncharacterized. While the role of the dehydrogenase is established, the activities of the C and D proteins remain contentious. Furthermore, the mechanism of azoline formation, in particular the role that ATP plays during cyclodehydration, is not understood. Although other laboratories have attempted to address these outstanding questions in TOMM biosynthesis, the poor stability and solubility of previously characterized TOMM synthetases stymied progress. Using two novel TOMM synthetases from Bacillus sp. Al Hakam, I provide answers to these longstanding questions in TOMM biogenesis.
1.5 References


(71) Marsh, E. N.; Patterson, D. P.; Li, L. Chembiochem 2010, 11, 604.


CHAPTER 2: TOMM CYCLODEHYDRATASES UTILIZE ATP FOR CARBONYL OXYGEN ACTIVATION

This chapter is taken in part from Dunbar et al.,¹ Melby et al.,² and Dunbar & Mitchell³ and is reproduced with permission from Nature Publishing Group and the American Chemical Society.

Joyce Limm (Dixon Lab) cloned BalhA1, BalhA2, BalhB, BalhC and BalhD. Joel Melby cloned BalhA1 leader peptide, BalhNC, BalhNC-C40, and BcerB. Joel Melby performed the $^{31}$P-NMR experiment with the Balh enzymes and obtained the Michaelis-Menten kinetic parameters for BalhNC-C40. The microcin B17 biosynthetic enzymes were obtained from the Walsh laboratory.

Abstract

Thiazole/oxazole-modified microcins (TOMMs) encompass a recently defined class of ribosomally synthesized natural products with a diverse set of biological activities. Although TOMM biosynthesis has been investigated for over a decade, the mechanism of heterocycle formation by the synthetase enzymes remains poorly understood. Using a novel TOMM synthetase from an uncharacterized biosynthetic cluster in Bacillus sp. Al Hakam, substrate analogs and isotopic labeling, I demonstrate that adenosine 5'-triphosphate (ATP) is utilized to directly phosphorylate the peptide amide backbone during TOMM heterocycle formation. Such chemistry represents a novel utilization of ATP. Based on this information, I propose a mechanism for azoline formation.

2.1 Introduction

Since the enzymatic machinery responsible for TOMM biosynthesis was first characterized,⁴ it has become evident that this strategy for natural product biosynthesis has been extensively propagated in both bacteria and archaea.⁵,⁶ The diverse array of biological activities and pharmacological potential displayed by the characterized members of the TOMM family has led to extensive investigations into the molecular underpinnings of their biosynthesis.⁴,⁷-¹³ These
studies have provided the foundation to understanding the complex nature of substrate processing, but many of the finer details regarding heterocycle formation have remained elusive.

For example, the requirement for ATP hydrolysis with respect to azoline formation remains undescribed.\textsuperscript{7,10} Reports of the super-stoichiometric hydrolysis of ATP relative to azole formation,\textsuperscript{7} and continued ATP hydrolysis after complete substrate modification,\textsuperscript{10} have been used as evidence for the role of ATP as a dynamic regulator of the cyclodehydratase complex (C and D proteins; Fig. 1.3), yet direct evidence for such a mechanism has not been reported. Recently, this role for ATP utilization has been referred to as the “molecular machine” hypothesis (Fig. 2.1B).\textsuperscript{10} An alternative mechanism implicates ATP hydrolysis in the activation of the peptide backbone during cyclization (Fig. 2.1B).\textsuperscript{4,13} Such an activation event would result in the elimination of P\textsubscript{i} instead of water during the cyclization, and could potentially help drive the cyclodehydration reaction towards completion. Although isotopic labeling studies could readily differentiate between these mechanisms, the relative experimental intractability of previously characterized TOMM synthetases has precluded such a study.\textsuperscript{4,6} The discovery of a TOMM biosynthetic cluster from \textit{Bacillus} sp. Al Hakam (Balh; Fig. 2.1A)\textsuperscript{2} yielded robust heterocycle-forming enzymes for \textit{in vitro} characterization and allowed us to address this foremost question regarding TOMM biosynthesis. The findings presented in this report demonstrate that the TOMM synthetase utilizes ATP to phosphorylate the amide backbone of peptide substrates during the cyclodehydration reaction. As such, the data reported herein provide an improved framework to understand the factors that govern biosynthesis in this class of natural products that has garnered significant attention.\textsuperscript{14-26}
Figure 2.1 | TOMM biosynthetic gene clusters and possible mechanisms of ATP utilization during azole formation. (a) The TOMM clusters from *Bacillus* sp. Al Hakam (Balh) and *Escherichia coli* (Mcb) are depicted along with the percent amino acid identity for each of the three key proteins. Gene assignments are given below. Note that the dehydrogenase and the “cyclodehydratase” in the microcin B17 cluster are assigned as McbC and McbB, respectively. The sequences of the peptide substrates used in this study are shown. Color-coding: green, point mutations; orange, residues known to be cyclized *in vitro* (Balh) or *in vivo* (Mcb); blue hyphen, putative leader sequence cleavage site; blue caret, known leader sequence cleavage site. (b) The two leading hypotheses for ATP utilization. ATP hydrolysis could be used to control conformational dynamics (molecular machine) or to directly activate the peptidic substrate as shown. Reactions carried out in [18O]-H2O should give different products (shown), which can be used to distinguish between these mechanisms. X = S, O; R = H, CH3.
2.2 Minimal Requirements for Balh Cyclodehydratase Activity

In order to determine the minimal set of proteins required for azoline formation, reactions were initiated with one or more reaction components omitted with progress monitored by matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS, Fig. 2.2). Samples lacking ATP, BalhC or BalhD did not display a detectable level of BalhA1 substrate modification under the conditions employed. When the BalhA1 substrate was treated with both BalhC and BalhD, the majority of the peptide was converted to a penta-azoline species (-90 Da). Substrate treated with the full BCD synthetase (see note in experimental section) was converted to a penta-azole species (-100 Da) as previously reported. In congruence with previously characterized TOMM synthetases, this activity was dependent on the hydrolysis of ATP, as the omission of nucleotide triphosphate or the addition of a non-hydrolyzable ATP analog resulted in no modification (Fig. 2.2A).

Analogous to the microcin B17 and cyanobactin synthetases, the Balh synthetase also hydrolyzed ATP to ADP and P\textsubscript{i} during substrate processing (Fig. 2.2B). The rate at which P\textsubscript{i} was generated under varying reaction conditions was monitored by a purine nucleoside phosphorylase (PNP)-coupled assay. Robust P\textsubscript{i} generation was observed only when the C and D proteins were added to samples containing precursor peptide (BalhA1/BalhA2) and ATP (Fig. 2.2C). While the $K_M$ for both precursor peptides was 16 µM, the $k_{obs}$ for BalhA1 was approximately 4 times higher than for BalhA2 (Table 2.1). Addition of the B-protein did not significantly alter the rate of ATP hydrolysis, which is consistent with earlier reports of C-D fusion enzymes forming azoline heterocycles in the absence of the dehydrogenase. In contrast to the Balh and the C-D fusion enzymes, the microcin B17 synthetase was not functional in the absence of the dehydrogenase. The reason for this discrepancy is currently unknown but the
available data suggest that the Mcb dehydrogenase may play a pivotal role in the assembly of an active synthetase complex. Moreover, subsequent studies demonstrated that the Balh enzymatic machinery was able to utilize both ATP and GTP (analogous to the microcin B17 synthetase), although ATP was preferred over GTP by approximately a factor of 10 due to differences mainly in $K_m$ (Table 2.1).

**Table 2.1 | Substrate kinetic parameters for Balh synthetase.** Error represents the error from the curve fit. $^1$, Apparent $K_M$; $^2$, obtained with 3 mM ATP; $^3$, obtained with 110 µM BalhA1. This table was adapted from Melby *et al.*$^2$ and Dunbar *et al.*$^1$ with permission.

<table>
<thead>
<tr>
<th></th>
<th>$k_{obs}$, min$^{-1}$</th>
<th>$K_M$, µM$^4$</th>
<th>$k_{obs}/K_M$, M$^{-1}$ s$^{-1}$</th>
<th></th>
<th>$k_{obs}$, min$^{-1}$</th>
<th>$K_M$, µM$^4$</th>
<th>$k_{obs}/K_M$, M$^{-1}$ s$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^2$BalhA1</td>
<td>12.9 ± 0.4</td>
<td>16 ± 2</td>
<td>13000</td>
<td>$^3$ATP</td>
<td>12.2 ± 0.3</td>
<td>240 ± 20</td>
<td>850</td>
</tr>
<tr>
<td>$^2$BalhA2</td>
<td>4.4 ± 0.2</td>
<td>17 ± 2</td>
<td>4300</td>
<td>$^3$GTP</td>
<td>9.3 ± 0.2</td>
<td>1500 ± 80</td>
<td>100</td>
</tr>
</tbody>
</table>
Figure 2.2 | Minimal requirement for cyclodehydratase activity and ATP hydrolysis product of the Balh synthetase. (a) MALDI-TOF spectra monitoring heterocycle formation. Labels represent the components of the reaction (e.g. C contains BalhC). Apart from the indicated samples, all reactions contained ATP. The mass shift for the major species relative to the unmodified substrate is displayed above the appropriate peak. NH-ATP represents β,γ-imido-ATP (a non-hydrolyzable ATP-analog) (b) Rates of phosphate production in Balh synthetase reactions with and without the addition of 1 unit of pyrophosphatase (PPase) are shown. Error represents the standard deviation from the mean (n=2). (c) Rates of ATP hydrolysis in reactions lacking one or more components of the synthetase complex are displayed. Activity was monitored with the PNP-coupled phosphate assay and activity was normalized to a reaction containing both substrates (ATP and BalhA1) and the BCD complex. Error bars represent the standard deviation from the mean (n=3). This figure is adapted with permission from Dunbar et al.1
2.3 A Cyclizable Residue Stimulates Rapid ATP Hydrolysis

The “molecular machine” hypothesis asserts that ATP hydrolysis regulates the conformational dynamics/enzymatic activity of the synthetase complex. While the means by which this may work has not yet been explicitly stated, one possible route could be that upon peptide substrate binding, the synthetase undergoes a conformational change that leads to an active complex (Fig. 2.1). Previous studies on microcin B17 biosynthesis have demonstrated that the leader peptide and a severely truncated (non-processed) precursor peptide substrate analog were incapable of stimulating ATP hydrolysis. These results suggested that the core sequence of TOMM precursor peptides must be important for the ATPase activity of the synthetase. Studies conducted in the absence of BalhA1 demonstrated that a complex of BalhC and BalhD had a measureable level of ATP-hydrolysis; roughly 100-fold lower than at saturating BalhA1 concentrations (Fig. 2.3). To test if a non-cyclizable (NC), full-length substrate analog could potentiate this basal ATPase activity of the TOMM enzymatic complex, we prepared BalhA1-NC, in which every heterocyclizable residue in the core region was mutated to a hydrophobic amino acid (Fig. 2.1). Interestingly, the presence of BalhA1-NC decreased the rate of ATP hydrolysis four-fold compared to samples containing only BalhC and BalhD (Fig. 2.3A). Addition of the predicted BalhA1 leader sequence (BalhA1-LS, Fig. 2.1) to similar reactions showed a slight, but statistically significant, reduction in the basal ATPase activity. The marked difference between the levels of inhibition provided evidence that the core sequence was primarily responsible for suppressing the ATPase activity and indicated that there may be a direct interaction between residues of the core sequence and the ATP binding site. In support of this, the reinstatement of a naturally occurring cysteine (BalhNC-C40, Fig. 2.1) yielded a substrate that was not only processed by the synthetase, but increased the rate of ATP hydrolysis close to wild-type levels.
These data provided substantial evidence that the ATP binding site and the residues of the core sequence directly communicate. Moreover, the ability of BalhA1-NC to inhibit the basal ATPase activity suggested that the ATP site had greater accessibility in the absence of the precursor peptide. While not definitive, these results are difficult to reconcile with respect to an ATP-utilizing mechanism consistent with the molecular machine hypothesis, which prompted us to explore alternative mechanistic possibilities.

Figure 2.3 | BalhA1-NC inhibition and BalhNC-C40 processing. (a) The rate of ATP hydrolysis with and without the addition of BalhA1-NC or BalhLS as measured by the PNP phosphate detection assay. Due to the extremely rapid rate of ATP hydrolysis when BalhC/D are at this high concentration (15 μM), it is not possible to accurately measure initial rates for rapidly processed substrates. As such, the rates of ATP hydrolysis with the addition of BalhA1 and BalhNC-C40 were extrapolated from the nucleotide kinetics (BalhC/D = 1 μM) assays based on the measured $k_{obs}$ for each substrate. Error bars represent the standard deviation from the mean (n=3). (b) A MALDI-TOF spectrum of a reaction carried out with BalhNC-C40 either with (red) or without ATP (black).
2.4 ATP Hydrolysis is Tightly Coupled to Azoline Formation

The microcin B17 synthetase has been shown to consume approximately 5 moles of ATP per mole of heterocycle formed. Although the super-stoichiometric consumption of ATP has not been observed for all TOMM biosynthetic enzymes, this observation cast doubt on the use of ATP to directly activate the substrate during processing and has been used as supportive evidence of the molecular machine. To evaluate the stoichiometry of products formed by the Balh synthetase, the rate of P$_i$ and azole production were monitored in parallel time course assays. An overlay of P$_i$ and azole synthesis illustrated that the formation of these products was tightly coupled during the reaction, and that the ATP to azole stoichiometry was essentially one-to-one over the entire six-hour time course (Fig. 2.4A). Furthermore, when BalhA1 was fully heterocyclized, the rate of ATP hydrolysis returned to approximately the basal rate. This result was consistent with an earlier report on trunkamide biosynthesis, which employs a C-D fusion protein.

Given the markedly different ATP/azole stoichiometry in the Balh and Mcb systems, I revisited the ATP stoichiometry of microcin B17 processing. I hypothesized that the use of a truncated substrate and a synthetase complex with an abnormal ratio of the B, C and D proteins (due to the purification strategy) could have led to an aberrant usage of ATP. Therefore, the Mcb ATP stoichiometry was repeated using the full-length substrate (McbA; Fig. 1a) and a completely tag-free synthetase complex added in at a 1:1:1 molar ratio. As with the Balh synthetase, the concentration of P$_i$ and azole heterocycles mapped almost perfectly over the time course assayed and an average stoichiometric ratio was calculated to be unity (Fig. 2.4B). Taken together, these data demonstrated that ATP hydrolysis is tightly linked to heterocycle formation, which appears to be a general feature of TOMM heterocycle formation.
Figure 2.4 | ATP/azole stoichiometry for substrate processing. The stoichiometry of ATP hydrolysis to azole heterocycle formation in (a) the Balh TOMM and (b) the Mcb TOMM is shown. Phosphate (squares), heterocycles (circles), BCD background ATPase activity (solid line, PNP assay; dashed line, extrapolated PNP assay; triangles, malachite green assay), maximum product produced based on the concentration of substrate (solid horizontal line). Error bars represent the standard deviation from the mean (n=3). The ATP/azole stoichiometric ratio is displayed and is the average of all time points.
2.5 ATP is Hydrolyzed by a Non-Solvent Oxygen Nucleophile

Intrigued by the inability to artificially activate ATP hydrolysis and the 1:1 stoichiometry between ATP consumption and ring formation, isotopic labeling was used to determine the source of the oxygen in the P$_i$ generated during the course of the reaction. If ATP were used to directly activate the substrate, the oxygen incorporated into the P$_i$ byproduct would originate from the peptide backbone. Thus, running the reaction in [¹⁸O]-H$_2$O would result in $^{16}$O incorporation into phosphate, producing [¹⁶O$_4$]-P$_i$ (Fig. 2.1). Alternatively, the molecular machine hypothesis demands that the oxygen incorporated into the P$_i$ would originate from bulk solvent (water) in the same manner as with any ATPase (Fig. 2.1). The molecular machine mechanism, if acting, would introduce a single $^{18}$O label in the P$_i$ produced when the reaction was run in [¹⁸O]-H$_2$O, which is readily detectable by a roughly 0.02 ppm upfield shift in $^{31}$P-NMR compared to [¹⁶O$_4$]-P$_i$.\(^{30}\)

Reactions of the Balh TOMM biosynthetic enzymes with the BalhA1 substrate in [¹⁸O]-H$_2$O resulted in the production of primarily [¹⁶O$_4$]-P$_i$ (Fig. 2.5A). As would be predicted from the earlier stoichiometry assay (Fig. 2.4) and the basal rate of ATP hydrolysis (Fig. 2.3), these samples also contained a small amount (~15%) of [¹⁸O$_1$¹⁶O$_3$]-P$_i$. The majority production of [¹⁶O$_4$]-P$_i$ was a consequence of substrate processing and was neither a contaminant of the ATP buffer, nor the proteinaceous components of the synthetase reaction as assessed by the malachite green assay. While the direct activation of the amide carbonyl is the most logical explanation for this data, the possibility existed that [¹⁶O$_4$]-P$_i$ was generated as a consequence of nucleophilic attack at the β-phosphate of ATP or through isotope scrambling with the ADP byproduct. Thus, a liquid chromatography (LC)-MS analysis of the ADP formed during the reaction was carried out to determine if this byproduct contained $^{18}$O. The trace clearly demonstrated that $^{18}$O was not
present in the ADP and provided strong evidence that ATP consumption occurred due to the action of a non-solvent, oxygen nucleophile (Fig. 2.6). This assay was repeated with the microcin B17 synthetase, which yielded the same result (Fig. 2.5B), though the requirement for a lower pH buffer gave a slight perturbation in the $^{31}\text{P}$-phosphate chemical shift. Taken together, these results strongly suggest that the molecular machine hypothesis for ATP utilization is incorrect.

**Figure 2.5** | **ATP is utilized to directly activate the amide backbone of the substrate.** $^{31}\text{P}$-NMR spectra of reactions with (a) the Balh or (b) the Mcb synthetase. The identity of the major product, $[^{16}\text{O}]-\text{P}_\text{i}$, was confirmed by spiking with authentic $[^{16}\text{O}]-\text{P}_\text{i}$. Slight pH differences between the Balh and Mcb samples (8.5 and 8.0, respectively) account for the altered chemical shifts. This figure was adapted from Dunbar et al.\textsuperscript{1} with permission.

**Figure 2.6** | **Isotope composition of ADP produced in $[^{18}\text{O}]$-water.** An LCMS trace for the ADP byproduct of a reaction carried out in $^{18}\text{O}$-water. LCMS traces are single ion monitoring (SIM) measurements of [M – H] for ADP (m/z 426) and for the putative $[^{18}\text{O}_1]$-ADP (m/z 428). A reference line (gray dashed line) is displayed which indicates the m/z 428 signal expected from the natural 1.1% $^{13}\text{C}$ isotope abundance (approximately 3.7%).
2.6 TOMM Cyclodehydrations Occur via an Activated Carbonyl Intermediate

While the above results are most consistent with the direct activation mechanism for ATP utilization, these data do not directly confirm the proposed amide phosphorylation event (Fig. 2.7A). A more direct approach would be to perform cyclodehydration reactions with a substrate containing $^{18}$O labels at every cyclized amide. In the event that direct activation was occurring, these $^{18}$O labels would be found in the phosphate byproduct. However, due to the size of BalhA1 (57 residues; Fig. 2.1A), solid phase synthesis is impractical. While the generation of an appropriately labeled BalhA1 derivative could be achieved by native chemical ligation, this strategy would involve the synthesis and purification of two $^{18}$O-labeled large peptide fragments (27 and 30 residues) and would be a tedious endeavor. Conversely, azoline-mediated peptide backbone labeling (AMPL; see Chapter 3) provides a convenient route to access labeled peptides such as $[^{18}O_5]$-BalhA1. Trx-tagged $[^{18}O_5]$-BalhA1, generated from azoline hydrolysis, was treated with BcerB/BalhCD and the isotopic constitution of the phosphate byproduct was analyzed by $^{31}$P-NMR. A peak corresponding to $[^{16}O_3^{18}O]$-phosphate (P$_i$) was observed only in reactions with $[^{18}O_5]$-BalhA1 (Fig. 2.7B). Importantly, $[^{16}O_3^{18}O]$-P$_i$ was not detected in control reactions lacking enzyme or substrate.

While the detection of $[^{16}O_3^{18}O]$-P$_i$ provided strong support for the proposed phosphorylation mechanism, the detected amount of $[^{16}O_4]$-P$_i$ was seemingly inconsistent with direct backbone activation being the sole mechanism of ATP utilization. I originally hypothesized that isotope scrambling with the ADP byproduct may be responsible for the $[^{16}O_4]$-P$_i$ peak; however, a mass spectrum of ADP generated during the modification of $[^{18}O_5]$-BalhA1 demonstrated that this was not the case (Fig. 2.7C). Although the ATP/azole stoichiometry of the Balh cyclodehydratase is 1:1 (vide supra), an alternative explanation was that ATP/azole
stoichiometry was dysregulated for Trx-tagged $[^{18}O_5]$-BalhA1 as a result of the requisite, extensive sample preparation. To test this, an identical reaction was carried out but the ATP/azole stoichiometry was measured rather than the P$_i$ isotope distribution. Indeed, the stoichiometry assay showed that the 1:1 ratio of ATP/azole increased to 1.9:1 following the hydrolysis procedure. Taking into account the isotopic enrichment of $[^{18}O_5]$-BalhA1 (Fig. 2.7D), the predicted $[^{16}O_4]$-P$_i$:[$^{16}O_3^{18}O$]-P$_i$ ratio is approximately 1.3:1. This result is in accord with the ratio observed by $^{31}$P-NMR. Collectively, these data clearly demonstrate that TOMM cyclodehydratases utilize ATP to activate the carbonyl oxygen of the amide involved in azoline formation.
Figure 2.7 | $^{31}$P-NMR of $^{18}$O$_5$-BalhA1 reactions. (a) The putative carbonyl oxygen phosphorylation mechanism for ATP utilization is displayed along with $^{31}$P-NMR spectra from both $^{18}$O$_5$-BalhA1 and $^{16}$O$_5$-BalhA1 reactions (b). The identity of the $^{16}$O$_4$-P$_i$ peak was verified by spiking the sample with a P$_i$ standard (spectrum 3). (c) A negative mode MALDI-TOF spectral overlay of an ADP standard (black) and the ADP produced during $^{18}$O-BalhA1 (red) processing is shown. If $^{16}$O$_4$-P$_i$ production is due to $^{18}$O incorporation into ADP, the second isotope peak of ADP should be approximately the same height as the monoisotopic peak. (d) A high-resolution mass spectrum of $^{18}$O$_5$-BalhA1 is displayed below. The peaks are colored according to the number of $^{18}$O labels in the peptide (green, five; purple, 4; orange, 3; blue, 2; red, 1; black, 0). The calculated and measured masses of the $^{18}$O$_5$-BalhA1 peak are displayed in the 3+ charge state along with the ppm error of the measurement. The level of $^{18}$O enrichment was determined based on the intensities of the monoisotopic peaks for each of the $^{18}$O-labeled BalhA1 derivatives. The theoretical maximum $^{18}$O enrichment of 92% was determined based on the final atom % of $^{18}$O during the hydrolysis reaction. This figure was adapted with permission from Dunbar & Mitchell. 3
2.7 Summary and Outlook

Prior to this study, the utilization of ATP hydrolysis for cyclodehydration was poorly understood. Although it had been noted that the initial steps of this transformation could mimic the mechanism of protein splicing by intein domains (an ATP-independent process; Fig. 2.8), the absolute requirement for nucleotide triphosphate hydrolysis had never definitively been factored into the reaction.4,10 These findings make the TOMM cyclodehydratases only the third enzyme class known that can utilize ATP to directly activate an amide carbonyl oxygen, the other two being the PurM and adenine nucleotide α-hydrolase (α-ANH) superfamilies. Characterized members of the PurM and α-ANH families conduct similar chemistry on non-peptide substrates (Fig. 2.9).31-33 As such, the mechanism of azoline formation in TOMM biosynthesis represents a novel utilization of this ubiquitous cofactor. While members of the α-ANH superfamily contain a canonical ATP-binding fold,33 structurally characterized members of the PurM family all contain a non-canonical ATP-binding domain.34,35 Despite the similarity in ATP usage between TOMM cyclodehydratases, PurM proteins, and α-ANH family members, the three enzyme families share no sequence similarity and represent an example of convergent evolution towards a common mode of ATP utilization.

Members of the PurM and α-ANH families are proposed to act via the formation of an iminophosphate and acyl-adenylate intermediate, respectively.33,36 In both mechanisms, carbonyl oxygen activation (phosphorylation in PurM and adenylation in α-ANH) occurs prior to nucleophilic attack (Fig. 2.9). While a similar mechanism can be drawn for azoline formation by the TOMM synthetase (Fig. 2.10), it would require a disfavored 5-endo-trig cyclization37 and is inconsistent with the suppression of ATP hydrolysis in the presence of BalhA1-NC. Alternatively, I hypothesize that azoline formation by the TOMM synthetase occurs through a
hemi-orthoamide intermediate analogous to that implicated in protein autoproteolytic pathways.\textsuperscript{38-40} During intein splicing and other autoproteolytic events, the hemi-orthoamide is resolved by $N$-protonation and (thio)ester formation (Fig. 2.8).\textsuperscript{39} I assert that the phosphorylation of the amide oxygen (in lieu of $O$-protonation) would direct the hemi-orthoamide towards azoline formation and prevent the non-productive breakdown of the intermediate (Fig. 2.8). Phosphorylation would not only accelerate the elimination reaction (based on the lower pK$_a$ of phosphate relative to water), but would directly couple ATP hydrolysis to cyclodehydration, providing a thermodynamic drive for the reaction. The resolution of the hemi-orthoamide via the inclusion of a thermodynamically favorable step is seen in all autoproteolytic pathways.\textsuperscript{39} Further evidence supporting a mechanism where cyclization precedes phosphorylation (\textit{i.e.} the intein-like mechanism) comes from a recent report of ester formation during microcin B17 biosynthesis\textsuperscript{41} and the discovery that engineered intein domains can catalyze azoline formation.\textsuperscript{42} In light of earlier reports, and the data presented here, TOMM cyclodehydration and intein splicing proceed through a common intermediate, which I propose is the hemi-orthoamide depicted in Figure 2.8.

The mechanism for thiazole/oxazole formation presented in this report represents the fourth strategy identified to date to generate these moieties. Two such strategies have been uncovered in the biosynthesis of the primary metabolite thiamine,\textsuperscript{43} while the third is found in nonribosomal natural products biosynthesis (Fig. 2.11).\textsuperscript{44} Although the TOMM cyclodehydratase is not homologous to the nonribosomal peptide synthetase counterpart, the TOMM C protein is homologous to ThiF,\textsuperscript{6,45} the bacterial protein involved in sulfur mobilization in thiamine biosynthesis. Rather than perform amide activation, ThiF carries out the canonical carboxylate adenylation catalyzed by all members of the E1-superfamily.\textsuperscript{46} Further study will be necessary to
elucidate the component of the TOMM cyclodehydratase complex (C/D proteins) responsible for carrying out this unique transformation.

Figure 2.8 | TOMM azoline installation is reminiscent of intein-mediated protein splicing.  
(a) The mechanism of protein splicing and autoproteolysis follows the N-protonation pathway to generate a (thio)ester. An azoline heterocycle could also be formed from the hemi-orthoamide intermediate if O-protonation occurred.  
(b) In TOMM biosynthesis, the use of ATP to phosphorylate the substrate could be used to drive the reaction down an O-elimination pathway to generate an azoline. X = S, O; R = H, CH₃.
Figure 2.9 | The proposed enzymatic mechanism of PurM and ThiI. (a) The fifth enzyme in the purine biosynthesis pathway in prokaryotes is aminoimidazole ribonucleotide (AIR) synthetase, PurM. This enzyme is responsible for the heterocyclization of formylglycinamide ribonucleotide (FGAM) to afford AIR. ATP has been implicated in the phosphorylation of the backbone amide oxygen using isotopically labeled substrate. Although a mechanism wherein activation precedes nucleophilic attack is invoked for the transformation, evidence has not been provided to demonstrate this order of events. Note that the phosphorylation of an amide oxygen has been implicated in the transformations catalyzed by the PurM homologs, PurL, HypE, and SelD. (b) The mechanism of 4-thiouridine formation by ThiI is displayed. The ThiI homologs MnmA and TtuA use identical chemistry in the biosynthesis of 2-thiouridine and 2-thiocytosine. This figure was adapted from Dunbar et al. with permission.

Figure 2.10 | Alternative mechanism of azoline formation. An alternative mechanism for azoline formation is presented involving phosphorylation prior to nucleophilic attack. While the mechanism is reminiscent of Vilsmeier-Haack dehydration chemistry, the transformation involves a disfavored 5-endo trig cyclization step.
Figure 2.11 | Strategies for azole biosynthesis outside of TOMMs. (a) While thiamine biosynthesis occurs via different mechanisms in prokaryotes and eukaryotes, neither pathway utilizes the same chemical logic as TOMM synthetases. ThiF is involved in the installation of the sulfur atom during bacterial thiamine production. (b) Thiazole/oxazole biosynthesis in non-ribosomal peptides is hypothesized to occur through a hemi-orthoamide intermediate; however, resolution of the intermediate is believed to occur without backbone activation.
2.8 Experimental

2.8.1 Substitution of BcerB for BalhB. Because the dehydrogenase from the Balh cluster (BalhB) purified without FMN bound, a highly similar (78% identical/94% similar) dehydrogenase from *Bacillus cereus* 172560W (BcerB) was used as a surrogate. This technique has been successfully employed in the study of other TOMM synthetases.

2.8.2 Protein overexpression and purification. All proteins were overexpressed as tobacco etch virus (TEV) or thrombin protease-cleavable fusions to maltose binding protein (MBP) and purified by amylose affinity chromatography by methodology similar to that previously reported. BL21(DE3-RIPL) cells were transformed with pET28b plasmids containing the desired MBP fusion gene (BalhA-D, BcerB, McbD). Cells were grown overnight on Luria-Bertani (LB) plates with 50 µg/mL kanamycin. Single colonies were picked for starter cultures containing 50 µg/mL kanamycin and 34 µg/mL chloramphenicol and were grown at 37 °C. A 10 mL overnight culture was used to inoculate 1 L of LB containing 50 µg/mL kanamycin and 34 µg/mL chloramphenicol. For the overexpression of the modification enzymes, cells were grown to an optical density at 600 nm (OD$_{600}$) of 0.6 before a 10 min cold shock and induction with 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 16 h at 22 °C. For the overexpression of precursor peptides, cells were grown to an OD$_{600}$ of 0.8 before induction with 1 mM IPTG for 1.5 h at 22 °C. After the indicated times, cells were harvested at 3000 x g for 15 min, washed with TBS (tris buffered saline; 10 mM Tris pH 7.5, 150 mM NaCl) and stored at -20 °C for up to one week before use. The overexpression of pET15b constructs containing McbA-C were carried out as above except ampicillin (100 µg/mL) was used instead of kanamycin in the relevant steps. Cell pellets were resuspended in lysis buffer [50 mM Tris pH 7.5, 500 mM NaCl, 2.5% glycerol (v/v), 0.1% triton X-100 (v/v)] containing 4 mg/mL lysozyme, leupeptin (2 µM), PMSF (200
μM), benzamidine (2 mM), and E64 (2 μM). After a 30 min incubation at 4 °C, cells were disrupted via sonication 3 x 30s with 10 min breaks at 4 °C. The insoluble debris was removed from the sample via centrifugation at 35,000 x g for 45 min. The resulting supernatant was applied to pre-equilibrated amylose resin (5 mL resin per L of cells). The column was washed with 10 column volumes of lysis buffer, followed by 5 column volumes of wash buffer (lysis buffer lacking triton X-100 and with 400 mM NaCl). The MBP-tagged proteins were eluted using 4 column volumes of elution buffer (wash buffer with 300 mM NaCl and 10 mM maltose) and the eluent was concentrated using an appropriate Amicon Ultra centrifugal filter (Millipore). A 10-fold buffer exchange with storage buffer [50 mM HEPES pH 7.5, 300 mM NaCl, 2.5% glycerol (v/v)] was performed in the filtration device before a final concentration and storage. After loading onto the column, all buffers used contained 1 mM tris-(2-carboxylethyl)-phosphine. Protein concentration was assessed by both the 280 nm absorbance and a Bradford colorimetric assay (Thermo Scientific). Purity was determined by Coomassie-stained SDS-PAGE gel.

Unless otherwise stated, all proteins and substrates were used as MBP-fusions. The Balh TOMM synthetase proteins are highly active while MBP-tagged and largely insoluble when untagged.

2.8.3 Purine nucleotide phosphorylase-based phosphate detection assay. In general, the indicated synthetase proteins were added to a cuvette for a final concentration of 1 μM each. Reactions were initiated via the addition of a room temperature mixture of peptide substrate, 200 μM 2-amino-6-mercaptop-7-methylpurine riboside (Berry and Associates), and 0.2 units of PNP in synthetase buffer [50 mM Tris pH 7.5, 125 mM NaCl, 20 mM MgCl₂, 10 mM dithiothreitol (DTT) and 3 mM ATP]. Reaction progress was monitored by the change in absorbance at 360
nm on a Cary 4000 UV-Vis spectrophotometer (Agilent). Initial rates of phosphate production were calculated based on the linear absorbance change during the first 3 min of the reaction and the extinction coefficient of the resulting guanine analog (11,000 M$^{-1}$cm$^{-1}$). For determining the nucleotide triphosphate kinetic constants, a synthetase buffer was used that lacked ATP, and the reactions were initiated by the addition of the indicated nucleotide. Given the slow rate of ATP hydrolysis without the presence the precursor peptide, the background ATPase activity was measured using 15 µM MBP-tagged BalhC/D instead of 1 µM. Reactions were carried out as above (n ≥ 3). Regression analyses to obtain the kinetic parameters were carried out using IGOR Pro version 6.12.

2.8.4 Subtractive activity studies. The effect of each protein on heterocyclization was monitored by MALDI-TOF-MS. Reactions were carried out for 16 h at 23 °C with 1 µM of the indicated components, 50 µM MBP-BalhA1 and the synthetase buffer listed above. MBP was proteolytically removed from the substrate by the addition of 2 µg/ml recombinant TEV protease and a 30 min incubation at 30 °C. Samples were desalted via C$_{18}$ ZipTip (Millipore) according to the manufacturer’s instructions and analyzed on a Bruker Daltonics UltrafleXtreme MALDI-TOF spectrometer. Spectra were obtained in positive reflector mode using α-cyano-hydroxycinnamic acid as the matrix.

2.8.5 Substrate analog assays. Inhibition studies were performed with 15 µM MBP-BalhCD and either 50 µM MBP-BalhA1-NC or MBP-BalhA1-LS. Rates were measured with the PNP assay as described above following a 15 min equilibration at 23 °C. MBP-BalhNC-C40 reactions were carried out with 1 µM MBP-tagged BeerB, BalhC, BalhD and 100 µM substrate for 16 h at 23 °C. Products were analyzed via MALDI-TOF-MS as described above.
2.8.6 Stoichiometry of ATP utilization. Samples were set up with 2 μM MBP-BalhC/D, 10 μM MBP-BcerB and 50 μM MBP-BalhA1 in synthetase buffer. Aliquots were removed at the indicated time points and frozen in liquid nitrogen. The samples were analyzed by LCMS to detect azole heterocycles and a malachite green assay to detect phosphate (Supplementary methods). The Mcb reactions were carried out in a similar fashion with the following exception: samples containing 1 μM MBP-tagged McbBCD and 20 μM MBP-McbA were digested with thrombin for 4 h before the reaction was initiated with the addition of ATP. The background ATPase activities of both synthetase complexes were analyzed using both the PNP and malachite green phosphate assays.

To determine the ATP/azole stoichiometry following azoline hydrolysis, an azoline hydrolysis reaction was carried out on Trx-BalhA1 in an identical fashion to the 31P-NMR experiment (see below) except [16O]-H2O was used in place of [18O]-H2O. Following buffer exchange to remove ADP and inorganic phosphate (P_i), a reaction was carried out with 50 μM BalhA1 and 2 μM BCD in low-salt synthetase buffer (50 mM Tris pH 8.5, 25 mM NaCl, 5 mM MgCl_2, 10 mM DTT, and 2 mM ATP) for 3 h at 25 °C. The sample was then divided into two aliquots and frozen in liquid nitrogen. Azole formation and ATP hydrolysis were measured as indicated above.

2.8.7 31P-NMR analysis of phosphate isotope incorporation of 18O-water reactions. Two samples containing either the MBP-tagged BCD (4 μM) synthetase components or MBP-BalhA1 substrate (100 μM) with low-salt synthetase buffer (50 mM Tris pH 8.5, 25 mM NaCl, 5 mM MgCl_2, 10 mM DTT, 2 mM ATP) were lyophilized separately for 16 h. The resulting solid was reconstituted in 500 μL of 97 atom% [18O]-H2O (Cambridge Isotope Laboratories) and incubated at 23 °C with rocking for 8.5 h, before quenching the reaction with 5.5 μL of 500 mM
ethylenediaminetetraacetic acid (EDTA; pH 8.0). 50 µL of D$_2$O was added and the sample was transferred to a standard 5 mm NMR tube. The $^{31}$P-NMR spectrum was obtained on a 600 MHz Varian Unity Inova NMR with a 5 mm Varian AutoTuneX probe, using 512 transients, 32768 points, and a spectral window of -15 to 5 ppm.

The MBP-tagged McbBCD proteins were pooled at a 1:1:1 molar ratio and concentrated to 100 µM using a 3 kDa amicon centrifugal filter (Millipore). To this sample, 0.2 µg of thrombin (from bovine plasma) was added and the sample was proteolytically digested for 4 h at 23 °C to remove the MBP tags. A 600 µL sample of 20 µM MBP-McbA and low-salt synthetase buffer (pH 8.0) was lyophilized for 16 h. The lyophilized protein was reconstituted in 600 µL of [¹⁸O]-H$_2$O and the above mixture of McbBCD was added to a final concentration of 5 µM to initiate the reaction. After a 2 h incubation at 23 °C, the sample was frozen in liquid nitrogen and lyophilized to dryness. The sample was reconstituted in 220 µL of D$_2$O with 15 mM EDTA and placed in a D$_2$O-matched Shigemi NMR tube. Analysis by $^{31}$P-NMR was carried out as before with 2000 transients.

For all reactions, a [¹⁶O]-H$_2$O control and a buffer only control were run. For the BcerB/BalhC/D synthetase reaction, a substrate-free control was also performed in [¹⁸O]-H$_2$O in an identical fashion as the BalhA1 modification reaction. All controls were carried out in an identical fashion to the corresponding assay. The identity of the $^{16}$O$_4$-P$_i$ peak in sample was verified by spiking the sample with a P$_i$ standard and reacquiring the spectra under identical conditions.

2.8.8 $^{31}$P-NMR analysis of phosphate isotope incorporation in [¹⁸O$_5$]-BalhA1 reactions. [¹⁸O$_5$]-BalhA1 was generated in an identical strategy to the one described in Chapter 3 except that 100 µM Trx-BalhA1 was used instead of 100 µM MBP-BalhA1 and the fusion tags were not
removed. Following hydrolysis in [\(^{18}\text{O}\)]-\(\text{H}_2\text{O}\), all contaminating ADP and \(\text{P}_i\) was dialyzed from the sample by a 30,000-fold buffer exchange into storage buffer [50 mM Tris pH 7.5, 150 mM NaCl, 2.5 % glycerol (v/v)] using a 10 kDa Amicon Ultra centrifugal filtration unit (Millipore). Next, the phosphate concentration in the sample was measured by a malachite green phosphate detection assay (BioAssay Systems) to ensure that the concentration of contaminating \(\text{P}_i\) was below 4 \(\mu\text{M}\). The concentration of Trx-[\(^{18}\text{O}_3\)]-BalhA1 was determined both by the 280 nm absorbance and the Bradford colorimetric assay (Thermo Scientific). Reactions for \(^{31}\text{P}\)-NMR were carried out with 50 \(\mu\text{M}\) Trx-[\(^{18}\text{O}_3\)]-BalhA1 and 2 \(\mu\text{M}\) MBP-tagged BCD in low-salt synthetase buffer for 3 h at 25 °C. Subsequently, the sample was quenched by the addition of 500 mM EDTA (pH 8.5) to a final concentration of 5 mM. \(\text{D}_2\text{O}\) was added to a final concentration of 10 % and the sample was transferred to a \(\text{D}_2\text{O}\)-matched Shigemi NMR tube. The \(^{31}\text{P}\)-NMR spectrum was obtained on a 600 MHz Varian Unity Inova NMR with a 5 mm Varian AutoTuneX probe, 1000 transients, 32768 points, and a spectral window of -15 to 5 ppm.

The identity of the \([^{16}\text{O}_4]\)-\(\text{P}_i\) peak was verified by spiking the reaction with \(\text{K}_2\text{PO}_4\) (pH 8.5) at a final concentration of 200 \(\mu\text{M}\) and collecting a \(^{31}\text{P}\)-NMR spectrum under identical conditions. A control reaction with unlabeled Trx-BalhA1 was carried out in an identical manner, except that the initial azoline hydrolysis event was carried out in [\(^{16}\text{O}\)]-\(\text{H}_2\text{O}\).

2.8.9 LCMS detection of ring formation. Reactions were trypsin (sequencing grade, Promega) digested in 50 mM \(\text{NH}_4\text{HCO}_3\) (pH 8.0) for 30 min at 30 °C before quenching with the addition of formic acid to a final concentration of 10%. Precipitate was removed by centrifugation and the sample was separated on an Agilent 1200 series HPLC with a Vydac C\textsubscript{18} 250 x 4.6 mm protein/peptide column. Samples were loaded at 5% solvent B [acetonitrile with 0.1% formic acid (v/v)] and were separated with a gradient of 2% solvent B per min for 20 min. Masses were
analyzed on a single quadrupole mass analyzer (G1956B) in positive ion scan mode using the manufacturer’s suggested parameters. Data was analyzed with ChemStation. The intensities of each of the ring states were calculated and used to determine the overall level of modification (n=3).

2.8.10 Malachite green phosphate detection assay. Samples were diluted 1:8 in water to a final volume of 80 µL to remove any background signal from unreacted ATP. This solution was transferred to a 96-well plate where the reaction was initiated by the addition of 20 µL malachite green working reagent (Bioassay Systems). Reactions were allowed to develop for 30 min before the absorbance was recorded at 620 nm. A standard curve was made from a known concentration of inorganic phosphate in appropriately diluted synthetase buffer. The absorbance for each of the samples was corrected for any background originating from the buffers.

Alternatively, the level of contaminating Pᵢ in the buffer exchanged TrxA-[¹⁸O]₅-BalhA1 sample was determined by conducting the malachite green assay with 80 µL of undiluted sample. In both cases the analysis of all samples was performed in triplicate.

2.8.11 LCMS detection of [¹⁶O]-ADP from ¹⁸O-water reactions. Reactions were analyzed on a 5 µm, 4.6 x 250 mm Luna HILIC column with a gradient from 90% solvent B (acetonitrile) and 10% solvent A (10 mM NH₄HCO₃, pH 9) at 5 min to 50:50 A and B at 30 min. Nucleotides were detected using a diode array detector at 260 nm. Single ions were monitored at m/z 426 and 428 on a single quadrupole mass analyzer (G1956B) in negative ion mode using the manufacturer’s suggested parameters to determine the ¹⁶O/¹⁸O ratio.

2.8.12 ADP isotope composition from [¹⁸O₅]-BalhA1 reactions. A portion (1 µL) of the of the TrxA-[¹⁸O]₅-BalhA1 sampled prepared for ³¹P-NMR analysis was reserved for MALDI-MS. This sample was spotted onto the MALDI target and overlaid with 1 µL of a saturated solution of 9-
aminoacridine in 50% aqueous acetonitrile. As a control, 1 mM ADP in ATP-free synthetase
buffer was spotted in an analogous fashion. Both samples were analyzed on a Bruker Daltonics
UltrafleXtreme MALDI-TOF in negative reflector mode.

2.8.13 Bioinformatics. Alignments were made using ClustalW2 using the standard parameters.\textsuperscript{47}
The percent identity between the Balh and Mcb biosynthetic enzymes were obtained from the
ClustalW2 alignment.

2.9 References


5309.


(29) Roy, R. S.; Belshaw, P. J.; Walsh, C. T. *Biochemistry* 1998, 37, 4125.


(34) Li, C.; Kappock, T. J.; Stubbe, J.; Weaver, T. M.; Ealick, S. E. *Structure* 1999, 7, 1155.


CHAPTER 3: AZOLINE-MEDIATED PEPTIDE BACKBONE LABELING AS A TOOL FOR STUDYING TOMM BIOSYNTHESIS

This chapter was taken from Dunbar & Mitchell\(^1\) and is reproduced with permission from the American Chemical Society.

I am grateful to Joel Melby for assistance with collecting MS/MS data and Eric Schmidt (University of Utah) for the generous gift of ulithiacyclamide and lissoclinamide 4. Trx-BalhA1 was cloned by Ambika Nadkarni.

Abstract

Current strategies for generating peptides and proteins bearing amide carbonyl derivatives rely on solid-phase peptide synthesis for amide functionalization. Although such strategies have been successfully implemented, technical limitations restrict both the length and sequence of the synthetic fragments. Herein I report the repurposing of a thiazole/oxazole-modified microcin (TOMM) cyclodehydratase to site-specifically install amide backbone labels onto diverse peptide substrates, a method I refer to as azoline-mediated peptide backbone labeling (AMPL). This convenient chemoenzymatic strategy can generate both thioamides and amides with isotopically labeled oxygen atoms. Moreover, I demonstrate the first leader peptide-independent activity of a TOMM synthetase, circumventing the requirement that sequences of interest be fused to a leader peptide for modification. Through bioinformatics-guided site-directed mutagenesis, I also convert a strictly dehydrogenase-dependent TOMM azole synthetase into an azoline synthetase. This vastly expands the spectrum of substrates modifiable by AMPL by allowing any *in vitro* reconstituted TOMM synthetase to be employed. To demonstrate the utility of AMPL for mechanistic enzymology studies, I demonstrate that AMPL is a useful tool for establishing the location of azolines on *in vitro* modified peptides. Furthermore, AMPL has utility in localizing and determining the number of azoline heterocycles in TOMM natural products.
3.1 Introduction

Stable amide labels are gaining prominence in the field of chemical biology as probes that minimally perturb polypeptide/structure and electronics at the site of installation. For example, $^{17}$O- and $^{18}$O- labeled amides are useful probes for the study of polypeptide dynamics by NMR and time-resolved 2D-IR, respectively.\(^2\)\(^-\)\(^8\) Thioamides, on the other hand, have garnered significant attention due to their ability to increase the protease stability of peptides, restrict conformational flexibility around the amide bond, photoisomerize from trans to cis upon excitation with 260 nm light, and act as minimal fluorescence quenchers for protein folding studies.\(^9\)\(^-\)\(^13\) While the utility of such labels is continually expanding, the backbone-incorporation techniques for these labels are currently limited to solid-phase peptide synthesis (SPPS).\(^2\)\(^-\)\(^9\),\(^11\)\(^-\)\(^14\) This process is expensive and time-consuming, as the generation of labeled residues must be done prior to SPPS, and also suffers from all the limitations of SPPS (e.g. peptide length and sequence restrictions). An alternative approach that would obviate the requirement for SPPS would be heterologous expression and purification of the desired polypeptide/protein, followed by the posttranslational installation of the backbone label. My previous work in peptide heterocyclization chemistry has suggested a route for installing oxygen isotopes and thioamides into polypeptide/proteins using the azoline heterocycle as an entry point.

Thiazoline and (methyl)oxazoline heterocycles are posttranslationally installed in the thiazole/oxazole-modified microcin (TOMM) class of peptidic natural products. These rings are formed from the cyclization of Cys, Ser and Thr residues onto the preceding amide carbonyl. Although the precise mechanism has yet to be fully elucidated, it is known that a two-protein cyclodehydratase complex (C and D proteins) performs the ATP-dependent cyclization of select
residues to afford azoline heterocycles.\textsuperscript{15,16} In the majority of TOMM natural products, a flavin mononucleotide (FMN)-dependent dehydrogenase (B protein) catalyzes the two-electron oxidation of the nascent rings to afford the aromatic azoles.\textsuperscript{15,17} In cases where oxidation does not occur, the azolines are often either sterically or electronically protected from nucleophiles (\textit{e.g.} thiostrepton, ulithiacyclamide, and plantazolicin; Fig. 3.1A).\textsuperscript{18-20} As azolines lack the aqueous stability of their aromatic counterparts,\textsuperscript{21-23} the dearth of unprotected azoline heterocycles in characterized TOMMs likely represents an evolutionary bias for compounds with increased stability. Even so, azoline-containing TOMM natural products have been reported to undergo partial hydrolysis during purification.\textsuperscript{20} Based on the susceptibility of azoline heterocycles towards nucleophilic attack, I hypothesized that azoline heterocycles could be used as moieties for the site-selective modification of peptide backbones.

Recently, we reported on the characterization of a TOMM synthetase from \textit{Bacillus} sp. Al Hakam (Balh), and demonstrated that the Balh cyclodehydratase is catalytically competent in the absence of the associated dehydrogenase.\textsuperscript{24,25} This enzyme complex had exceptional \textit{in vitro} activity and a predictable set of rules for the modification of diverse substrates.\textsuperscript{24} By coupling the promiscuous, yet highly predictable, nature of heterocyclization by the Balh cyclodehydratase with the lability of azoline heterocycles, I report a convenient and robust chemoenzymatic strategy for the site-selective installation of thioamides and isotopically labeled amides \textit{in vitro}, an approach that I refer to as azoline-mediated peptide backbone labeling (AMPL; Fig. 3.1B). The AMPL method has a broad substrate scope, including peptides lacking an N-terminal leader peptide, and can even be utilized to install backbone labels on large proteins with an appropriate substrate sequence. Furthermore, I demonstrate that this method is generalizable and can be
carried out with divergent TOMM cyclodehydratases, facilitating the expansion of the methodology to diverse peptide scaffolds.

To explore the potential value of the AMPL method, I applied this labeling strategy to study substrate processing by TOMM cyclodehydratases. Previously I provided substantial, but indirect, support for a mechanism of cyclodehydration in which ATP is used to phosphorylate the carbonyl oxygen of the amide preceding the cyclized residue.\textsuperscript{24,25} Through AMPL, I generated a cyclodehydratase substrate containing \textsuperscript{18}O labels at these amides and used this substrate to provide direct evidence of a phosphorylated intermediate during TOMM cyclodehydrations (see Chapter 2). The AMPL method also proved generally useful for localizing azolines on cyclodehydratase-treated peptides. The lability of azoline heterocycles renders their detection by mass spectrometry challenging. While this can be overcome via the addition of a dehydrogenase to convert the azoline heterocycles to stable azoles, TOMM clusters have been identified that lack a dehydrogenase,\textsuperscript{26} preventing the universal application of this solution. Further, dehydrogenases have an inherent regio- and chemoselectivity, which can differ from the selectivity of the cyclodehydratase (see Fig. 3.1A).\textsuperscript{20,24,27} The selective incorporation of \textsuperscript{18}O labels into peptide backbones at the sites of azoline heterocycles using the AMPL method circumvents this pitfall. As a proof of principle, I utilized AMPL to localize azoline heterocycles on diverse \textit{in vitro} cyclized substrates. More generally, I demonstrate that AMPL can be utilized for azoline heterocycle detection on natural products, which has the potential to assist in structure determination efforts.
Figure 3.1 | Azolines in natural products and AMPL overview. (a) Structures of select azoline-containing natural products are displayed with azolines colored red. The class of natural product to which the compound belongs is shown, along with the known bioactivity of the compound in parentheses. In each instance, the azoline heterocycles are either sterically (e.g. ulithiacyclamide, yersiniabactin) or electronically (e.g. plantazolicin, telomestatin) protected from nucleophilic attack. RiPP, ribosomally synthesized posttranslationally modified peptide; NRP, nonribosomal peptide. (b) A general scheme for TOMM heterocycle biogenesis is displayed. The electrophilic azoline heterocycle can be exploited for the installation of oxygen isotopes and thioamides using \([^{18}\text{O}]\text{-H}_2\text{O}\) and potassium hydrosulfide (KHS) as nucleophilic ring-opening agents, respectively. The mass shift relative to unmodified peptide is displayed underneath each product. C/D: ATP-dependent cyclodehydratase; B: FMN-dependent dehydrogenase. This figure is adapted with permission from Dunbar & Mitchell.\(^1\)
3.2 Oxygen Isotope Incorporation into Peptide Backbones Through Azoline Hydrolysis

I reasoned that since azolines hydrolyze under even mildly acidic conditions to yield an amide bearing a solvent-derived oxygen,$^20$ performing the hydrolysis in $[^{18}\text{O}]-\text{H}_2\text{O}$ would afford a peptide containing an $^{18}\text{O}$-labeled carbonyl directly N-terminal to every hydrolyzed azoline heterocycle (Fig. 3.1B). During the initial characterization of the Balh cyclodehydratase (complex of BalhC and BalhD), it was discovered that cyclodehydration occurred unabated in the absence of the dehydrogenase.$^{25}$ I surmised that it would be possible to generate appropriately labeled substrates by first cyclizing the BalhA1 substrate and subsequently hydrolyzing the nascent rings in mildly acidic $[^{18}\text{O}]-\text{H}_2\text{O}$. To test this hypothesis, thioredoxin tagged-BalhA1 (Trx-BalhA1; Fig. 3.2) was treated with BalhCD to afford the previously reported penta-azoline species (-90 Da; Fig. 3.3A).$^{25}$ Following lyophilization to remove $[^{16}\text{O}]-\text{H}_2\text{O}$, the penta-azoline peptide was hydrolyzed in $[^{18}\text{O}]-\text{H}_2\text{O}$ containing 0.5% (v/v) formic acid. As suspected, this treatment afforded a species with a mass 10 Da heavier than the starting material, suggesting that the substrate contained five $^{18}\text{O}$-labels, $[^{18}\text{O}_5]$-BalhA1. Moreover, the mass increase was predicated on azoline formation, as both an unmodified peptide and a peptide containing oxidized azole heterocycles were inert under the hydrolysis conditions employed (Fig. 3.4). This result was consistent with previous reports demonstrating that acid-catalyzed isotope exchange of carboxylic acids occurs slowly.$^{28,29}$ Following the removal of spent synthetase buffer, $[^{18}\text{O}_5]$-BalhA1 was subjected to a second treatment with BalhCD. As before, the peptide was modified to a penta-azoline state; however, ring formation was accompanied by a -100 Da mass shift, consistent with the loss of all five $^{18}\text{O}$ labels (Fig. 3.3A). Fourier transform (FT) MS/MS was utilized to definitively localize the $^{18}\text{O}$-labeled positions to the amides immediately N-terminal to the cyclized cysteines (Fig. 3.3B).
**Figure 3.2 | Peptide substrate sequences.** The substrates used in this chapter are listed above. Known heterocyclization sites are highlighted in red, with the exception of the first bis-heterocycle site in McbA, which is highlighted in orange. In McbA, the red underlined serine is the site of the ninth heterocycle that is installed \textit{in vitro} and found as a minor species during the microcin B17 heterologous expression.\textsuperscript{30} The predicted site of the azoline heterocycle in BalhA2 from previous studies is colored blue.\textsuperscript{24} Putative and known leader peptide cleavage sites are marked with a caret and an asterisk, respectively. As SagX is an unnatural TOMM analog (see reference 31 for description), no residues are colored.

![Diagram](image)

**Figure 3.3 | Site-specific incorporation of $^{18}$O labels into BalhA1. (a)** An overlay of MALDI-TOF MS spectra displaying the $^{18}$O exchange process. Mass labels correspond to the 1+ charge state and all mass shifts are reported relative to the pertinent non-cyclized species (spectra 1 and 3). The loss of the $^{18}$O labels in $[^{18}$O$_5$]-BalhA1 upon retreatment with BalhCD is evidenced by the loss of 100 Da (compared to -90 Da for BalhA1). $^{18}$O labels are represented by filled circles. **(b)** The FT-MS/MS spectrum of trypsin digested $[^{18}$O$_5$]-BalhA1. The b and y ions are colored based on the number of $^{18}$O labels present in the fragment (green, 5; purple, 4; orange, 3; blue, 2; red, 1; black, 0). Stars represent the location of the $^{18}$O labels.
Figure 3.4 | $^{18}$O labeling of BalhA1 is dependent on azoline heterocycles. To ensure that the installation of $^{18}$O-labels into BalhA1 was dependent on the presence of azoline heterocycles, both unmodified (1) and the penta-azole (2) forms of BalhA1 were treated with 0.5% formic acid in $[^{18}$O]-H$_2$O for 18 h. The resultant MALDI-TOF mass spectra demonstrated that in both cases, the peptides were insensitive to the $^{18}$O exchange conditions. These data suggested that this strategy for substrate labeling would be selective for azoline-containing peptides. All listed masses are for the 1+ charge state. The filled circle represents an $^{18}$O label.
3.3 Thiolysis of Oxazoline Heterocycles Allows for Site-Selective Thioamide Incorporation

The conversion of amides into thioamides is commonly achieved through the use of thionation reagents (e.g. Lawesson’s reagent).\textsuperscript{12,32} Such treatments suffer from exceedingly poor site-selectivity and are not suitable for the modification of larger peptides.\textsuperscript{33,34} An alternative strategy entails the thiolysis of an oxazoline heterocycle using H$_2$S (\textit{i.e.} sulfur nucleophile) for the site- and chemoselective formation of thioamides.\textsuperscript{35} Based on this, I hypothesized that AMPL could be extended accordingly. For initial tests, I chose a simplified BalhA1 substrate derivative designed to form only a single oxazoline heterocycle, BalhA1$_{NC}$-A40T (Fig. 3.2). All other Cys/Ser and Thr residues in the core peptide of BalhA1 have been mutated to Ala and Val, respectively, and are thus non-cyclizable (BalhA1$_{NC}$). Following cyclodehydration by BalhCD, BalhA1$_{NC}$-A40T was treated with potassium hydrosulfide (KHS; a solid, readily-handled alternative to H$_2$S) and analyzed by mass spectrometry (Fig. 3.5A). KHS treatment resulted in a 32 Da mass increase from the cyclized peptide, consistent with oxazoline thiolysis. Importantly, no modification was seen if the oxazoline was not formed prior to KHS treatment (Fig. 3.6).

While these data indicated that thioamide formation had occurred, attempts to locate the modification site by LC-MS/MS failed due to incomplete fragmentation of the peptide (data not shown). To demonstrate selective thioamide incorporation, a minimal substrate, BalhX (Fig. S2, \textit{vide infra}), was subjected to thiolysis and subsequent LC-MS/MS analysis. As expected, the methyloxazoline on BalhX efficiently converted to a thioamide (Fig. 3.7), and the modification could be definitively localized to the threonine heterocyclization site (Fig. 3.8).

Previous reports on the relative reactivity of thiazolines and oxazolines towards nucleophiles demonstrated that the former is substantially less reactive.\textsuperscript{21-23} To determine if this attenuated reactivity would preclude thiolysis of thiazoline heterocycles, an analogous study was
performed with BalhA1_{NC}-A40C (Fig. 3.2). The resultant spectra demonstrated that thioamide formation did not occur upon KHS treatment of the thiazoline heterocycle (Fig. 3.5B). Attempts to promote thiolysis with higher concentrations of KHS, higher temperature, pH modulation, and longer reaction times all failed to generate the thioamide product (data not shown). Thus, under the conditions employed, the AMPL-based thiolysis is remarkably chemoselective for (methyl)oxazoline heterocycles. As a further illustration of chemoselectivity, BalhA2 was converted to a tri-thiazoline, mono-oxazoline species by BalhCD and treated with KHS. As predicted, only a single thiolysis event was observed (Fig. 3.5C). Previously, we have demonstrated that BcerB (dehydrogenase from a closely related TOMM cluster in B. cereus) was unable to oxidize oxazoline heterocycles. As such, I predicted that treatment of BalhA2 with BcerB and BalhCD, coupled with subsequent thiolysis, would afford the tri-thiazole, mono-thioamide substrate. This was indeed the case, indicating that AMPL can utilized for the synthesis of differentially functionalized peptides (Fig. 3.5D).
Figure 3.5 | Thiolysis is specific for oxazoline heterocycles. MALDI-TOF MS overlays showing the thiolysis of (a) BalhA1-NC-A40T, (b) BalhA1-NC-A40C, (c) BalhA2 treated with the cyclodehydratase (CD), and (d) BalhA2 treated with both the cyclodehydratase and the dehydrogenase (BCD) are displayed. Mass labels correspond to the 1+ charge state and all mass shifts are reported relative to the pertinent unmodified species (spectrum 1).
Figure 3.6 | Potassium hydrosulfide (KHS) treatment of unmodified BalhNC-A40T. To ensure that the thiolysis of BalhNC-A40T is dependent on the presence of azoline heterocycles, unmodified substrate was treated with 100 mM KHS for 18 h at 25 °C. The resultant MALDI-TOF spectra demonstrate that the peptide is insensitive to the thiolysis conditions. These data suggest that this strategy for thioamide formation will be selective for azoline-containing peptides. All listed masses are for the 1+ charge state.

Figure 3.7 | BalhX thiolysis reaction and localization. (a) A MALDI-TOF spectrum of BalhCD-treated BalhX (a substrate lacking a leader peptide, Fig. 3.2) is displayed. The spectra show that upon treatment with BalhCD, the single cyclizable residue is converted to an azoline heterocycle. Following azoline formation, BalhX was subjected to KHS treatment. The last spectrum shows that the methyloxazoline on BalhX is converted to a thioamide. The masses correspond to the 1+ charge state of the peptide and the mass shifts relative to the unmodified peptide are displayed under each label. (b) An ion trap-MS/MS spectrum of the BalhX peptide following thiolysis is shown. The b and y ions along with the corresponding b and y ion symbols are colored based on the presence of the sulfur atom (+16 Da) in the fragment (red, contains the sulfur; black, does not). Based on the fragmentation, it is possible to definitively localize the thioamide moiety to the Gly (red) immediately upstream of the Thr that was cyclized in the starting material.
3.4 Leader Peptide-Independent Peptide Backbone Labeling

An intriguing feature of ribosomally synthesized posttranslationally modified peptide (RiPP) natural product biosynthesis is the broad substrate tolerance of the modification enzymes.\textsuperscript{36-39} While the basis of this promiscuity is under active investigation, it is well understood that the bipartite nature of the precursor peptides is a major contributor. RiPP modification enzymes recognize substrates primarily through binding an N-terminal leader peptide. In contrast to the extensively modified C-terminal core region, the leader peptide does not undergo enzymatic processing.\textsuperscript{40} Nonetheless, reports of leader peptide-independent RiPP processing enzymes have emerged.\textsuperscript{41} For example, some lanthipeptide biosynthetic enzymes can process their substrates without the presence of the leader peptide, although the enzymatic activity is always higher when the leader peptide is supplied \textit{in trans}.\textsuperscript{42-44} Although a TOMM processing enzyme had not yet been reported to modify peptides without a leader peptide, I hypothesized that leader peptide-independent processing might be achievable with the substrate tolerant Balh synthetase.\textsuperscript{24} To this end, a BalhA1 substrate derivative was generated lacking the leader peptide (BalhA1\textsubscript{core}; Fig. 3.2) and modification by the Balh synthetase complex was assessed. Although modification by BalhCD alone generated the expected penta-azoline species, the addition of the dehydrogenase (BcerB) resulted in a complex mixture of oxidation products (Fig. 3.8). This was unexpected, as previous studies had demonstrated that cyclodehydration and oxidation were tightly coupled during BalhA1 processing,\textsuperscript{24,25} and leader peptide binding was thought to be governed by the cyclodehydratase, not the dehydrogenase.\textsuperscript{17,31} In an effort to promote more complete oxidation, the dehydrogenase concentration was doubled in a second reaction. The primary product, a tetra-azole, mono-azoline species (Fig. 3.9), was subjected to MS/MS sequencing to localize the heterocycles. The fragmentation data demonstrated that the cyclized positions were identical to
the wild-type BalhA1 substrate and localized the single azoline to Cys20 (Cys45 in full-length BalhA1; Fig. 3.2, 3.9). During BalhA1 processing, Cys45 is known to be the penultimate azole to be installed, thus removing the leader peptide from the substrate appeared to disrupt both the order and efficiency of oxidation by the dehydrogenase.

While a decrease in oxidation rate suggested that substrate processing was dysregulated upon removal of the leader peptide, oxidation is not required for AMPL. To determine the effect of leader peptide removal on the rate of BalhA1$_{\text{core}}$ cyclodehydration, I carried out a purine nucleoside phosphorylase-coupled phosphate detection assay as previously described. Attempts to determine the kinetic parameters for BalhA1$_{\text{core}}$ were unsuccessful, as it was not possible to saturate the enzyme with substrate. This result was in agreement with the well-established model of leader peptide-mediated substrate recognition. As such, the initial velocity of BalhA1$_{\text{core}}$ processing at high substrate concentrations was compared to previously reported values for various substrate derivatives. The data demonstrated that BalhA1$_{\text{core}}$ was modified at a slower rate than BalhA1, but at a rate comparable to the naturally occurring BalhA2 substrate (Fig. 3.8B). The origin of the different processing rates for the naturally occurring BalhA1 and BalhA2 peptides derives from the presence of a Ser (Ser36), rather than the preferred Gly, immediately preceding the first cyclized residue (Cys37) in BalhA2. As the $K_M$ for BalhA1$_{\text{core}}$ was significantly larger than for BalhA2 (16 $\mu$M), the similar processing reaction velocity between the two substrates requires that the intrinsically faster processing rate of BalhA1 ($k_{\text{obs}} = 12.9$ min$^{-1}$) be retained upon the leader peptide removal.

To explore the generality of leader peptide-independent processing by the Balh synthetase, the modification of unnatural substrates was explored. Although not explicitly stated in the previous section, BalhX is a minimal substrate lacking a leader peptide (Fig. 3.2). This 17-
mer substrate contains a single heterocyclizable site that is cyclized by the Balh cyclodehydratase in route to thioamide formation (Fig. 3.7). I observed that the modification efficiency of BalhX greatly improved upon the addition of leader peptide \textit{in trans} (data not shown). This is reminiscent of studies with lanthipeptide modification machineries in which the leader peptide has been proposed to place the enzyme into an active conformation for substrate modification.\textsuperscript{42,43,46} In the case of the Balh cyclodehydratase, it remains unclear whether the addition of leader peptide \textit{in trans} fulfills a similar role.

While modification of this simple 17-mer substrate demonstrated that leader peptide-independent processing was achievable on unnatural sequences, BalhX was still derived from the BalhA1 core peptide (see Fig. 3.2). As such, the possibility remained that modification was due to inherent recognition of the core peptide by the Balh cyclodehydratase. Previous work demonstrated that the Balh synthetase can install azoles onto unnatural core peptides fused to the BalhA1 leader peptide.\textsuperscript{24} One such chimeric substrate was the fusion of a truncated version of the McbA (microcin B17 precursor peptide) core peptide to the BalhA1 leader peptide (BalhA1-McbA). In order to test for leader peptide-independent conversion, wild-type McbA (Fig. 3.2) was treated with BcerB and BalhCD. The resultant mass spectrum demonstrated that the Balh synthetase was able to modify the non-cognate McbA substrate (Fig. 3.10A), albeit with a decreased efficiency compared to reactions carried out with the chimeric BalhA1-McbA substrate.\textsuperscript{24} In order to further assess the generality of the BalhCD leader peptide-independent activity, a similar reaction was carried out with a streptolysin S-like precursor peptide (SagX; Fig. 3.2). As with the other substrates, treatment with the Balh synthetase resulted in modification of SagX (Fig. 3.10B). In the three aforementioned cases, the substrates contained heterocyclization sites that adhered to the previously reported Balh synthetase modification rules.
(Gly-Cys/Ser/Thr-Xaa, where Xaa ≠ Pro).\textsuperscript{24} To determine if these rules were retained upon leader peptide removal, an additional three leader peptide-free substrates lacking Balh cyclodehydratase cyclization sites were reacted with BalhCD (Fig. 3.11). In all cases, cyclodehydration was not observed indicating that the predictability of the Balh synthetase is retained upon leader peptide removal.

Figure 3.8 | BalhA\textsubscript{1}\textsubscript{core} processing. (a) An overlay of MALDI-TOF MS spectra displaying BalhCD- and BeerB/BalhCD-treated BalhA\textsubscript{1}\textsubscript{core}. Mass labels correspond to the 1+ charge state and all mass shifts are reported relative to the unmodified species (spectrum 1). (b) Relative processing rates for previously reported leader peptide-containing substrates of BalhCD are displayed alongside the processing rate for BalhA\textsubscript{1}\textsubscript{core}. Error bars represent standard deviation from the mean (n = 3).
Figure 3.9 | Azol(in)e localization on BCD-treated BalhA1core. (a) An FT-MS spectrum of BCD-treated BalhA1core is displayed. The calculated and observed masses for the tetra-azole, mono-azoline form of the peptide in the 4+ charge state are shown along with the ppm error of the measurement. (b) An FT-MS/MS spectrum of the species shown in panel (a). The b and y ions are colored based on the number of heterocycles found in the fragment (green, 5; purple, 4; orange, 3; blue, 2; red, 1; black, 0). Asterisks denote ions with a neutral loss of water. The yellow stars denote the sites of azole formation while the blue star denotes the lone azoline site. This data shows that the same positions in the BalhA1 core peptide are modified when the leader peptide is removed, and that the azoline oxidation is dysregulated on substrates lacking a leader peptide.

Figure 3.10 | McbA and SagX processing by a non-cognate BCD complex. A MALDI spectrum of BcerB/BalhCD-treated McbA (a) and SagX (b) is shown. All displayed masses are for the 1+ charge state. The mass shift relative to the unmodified species is shown below the mass of the peak in the BCD-treated sample. These spectral overlays demonstrate that the Balh synthetase was able to modify non-cognate substrates irrespective of the presence of a Balh leader peptide.
Figure 3.11 | Leader peptide-free processing follows previously established rules. (a) Three peptides lacking this optimal heterocyclization sequence are displayed with heterocyclizable residues colored blue. (b) The MALDI-TOF spectra of the BalhCD-treated peptides are shown. In all cases, treatment with the Balh cyclodehydratase failed to result in substrate processing, even at substrate:enzyme ratios of 1:1. Apart from the terminal Ser on ClosA-LP, all of the heterocyclizable residues are downstream of a non-Gly residue and are not processed. The C-terminal Ser on ClosA-LP is likely not processed due to the lack of a +1 residue. All masses correspond to the 1+ charge state.
3.5 Engineering of Dehydrogenase-Dependent Synthetases for Backbone Labeling

In principle, AMPL could be carried out with any TOMM synthetase. However, of the six TOMM synthetases that have been reconstituted in vitro, only the cyanobactin (PatD, patellamide; TruD, trunkamide) and Balh modification enzymes have been shown to function without the addition of the dehydrogenase. The AMPL methodology is not inherently compatible with TOMM cyclodehydratases that function in a dehydrogenase-dependent fashion, as the azoline heterocycles synthesized by these synthetases are typically transformed to hydrolytically inert azoles. In the absence of a crystal structure of a TOMM synthetase, I posited that such dehydrogenase-dependent cyclodehydratases would likely require precise quaternary interactions in order to be competent catalysts; however, I reasoned that it would be quite unlikely that downstream azoline oxidation was required for efficient cyclodehydration. In order to co-opt a dehydrogenase-dependent cyclodehydratase for the production of only azoline heterocycles, I used bioinformatics to aid in the identification of conserved residues in TOMM dehydrogenases that were important for azoline oxidation. In an attempt to avoid mutations that may structurally destabilize the protein, positions predicted to be directly involved in FMN-binding were ruled out. By aligning divergent dehydrogenases, two highly conserved residues were identified that I hypothesized to be important for catalysis but not for FMN-binding (Fig. 3.12). To test this hypothesis, I prepared two single-point mutants (K201A and Y202A) of the microcin B17 dehydrogenase (McbC), which has been long known to be required for cyclodehydration during microcin B17 processing. Following purification, McbC-K201A and -Y202A were found to co-purify with FMN at levels similar to wild-type (Fig. 3.13). As designed, the McbC-K201A and -Y202A permitted the production of azoline, but not azole, heterocycles when co-reacted with the Mcb cyclodehydratase (Fig. 3.14). However, instead of
converging to the known 9-azole form of McbA,\textsuperscript{30} reactions carried out with the mutant dehydrogenases prematurely terminated after installing just two azolines. Due to the aforementioned lability of azolines, FT-MS/MS localization was not possible using standard methods (\textit{vide infra}). Regardless, the use of this straightforward bioinformatics-guided strategy for conversion of dehydrogenase-dependent cyclodehydratases to azoline synthetases has the possibility to be a general strategy for synthetase repurposing.

**Figure 3.12 | TOMM dehydrogenase alignments.** A Clustal Omega alignment of dehydrogenases from diverse TOMM clusters is displayed (for a larger alignment see reference \textsuperscript{47}). Residues predicted to be involved in coordinating the phosphate moiety of FMN (based on the crystal structure of a non-TOMM “nitroreductase”, PDB entry 3EO7) are colored blue. The residues highlighted in red were chosen for the mutagenesis due to their high level of conservation and the prediction that they were not directly involved in FMN binding. In McbC, these residues are K201 and Y202.

| BcerB       | LNRKSVEE----------LEIRTRFETLSNLHFSGYINKP----------H |
| ClosB       | IKRSVRR------------YSS--KSMLNDVANIFYQTGICDEVEYNLENGKFKKL |
| McbC        | VINISSHNFSERLPGFCDKKNLSRITIEKLKVNAF----------SSPDGVR |
| LlsB        | LEQRHSTR-----------FYV--ETMDLSFTSNIQFSGLSTKLKVYN------DLQSTT |
| PdgB        | IVTRRIRT---------FSY--EPIKLEQSVLKLSSGVVLQD-----------EHSYIH |
| PatG        | IETRQSIRE--------YDD--YPIITIEQGELYRCAVTEVY------QMEEGEVSR |
| PznB        | IONRSLQ-----------FSG--GSTTQLSTILQSYGLER------EGPR |
| SagB        | IYKRSHRQ----------FSD--RQMLQDLSNIYYACGVSQSAS--IRDGASDKitL |

**BcerB**

|       | SAAPSGKGYPIINYIAVFN--VENL--EIQYIYDREQQVLDMLRRG----------F |
| ClosB | RANPSAGGLYPIELYVMKS--IKDL--EDGITYYYPYSHGLPKVNEALKIENFAEF |
| McbC  | RYPSSGGAGYPIEVLCLSENTEQAGTNYHLYLPLQALEPVTACNQSLRSGLS-- |
| LlsB  | RHYSSGGGLYPIDVFLYINN--ISGI--AKGIIYKQPYYTHLSLHPLDVK--IDVESFFVG |
| PdgB  | RSFPTAGGLNSCHVYLISN--VDDL--PGFSSYDYDTHELIKIEEYQI-----SQKNEF |
| PatG  | RYPYGCGARYELIQPVQQ--CEGL--DAGLYHYDPNQLQIEQIADNYPEVAA-----L |
| PznB  | RPIPSGGALPYDLYVSNK--VDSL--EKGHGFDPYRKLVLHGEYSE------EDFRG- |
| SagB  | RNACSSGGGLYPILLYFYARN--ISKL--IDGFYELPYQHALRCYRHSS--EENVRDFAEY *

**BcerB**

|       | RESINNLVNDTHISSLSSIFMFHAANLQTSSKYOARGYKLHLMGHSQNLNYLSSAQ |
| ClosB | -------------GVIGNANLNVYFVNVPLKNSRKYGDAGFSGYALITEMQALQLVSTAL |
| McbC  | --------------GDSERLGPHALCYCIFKEALKFKYRHYRGMETGSMQNAVLDADQ |
| LlsB  | NDITSDSMMFCVFPGYSINKNYKYGELSLNTVFELVGISHNFDVLCHSV |
| PdgB  | LTDLVKVLGQEWIRTAGLILIIYTDYSMKIRLKYGDGRYRLLLLEAGHIMQFNYIASL |
| PatG  | TADARLSSGQEQDTPQVILLTARTFGLRCFYYKSKLAYALVKHKGVLEYNFLVATDM |
| PznB  | --------------IMLQREEAVKDFSAVIISASPWRSRFKYGHRSYFIFIEAGHLMQNMILLATAQ |
| SagB  | --------------GAINTACNIIIIIYVYHKNTKGYGNQATAFIESGEIAQNIQLTATAL |

---

81
Figure 3.13 | Stability and FMN loading of McbC-K201A and -Y202A. (a) A Coomassie-stained SDS-PAGE gel of amylose resin purified WT, K201A, and Y202A MBP-McbC is displayed along with a table listing the yield of each protein. While the Y202A mutant was lower yielding, the SDS-PAGE gel indicated that the protein was not structurally destabilized to any significant extent, as would be evidenced by faster migrating proteolytic degradation bands. (b) In order to determine the effect, if any, the K201A and Y202A mutations had on FMN-binding in McbC, a UV-vis absorbance spectrum was acquired for each protein. The ratio of the absorbance at 450 and 280 nm was used to evaluate FMN-loading. While Y202A has a slightly lower molar absorptivity at 280 nm than WT McbC, the difference is less than 2% and within the error of the measurement. The graph shows the loading for each of the mutants has not been drastically effected by the mutations, relative to WT McbC. Error bars represent the standard deviation from the mean (n = 3).

Figure 3.14 | Azoline formation by the microcin B17 synthetase. An overlay of MALDI-TOF spectra displaying McbA processing by the Mcb cyclodehydratase (McbBD) with either wild-type (WT) or the mutant (K201A, Y202A) Mcb dehydrogenases (McbC) is shown. Mass labels correspond to the 1+ charge state and all mass shifts are reported relative to the unmodified species (spectrum 1). In the WT McbC reaction (spectrum 3), partially processed intermediates with 5-8 azoles (labeled -20x) are seen in addition to the fully modified, 9-azole substrate.
3.6 Azoline Localization Provides Insights into Cyclodehydratase Substrate Processing

The acid-labile nature of azoline heterocycles has hindered their localization due to the acidic conditions often employed for efficient detection by mass spectrometry. As such, AMPL has the potential to aid in a variety of azoline biosynthetic studies by providing a rapid and robust method for azoline localization that takes advantage of acid lability, rather than suffering from it. To demonstrate the usefulness of such a strategy, I began by investigating azoline localization on BalhA2 (Fig. 3.2). Upon BcerB/BalhCD treatment, BalhA2 was converted to a tri-azole, mono-azoline form (Fig. 3.5D). While previous efforts localized the three azoles to the three N-terminal Cys residues, the azoline heterocycle could only be generally localized to a region of BalhA2 containing three heterocyclizable sites; Thr34, Thr35 and Ser36 (see Fig. 3.2). Through AMPL, I generated $[^{18}O]_{4}$-BalhA2 and was able to definitively localize the fourth heterocycle to Thr34 using FT-MS/MS (Fig. 3.15). This result is in accord with the known Balh cyclodehydratase selectivity rules (the residue N-terminal to Thr34 is Gly33).

Encouraged by the successful utilization of AMPL for identifying previously indeterminable sites of azoline formation, I returned my attention to the di-azoline McbA species generated from reactions containing McbBD and either McbC-K201A or -Y202A (Fig. 3.14). Previous work on the order of ring formation in microcin B17 biogenesis demonstrated that the first two rings to be cyclized are Cys41 and Ser40, respectively. With this in mind, I predicted that the two azolines installed on McbA in the mutant dehydrogenase assays would be located at the site of this bisheterocycle. A combination of iodoacetamide labeling (Fig. 3.16) and the aforementioned $^{18}$O-labeling strategy coupled with FT-MS/MS, demonstrated that one of the cyclized residues was indeed Cys41, but that the other azoline was installed at either Cys48 or Cys51 (incomplete fragmentation between the two Cys precluded exact localization; Fig. 3.17,
3.18). While it has been shown that both Cys48 and Cys51 can be installed on mutant substrates lacking the Ser of the first bisheterocycle site (Ser40A),\textsuperscript{50} never before has a substrate with an intact, wild-type bisheterocycle site been observed to stop with only one of those heterocycles formed.\textsuperscript{49,51,52} These results suggest that in microcin B17 biosynthesis, either the dehydrogenase controls the regioselectivity of the cyclodehydratase or that azoline oxidation is required for correct substrate processing. While the data cannot distinguish between these two possibilities, it has been previously demonstrated that proper installation of the earlier rings is required for formation of the later rings, potentially through conformational control of the substrate.\textsuperscript{24,50} As such, I favor a mechanism where azoline oxidation induces the requisite conformational restraints to facilitate modification of downstream residues. In this model, Cys41 would be heterocyclized first, but without oxidation, Ser40 is unable to be cyclized. This “roadblock” takes the cyclodehydratase off-pathway and results in the cyclization of a second Cys (either 48 or 51), at which point processing stalls. Although further study will be required to definitively determine the collaborating role that the dehydrogenase plays in cyclodehydration, these data provide yet another glimpse into the exquisite orchestration of substrate handling during TOMM biosynthesis and demonstrate the utility of the AMPL method.
Figure 3.15 | Azoline localization on BalhA2 via $^{18}$O labeling. Following treatment with BalhCD and $^{18}$O-hydrolysis, BalhA2 was subjected to FT-MS/MS. (a) An intact mass spectrum for [${^{18}}$O$_4$]-BalhA2 in the 4+ charge state is displayed. The monoisotopic masses for the most prominent $^{18}$O-labeled species are displayed and colored based on the number of $^{18}$O labels (purple, 4; orange, 3; blue, 2). Note that the isotope distribution of the [${^{18}}$O$_4$]-peak is skewed to higher m/z values, indicating that a [${^{18}}$O$_5$]-labeled species also exists. The ppm error for the major species is displayed. (b) A MS/MS spectra of the [${^{18}}$O$_4$]-BalhA2 species from panel (a) demonstrates that the oxazoline site is the first Thr, as was previously hypothesized. The b and y ions are colored based on the number of $^{18}$O present in the fragment (purple, 4; orange, 3; blue, 2; red, 1; black, 0). Stars represent the sites of $^{18}$O labels from hydrolysis of the azoline heterocycles and asterisks label peaks with neutral water loss. Based on the y ion series, the C-terminal carboxylate bears a single $^{18}$O label (denoted by red residue). This is a consequence of the sample preparation and is a known reaction occurring on acidified tryptic peptides.
Figure 3.16 | Iodoacetamide labeling of McbA following treatment with McbC variants. To determine if the azolines installed on McbA following treatment with McbBD and mutants of McbC were thiazolines or oxazolines, the products were subjected to iodoacetamide (IA) labeling. In the case of unmodified McbA (McbA + McbBD, top spectrum), up to five IA labels were installed. As McbA contains four Cys residues (see Figure S2), the presence of a low intensity peak corresponding to a substrate with five IA labels is likely due to the IA-dependent labeling of amines. Upon treatment with the all wild-type McbBCD (WT) complex, the peptide is no longer reactive towards IA, as all cysteines have been converted to thiazoles (Thz). In contrast, for samples treated with McbBD and either McbC-K201A or -Y202A, the di-azoline substrate undergoes two labeling events. As with the unclyclized McbA sample, both the WT and K201A/Y202A treated peptides have a low intensity peak corresponding to the non-specific IA-labeling of amines. Masses correspond to the 1+ charge state of the species and the number of IA labels, and Thz, oxazole (Oxz) and thiazoline (ThH) heterocycles are displayed beneath each mass.
Figure 3.17 | Azoline localization on McbA following treatment with McbC-K201A. Azoline heterocycles on McbA were subjected to $^{18}$O hydrolysis to form $[^{18}\text{O}_2]$-McbA, and the labeled peptide was analyzed by FT-MS/MS to locate the sites of azoline formation. The b and y ions are colored based on the number of $^{18}$O present in the fragment (blue, 2; red, 1; black, 0). Asterisks indicate peaks with a neutral loss of water. Based on the fragmentation pattern and the iodoacetamide labeling data (Figure S16), one of the thiazoline heterocycles can be definitively localized to Cys41 (red). The other thiazoline is formed at either Cys48 or Cys51 (blue). Incomplete fragmentation in this region precluded the precise localization of the modification site.
Figure 3.18 | Azoline localization on McbA following treatment with McbC-Y201A. Azoline heterocycles on McbA were subjected to $^{18}$O hydrolysis to form $[^{18}\text{O}_2]$-McbA, and the labeled peptide was analyzed by FT-MS/MS to locate the sites of azoline formation. The b and y ions are colored based on the number of $^{18}$O present in the fragment (blue, 2; red, 1; black, 0). Asterisks indicate peaks with a neutral loss of water. Based on the fragmentation pattern and the iodoacetamide labeling data (Figure S12), one of the thiazoline heterocycles can be definitively localized to Cys41 (red). The other thiazoline is formed at either Cys48 or Cys51 (blue). Incomplete fragmentation in this region precluded the precise localization of the modification site.
3.7 Azoline Stoichiometry of Natural Products Obtained via Backbone Labeling

The labeling of azoline heterocycles using selective hydrolysis is not restricted to *in vitro* reactions. In principle, the AMPL method for azoline labeling could be extended to determine the number and location of azoline rings on natural products. To demonstrate this, the azoline-containing TOMMs plantazolicin, ulithiacyclamide, and lissocinamide 4 (Fig. 3.1A) were subjected to the AMPL method for $^{18}$O-labeling. In all cases $^{18}$O labeling was achieved (Figure 3.19-3.21), demonstrating that AMPL can be utilized to derivatize both thiazolines and methyloxazolines on natural products. This suggests that the method could be generalized to provide the number and type (thiazole vs. oxazole) of azoline heterocycles in any natural product. Given that emerging natural products are most often available only in vanishingly small quantities, AMPL may prove valuable in assisting with the structural elucidation of these compounds.
Figure 3.19 | $^{18}\text{O}$-labeling of plantazolicin. A MALDI-TOF spectrum demonstrating that AMPL can be utilized to selectively $^{18}\text{O}$-label the TOMM natural product plantazolicin (Pzn; Fig. 3.1A) is displayed. As expected based on the structure of Pzn, exposure to acid in $[^{18}\text{O}]$-$\text{H}_2\text{O}$ results in the incorporation of a single $^{18}\text{O}$ label (+20 Da) into Pzn due to the hydrolysis of the single methyloxazoline ring. The ratio between the $^{18}\text{O}$- and $^{16}\text{O}$-labeled hydrolysis species (m/z 1356 and 1354, respectively) is due to the presence of ~10% $[^{16}\text{O}]$-$\text{H}_2\text{O}$ in the sample. All masses are for the 1+ charge state and the mass shift relative to the unmodified species (m/z 1336) is shown below the mass label.
Figure 3.20 | $^{18}$O-labeling of ulithiacyclamide. A MALDI-TOF spectrum demonstrating that AMPL can be utilized to selectively $^{18}$O-label the TOMM natural product ulithiacyclamide (Uli; Fig. 3.1A) is displayed. Based on the structure of Uli, exposure to acidified [$^{18}$O]-H$_2$O should result in the incorporation of two $^{18}$O-labels (m/z 803.2). Although this peak is not detected, both the +Na (m/z 825.2) and +K (m/z 841.2) species are observed. All masses are for the 1+ charge state and the mass shift relative to the unmodified species (m/z 763.2) is shown below the mass label.
Figure 3.21 | $^{18}$O-labeling of lissoclinamide 4. A MALDI-TOF spectrum demonstrating that AMPL can be utilized to selectively $^{18}$O-label the TOMM natural product lissoclinamide 4 (Lis; Fig. 3.1A) is displayed. While Lis contains two azoline heterocycles (a thiazoline and a methylloxazoline) a spectrum of the starting material (black) indicates that a majority of the species is hydrolyzed to begin with (+18 Da). As such, treatment with acidified [${}^{18}$O]H$_2$O is expected to generate a species containing both a $^{16}$O and an $^{18}$O label ($m/z$ 780.2). Indeed this dual labeled product is obtained. All masses are for the 1+ charge state and the mass shift relative to the unmodified species ($m/z$ 742.2) is shown below the mass label.
3.8 Summary and Outlook

Prior to this work, the primary strategy for generating thioamides and oxygen isotope-labeled amides was SPPS. The method reported herein, which I refer to as azoline-mediated peptide backbone labeling (AMPL), provides the first chemoenzymatic strategy for accessing such derivatives. AMPL is compatible with both large and small peptides and peptide substrates fused to large proteins. Conveniently, labeling can be performed in the same pot as cyclodehydration, minimizing the amount of sample handling required during derivatization. In exploring the scope of AMPL, I have not only identified leader peptide-independent processing by a TOMM cyclodehydratase, but also converted the naturally azole-producing Mcb synthetase into an azoline-only synthetase. Although my studies focused on the Balh and Mcb enzymes, AMPL can theoretically be applied to any TOMM synthetase, provided the enzymes can function independently of a dehydrogenase or the dehydrogenase can be inactivated without disrupting cyclodehydratase activity. The major limitation of the AMPL methodology is the inherent local sequence selectivity of the cyclodehydratase, which can be overcome by selecting the appropriate synthetase to modify a sequence of interest. As such, the in vitro reconstitution and characterization of additional TOMM cyclodehydratases will expand the sequence space accessible to the AMPL method.

In addition to providing a convenient strategy for generating polypeptides bearing unnatural amide derivatives, the utility of the AMPL method should also be of interest to enzymologists and natural products chemists, as I have used this labeling strategy to both investigate the mechanism of cyclodehydration (see chapter 2) and elucidate the localization of previously unresolvable heterocyclization sites. For the latter, I have focused the use of AMPL on identifying the location of enzymatically installed azoline heterocycles and through this
simple labeling strategy have provided new insights into substrate processing by divergent TOMM cyclodehydratases. Furthermore, I demonstrate that the AMPL method can be extended to the modification of natural products, which opens up the possibility of using AMPL for both azoline-containing natural product discovery and structure elucidation.

3.9 Experimental

3.9.1 General methods. Unless otherwise specified, all chemicals were purchased from Sigma or Fisher Scientific. DNA sequencing was performed by either the Roy J. Carver Biotechnology Center (UIUC) or ACGT Inc. Restriction enzymes were purchased from New England Biolabs (NEB), Pfu Turbo was purchased from Agilent, and dNTPs were purchased from either NEB or GenScript. Oligonucleotide primers were synthesized by Integrated DNA Technologies (IDT). Unless otherwise stated, all proteins and substrates were used as MBP fusions to circumvent solubility issues.

3.9.2 Protein overexpression and purification. Apart from the Trx-BalhA1 substrate, all proteins were overexpressed as tobacco etch virus (TEV) protease- or thrombin-cleavable fusions to maltose-binding protein (MBP) and purified by amylose affinity chromatography as previously reported. Unless otherwise stated, all reactions were carried out with tagged substrates and enzyme to avoid solubility issues.

3.9.3 Overexpression and purification of Trx-BalhA1. BL21(DE3-RIPL) cells were transformed with a pET32b plasmid containing Trx-BalhA1 (E. coli thioredoxin fusion protein that bears an N-terminal His-tag). Cells were grown overnight on Luria-Bertani (LB) plates with 100 µg/mL ampicillin. Single colonies were picked for starter cultures containing 100 µg/mL ampicillin and 34 µg/mL chloramphenicol and were grown at 37 °C. A 10 mL overnight culture
was used to inoculate 1 L of LB containing 100 µg/mL ampicillin and 34 µg/mL chloramphenicol. Cultures were grown to an optical density at 600 nm (OD$_{600}$) of 0.8 before induction with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 1.5 h at 22 °C. Subsequently, cells were harvested at 3000 × g for 15 min, washed with TBS (Tris buffered saline; 10 mM Tris pH 7.5, 150 mM NaCl) and stored at -20 °C for up to one week before use.

Cell pellets were resuspended in lysis buffer [50 mM Tris pH 8, 500 mM NaCl, 25 mM imidazole, 2.5% glycerol (v/v), 0.1% Triton X-100 (v/v)] containing lysozyme (4 mg/mL), leupeptin (2 µM), PMSF (200 µM), benzamidine (2 mM), and E64 (2 µM). After a 30 min incubation at 4 °C, cells were disrupted via sonication 3 × 30s with 10 min equilibration periods at 4 °C. The insoluble debris was removed from the sample via centrifugation at 20,000 × g for 45 min. The resulting supernatant was applied to pre-equilibrated Ni-NTA resin (Qiagen, 5 mL resin per L of cells). The column was washed with 10 column volumes of lysis buffer, followed by 5 column volumes of wash buffer (lysis buffer with NaCl concentration increased to 300 mM and Triton X-100 omitted). The His-tagged proteins were eluted using 4 column volumes of elution buffer (wash buffer with 150 mM NaCl and 200 mM imidazole) and the eluent was concentrated using an appropriate Amicon Ultra centrifugal filter (Millipore). A 100-fold buffer exchange with storage buffer [50 mM HEPES pH 7.5, 150 mM NaCl, 2.5% glycerol (v/v)] was performed in the filtration device before a final concentration and storage. After loading onto the column, all buffers used contained 1 mM tris-(2-carboxylethyl)-phosphine (TCEP). Protein concentration was determined both by the 280 nm absorbance and a Bradford colorimetric assay (Thermo Scientific). Purity was visually assessed by Coomassie-stained SDS-PAGE gel.

3.9.4 BalhA1 $^{18}$O-labeling studies. Reactions containing 100 µM MBP-BalhA1, 2 µM MBP-BalhC/D and 0.2 µg/mL TEV protease (to remove the MBP tags) were carried out in synthetase
buffer [50 µM Tris pH 7.5, 125 mM NaCl, 20 mM MgCl₂, 3 mM ATP, and 10 mM dithiothreitol (DTT)] at 25 °C for 18 h. Subsequently, the samples were frozen in liquid nitrogen, lyophilized to dryness to remove all [¹⁶O]-H₂O, and the resultant solids were reconstituted to half of the initial volume with 97 atom % [¹⁸O]-H₂O. To this solution, 10% formic acid in [¹⁶O]-H₂O (v/v) was added to a final concentration of 0.5 % (v/v). This resulted in a final isotopic enrichment of ~92 atom %. Azoline hydrolysis was allowed to proceed for 18 h at 25 °C. The sample was then desalted via C₁₈ ZipTip (Millipore) according to the manufacturer’s instructions and ¹⁸O incorporation was analyzed on a Bruker Daltonics UltrafleXtreme MALDI-TOF spectrometer. Spectra were obtained in positive reflector mode using α-cyano-hydroxycinnamic acid (CHCA) as the matrix. Control samples with either unmodified BalhA1 or fully oxidized BalhA1 (5 azoles instead of 5 azolines) were treated in an analogous fashion, except that MBP-BalhC/D were omitted or 2 µM MBP-BcerB was added, respectively.

For the remodification of [¹⁸O₅]-BalhA1, the labeled substrate was C₁₈ ZipTip desalted, to remove contaminating ADP and Pᵢ, and subjected to a second treatment with fresh BalhC/D under identical conditions to the initial reaction. Following an 18 h reaction at 25 °C, the sample was analyzed by MALDI-TOF MS as described above.

3.9.5 FT-MS/MS localization of ¹⁸O labels. Following azoline-mediated ¹⁸O label incorporation, peptides were trypsin (sequencing grade, Promega) digested in 50 mM NH₄HCO₃ (pH 8.0) for 30 min at 37 °C before quenching with formic acid at a final concentration of 10 % (v/v). Precipitate was removed via centrifugation at 11,000 × g and the samples were analyzed via reverse-phase, high-performance liquid chromatography-Fourier transform mass spectrometry (FT-MS). All FT-MS was carried out on an Agilent 1200 HPLC system with an autosampler connected directly to a Thermo Fisher Scientific LTQ-FT hybrid linear ion trap.
operating at 11 T. Separations were conducted using a 1 × 150 mm Jupiter C_{18} column (300 Å, 5 
µM, Phenomenex) with 0.1 % formic acid (A) and acetonitrile containing 0.1 % formic acid (B) 
as the mobile phases. The LC gradient was held at 5 % B for 2 min before being increased to 50 
% B over 43 min. All methods contained a full FT-MS scan followed by MS/MS of selected ions 
conducted in the FT. FT-MS scan parameters: minimal target signal counts: 5,000; resolution: 
25,000; $m/z$ range: variable. FT-MS/MS parameters: minimum target signal counts: 5,000; 
resolution: 25,000; $m/z$ range: dependent on target $m/z$; isolation width: 5 $m/z$; normalized 
collision energy: 35; activation q value: 0.25; activation time: 30 ms. Data analysis was 
conducted using the Qualbrowser application of Xcalibur v 2.2 (Thermo Fisher Scientific).

3.9.6 Processing of leader peptide-free substrates. 100 µM MBP-BalhA_{1/core} was reacted with 
10 µM MBP-tagged BalhC/D, either with or without 10 µM MBP-BcerB, and 0.2 µg/mL TEV 
protease in synthetase buffer for 18 h at 25 °C. Samples were then desalted and analyzed by 
MALDI-MS as above. Processing of the unnatural substrates, MBP-SagX, MBP-McbA, and 
BalhX, were carried out in an analogous fashion with the following minor variations: all 
substrate concentrations were reduced to 50 µM, proteolytic cleavage in the MBP-McbA sample 
was carried out using 0.02 µg/mL thrombin (from bovine plasma), and MBP-tagged BalhA_{1} 
leader peptide was added to the BalhX sample at a final concentration of 50 µM.

3.9.7 Rate of BalhA_{1/core} processing. The rate of MBP-BalhA_{1/core} processing was determined 
using slight modifications to a previously described method.\textsuperscript{25} Briefly, MBP-BalhC and MBP-
BalhD were added to a cuvette for a final concentration of 1 µM each. Reactions were initiated 
via the addition of a mixture of 100 µM MBP-BalhA_{1/core}, 200 µM 2-amino-6-mercapto-7-
methylpurine riboside (Berry and Associates), and 0.2 units of purine nucleotide phosphorylase 
in synthetase buffer at 25 °C. Reaction progress was monitored by the change in absorbance at
360 nm on a Cary 4000 UV-vis spectrophotometer (Agilent). Initial rates of phosphate production were calculated based on the linear absorbance change during the first 3 min of the reaction and the extinction coefficient of the resulting guanine analog (11,000 M\(^{-1}\)cm\(^{-1}\)). Error is reported as the standard deviation of the mean (n = 3).

### 3.9.8 Mcb synthetase reactions.

20 µM MBP-McbA was reacted with 1 µM MBP-tagged McbB/D and either WT, K201A or Y202A McbC in synthetase buffer for 18 h at 25 °C. Reactions were initiated by the removal of MBP with the addition of 0.02 µg/mL thrombin. As above, heterocycle formation was monitored after desalting by MALDI-MS. A control sample lacking McbC was treated in an identical fashion.

### 3.9.9 Thiolysis of azoline heterocycles.

Heterocyclization reactions of MBP-BalhA\(_{1\text{NC}}\)-A40T and -A40C were carried out with 50 µM substrate and 1 µM MBP-BalhC/D in synthetase buffer for 18 h at 25 °C. For BalhX modification, the enzyme concentration was increased to 10 µM and MBP-tagged BalhA1 leader peptide was added \textit{in trans} to a final concentration of 50 µM. In the A40T and A40C samples, MBP was proteolytically removed by the addition of 2 µg/mL TEV protease. To each of samples, the proper volume of a fresh 1 M solution of potassium hydrosulfide (KHS; Strem Chemicals) was added to achieve a final concentration of 100 mM KHS, and thiolysis was allowed to proceed for an additional 18 h at 25 °C. The samples were then desalted and analyzed by MALDI-MS as above. For each substrate, a control reaction lacking MBP-BalhC/D was also subjected to thiolysis.

For reactions with BalhA2, 50 µM MBP-tagged substrate was incubated with 1 µM MBP-BalhC/D with and without 1 µM MBP-BcerB. MBP was removed via the addition of 2 µg/mL TEV protease. Following an 18 h reaction at 25 °C, the samples were treated with KHS and analyzed by MALDI-TOF MS as above.
3.9.10 BalhX Synthesis: The synthetic peptide was ordered from GenScript (www.genscript.com) and received as a lyophilized powder (92 % purity).

3.9.11 Dehydrogenase Alignments. Alignments were made with Clustal Omega using the standard parameters.54

3.9.12 McbC mutagenesis. The K201A and Y202A mutants of McbC (dehydrogenase involved in microcin B17 biosynthesis) were generated via site-directed mutagenesis of a pET15b plasmid containing MBP-McbC using QuikChange as per the manufacturer’s instructions. The primers used are listed in table A.1.

3.9.13 18O labeling of BalhA2 and McbC-K201A/Y202A reactions. 100 µM MBP-BalhA2 was incubated with 1 µM MBP-BalhC/D in synthetase buffer for 18 h at 25 °C. The resultant tetra-azoline peptide was trypsin digested with 0.02 mg/mL trypsin (sequence grade, Promega) for 2 h at 25 °C before being lyophilized to dryness to remove all [16O]-H2O. The resultant solid was reconstituted with ½ of the initial volume of 97 atom % [18O]-H2O and 10% formic acid in [16O]-H2O (v/v) was added to a final concentration of 0.5% (v/v). This resulted in a final isotopic enrichment of approximately 92 atom %. Azoline hydrolysis was allowed to proceed for 18 h at 25 °C. Following hydrolysis, the localization of the 18O labels was carried according to Fourier-transform MS/MS procedure listed in the main text.

In the case of the McbC-K201A and -Y202A reactions, 20 µM MBP-McbA was incubated with 1 µM MBP-tagged McbB/D and either K201A or Y202A McbC in synthetase buffer for 18 h at 25 °C. Reactions were then initiated by the addition of 0.02 µg/mL thrombin (from bovine plasma) to remove the MBP tags. Unlike the Balh cyclodehydratase, the Mcb synthetase requires the removal of MBP to be catalytically active. Processing of McbA with
concurrent thrombin cleavage was allowed to proceed for 18 h at 25 °C. Following cyclization, the sample was handled as described above.

3.9.14 MS/MS localization thioamides and azol(in)e heterocycles. To locate the thioamide on BalhX and the heterocyclized sites of BalhA1core, ion trap (IT)-MS/MS was used instead of FT-MS/MS. In both cases, samples were directly infused onto a Thermo Fisher Scientific LTQ-FT hybrid linear ion trap operating at 11 T using an Advion Nanomate 100. A full FTMS scan was conducted on all samples followed by IT-MS/MS of selected ions. FTMS scan parameters: minimal target signal counts: 5,000; resolution: 100,000; m/z range: variable. IT-MS/MS parameters: minimum target signal counts: 5,000; m/z range: dependent on target m/z; isolation width: 5 m/z; normalized collision energy: 35; activation q value: 0.25; activation time: 30 ms. Data analysis was conducted using the Qualbrowser application of Xcalibur v 2.2 (Thermo Fisher Scientific).

3.9.15 Iodoacetamide labeling of McbA in McbC-K201A and -Y202A reactions. Reactions were carried out with 20 µM MBP-McbA, 1 µM MBP-McbB/D (cyclodehydratase), and 1 µM of either WT, K201A or Y202A McbC (dehydrogenase) in synthetase buffer. A reaction lacking McbC was also performed as a control. MBP tag removal was performed concurrently with cyclodehydration by the addition of 0.02 µg/mL thrombin (from bovine plasma). Following an 18 h reaction at 25 °C, the samples were C18 ZipTip (Millipore) purified according to the manufacturer’s instructions and eluted into 50% acetonitrile in 10 mM MOPS pH 8.0 with 50 mM iodoacetamide. Labeling proceeded at 25 °C for 8 h before analysis on a Bruker Daltonics UltrafleXtreme MALDI-TOF in positive reflector mode with α-cyano-hydroxycinnamic acid (CHCA) as the matrix.
3.9.16 $^{18}$O-labeling of TOMM natural products. Purified samples of plantazolicin, ulithiacyclamide and lissoclinamide 4 were dissolved in $[^{18}\text{O}]$-H$_2$O to a final concentration of 250 µg/mL and each sample was acidified via the addition of 1% formic acid. Azoline hydrolysis was allowed to proceed for 40 h at 25 °C before the samples were ZipTip desalted and analyzed on a Bruker Daltonics UltrafleXtreme MALDI-TOF as described above.

3.9.17 BalhA1 core cloning. BalhA1$_{\text{core}}$ was amplified from the previously prepared pET28-MBP-BalhA1 vector using the primers listed in table A.1. Polymerase reactions were performed with PFU Turbo and the amplified product was digested with BamHI and NotI following gel extraction. The digested gene was PCR-purified and ligated into an appropriately digested, modified pET28 vector containing a tobacco etch virus (TEV) protease cleavable, N-terminal MBP-tag.

3.10 References


(53) Petritis, B. O.; Qian, W. J.; Camp, D. G., 2nd; Smith, R. D. *J. Proteome Res.* **2009**, *8*, 2157.

CHAPTER 4: BIOCHEMICAL AND STRUCTURAL CHARACTERIZATION OF THE YCAO SUPERFAMILY

This chapter was taken in part from Dunbar et al.\textsuperscript{1} and Dunbar & Chekan et al.\textsuperscript{2} All material is reproduced with permission from Nature Publishing Group.

I am grateful to Jonathan Chekan for the Ec-YcaO crystal structure, Courtney Cox for the maximum likelihood tree and Cytoscape network, Joel Melby for assistance with collecting MS/MS data, and Kelley Taylor for assistance with BalhD mutagenesis. Joel Melby cloned BalhNC-CCG*. Jonathan Chekan and Satish Nair generated all crystal structure figures and the LigPlot diagram.

Abstract

Despite intensive research, the cyclodehydratase responsible for azoline biogenesis in thiazole/oxazole-modified microcin (TOMM) natural products remains enigmatic. The collaboration of two proteins, C and D, is required for cyclodehydration. The C protein is homologous to E1 ubiquitin-activating enzymes, while the D protein is within the YcaO superfamily. Recent studies have demonstrated that TOMM cyclodehydratases phosphorylate amide carbonyl oxygens to facilitate azoline formation. Here I present the first experimental evidence that the D protein component of the heterocycle-forming synthetase (YcaO/DUF181 family member), formerly annotated as a docking/scaffolding protein involved in complex formation and regulation, is able to perform the ATP-dependent cyclodehydration reaction in the absence of the other TOMM biosynthetic proteins. Using an X-ray crystal structure of an uncharacterized YcaO from Escherichia coli (Ec-YcaO), I demonstrate that Ec-YcaO harbors an unprecedented fold and ATP-binding motif. This motif is conserved among TOMM YcaOs and is required for cyclodehydration. By identifying the active site of the TOMM cyclodehydratase, I also demonstrate that the proline-rich C-terminus of the D protein is involved in C protein recognition and catalysis. Together, these data pave the way for the characterization of the
numerous YcaO domains not associated with TOMM biosynthesis and prompt a reclassification of the enzymes involved in azoline installation.

4.1 Introduction

The YcaO/DUF181 (domain of unknown function) family of proteins is currently comprised of nearly 5000 members distributed across the bacterial and archaeal domains. Disparate functions have been ascribed to members of the YcaO family. In *Escherichia coli*, the deletion or overexpression of the eponymous YcaO protein (Ec-YcaO; Fig. 4.1A) suggested that it potentiates the methythiolation of ribosomal protein S12 and influences biofilm formation, respectively.\(^3,4\) However, a molecular explanation for these observations is currently unavailable.

Another YcaO-associated activity is the ATP-dependent cyclodehydration of Ser, Thr and Cys residues to azoline heterocycles, which is the defining modification of thiazole/oxazole-modified microcin (TOMM) natural products (Fig. 1.3).\(^5\) TOMMs display diverse structures and activities,\(^5,6\) with some implicated in bacterial pathogenesis,\(^7\) making the ~1000 bioinformatically identifiable TOMM YcaO proteins noteworthy members of the larger superfamily.

Although the TOMM YcaO domain was first implicated in cyclodehydration reactions in the mid-1990s,\(^8\) its exact role remains unclear.\(^1,9\) The function of the TOMM YcaO (D protein) is intimately linked to members of the E1 ubiquitin-activating enzyme family (C protein) found in canonical TOMM biosynthetic clusters.\(^8,10,11\) Underscoring this linked function, roughly half of known TOMM clusters express C and D as a single polypeptide.\(^5,11\) Studies on both fused and unfused cyclodehydratases have demonstrated that these domains are necessary and sufficient for TOMM azoline formation.\(^1,12\) Consequently, the C-D complex is referred to as the TOMM cyclodehydratase (or alternatively, heterocyclase\(^9,11\)). However, a definitive function for each
domain could not discerned as early studies on multiple TOMM cyclodehydratases were unable to observe activity from either protein in isolation.\textsuperscript{8,10,13}

Despite the challenges in separating the enzymatic activities of the C and D proteins, functional assignments have been postulated. Based on the ATP dependence of the reaction\textsuperscript{14} and the primary sequence/structural homology of C to members of the E1 ubiquitin-activating superfamily,\textsuperscript{9-11} which includes other ATP-utilizing enzymes (\textit{e.g.} MccB, ThiF and MoeB),\textsuperscript{15} it was assumed that C was responsible for cyclodehydration, while the uncharacterized YcaO (D protein) played a regulatory or scaffolding role.\textsuperscript{10,11} On this basis, the D protein has previously been referred to as a docking/scaffolding protein.\textsuperscript{10}

These functional assignments are called into question for two reasons. Firstly, no characterized TOMM C protein contains a bioinformatically recognizable nucleotide-binding site. Although the D protein from the microcin B17 synthetase has been suggested to be responsible for ATP utilization because it contains sequences of weak resemblance to nucleotide binding motifs,\textsuperscript{14} these residues are not conserved in other D proteins, which casts doubt on their importance in TOMM biosynthesis. Secondly, although YcaO proteins lack recognizable ATP-binding motifs, the presence of one or more YcaOs in the bottromycin and trifolitoxin biosynthetic clusters, which lack recognizable C proteins, suggests that the TOMM YcaOs may perform the requisite ATP-dependent cyclodehydration (Fig. 4.2).\textsuperscript{16-20}

Using the highly active TOMM synthetase from \textit{Bacillus} sp. Al Hakam (Fig. 4.1A), I report the first \textit{in vitro} activity for a member of the YcaO superfamily, and demonstrate that the TOMM D protein is an ATP-dependent cyclodehydratase. In collaboration with the Nair laboratory, I report the structure of a non-TOMM YcaO from \textit{E. coli} in various nucleotide-bound and nucleotide-free forms and demonstrate that the most conserved residues in YcaOs comprise
a previously uncharacterized ATP-binding motif. I show that these ATP-binding residues are critical for catalysis in TOMM YcaOs using BalhD as a model cyclodehydratase. I further identified the active site of TOMM cyclodehydratases and demonstrated that the conserved, proline-rich C-termini are involved in active site organization and C protein binding. These results strongly support a model where ATP utilization is a universal feature of YcaOs (TOMM and non-TOMM) and where TOMM YcaOs catalyze backbone activation chemistry.

![Figure 4.1](image)

**Figure 4.1 | YcaO gene clusters and sequence of precursor peptides used in this study.** (a) The local genomic environment for the *E. coli* non-TOMM YcaO (Ec-YcaO), *E. coli* TOMM YcaO (Mcb) and the *Bacillus* sp. Al Hakam TOMM YcaO (BalhD) is depicted along with the percent amino acid identity between Ec-YcaO and BalhD. Gene assignments are shown. (b) The sequences of the peptide substrates used in this study are shown. Color-coding: green, point mutations; orange, residues known to be cyclized *in vitro* (Balh)\(^2\) or *in vivo* (Mcb)\(^8\), blue hyphen, putative leader sequence cleavage site; blue caret, known leader sequence cleavage site.\(^8\) The site of the ninth heterocycle installed on the McbA precursor peptide is underlined. This figure was adapted from Dunbar & Chekan *et al.*\(^2\) with permission.
Figure 4.2 | The biosynthetic gene clusters for trifolitoxin and bottromycin do not contain a recognizable TOMM C protein. (a) The gene clusters for bottromycin A2 and trifolitoxin biosynthesis are displayed along with the amino acid sequence of the precursor peptide. Although each cluster contains at least one YcaO homolog, neither cluster contains a recognizable TOMM C protein (homolog of the E1 ubiquitin-activating enzyme superfamily). Asterisks indicate the leader and follower peptide cleavage sites on the trifolitoxin and bottromycin, respectively. (b) The structure of bottromycin A2 and the partial structure of trifolitoxin are displayed. Post-translational modifications presumably installed by the YcaO proteins encoded in each cluster are colored blue. The red Xs in the trifolitoxin structure denote an uncharacterized post-translational modification involving Arg and Gln.
4.2 The TOMM D Protein Alone is Sufficient for Azoline Formation

With the elucidation of the role of ATP in heterocycle formation (see Chapter 2), I sought to more precisely assign functions to the individual BalhC and BalhD proteins, which collectively comprise the cyclodehydratase (see Fig. 2.2). Given the homology between the C protein, previously labeled as the cyclodehydratase, and MccB,9-11 an adenylating protein involved in microcin C7 biosynthesis,23 I hypothesized that BalhC catalyzed the direct activation of the peptidic substrate. BalhD was suspected to be involved with general substrate handling and enzymatic regulation, thus allowing BalhC to perform multiple turnovers. To test these putative assignments, I performed a pseudo-single turnover experiment with a stoichiometric amount of either BalhC or BalhD and the BalhA1 substrate (see Fig. 4.1B for sequence), under saturating ATP. Unexpectedly, no modifications were detected when BalhC alone was added to BalhA1, but up to 4 dehydrations (-90 Da) were observed on BalhA1 treated with BalhD alone (Fig. 4.3A). Subsequent studies showed that these modifications were ATP-dependent (Fig. 4.3A) and competent intermediates in the production of the penta-azole product (Fig. 4.3B). Similar reactions on substrates containing a reduced number of heterocyclizable residues (BalhNC-C40 and BalhNC-CCG*; Fig. 4.1B), subjected to either tandem MS spectroscopy or iodoacetamide labeling, proved that the BalhD-installed modifications were cyclodehydration products (to yield an azoline) and not dehydrations (to yield a dehydroalanine/dehydrobutyrine; Fig. 4.3C and 4.4). These data confirmed that BalhD, formerly assigned as a docking/scaffold protein, was necessary and sufficient for azoline formation. Importantly, the rate of BalhD-only catalyzed heterocyclization was nearly three orders of magnitude slower than when BalhC was present (Fig. 4.5). Analysis of the BalhD-only catalyzed reaction products revealed that the ATP/azoline
stoichiometric ratio significantly deviated from unity (Fig. 4.6). Thus, I recommend that the term cyclodehydratase should be reserved for describing the C-D complex.

Figure 4.3 | BalhD is an ATP-dependent cyclodehydratase. (a) MALDI-TOF-MS spectra of BalhA1 treated with a stoichiometric amount of either BalhC or BalhD are displayed. For the BalhD reaction, a control lacking ATP is also displayed. The numbers indicate the number of dehydrations on the substrate. (b) The addition 2 µM BcerB and BalhC to the BalhD-induced 2-dehydration sample (~36 Da species) showed conversion to mostly a 4-azole species after 2 hours. A control sample where buffer was added instead of BcerB and BalhC did not show a similar modification. Increasing the reaction time to 6 h resulted in complete processing (only the penta-azole species was observed at ~100 Da, data not shown). (c) MALDI-TOF MS spectra of reactions of BalhNC-C40 and BalhD with or without ATP and subsequent labeling with iodoacetamide are displayed. All masses are in the +1 charge state. This figure was adapted from Dunbar et al.\textsuperscript{1} with permission.
Figure 4.4 | MS/MS localization of the BalhD installed dehydration product. FT-ICR MS spectra of the BalhD-induced cyclodehydration product (top). The expected mass for the 5+ charge state is shown. The low intensity of the 1-ring species is caused by the rapid hydrolysis of the azoline ring in the acidic ESI solution. Note that there is a third species 2 Da lighter than the monoisotopic peak of the unmodified substrate (blue). This peak corresponds to the internal disulfide adduct. The ppm error for each of these species is displayed next to their respective theoretical mass. MS/MS fragmentation of the modified species (bottom). Red ions indicate the presence of an 18 Da mass loss indicative of a cyclodehydration. The fragment map indicates that the modification has occurred in the immediate vicinity of position 43. In BalhNC-CCG* this is Cys40 (red, underlined).
Figure 4.5 | BalhC potentiates the BalhD ATPase activity. (a) The addition of BalhA1 substrate (100 µM) to a sample containing BalhD (15 µM) did not result in a significant increase in the rate of ATP hydrolysis as measured by the purine nucleoside phosphorylase (PNP) phosphate detection assay. (b) The addition of BalhA1 (100 µM) to BalhC/D (1 µM) resulted in robust ATP hydrolysis (compare Y-axis values). Error bars represent the standard deviation from the mean (n=3).
Figure 4.6 | BalhD ATP/ring stoichiometry is dysregulated in the absence of BalhC. Product progress curves are shown for both phosphate (blue diamonds) and azoline heterocycles (red squares) at different concentrations of BalhD (a-d). The ATP/azoline stoichiometry at each of the time points for all four BalhD concentrations is displayed (e). The graph demonstrates that as the reaction progresses, and as the concentration of BalhD is increased, the stoichiometry increases significantly. A red box highlights the expected ATP/azoline stoichiometry for the BalhD-only 31P-NMR sample and predicts that roughly half of the ATP hydrolyzed will be non-productive and occur through the action of a solvent nucleophile. If BalhD was able to activate the BalhA1 substrate in an identical fashion as the BalhC-BalhD complex, this would correspond to a 50:50 ratio of [16O₄]-P₁: [18O₁,16O₃]-P₁ when the reaction is carried out in [18O]-water.
4.3 TOMM YcaOs Utilize ATP for Amide Backbone Activation

The discovery that BalhD could independently install azolines on BalhA1 in an ATP-dependent manner, but that ATP consumption was dysregulated in the absence of BalhC, led us to investigate if BalhD could independently activate the amide backbone of BalhA1. As before, a reaction was carried out in $^{18}\text{O}\cdot\text{H}_2\text{O}$ and the $^{16}\text{O}/^{18}\text{O}$ isotopic ratio in the resultant $\text{P}_i$ was analyzed by $^{31}\text{P}$-NMR spectroscopy. Samples containing BalhD, BalhA1 and ATP showed a significant $[^{16}\text{O}_4]\cdot\text{P}_i$ peak (Fig. 4.7). However, if any of these essential components were omitted, $[^{16}\text{O}_4]\cdot\text{P}_i$ was not detectable (Fig. 4.7). While these data clearly demonstrated that BalhD was responsible for the amide activation event, the nearly identical heights of the $[^{16}\text{O}_4]\cdot\text{P}_i$ and $[^{18}\text{O}_1^{16}\text{O}_3]\cdot\text{P}_i$ NMR signals provided further evidence that ATP consumption was dysregulated in the absence of BalhC. Unlike the BalhC-BalhD complex, which had a strong enhancement in the rate of ATP hydrolysis upon the addition of BalhA1, the ATPase activity of BalhD remained unchanged upon the addition of BalhA1 (Fig. 4.5A). These findings implicate the C protein as a potentiator of the D protein, which is the stark opposite of the original assignments.
Figure 4.7 | TOMM YcaOs activate carbonyl oxygens with ATP. $^{31}$P-NMR spectra of BalhD reactions conducted in $^{[18]O}$-water are displayed. Peak identities were confirmed by spiking the samples with authentic $^{[16]O_4}$-P$_i$. Buffer control, ATP-synthetase buffer alone; precipitation control, control in which both BalhA1 and BalhD were precipitated by the addition of 25% acetonitrile prior to adding ATP; BalhD control, BalhA1 was omitted from the sample. This figure was adapted from Dunbar et al.\textsuperscript{1} with permission.
4.4 Non-TOMM YcaO Proteins Also Utilize ATP

With the ATP-dependent cyclodehydratase activity of BalhD established, I attempted to locate the ATP-binding site in BalhD by 8-azido-ATP crosslinking; however, these experiments were unsuccessful. Furthermore, the TruD crystal structure (a CD fusion protein involved in trunkamide biosynthesis) did not reveal an obvious ATP-binding site and BalhD was refractory to numerous crystallization attempts. Attempts to crystallize other TOMM YcaOs were equally unsuccessful (see appendix 3). I reasoned that since TOMM YcaOs evolved to interact with their cognate C proteins, working with a non-TOMM (with no C protein partner) might alleviate the previously encountered challenges. The local genomic environment of Ec-YcaO does not contain an E1 homolog (i.e. TOMM C protein, Fig. 4.1A), nor is Ec-YcaO known to interact with an E1 homolog, making it an attractive candidate for structural and biochemical characterization. Ec-YcaO was cloned into a tobacco etch virus (TEV) protease-cleavable maltose-binding protein (MBP)-fusion vector and expressed in E. coli. While the function of Ec-YcaO was unknown, characterized cyclodehydratases hydrolyze ATP in the absence of peptide substrates. Consequently, I measured the ATPase activity of Ec-YcaO using an established purine nucleoside phosphorylase assay. This assay revealed that Ec-YcaO indeed hydrolyzed ATP, preferentially generating AMP and PP$_i$ (Fig. 4.8). Although ATP hydrolysis was slow, perhaps because the native substrate was not present, Ec-YcaO displayed a $K_M$ for ATP of ~80 µM (Fig. 4.8), comparable to several characterized cyclodehydratases.
Figure 4.8 | Ec-YcaO hydrolyzes ATP to AMP and pyrophosphate. (a) The ATPase activity of Ec-YcaO was screened using the purine nucleoside phosphorylase (PNP)-coupled assay\(^{24}\) with and without 1 unit of pyrophosphatase (PPase). The addition of PPase increased the rate of chromophore production in the assay by 5-fold after the signal amplification achieved during PP\(_i\) cleavage is taken into account. Error bars represent the standard deviation from the mean (n=3). (b) An ATP kinetic curve was obtained for Ec-YcaO using the PPase-supplemented PNP assay. Error bars represent the standard deviation from the mean (n=3). Regression analyses to obtain Michaelis-Menten kinetic parameters were carried out in IGOR Pro version 6.12 (Wavemetrics). The error on the kinetic parameters represents the standard deviation from the curve fitting.
4.5 Structural Characterization of ATP-binding in Ec-YcaO

The structure of nucleotide-free Ec-YcaO (contains a mercurial salt for phasing) was determined to a Bragg limit of 2.63 Å and revealed a circularly symmetric homodimer in the asymmetric unit. The overall structure consists of an N-terminal YcaO domain of ~400 residues and a 150 residue C-terminal domain resembling a tetratricopeptide repeat (TPR) that mediates dimerization. A structure-based comparison against the PDB revealed similarity solely with TruD, the only other solved YcaO structure (RMSD of 3.1 Å over 279 aligned Ca atoms), confirming that YcaOs constitute a new structural fold (Fig. 4.9). To identify the ATP-binding site, the structure of Ec-YcaO in complex with multiple nucleotides was also determined. Co-crystallization of Ec-YcaO with ATP produced an AMP-bound structure (2.25 Å), suggesting that in situ hydrolysis had occurred (Fig. 4.10). To clarify the residues involved in ATP binding, the co-structure of Ec-YcaO with α, β-methyleneadenosine 5’-triphosphate (AMPCPP, a non-hydrolyzable ATP analog) was also solved.

Analysis of the 2.25 Å resolution AMP-bound and 3.29 Å resolution AMPCPP-bound co-crystal structures revealed that the adenine ring is recognized by Glu191 and Asn187 via interactions through the N7 nitrogen and between Ser16 and the exocyclic-N6 (Fig. 4.11A). Additionally, Lys9 resides above one face of the adenine ring, while Ala70, Ser71, and Gly74 are found within an α-helix that extends below the adenine and ribose rings, forming a hydrophobic surface. The ribose within the AMP- and AMPCPP-bound structures is oriented perpendicular to the adenine ring, with Ser184 and Glu78 coordinating the 2’- and 3’-hydroxyls, respectively (Fig. 4.11B and C). While Ser71 coordinates the α-phosphate in both structures, Arg286 coordinates to the α-phosphate in only the AMP-bound form (Fig. 4.11B and C). Surprisingly, two Mg\(^{2+}\) ions are found in the nucleotide-binding pocket in both structures. In the
AMP structure, Glu199 and Glu78 ligate one Mg$^{2+}$ ion while Glu290 and Glu75 bind the second (Fig. 4.11B). The Mg$^{2+}$ ions are coordinated in a similar fashion in the AMPCPP structure with the subtle difference that Glu202, rather than Glu75, coordinated the second Mg$^{2+}$ (Fig. 4.11C). This slight change in the coordination of the second Mg$^{2+}$ ion positions the metal ions on opposite sides of the β- and γ-phosphates. Furthermore, Arg203 coordinates the γ-phosphate of AMPCPP (Fig. 4.11C). The interactions between Ec-YcaO and AMPCPP are summarized in Figure 4.11D.
Figure 4.9 | Structural homology between Ec-YcaO and TruD and structure of AMP-bound Ec-YcaO. (a) Structural alignment of TruD (green) and Ec-YcaO (purple). The alignment demonstrates that the ATP-binding region of Ec-YcaO is also conserved in TruD. (b) TruD structure with the conserved domain shown in green. (c) Ec-YcaO structure with the conserved region shown in purple.
Figure 4.10 | Structure of AMP-bound Ec-YcaO. Structure of the Ec-YcaO homodimer with one monomer colored in purple (ATP-binding domain) and green (TPR domain) and the other monomer colored in gray. AMP is displayed as spheres in both monomers. This figure is adapted with permission from Dunbar & Chekan et al.²
Figure 4.11 | ATP coordination in Ec-YcaO. Orthogonal views of the Ec-YcaO AMP-bound structure in the vicinity of the active site, showing residues responsible for adenine/ribose binding (a) and Mg\(^{2+}/P_i\) binding (b). AMP and Mg\(^{2+}\) ions are colored cyan and gray, respectively. The superimposed difference Fourier maps are contoured at levels of 2.0\(\sigma\) (blue) and 8.0\(\sigma\) (red). (c) AMPCPP (cyan) and Mg\(^{2+}\) (grey) bound in the Ec-YcaO active site coordinates with superimposed difference Fourier maps contoured at levels of 2.0\(\sigma\) (blue) and 6.0\(\sigma\) (red). Residues responsible for \(P_i\), ribose and Mg\(^{2+}\) binding are indicated. (d) A ligand interaction diagram for the AMPCPP-bound structure is shown. Putative hydrogen bonds are displayed in orange with distances indicated, while red arcs denote hydrophobic interactions. Due to slight differences in residue orientation in the monomer subunits, only a subset of the interactions is displayed for clarity. This figure was adapted with permission from Dunbar & Chekan et al.\(^2\)
4.6 The Ec-YcaO ATP-Binding Site is Conserved in TOMM YcaOs

Utilizing the nucleotide-bound structures of Ec-YcaO, I established the conservation of the ATP-binding residues across the superfamily. First, a Cytoscape sequence similarity network of all YcaO members in InterPro (IPR003776) was generated. During assembly, redundant sequences were removed, leaving ~2000 sequences in the network. While the sequence of TOMM precursor peptide dictates the structure of the natural product, there is also a strong correlation between TOMM structure and the sequence similarity of the cognate YcaO. For the network, all YcaO sequences were manually annotated based on neighboring genes. YcaOs were categorized as being involved in TOMM biosynthesis if there was a gene encoding for a recognizable C protein in the local region (~10 kb on either side of the ycaO gene) or if the protein had an experimentally verified link to a known TOMM (e.g. bottromycin and trifolitoxin). The remaining YcaOs were separated into two other categories, non-TOMM YcaOs (e.g. Ec-YcaO) and TfuA-associated non-TOMM YcaOs. The latter were found within 10 kb of a gene encoding for the protein TfuA, which is implicated in trifolitoxin biosynthesis. Whenever possible, TOMM YcaOs were further subdivided by expected structural class. Based on these classifications, I determined that an expectation value of $10^{-80}$ gave an optimal separation of YcaO sequences into isofunctional clusters (Fig. 4.12).

Using the sequence similarity network as a guide, 349 of the ~2000 members in the non-redundant network were selected from across all clusters and a maximum likelihood tree was generated (Fig. 4.13A). Among the 349 were all singletons, defined as divergent family members not grouping with any other YcaO at an e-value of $10^{-80}$. Using this diversity-maximized tree, a sequence logo for each of the regions involved in ATP binding was generated using WebLogo (Fig. 4.13B). The logos clearly demonstrate that the ATP-binding pocket is highly conserved in
the YcaO family across all three groups (*i.e.* TOMM, non-TOMM, and TfuA-associated non-TOMM). The ATP-binding residues were found to be the most conserved feature in the YcaO superfamily (Fig. 4.13C). This is in stark contrast to TOMM C proteins, which do not harbor the ATP-binding residues conserved in all characterized non-TOMM E1 ubiquitin-activating superfamily members (Fig. 4.14). Furthermore, the conservation in the YcaO ATP-binding residues is maintained in all characterized TOMMs (Fig. 4.15), suggesting that the previously reported carbonyl activation mechanism is likely a universal biosynthetic feature.\textsuperscript{1,30}
**Figure 4.12 | Cytoscape sequence similarity network for the YcaO superfamily.** A similarity network for non-redundant YcaO sequences is displayed. Each node represents a unique YcaO, while lines between nodes exist if the proteins bear significant similarity (in this case, a BLAST expectation value lower than $10^{-80}$). At this e-value, YcaO proteins from characterized TOMM biosynthetic clusters primarily form isofunctional clusters. TOMM YcaOs are defined as having a bioinformatically identifiable C protein within 10 kb of the YcaO protein. Stand-alone TOMM YcaOs are proteins found in a characterized TOMM biosynthetic clusters that lack a C protein (e.g. bottromycin and trifolitoxin). It is noteworthy that many non-TOMM YcaOs appear to exist in RiPP biosynthetic gene clusters, indicating that additional stand-alone YcaO cyclodehydratases may exist. Nodes are colored according to the legend and the clusters containing Ec-YcaO and BalhD are indicated.
Figure 4.13 | Conservation of the Ec-YcaO ATP-binding residues in the superfamily. (a) A diversity-maximized tree for the YcaO superfamily is displayed. Red asterisks denote YcaOs associated with clusters of functionally or structurally characterized natural products, while the proteins used in this study are circled and colored blue. (b) WebLogo frequency plots for the ATP- and Mg$^{2+}$-binding residues of YcaO domains for each subclass (TOMM, non-TOMM, TfuA non-TOMM). Due to the high level of diversity in the sequences, WebLogos for the N-terminal ATP-binding residues were not generated. The ATP-binding motif identified in Ec-YcaO is displayed above each of the ATP-binding regions with the number representing the residue in Ec-YcaO, and conserved residues colored orange. (c) YcaO superfamily sequence conservation was mapped onto the structure of Ec-YcaO, which highlights strong conservation of the ATP-binding region. This figure is adapted with permission from Dunbar & Chekan et al.$^2$
Figure 4.14 | The canonical E1 domain ATP-binding site is not conserved in TOMM C proteins. Alignments of TOMM C proteins associated with cyanobactin biosynthesis to E. coli MoeB, ThiF and MccB are displayed. Residues responsible for ATP- and Mg\(^{2+}\)-binding in E1 superfamily members were identified in the nucleotide-bound crystal structures for the following E1 superfamily members: ThiF (1ZFN), MoeB (1JWA), and MccB (3H5N). The ATP- and Mg\(^{2+}\)-binding residues identified in these crystal structures were highlighted in orange. The two CxxC motifs responsible for coordination of the structural Zn\(^{2+}\) found in all characterized E1 ubiquitin-activating family members are highlighted in blue.\(^{10,15,31}\) To improve the alignments, the highly variable N-terminal “peptide clamp” domain found in MccB homologs was manually removed. Similarly, the highly variable N-terminal residues of each C protein and the YcaO domain of each CD fusion were removed. While non-TOMM E1 sequences are identified by GI number and the subclass of the E1 superfamily they belong to, TOMM E1 sequences are identified by the natural product they produce.
**Figure 4.15** | The ATP-binding pocket is conserved in characterized TOMM YcaOs. WebLogo frequency plots of all of the YcaO sequences from the indicated TOMM subclasses are displayed. The ATP-binding motif identified in Ec-YcaO is displayed above each of the ATP-binding regions. Residues that are similar to the ATP-binding motif identified in Ec-YcaO are colored orange. Due to the high level of diversity in the sequences, WebLogos for the N-terminal ATP-binding residues could not be generated. The number of sequences represented in each WebLogo is displayed in parentheses.
4.7 The Ec-YcaO ATP-Binding Site is Vital for BalhD Cyclodehydratase Activity

Because the native substrate of Ec-YcaO is unknown, I validated the ATP-binding residues by conducting structure-function studies on BalhD. An alignment of BalhD and Ec-YcaO permitted the mapping of the nucleotide- and Mg$^{2+}$-binding residues onto BalhD (Fig. 4.16). Subsequently, an alanine mutagenesis scan was performed on the polar residues of BalhD predicted to bind ATP. Every mutation was well tolerated in terms of protein yield/stability. The effect on heterocycle formation on BalhA1 (the peptide substrate) by the mutant BalhD proteins, in the presence of BalhC, was monitored in a 16 h endpoint assay (Fig. 4.16). Of the 11 mutated residues in the ATP-binding pocket, four were able to convert BalhA1 to the previously reported penta-azoline species,1 three showed intermediate levels of processing (2-4 heterocycles), while the remaining four generated no heterocyclic products within the limit of detection (Table 4.1). To quantify the effect of each mutation to BalhC/D activity, the rate of ATP hydrolysis was monitored using the $K_M$ concentration of BalhA1 (15 µM) and a concentration of ATP that would be saturating for wild-type BalhD (3 mM). Mutants unable to cyclize BalhA1, even after extended reaction times, displayed no detectable ATP hydrolysis over the assay background (Fig. 4.17). Likewise, mutants that installed five azolines on BalhA1 in the endpoint assay had the highest ATP hydrolysis rates. These data are consistent with my earlier work showing that ATP hydrolysis is tightly coupled to heterocycle formation.1 The YcaO mutations examined here did not appear to disrupt this feature of TOMM cyclodehydration.

While mutation of the BalhD ATP-binding pocket reduced cyclodehydratase activity, an alternative interpretation of the above data could be that these mutations interfered with the association of BalhC and BalhD. A unique feature of the Balh cyclodehydratase is that BalhD is catalytically active in the absence of BalhC.1 This permitted the use of BalhD-only activity
measurements to determine if the alanine mutations affected the intrinsic cyclodehydratase activity. Heterocycle formation endpoint assays (16 h) were again conducted, but this time with 50 μM BalhA1 and 25 μM BalhD mutant to account for the expected ~10³ drop in catalytic activity in the absence of BalhC.¹ The resultant mass spectra confirmed that the decrease in cyclodehydration arose from a perturbation in BalhD activity (Table 4.2 and Fig. 4.18).

For all BalhD mutants with measureable cyclodehydratase activity, I obtained the Michaelis-Menten kinetic parameters for BalhA1 and ATP (Table 4.1 and Fig. 4.19). Every mutation negatively affected the observed $k_{\text{cat}}$ ($k_{\text{obs}}$), indicating that the selected residues were of catalytic importance. Apart from K281A, and to a lesser extent S72A, all ATP-binding site mutations of BalhD substantially increased the $K_M$ for ATP. In contrast, the only mutant in this series to substantially raise the $K_M$ for BalhA1 was R198A (Table 4.1).

Four BalhD mutants (i.e. E76A, E79A, E194A, E197A) did not exhibit detectable cyclodehydratase activity. Potential explanations include an inability to bind the substrates (BalhA1 and/or ATP), hydrolyze ATP, or a structural perturbation with these mutants. Previous work demonstrated that BalhC and BalhD hydrolyze ATP slowly in the absence of BalhA1.¹ In reactions with wild-type BalhD, addition of BalhC increases the rate of ATP hydrolysis by 2.5-fold over an additive rate of both proteins; however, when the four inactive mutants were assayed, no potentiation was observed (Fig. 4.20). This suggested that the lack of BalhD activity was due to a structural perturbation or the inability to bind/hydrolyze ATP. Unfortunately, attempts to directly measure ATP binding or a secondary structure perturbation of BalhD with isothermal titration calorimetry or circular dichroism spectroscopy, respectively, were problematic owing to the solubility characteristics of BalhD. However, I reasoned that the latter could be assayed indirectly by monitoring the interaction between BalhC and a mutant BalhD.
through a competition assay and size-exclusion chromatography. While all of the BalhD mutants were able to associate with the BalhC (Fig. 4.21), E76A and E194A did so with reduced affinity, suggesting that these mutations affected the BalhC-BalhD interaction surface. Conversely, the wild-type-like affinity that BalhD E79A and E197A displayed for BalhC suggested that these mutants were inactive due to inability to bind/hydrolyze ATP.

Table 4.1 | Mutations to the ATP-binding pocket of BalhD decrease cyclodehydratase activity. a% processing = (5·P₅ + 4·P₄ + 3·P₃ + 2·P₂ + 1·P₁)/(5); where Pₓ is the percentage of the substrate with x number of azolines. The number of heterocycles formed in the assay is listed in parentheses. b Apparent K_M. c Not determined. Error on the Michaelis-Menten parameters represents the standard deviation from the regression analysis.

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Ring Formation</th>
<th>BalhA Kinetics</th>
<th>ATP Kinetics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CD D-only</td>
<td>kₚ, min⁻¹</td>
<td>K_M, µM</td>
</tr>
<tr>
<td>WT</td>
<td>(5); 100% (2-3); 45%</td>
<td>12.9 ± 0.4</td>
<td>16 ± 2</td>
</tr>
<tr>
<td>S72A</td>
<td>(5); 100% (2); 40%</td>
<td>8.3 ± 0.2</td>
<td>11 ± 1</td>
</tr>
<tr>
<td>E76A</td>
<td>(0); 0% (0); 0%</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>E79A</td>
<td>(0); 0% (0); 0%</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>R80A</td>
<td>(2-4); 63% (0-1); 10%</td>
<td>0.79 ± 0.05</td>
<td>16 ± 5</td>
</tr>
<tr>
<td>Q186A</td>
<td>(5); 100% (0-2); 10%</td>
<td>3.9 ± 0.1</td>
<td>12 ± 1</td>
</tr>
<tr>
<td>N190A</td>
<td>(5); 100% (0-2); 15%</td>
<td>5.1 ± 0.2</td>
<td>27 ± 3</td>
</tr>
<tr>
<td>E194A</td>
<td>(0); 0% (0); 0%</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>E197A</td>
<td>(0); 0% (0); 0%</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>R198A</td>
<td>(2-4); 40% (0); 0%</td>
<td>3.3 ± 0.2</td>
<td>50 ± 8</td>
</tr>
<tr>
<td>K281A</td>
<td>(5); 100% (2-3); 45%</td>
<td>7.9 ± 0.2</td>
<td>16 ± 2</td>
</tr>
<tr>
<td>E286A</td>
<td>(2-4); 64% (0); 0%</td>
<td>0.49 ± 0.03</td>
<td>27 ± 3</td>
</tr>
</tbody>
</table>
Figure 4.16 | Mapping ATP-binding residues from Ec-YcaO to BalhD. Structure-based sequence alignment of YcaO domain-containing proteins with the secondary structure of Ec-YcaO superimposed. Residues mutated in BalhD for this study are denoted by black triangles. A summary of the BalhD mutants made, along with their predicted roles in ATP recognition is provided. This figure was adapted with permission from Dunbar & Chekan et al.²
Figure 4.17 | Mutations to the BalhD ATP-binding site affect heterocycle formation. (a) A MALDI-TOF MS spectral overlay for BalhA1 treated with BalhC and either wild-type (WT) or mutant BalhD is displayed. The level of processing is summarized in Table 4.1. †, laser-induced deamination. A loss of 90 Da indicates the formation of 5 azolines, a full in vitro processed BalhA1 substrate.²¹ (b) The rate of ATP hydrolysis was measured with BalhC and either wild-type (WT) or mutant BalhD using the PNP assay. Error represents the standard deviation from the mean (n ≥ 3).
Figure 4.18 | Mutations to the ATP-binding site that decrease cyclodehydratase activity also affect BalhD-only activity. MALDI-TOF MS spectra for BalhA1 treated with wild-type (WT) or mutant BalhD is displayed. Apart from BalhD S72A, all of the mutants displayed a decreased level of D-only processing consistent with the rate of ATP hydrolysis measured in the presence of the C protein. The level of processing is summarized in Table 4.1. †, laser-induced deamination.
Figure 4.19 | The \( K_m \) for ATP does not depend on the concentration of \( \text{BalhA1} \). Due to solubility limitations, saturating concentrations of \( \text{BalhA1} \) were not obtainable for all of the \( \text{BalhD} \) mutants. To determine if the \( K_m \) for ATP changed at non-saturating concentrations of \( \text{BalhA1} \), an ATP kinetic curve was carried out at the \( K_m \) for \( \text{BalhA1} \) (15 \( \mu M \)). The resultant \( K_m \) is within error of the previously reported \( K_m \) for ATP, 240 ± 20 \( \mu M \), obtained with a saturating concentration of \( \text{BalhA1} \).\(^1\) Error on the Michaelis-Menten parameters represents the standard deviation from the regression analysis.

Figure 4.20 | Mutations of the Mg\(^{2+}\)-binding residues prevent ATPase potentiation by \( \text{BalhC} \). The background ATP hydrolysis rate of mutant \( \text{BalhD} \) proteins lacking the ability to modify \( \text{BalhA1} \) was measured with and without \( \text{BalhC} \) using the PNP assay. Although the ATPase activity of wild-type \( \text{BalhD} \) was potentiated 2.5-fold by the addition of \( \text{BalhC} \), none of the mutants displayed similar increases in ATP hydrolysis. For reactions with \( \text{BalhC} \), the horizontal lines indicate the expected rate of ATP hydrolysis after accounting for the slight ATPase rate of \( \text{BalhC} \). Error bars represent the standard deviation to the mean (\( n \geq 3 \)).
Figure 4.21 | Effect of the ATP-binding site mutations on BalhC binding. (a) Size-exclusion chromatography was used to monitor the effect the ATP-binding site mutations had on complex formation. Although BalhD retained as a monomer (3), BalhC had a retention volume consistent with a dimer (2). When the proteins were mixed, the intensity of the BalhC dimer peak decreased and a new peak corresponding to the BalhC/D complex appeared (1). Absorbance values were normalized to the peak at ~3.5 mL. (b) To better quantify the BalhC binding affinity of the mutant BalhD proteins, ATP hydrolysis rates were monitored for wild-type cyclodehydratase reactions with the addition of variable concentrations of inactive BalhD mutants. As the reactions were carried out with 1 µM wild-type BalhD, mutants with an unperturbed BalhC interaction would display IC₅₀ values of approximately 1 µM. Error bars represent the standard deviation from the mean (n ≥ 3). Regression analysis to determine the IC₅₀ values for each inhibition curve was determined in IGOR Pro version 6.12 (Wavemetrics). The error on the IC₅₀ values represents the standard deviation from the curve fitting.
4.8 The Proline-Rich C-terminus of TOMM YcaOs Affects C Protein Binding and Catalysis

In addition to the conserved ATP-binding site, TOMM YcaO proteins have a highly conserved, proline-rich C-terminus. In the most pronounced cases, the final five residues of the YcaO are PxPxP (Fig. 4.22). The proline-rich C-terminus is not conserved in non-TOMM or TfuA-associated YcaO domains (Fig. 4.22), implicating the motif in either C protein recognition or cyclodehydratase activity. This hypothesis is supported by the observation that the C-terminus of TruD is in close proximity to the YcaO ATP-binding site and is surface-accessible (Fig. 4.22). I first interrogated the importance of this motif by truncating 5 residues from the BalhD C-terminus. This minor perturbation abolished the catalytic activity of BalhC/D (Table 4.2 and Fig. 23). Removing the C-terminal three residues of BalhD produced an identical result, while removal of the C-terminal residue of BalhD (BalhD P429*; * stop codon) decreased activity by >100-fold (Table 4.2 and Fig. 4.23). Similarly, extending the C-terminus by a single amino acid (BalhD PxPxPG), or deleting two amino acids upstream of the PxPxP motif (BalhD ∆418-419), also resulted in inactive cyclodehydratases (Table 4.2 and Fig. 4.23).

To establish if altering the BalhD C-terminus affected the interaction with BalhC, I assessed the ability of BalhC to potentiate the background ATPase activity of BalhD mutants lacking detectable activity. Potentiation was not observed in any case (Fig. 4.24), indicating that mutants of the BalhD PxPxP motif had lost the ability to bind/hydrolyze ATP or to bind BalhC. I next assessed the ability of select PxPxP mutants to bind BalhC. Utilizing a combination of size-exclusion chromatography and a competition assay, all truncations to the BalhD PxPxP motif were shown to have decreased affinity to BalhC (Fig. 4.25). Moreover, the order of heterocycle formation was dysregulated in BalhD P429*, reminiscent of wild-type BalhD reactions lacking BalhC (Fig. 4.26; see Chapter 5).
Intrigued by the loss of activity observed upon extending or truncating the C-terminus, I next investigated the importance of the amino acid composition of the BalhD PxPxP motif (PHPFP_{429}). As with the truncations, any mutation to the C-terminal 5 residues of BalhD decreased cyclodehydratase activity (Table 4.2 and Fig. 4.27). The decrease in activity ranged from ~2.5-fold (P429G) to 100-fold (H426A), with the severity diminishing the closer the mutation was to the C-terminus. This result was consistent with the observation that the C-terminal residues of TruD are located in a channel leading to the active site. As with the PxPxP truncations, every mutant tested, apart from F428A, displayed a decreased affinity for BalhC (Fig. 4.28). Consistent with this observation, mutation of the terminal amino acid (P429G) resulted in an aberrant order of heterocycle formation (Fig. 4.29). Increasing the flexibility of the PHPFP motif, by substituting it with GHGFG, yielded an inactive cyclodehydratase (Table 4.2 and Fig. 4.27).

While these results implicated the C-terminus of BalhD in BalhC recognition, a decrease in BalhC affinity could not explain the data in its entirety. For example, both BalhD P425* and P429* displayed reduced interactions with BalhC, but only P425* was catalytically inactive (Table 4.2 and Fig. 4.25). Furthermore, BalhD PxPxPG displayed a wild-type level of interaction with BalhC, despite being catalytically inactive. Based on these results, the activity of each mutant was tested in the absence of BalhC. Analogous to the mutations to the ATP-binding pocket, mutations to the PxPxP motif affected the intrinsic activity of BalhD (Fig. 4.30). For all tractable BalhD mutants, BalhA1 and ATP Michaelis-Menten kinetic curves were obtained for the mutant BalhC-D complexes. While the largest effects were on $k_{\text{obs}}$, the mutations also affected the $K_M$ for ATP and BalhA1 (Table 4.2). Unlike the ATP-binding mutants, the changes to $K_M$ for the two substrates were similar in the PxPxP mutants, suggesting that the C-terminus is
involved in active site organization/catalysis, not substrate binding. Furthermore, the importance of the YcaO C-terminus appears to be general for TOMM biosynthesis, given that the cyclodehydratase activity of McbD (microcin B17 YcaO protein) was also abolished when the C-terminus was truncated (Fig. 4.31).

**Table 4.2 | Mutations to the C-terminus of BalhD disrupt catalysis.**  

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Ring Formation*</th>
<th>ATP Kinetics</th>
<th>ATP Kinetics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D-only</td>
<td>C+D</td>
<td>BalhA1 Kinetics</td>
</tr>
<tr>
<td>WT</td>
<td>(1-3); 45%</td>
<td>(5); 100%</td>
<td>12.9 ± 0.4</td>
</tr>
<tr>
<td>P425*</td>
<td>(0); 0%</td>
<td>(0); 0%</td>
<td>n.d.</td>
</tr>
<tr>
<td>P427*</td>
<td>(0); 0%</td>
<td>(0); 0%</td>
<td>n.d.</td>
</tr>
<tr>
<td>P429*</td>
<td>(0); 0%</td>
<td>(1-2); 36%</td>
<td>0.19 ± 0.03</td>
</tr>
<tr>
<td>Δ2 AA</td>
<td>(0); 0%</td>
<td>(0); 0%</td>
<td>n.d.</td>
</tr>
<tr>
<td>PxPxPG</td>
<td>(0); 0%</td>
<td>(0); 0%</td>
<td>n.d.</td>
</tr>
<tr>
<td>H426A</td>
<td>(0); 0%</td>
<td>(0-2); 25%</td>
<td>n.d.</td>
</tr>
<tr>
<td>P427G</td>
<td>(0); 0%</td>
<td>(3-5); 80%</td>
<td>0.43 ± 0.04</td>
</tr>
<tr>
<td>F428A</td>
<td>(0); 0%</td>
<td>(3-4); 80%</td>
<td>0.70 ± 0.02</td>
</tr>
<tr>
<td>P429G</td>
<td>(0-2); 6%</td>
<td>(4); 80%</td>
<td>4.7 ± 0.1</td>
</tr>
<tr>
<td>GxGxG</td>
<td>(0); 0%</td>
<td>(0-3); 45%</td>
<td>n.d.</td>
</tr>
</tbody>
</table>
Figure 4.22 | The C-terminal PxPxP motif conserved in TOMM YcaOs is near the ATP-binding site in TruD. (a) Alignment of C-termini of select TOMM YcaO proteins. Prolines are colored blue and CD fusion proteins are underlined. Numbers denote the location of the residue in BalhD. McbD: microcin B17, SagD: streptolysin S, TruD: trunkamide, BalhD: Bacillus sp Al Hakam, Pzn: plantazolicin, TsrH: thiostrepton (b) Graphs displaying the proline content of the C-terminus of all members of the YcaO superfamily (TOMM and non-TOMM) are shown. (c) The structure TruD<sup>5</sup> (PDB code 4BS9) is displayed. Residues implicated in ATP-binding are colored cyan while the PxPxP motif is colored orange. A zoom in of the YcaO active site (red circle) is displayed in panel d. The C-terminus of TruD is proximal (<8 Å) to the conserved ExxERD and RxxE motifs.
Figure 4.23 | The PxPxP motif of BalhD is critical for cyclodehydratase activity. (a) A MALDI-TOF MS spectral overlay for BalhA1 treated with BalhC and either wild-type (WT) or mutant BalhD is displayed. The level of processing is summarized in Table 4.2. (b) The rate of ATP hydrolysis was measured with BalhC and either wild-type (WT) or mutant BalhD using the PNP assay. Error represents the standard deviation from the mean (n ≥ 3). *, stop codons; †, laser-induced deamination; Δ2 AA, BalhD (ΔA418, ΔK419).
Figure 4.24 | Mutations that relocate the C-terminus of BalhD prevent ATPase potentiation by BalhC. The background ATP hydrolysis rate of BalhD C-terminus mutants lacking the ability to modify BalhA1 was measured with and without BalhC. Although the mutants displayed a higher rate of ATP hydrolysis than wild-type BalhD, none of the mutants could be potentiated by the addition of BalhC (P > 0.1). Error bars represent the standard deviation to the mean (n ≥ 3). For reactions with BalhC, the horizontal lines indicate the expected rate of ATP hydrolysis after accounting for the slight ATPase rate of BalhC. *, stop codons; Δ2 AA, BalhD (∆A418, ∆K419).
Figure 4.25 | C-terminal truncations of BalhD affect BalhC binding. (a) Size-exclusion chromatography was used to monitor the effect the C-terminal truncations had on complex formation. Although BalhD retained as a monomer (3), BalhC had a retention volume consistent with a dimer (2). When the proteins were mixed, the intensity of the BalhC dimer peak decreased and a new peak corresponding to the BalhC/D complex appeared (1). Absorbance values were normalized to the peak at ~3.5 mL. (b) ATP hydrolysis rates were monitored for wild-type cyclodehydratase reactions with the addition of variable concentrations of a C-terminal BalhD truncated protein. As the reactions were carried out with 1 µM wild-type BalhD, mutants with an unperturbed BalhC interaction would display IC₅₀ values of approximately 1 µM. *, stop codon. Error bars represent the standard deviation from the mean (n ≥ 3). Regression analysis to determine the IC₅₀ values for each inhibition curve was determined in IGOR Pro version 6.12 (Wavemetrics). The error on the IC₅₀ values represents the standard deviation from the curve fitting.
Figure 4.26 | Removal of the C-terminal residue of BalhD dysregulates the order of azole formation. BalhA1 was treated with BcerB, BalhC, and BalhD P429* for 16 h at 25 °C. (a) A high-resolution, intact mass spectrum of the processed substrate is displayed. The error (ppm) associated with each measurement from the calculated m/z is shown. (b) MS/MS fragmentation of the 0, 1 and 2 ring species was performed and the resultant fragmentation maps are displayed. The b- and y-ions are colored based on the number of azoles in the fragment (black = 0, orange = 1, purple = 2). Stars indicate the location of the azole heterocycles. In the di-azole species, the second heterocycle is found at both Cys45 and Cys31. The MS/MS spectrum for the di-azole species is shown in (c). Cys45 is the penultimate residue cyclized by a wild-type Balh cyclodehydratase. As such, these data indicate that the ring order of the complex has been perturbed in the P429* mutant.
Figure 4.27 | Mutagenesis of the C-terminus of BalhD affects cyclodehydratase activity. (a) MALDI-TOF MS spectra for BalhA1 treated with BalhC and either wild-type (WT) or mutant BalhD are displayed. These data indicate that the sequence of the C-terminus of BalhD is critical for cyclodehydratase activity. The level of processing is summarized in Table 4.2. (b) The rate of ATP hydrolysis was measured with BalhC and either wild-type (WT) or mutant BalhD using the PNP assay. Error represents the standard deviation from the mean (n ≥ 3). These data provided further evidence that the C-terminal residues of BalhD are critical for cyclodehydratase activity. †, laser-induced deamination. The wild-type sequence for the C-terminal PxPxP is 425PHPFP429.
Figure 4.28 | The sequence of the C-terminus of BalhD is important for binding BalhC. (a) Size-exclusion chromatography was used to monitor the effect the C-terminal mutations had on complex formation. Although BalhD retained as a monomer (3), BalhC had a retention volume consistent with a dimer (2). When the proteins were mixed, the intensity of the BalhC dimer peak decreased and a new peak corresponding to the BalhC/D complex appeared (1). Absorbance values were normalized to the peak at ~3.5 mL. (b) ATP hydrolysis rates were monitored for wild-type cyclodehydratase reactions with the addition of variable concentrations of a catalytically deficient BalhD mutant. As the reactions were carried out with 1 µM wild-type BalhD, mutants with an unperturbed BalhC interaction would display IC\textsubscript{50} values of approximately 1 µM. All of the BalhD mutants had a disrupted BalhC affinity. (c) As the competition assay is only compatible with BalhD mutations displaying almost no activity, the interaction between F428A and P429G could not be determined with this assay. For these active mutants, the interaction between BalhC and mutant BalhD was interrogated by measuring an apparent “K\textsubscript{M}” for BalhC. A BalhC “K\textsubscript{M}” was obtained by holding the concentration of mutant BalhD and BalhA1 constant and adding variable concentrations of BalhC. The rate of processing was monitored using the PNP-phosphate detection assay. For clarity, the initial processing rates were normalized to the rate obtained with a 1:1 ratio of BalhC and mutant BalhD. Together, these data indicated that the prolines in the PxPxD motif are more important for BalhC recognition than the intervening “x” residues and that residues closer to the active site (more C-terminal) are less important for BalhC binding. Error bars represent the standard deviation from the mean (n ≥ 3). Regression analysis to determine the IC\textsubscript{50} and K\textsubscript{M} values for each inhibition curve was determined in IGOR Pro version 6.12 (Wavemetrics). The error on the IC\textsubscript{50} and K\textsubscript{M} values represents the standard deviation from the curve fitting.
Figure 4.29 | BalhD P429G displays an altered order of heterocycle formation. Although BalhD P429G had an initial rate of substrate processing that is 1/3 of wild-type BalhD, the mutant produced primarily a tetra-ring species following a 16 h reaction. Based on the aberrant ring order observed in BalhD P429* reactions, I hypothesized that the inability to generate the penta-azole species could be due to perturbation of ring order. To locate the sites of heterocyclization, BalhA1 was treated with BcerB, BalhC, and BalhD P429G for 16 h at 25 °C. (a) A high-resolution, intact mass spectrum of the processed substrate is displayed. The error (ppm) associated with each measurement from the calculated m/z is shown. (b) MS/MS fragmentation of the 5, 4 and 3 ring species was performed and the resultant fragmentation maps are displayed. The b- and y-ions are colored based on the number of azoles heterocycles in the fragment (black = 0, orange = 1, purple = 2, blue = 3, pink = 4, green = 5). Stars indicate the location of the azoles. In the triazole species, the third heterocycle was found at both Cys45 and Cys34. The MS/MS spectrum for the tri-azole species is shown in (c). Based on the b- and y-ion intensities, both positions appeared to be cyclized to approximately the same extent. In wild-type reactions, the tri-azole species has rings at Cys31, Cys34, and Cys40.21,32 The deviation from the wild-type processing order was not observed in the tetra- or penta-azole species. Although less severe than the dysregulation in cyclization order found with BalhD P429*-treated BalhA1, the identification of partial processing at Cys45 in the tri-azole species suggested that the BalhC/D interaction was perturbed.
Figure 4.30 | Mutations to the C-terminus of BalhD also affect BalhC-independent catalysis. A MALDI-TOF MS spectral overlay for BalhA1 treated with either wild-type (WT) or mutant BalhD is displayed. Only BalhD P429G showed detectable levels of azoline formation. These data indicated that the C-terminus of BalhD is important for catalysis in addition to BalhC recognition. The level of processing is summarized in Table 2. *, stop codon; †, laser-induced deamination; Δ2 AA, BalhD ΔA418-K419.
Figure 4.31 | The conserved TOMM YcaO C-terminus is critical for activity in the microcin B17 cyclodehydratase. A MALDI-TOF MS spectral overlay for McbA treated with McbB/C and either wild-type (WT) or mutant McbD lacking the C-terminal 5 amino acids is displayed. *, stop codon; †, laser-induced deamination.
4.9 Summary and Outlook

While previous studies on fused C-D proteins implicated the YcaO/DUF181 domain in the cyclodehydration reaction, the exact role of the D domain in this complex was never elucidated. Besides the TOMM D protein, only one other YcaO/DUF181 family member has had a biological function reported. A recent study linked the function of E. coli YcaO to ribosomal methylthiolation, but did not reveal any mechanistic details or physiological ramifications of this process. Thus, these data assign the first definitive function to a member of this uncharacterized protein family and necessitates a reclassification of the roles of the C/D proteins.

I have discovered that Ec-YcaO contains a novel ATP-binding fold. Based on steric and electrostatic complimentary requirements, the YcaO strategy for binding ATP is reminiscent of other structurally characterized ATP-binding proteins. For example, the Lys/α-helix “sandwich” involved in adenine recognition is similar to the conserved Arg and Gly motif found in class I amino acyl tRNA synthetases. Furthermore, select members of the ATP-grasp and PurM families have been shown to bind ATP through the use of multiple divalent cations; however, in these proteins, the Mg$^{2+}$ ions are coordinated to all three phosphates, not just the β- and γ-phosphates. Since these similarities in ATP-binding occur despite a lack of structural and primary homology between YcaO and all other known ATP-binding proteins, this represents an example of convergent evolution in ATP-binding domains.

The ATP-binding residues are the most highly conserved motifs in the YcaO superfamily and, appropriately, represent a prominent signature for the hidden Markov model that bioinformatically defines the YcaO/DUF181 family (IPR003776). The extensive bioinformatics analysis, X-ray crystallographic data on Ec-YcaO, and biochemical characterization of BalhD presented in this report confirm that ATP-utilization is a conserved feature in the superfamily. In
spite of the low level of overall similarity between Ec-YcaO and BalhD, I was able to demonstrate that the YcaO ATP-binding motif was critical for cyclodehydratase activity. While the mutations affected BalhD activity to differing extents, the impact of mutating a particular residue on Balh cyclodehydratase activity was proportional to the level of conservation within the YcaO family (Fig. 4.13B).

During my analysis of the YcaO ATP-binding motif, I observed a striking difference between the TOMM and non-TOMM YcaO domains. TOMM YcaOs (D proteins) almost invariably harbor proline-rich C-termini, with PxPxP most often serving as the terminal five residues of the protein. While the widespread nature of the PxPxP motif had been previously recognized\textsuperscript{10}, prior to this work, it was unclear if this motif played any role in TOMM biogenesis. My data indicate that the C-terminus of TOMM YcaOs assists in both C protein recognition and cyclodehydration. It is rare for the C-terminus of an enzyme to be important for catalysis.\textsuperscript{36-40} In fact, the terminal regions are often highly sequence variable within a protein family. Intriguingly, the C-terminal proline content of a YcaO protein has powerful predictive value. If present, the YcaO is quite probably involved in TOMM biosynthesis. This tentative assignment can be confidently made even without knowledge of the flanking genes. As such, I hypothesize that a subset of the 249 (~8%) non-TOMM YcaOs that contain a proline-rich C-terminus may actually be stand-alone TOMM YcaOs (akin to the bottomycin YcaOs).

In addition to providing significant insight into the mechanism of TOMM cyclodehydration, the results presented here provide an initial framework to explore the elusive functions of the 4,000 uncharacterized non-TOMM YcaOs currently in GenBank. While I cannot confidently comment on the function of the remaining 4000 genes, the capacity to bind ATP (or possibly other nucleotide triphosphates) appears to be ubiquitous in the YcaO superfamily.
However, it remains unclear whether the TOMM cyclodehydratase-like direct activation of carbonyls is a universal feature. It is intriguing that YcaOs have recently been implicated in the formation of thioamides\textsuperscript{41} and macroamidine rings,\textsuperscript{16-19} as both of these modifications could conceivably occur through carbonyl activation.

My dissection of the cyclodehydratase complex and reclassification of the roles the C/D proteins poses an interesting question as to the role of the C protein in TOMM biosynthesis. While I have shown that BalhC acts cooperatively with BalhD to accelerate azoline formation and govern the proper utilization of ATP, this study did not allow me to conclusively deduce the role of BalhC. Nonetheless, it is conceivable that the C protein could partake in one of three plausible functions: (1) BalhC activates BalhD through an allosteric mechanism, increasing the rate of heterocyclization, (2) BalhC catalyzes the nucleophilic attack of the proceeding side chain by providing the requisite general acid/base residues, or (3) BalhC forms key contacts with both the substrate and BalhD to facilitate the interaction of the core region of BalhA1 and the BalhD active site. This topic will be discussed further in Chapter 5.

\textbf{4.10 Experimental}

\textbf{4.10.1 General methods.} Unless otherwise specified, all chemicals were purchased from Sigma or Fisher Scientific. DNA sequencing was performed by either the Roy J. Carver Biotechnology Center (UIUC) or ACGT Inc. Restriction enzymes were purchased from New England Biolabs (NEB), Pfu Turbo was purchased from Agilent, and dNTPs were purchased from either NEB or GenScript. Oligonucleotide primers were synthesized by either Integrated DNA Technologies (IDT) or Eurofins MWG Operon. Unless otherwise stated, all proteins and substrates were used as MBP fusions to circumvent solubility issues.
4.10.2 Single-turnover activity studies. Individual MBP-tagged synthetase components (45 µM) were set up with MBP-BalhA1 substrate at a 1:1 ratio in synthetase buffer [50 mM Tris (pH 7.5), 125 mM NaCl, 10 mM DTT, 20 mM MgCl₂ and 3 mM ATP]. After an 18 h reaction at 23 °C, MBP was proteolytically removed from the substrate by the addition of 0.2 mg/ml recombinant TEV protease and a 30 min incubation at 30 °C. Samples were desalted via C₁₈ ZipTip (Millipore) according to the manufacturer’s instructions and analyzed on a Bruker Daltonics UltrafleXtreme MALDI-TOF spectrometer. Spectra were obtained in positive reflector mode using α-cyano-hydroxycinnamic acid as the matrix.

4.10.3 Iodoacetamide labeling. Reactions were set up with 45 µM MBP-BalhD, 45 µM MBP-BalhNC-C40 (see Fig. 4.1B for amino acid sequence), and 0.2 mg/mL TEV protease in TCEP synthetase buffer [50 mM Tris (pH 7.5), 125 mM NaCl, 10 mM TCEP, 20 mM MgCl₂ and 3 mM ATP] and allowed to react for 18 h at 23 °C. Next, 50 mM iodoacetamide was added for 1 h at 23 °C before ZipTip (Millipore) desalting and analysis on a Bruker Daltonics UltrafleXtreme MALDI-TOF mass spectrometer in positive reflector mode with α-cyano-hydroxycinnamic acid used as the matrix.

4.10.4 Heterocycle localization via FT-MS/MS. For BalhD-only ring localization, reactions were set up with 45 µM MBP-BalhD, 45 µM MBP-BalhNC-CCG* (see Fig. 4.1B for amino acid sequence), and 0.2 mg/mL TEV in synthetase buffer and allowed to react for 18 h at 23 °C. Next the large proteins were precipitated via the addition of 25% acetonitrile, the precipitate was removed from the sample by centrifugation, and the soluble fraction was dried in a Savant SpeedVac (Thermo Fisher). The resultant solid was resuspended in 10 µL water and desalted by ZipTip according to the manufacturer’s instructions. The desalted sample was diluted 2-fold in 80% acetonitrile with 1% acetic acid and was analyzed on a ThermoFisher Scientific LTQ-FT
hybrid linear ion trap-FTMS system, operating at 11 T. Ions were introduced by direct infusion using an Advion Nanomate 100. An accurate mass was collected using the FTMS set at a resolution at 100,000. Fragmentation data was collected in the ion trap with the following settings: isolation width: 3 m/z, normalized collision energy (NCE): 35; activation q value: 0.25; activation time: 30 ms.

The localization of heterocycles on BalhA1 treated with a mutant BCD complex was performed in a similar fashion. 50 µM MBP-BalhA1 was modified by 2 µM MBP-tagged BcerB, BalhC, and BalhD in synthetase buffer for 18 h at 25 °C. Proteins were digested with 0.02 mg/mL sequencing grade trypsin (Promega) in 50 mM NH₄CO₃ (pH 8.0) for 30 min at 37 °C before the sample was quenched via the addition of formic acid to a final concentration of 10% (v/v) and precipitate was removed via centrifugation at 17,000 x g. FT-MS/MS analysis was carried out as described above.

4.10.5 BalhD-only ATP/ring stoichiometry. Reactions were initiated via the addition of 60 µM TEV cleaved MBP-BalhA1 to a sample containing the indicated amount of MBP-BalhD in synthetase buffer. Aliquots were removed at the indicated time points and frozen in liquid nitrogen. The samples were then analyzed by MALDI-TOF MS to determine an approximate azoline concentration (calculated by the relative peak heights of the different ring states) and malachite assay was used to quantify the amount of phosphate produced in the samples.

4.10.6 Malachite green phosphate detection assay. Samples were diluted 1:8 in water to a final volume of 80 µL to remove any background signal from unreacted ATP. This solution was transferred to a 96-well plate where the reaction was initiated by the addition of 20 µL malachite green working reagent (Bioassay Systems). Reactions were allowed to develop for 30 min before the absorbance was recorded at 620 nm. A standard curve was made from a known concentration
of inorganic phosphate in appropriately diluted synthetase buffer. The absorbance for each of the samples was corrected for any background originating from the buffers.

**4.10.7 $^{31}$P-NMR analysis of phosphate isotope incorporation of $^{18}$O-water reactions.** Two samples containing either the MBP-tagged BalhD (20 µM) synthetase components or MBP-BalhA1 substrate (60 µM) with low-salt synthetase buffer (50 mM Tris pH 8.5, 25 mM NaCl, 5 mM MgCl$_2$, 10 mM DTT, 2 mM ATP) were lyophilized separately for 16 h. The resulting solid was reconstituted in 500 µL of 97 atom% $[^{18}$O]-H$_2$O (Cambridge Isotope Laboratories) and incubated at 23 °C with rocking for 3 h before quenching the reaction with the addition of acetonitrile to a final concentration of 25%. The sample was dried using a Savant SpeedVac (Thermo Fisher), the non-proteinaceous components were reconstituted in 220 µL D$_2$O with 5 mM EDTA, and reconstituted sample was placed in a D$_2$O-matched Shigemi NMR tube. The $^{31}$P-NMR spectrum was obtained on a 600 MHz Varian Unity Inova NMR with a 5 mm Varian AutoTuneX probe, 1000 transients, 32768 points, and a spectral window of -15 to 5 ppm.

A $[^{16}$O]-H$_2$O control, an enzyme only control in $[^{18}$O]-H$_2$O, a buffer only control, and a control in which the BalhD and BalhA1 were precipitated with acetonitrile prior to the addition of ATP synthetase buffer was also performed. $^{31}$P-NMR spectra for each control were obtained as indicated above.

**4.10.8 Cloning of MBP-YcaO.** Ec-YcaO was amplified by PCR from *Escherichia coli* BL21 cells using the forward and reverse primers listed in table A.2. Polymerase reactions were carried out with PFU Turbo and the amplified product was digested with BamHI and NotI following a gel extraction. The digested gene was PCR-purified and ligated into an appropriately digested, modified pET28 vector containing a tobacco etch virus (TEV) protease cleavable, N-terminal MBP-tag.
4.10.9 Preparation of BalhD PxPxP mutants. BalhD was amplified by PCR from the previously described pET28-BalhD plasmid\textsuperscript{21} using the primers listed in table A.2. Polymerase reactions were performed with PFU Turbo and the amplified product was digested with BamHI and NotI following gel extraction. The digested gene was PCR-purified and ligated into an appropriately digested, modified pET28 vector containing a tobacco etch virus (TEV) protease cleavable, \textit{N}-terminal MBP-tag.

4.10.10 Site-directed mutagenesis. Site-directed mutagenesis of BalhD and McbD was carried out using the QuikChange method according to the manufacturer’s instructions. All mutagenesis primers are listed in table A.2.

4.10.11 Overexpression and purification of MBP-tagged proteins. All proteins were purified with amylose resin (NEB) according to previously described procedures.\textsuperscript{1}

4.10.12 Multiple sequence alignments. Alignments were made with Clustal Omega using the standard parameters.\textsuperscript{42}

4.10.13 Cytoscape sequence similarity network. A sequence similarity network was created using the Enzyme Function Initiative Enzyme Similarity Tool (EFI-EST, enzymefunction.org).\textsuperscript{43} Sequences from the YcaO superfamily (InterPro number IPR003776) were used for the analysis. The network was constructed at an expectation-value (e-value) of $10^{-80}$. Networks were visualized by Cytoscape using the organic layout.\textsuperscript{26} Sequences with 100% identity were visualized as a single node in the network and all nodes were manually annotated.

4.10.14 Maximum likelihood phylogenetic analysis. A set of YcaO sequences representing the full diversity of the YcaO family were selected based on the Cytoscape sequence similarity network (Fig. 4.12). This included at least one protein from each cluster and all singletons (proteins that fail to cluster with any other YcaO) for a total of 349 proteins (~17\% of the non-
redundant Cytoscape sequence similarity network). The phylogenetic analysis was performed with Molecular Evolutionary Genetics Analysis (MEGA5). An amino acid sequence alignment was created using the standard parameters of ClustalW and a maximum likelihood phylogenetic tree was created using standard parameters in MEGA5.

4.10.15 ATP-binding site conservation. The ligand interaction network for AMPCPP was generated using LigPlot plus using the standard parameters. A WebLogo frequency plot for the ATP-binding motif was generated from a Clustal Omega alignment of all of the sequences from the specified family using the standard parameters. The conservation map for the YcaO family was generated by aligning 150 unique YcaO sequences with at least 35% sequence similarity to Ec-YcaO (e-value < 10^-5) and mapping the resulting conservation data onto the Ec-YcaO structure using ConSurf. The structure based YcaO alignment was generated using Clustal Omega and ALINE, while the structural overlay of YcaO and TruD was generated using PyMOL version 1.5 (Schrödinger).

4.10.16 Proline-rich C-terminus analysis. The proline content of the C-termini of all YcaOs in the Cytoscape sequence similarity network was determined. Proteins were deemed to have a proline-rich (P-rich) C-terminus if at least 4 of the final 30 residues were Pro. In the most pronounced cases, the terminal 6 residues of the YcaO contain a PxPxP motif. Proteins were identified as containing a PxPxP C-terminus if they contained a PxP motif in the final 6 residues and at least 3 Pro in the final 30 residues.

4.10.17 Ec-YcaO crystallization and structural solution. Crystallization and structural solution was carried out by Jonathan Chekan (Nair lab). See Dunbar & Chekan et al. for detailed methods. Coordinates for the APO YcaO, AMP bound YcaO, and AMPCPP Bound

158
YcaO structures were deposited to the Protein Data Bank as codes 4Q84, 4Q86, and 4Q85 respectively.

4.10.18 Endpoint heterocycle formation assays. For the reactions with the BalhD mutants, 50 µM MBP-BalhA1 was mixed with either 2 µM MBP-BalhC/D (CD activity) or 25 µM MBP-BalhD (D-only activity) in synthetase buffer [50 mM Tris (pH 7.5), 125 mM NaCl, 10 mM DTT, 20 mM MgCl$_2$ and 3 mM ATP]. The MBP tags were removed using 0.05 mg/mL of TEV protease and reactions were carried out for 18 h at 25 °C. Reactions with the Mcb enzymes were carried out identically except that the dehydrogenase (McbC) was also added to the reaction and proteins were cleaved with 0.1 µg/mL thrombin protease (from bovine plasma). Processing was monitored by MALDI-TOF MS.

4.10.19 BalhD mutant kinetic studies. Substrate processing kinetics for the BalhD mutants were determined using a previously described purine nucleoside phosphorylase (PNP)-coupled assay.$^{1,24}$ For BalhA1 kinetic experiments, variable concentrations of MBP-BalhA1 (1-120 µM) were reacted with 1-10 µM MBP-tagged BalhC/D while ATP was held constant at 3 mM. ATP kinetic experiments were carried out in an identical fashion except that MBP-BalhA1 was fixed at 80 µM and variable concentrations of ATP (0.1- 5 mM) were used. Although this does not provide a saturating level of BalhA1 for all mutants, the $K_M$ for ATP does not change with varied BalhA1 concentration (Fig. 4.19). Reactions were carried out in triplicate. Regression analyses to obtain the kinetic parameters for both substrates were carried out with IGOR Pro version 6.12 (WaveMetrics).

4.10.20 Size-exclusion chromatography. A 200 µL sample containing 25 µM MBP-tagged BalhC/D was prepared in cleavage buffer [20 mM Tris (pH 7.5), 500 mM NaCl, 10 % glycerol (v/v), 0.5 mM tris(2-carboxyethyl)phosphine (Gold Biotechnology)] and treated with 0.05
mg/mL TEV protease for 12 h at 4 °C. The amount of protein in a BalhC/BalhD complex was assessed on a Flexar HPLC (Perkin Elmer) equipped with an analytical Yarra SEC-3000 (300 x 4.6 mm, Phenomenex) equilibrated with cleavage buffer. Peaks of interest were collected and their composition was determined via a Coomassie stained 12% SDS-PAGE gel. The approximate molecular weights were determined by generating a standard curve with a 12-200 kDa molecular weight standard kit (Sigma). Control runs were also performed in which one of the two proteins was omitted. Chromatograms were exported using Flexar Manager (Perkin Elmer) and analyzed using Microsoft Excel.

4.10.21 Mutant BalhD IC\textsubscript{50} determination. 1 µM MBP-tagged BalhC and BalhD were mixed in synthetase buffer [50 mM Tris (pH 7.5), 125 mM NaCl, 10 mM DTT, 20 mM MgCl\textsubscript{2} and 3 mM ATP] with 0.25-20 µM mutant BalhD protein. Reactions were initiated by the addition of 15 µM MBP-BalhA1 and progress was measured using the PNP phosphate detection assay. All reactions were performed in triplicate. IC\textsubscript{50} values were calculated with IGOR Pro version 6.12 (WaveMetrics).

4.10.22 BalhC “K\textsubscript{M}” for active BalhD mutants. The affinity of catalytically active BalhD mutants for BalhC was determined using the PNP phosphate detection assay and a previously described procedure.\textsuperscript{13} 25 µM MBP-BalhA1 was mixed with 1 µM MBP-BalhD in synthetase buffer and reactions were initiated via the addition of 0.15-4 µM BalhC. All reactions were performed in triplicate. Regression analyses to obtain kinetic parameters for BalhC were carried out with IGOR Pro version 6.12 (WaveMetrics).
4.11 References


(22) Herlache, T. C.; Triplett, E. W. 2006.


CHAPTER 5: THE ROLE OF TOMM E1-HOMOLOGS IN AZOLINE BIOSYNTHESIS

Sections of this chapter were published in Dunbar & Mitchell\(^1\) and Dunbar & Chekan \(\text{et al.}^2\) and are reproduced with permission from the American Chemical Society and Nature Publishing Group, respectively.

I am grateful to Joel Melby for assisting with the collection of MS/MS data and Nhat Trinh assisting with Balh3A cloning. Brandon Burkhart carried out the fluorescence polarization experiment.

Abstract

Although the TOMM C protein was originally thought to be responsible for performing the ATP-dependent cyclodehydration of TOMM precursor peptides, this activity has recently been assigned to TOMM YcaO proteins (see chapter 4). Rather the role of the C protein appears to be in the potentiation of the YcaO cyclodehydratase activity; however, the basis for this activity increase remained enigmatic. Using diverse TOMM synthetases, I demonstrate that TOMM E1-homologs provide substrate recognition and potentiate catalysis of their cognate YcaO. Together these results allow us to assign a revised function to over 1,000 proteins in GenBank and propose an updated biosynthetic scheme for TOMM natural products.

5.1 Introduction

The TOMM C protein, found in all canonical TOMM biosynthetic clusters, is one of two domains required for azoline formation during TOMM biogenesis (Fig. 1.3).\(^3,4\) TOMM C proteins are members of the E1 ubiquitin-activating superfamily. As all characterized E1 superfamily members are ATP-utilizing enzymes,\(^5\) TOMM C proteins were originally thought to catalyze azoline formation and be the ATP-utilizing component of the TOMM synthetase.\(^4\)
Consequently, the TOMM C protein was annotated as the TOMM cyclodehydratase. However, recent work demonstrated that the D protein (TOMM YcaO) is responsible for carrying out this transformation and that the C protein potentiates the activity of the D protein via an unknown mechanism. Although the basis for this potentiation was not investigated in these studies, it was hypothesized that the TOMM C protein might provide peptide substrate recognition for TOMM YcaO proteins. Precursor peptide recognition in the majority of ribosomally synthesized post-translationally modified peptides (RiPPs) occurs in a bipartite fashion. An N-terminal sequence (leader peptide), serves as the recognition sequence by the modification enzymes, while the C-terminal sequence (core peptide) contains the sites of post-translational modification. Previous work on a TOMM C protein involved in streptolysin S biosynthesis (SagC), demonstrated that C proteins have the ability to bind leader peptide; however, it remains unclear whether this function is conserved within the protein family. Complicating matters further, previous work on the microcin B17 synthetase demonstrated that the YcaO protein (McbD), not the C protein, was responsible for leader peptide recognition.

In addition to the canonical TOMM C protein, a subset of TOMM biosynthetic clusters contains an additional E1 homolog. These auxiliary proteins, annotated as ocin-ThiF-like proteins in GenBank, are much larger than a standard TOMM C protein (~600 vs ~325 amino acids), and serve an as of yet uncharacterized role in TOMM biosynthesis. These ThiF-like proteins are found in all heterocycloanthracin and a diverse subset thiopeptide biosynthetic clusters. Notably these ThiF homologs are found in clusters that contain truncated CD fusion proteins, in which the C domain is truncated. Based on homology modeling to the X-ray crystal structure of the CD fusion protein TruD and the non-TOMM RiPP biosynthetic enzyme MccB, this truncation appears to remove an N-terminal capping domain that has been...
implicated in peptide recognition in MccB.\textsuperscript{13} Due to the confusion surrounding the role of the TOMM C protein, the roles of each E1-homolog and the effect of the truncation on azoline formation are not understood.

In an attempt to assign a definitive role to TOMM C proteins, the E1-homologs from two divergent TOMM biosynthetic clusters in \textit{Bacillus} sp. Al Hakam (Balh) were investigated. The results demonstrate that in addition to providing substrate recognition, TOMM E1-homologs also increase the catalytic efficiency of their cognate YcaO. Although the molecular details for these interactions remain unresolved, the data presented in this report allow us to ascribe a function to this previously mischaracterized protein.
5.2 TOMM C Proteins Direct YcaO Cyclodehydratase Activity

Previously, I demonstrated that BalhD is able to catalyze cyclodehydration in the absence of BalhC, albeit with a drastically reduced efficiency.\(^3\) Although the role of BalhC was not explicitly interrogated, previous work carried out on SagC demonstrated that C proteins play an important role in substrate recognition.\(^6\) If the C protein was required for proper substrate recognition, I hypothesized that the highly regulated order of azoline installation would be disrupted in reactions lacking BalhC. In order to test this hypothesis, BalhD-treated BalhA1 was subjected to the AMPL methodology (see Chapter 3) to afford \([^{18}\text{O}_2]\)-BalhA1, and the isotope labels were localized by FT-MS/MS. While one of the azolines was localized to Cys45, the second \(^{18}\text{O}\) label was found at Gly37 and Gly39, indicating that the second azoline was installed at Thr38 and Cys40, but not both simultaneously (Fig. 5.1). In reactions containing BalhC, Cys40 is the first ring cyclized, Cys45 is the penultimate ring, and the cyclization of Thr38 is not detected.\(^15\) Thus, these data suggest that BalhC is at least partially responsible for dictating the regio- and chemoselectivity of substrate processing by the Balh cyclodehydratase.
Figure 5.1 | Azoline localization on BalhD-treated BalhA1 via $^{18}$O-labeling. Following treatment with BalhD and $^{18}$O-hydrolysis, BalhA1 was subjected to FT-MS/MS. (a) An intact mass spectrum for $[^{18}O_3]$-BalhA1 in the 4+ charge is displayed. The monoisotopic masses for the most prominent $^{18}$O-labeled species are given and colored based on the number of $^{18}$O labels (orange, 3; blue, 2; red, 1; black, 0). Prior to hydrolysis the peptide had two azoline heterocycles. The presence of a third $^{18}$O label indicates that a non-specific labeling event occurred. The ppm error for the major species is displayed. (b) A MS/MS spectra of the $[^{18}O_3]$-BalhA1 species from panel (a) demonstrates that the two azolines are spread across three sites: Thr38, Cys40 and Cys45. Based on the b and y ion intensities, Cys45 was always found as a thiazoline, while azoline formation at Thr38 and Cys40 occurred in equal amounts. The b and y ions are colored based on the number of $^{18}$O present in the fragment (orange, 3; blue, 2; red, 1; black, 0). Stars represent the sites of $^{18}$O labels from hydrolysis of the azoline heterocycles and asterisks label peaks with neutral water loss. The y ion series indicates that the C-terminal carboxylate has a single $^{18}$O label (denoted by red residue).
5.3 TOMM C Proteins are Responsible for Leader Peptide Recognition

The above results are consistent with the hypothesis that TOMM C proteins are responsible for peptide substrate recognition. To directly assess the role of the C and D proteins in peptide substrate recognition, a fluorescein-labeled BalhA1 leader peptide was employed to monitor binding to BalhC and BalhD by fluorescence polarization. While BalhD was found to not bind the BalhA1 leader peptide to any significant extent, BalhC displayed a $K_d$ of 11 $\mu$M (Fig. 5.2), near the previously measured $K_M$ for BalhA1 of 16 $\mu$M. Moreover, the addition of BalhD did not significantly alter BalhC’s affinity for the BalhA1 leader peptide ($p > 0.05$). Consequently, these data support a model where TOMM C proteins bind leader peptide and present the core peptide to the YcaO domain for modification.

Figure 5.2 | BalhC is responsible for leader peptide recognition. A fluorescent polarization curve for fluorescein-labeled BalhA1 leader peptide recognition by BalhC and BalhD is displayed. Error bars represent the standard error of the mean of three independent titrations. Errors on the $K_d$ values represent the error from curve fitting.
5.4 TOMM C Proteins Potentiate YcaO Cyclodehydratase Activity

During the mutagenesis of the proline-rich C-terminus of BalhD (see Chapter 4), a derivative containing a C-terminal His$_6$ tag was generated (PA$_4$LEH$_6$; where P is the last residue of wild-type BalhD). As expected from earlier experiments with BalhD PxPxPG (Table 4.2), addition of the longer tag abolished heterocycle formation (Fig. 5.3). Although heterocycle formation is stoichiometric with ATP hydrolysis for wild-type cyclodehydratase, the C-terminal His$_6$ tagged BalhD displayed robust ATPase activity, irrespective of the presence of BalhA1 (Fig. 5.4A). Moreover, this high level of unproductive ATP hydrolysis was potentiated by the addition of BalhC to the same extent observed with wild-type BalhD (2.5-fold increase; see Fig. 4.20), indicating that the His$_6$ tag did not interfere with BalhC recognition (Fig. 5.4A). With a BalhD derivative displaying robust BalhC-independent ATPase activity in-hand, I evaluated the role of BalhC on ATP utilization by BalhD by obtaining ATP Michaelis-Menten kinetics parameters for BalhD-A$_4$LEH$_6$ alone and in complex with BalhC (Fig. 5.4B). These data indicated that the addition of BalhC modulated BalhD activity by increasing the $k_{obs}$ and decreasing the $K_M$ for ATP.
Figure 5.3 | A C-terminal His\textsubscript{6} tag on BalhD abolishes cyclohydratase activity. A MALDI-TOF MS spectral overlay for BalhA1 treated with BalhD-A\textsubscript{4}LEH\textsubscript{6} either (a) with or (b) without BalhC is displayed. A reaction carried out with wild-type (WT) BalhD is shown for a comparison. †, laser-induced deamination.

Figure 5.4 | BalhC modulates ATP binding and hydrolysis by BalhD. (a) ATP hydrolysis rates measured by the PNP assay. (b) ATP kinetic curves for BalhD-A\textsubscript{4}LEH\textsubscript{6} with and without BalhC. Error bars represent the standard deviation from the mean (n=3), while error on the Michaelis-Menten parameters represents the standard deviation from the regression analysis.
5.5 The Role of Auxiliary E1-Homologs in Azoline Biogenesis

There are two TOMM biosynthetic clusters in *Bacillus* sp. Al Hakam (Fig. 5.5). The enzymes characterized in Chapters 2-4 of this thesis were from the canonical TOMM cluster (BalhA1, BalhC, etc.). The second cluster is a member of the uncharacterized heterocycloanthracin subfamily of TOMM natural products. Akin to all other heterocycloanthracin biosynthetic clusters, the second TOMM biosynthetic cluster in *Bacillus* sp. Al Hakam (Balh2) contains a dehydrogenase (Balh2B), a CD fusion protein (Balh2CD), and an enigmatic ThiF-like protein (Balh2F). Unlike Balh cluster 1, the Balh2 precursor peptide (Balh2A; see Fig. 5.5 for sequence) is located elsewhere on the genome and was identified based on similarity to other heterocycloanthracin precursor peptides.

The genes for the Balh2 modification enzymes and the precursor peptide were cloned from the Balh genome and expressed in *E. coli* as His$_6$- or MBP-fusion proteins, respectively. Although Balh2CD and Balh2F expressed well, Balh2B was heavily truncated and purified without the necessary FMN cofactor. In order to determine the minimal set of proteins required for azoline formation, reactions were initiated with Balh2F alone, Balh2CD alone or both Balh2CD and Balh2F, and processing of Balh2A was monitored matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS, Fig. 5.6). Samples lacking either of the enzymes or ATP did not display a detectable level of Balh2A substrate modification under the conditions employed. Increasing the concentrations of the individual components to 25 µM, similarly resulted in no modification. However, when the Balh2A substrate was treated with both Balh2CD and Balh2F, 13 azoline heterocycles (resulting in a 234 Da mass shift) were installed on the substrate. Using *N*-ethyl maleimide labeling, these modifications were localized to the 13 cysteines in the core peptide (Fig. 5.7). Although studies
carried out on other TOMM synthetases have demonstrated that cyclodehydratase activity can be dependent of the presence of the dehydrogenase,\textsuperscript{17-19} no additional domains have been shown to be required for azoline formation. As such, the absolute requirement for the presence of the ThiF-like protein is unprecedented. Given the size and cysteine rich nature of the substrate, the ability to convert the azolines to stable azoles would greatly assist in the characterization of substrate processing. Unfortunately attempts to obtain a full synthetase by using a Balh2B homolog were stymied by the inability to find a homolog that purified with FMN (see table A.4 for list of other dehydrogenases cloned).

In order to better understand the role of Balh2F in the cyclodehydration of Balh2A, the ability of Balh2F to associate with Balh2CD and the ability of each protein to bind Balh2A was interrogated by analytical size-exclusion chromatography (SEC; Fig. 5.8). Under the conditions tested, Balh2F and Balh2CD formed a heterodimeric complex. The addition of MBP-tagged Balh2A leader peptide (see Fig. 5.4) to this sample generated a new peak with a retention time consistent with a 2:1:1 complex of Balh2A:2F:2CD. In order to determine which protein was responsible for Balh2A leader peptide recognition, the control SEC traces were obtained in which either Balh2CD or Balh2F was omitted. While Balh2CD did not appear to interact with the Balh2A leader peptide, incubation of Balh2F and Balh2A leader peptide resulted in the appearance of a new peak with a retention time consistent with a 2:1 complex of Balh2A:2F (Fig. 5.8). Importantly, this association was unperturbed upon the addition of Balh2CD. These data provide definitive proof that Balh2F is responsible for leader peptide recognition in Balh cluster 2. As Balh2F homologs are found in all heterocycloanthracin biosynthetic clusters, it is likely that this strategy for precursor peptide recognition is universal in this TOMM subfamily.
**Figure 5.5 | Overview of the Balh TOMM biosynthetic clusters.** The gene clusters for each of the Balh TOMM clusters is displayed. Gene assignments are provided in the legend. The sequences of the precursor peptides for each cluster are shown along with the sequence of the Balh2A leader peptide used in this report. A blue hyphen denotes the putative leader peptide cleavage site, while the orange residues are positions known to be cyclized *in vitro*.
Figure 5.6 | Heterocycle formation on Balh2A is Balh2F and Balh2CD dependent. An uncalibrated linear positive mode MALDI-TOF MS spectral overlay for the Balh2A endpoint modification assays is displayed. Apart from the final trace, ATP is included in each sample. The mass of the major peaks is labeled along with the mass difference for the modified form. The theoretical mass for the peptide is displayed along with the ppm error for the measurement.
**Figure 5.7 | The cyclized Balh2A product contains 13 thiazolines.** A linear mode MALDI-TOF MS spectral overlay for N-ethylmaleimide (NEM) reacted Balh2A and Balh2CD/2F-treated Balh2A is displayed. The masses of the unmodified and cyclized peaks are displayed along with the mass shift relative to the unmodified species. The number of NEM labels on the peptide is indicated for the relevant peaks. The spectra were internally calibrated to the unmodified Balh2A peak.
Figure 5.8 | Balh2F is responsible for leader peptide recognition in Balh cluster 2. Analytical size-exclusion chromatography 280 nm absorbance traces are displayed. Numbers are assigned to each of the peaks in the traces and the calculated molecular weight for each numbered peak is displayed along with the calculated oligomeric state. For peaks 5 and 6, the oligomeric state in parenthesis is for a complex containing two molecules of MBP-Balh2A LP. The remained of the oligomeric states are calculated based off a 1:1 ratio of the binding partners. Asterisks denote the void peak.
5.6 Summary and Outlook

The data presented in this report demonstrate that TOMM E1 proteins play two distinct roles in azoline formation. First, the results demonstrate that TOMM E1 domains are responsible for leader peptide recognition, thus efficiently bringing the substrate in close proximity to the TOMM YcaO active site. These results are in accord with previous work implicating SagC in leader peptide recognition during streptolysin S biosynthesis, and the recent observation that TOMM C proteins house a MccB-like N-terminal “peptide clamp”, which is responsible for leader peptide binding in microcin C7 biosynthesis. Secondly, the serendipitous identification of a C-terminal His-tagged construct of BalhD with robust ATP hydrolysis (BalhD-A4LEH6) allowed us to show that the presence of BalhC increases $k_{obs}$ and lowers the $K_M$ for ATP. Although these data suggest that C protein potentiation occurs via allosteric activation, follow-up studies will be required to validate this hypothesis. Combined with the fact that TOMM E1 proteins lack the ATP-binding site that is conserved in all characterized non-TOMM E1 ubiquitin-activating family members, all lines of evidence suggest that TOMM C proteins engage the leader peptide while simultaneously potentiating the carbonyl activation chemistry of their cognate YcaO domain (D protein; Fig. 5.9A).

Based on the extensive bioinformatics that I have performed with TOMM synthetases, the above strategy is likely utilized in approximately 800 of the 1000 identifiable TOMM biosynthetic clusters in GenBank. In the remainder of the clusters, the canonical C protein is truncated and appears to have lost the N-terminal “peptide clamp”. My work with Balh cluster 2 demonstrates that in such clusters an auxiliary and highly divergent TOMM E1 homolog is responsible for leader peptide recognition. Although my data do not address the role that the truncated C domain plays on azoline formation, the simplest model is one in which the truncated
C domain is responsible for YcaO activation, while the ThiF-like protein provides leader peptide recognition (Fig. 5.9B). Further study is required to test this hypothesis.

Based on the data presented in this report, it is not yet clear how stand-alone TOMM YcaO proteins (*i.e.* for bottromycin and trifolitoxin production) perform cyclodehydrations in the absence of a C protein. Given the diversity between these stand-alone and canonical (C protein-containing) TOMM YcaOs, I envision that multiple solutions to the substrate recognition problem could exist. For example, it is possible that these biosynthetic pathways utilize an unidentified companion protein to bind the precursor peptide. Alternatively, these YcaO proteins may have evolved to bind a specific motif within the core peptide and modify the substrate a single time. Accordingly, the bottromycin and trifolitoxin stand-alone YcaO domains are each predicted to install a single heterocycle. This is in stark contrast to canonical TOMMs that process a wide array of core peptides, often at numerous locations. Such promiscuity is common in other RiPPs and likely accounts for the existence of leader peptides (*i.e.* the modification enzymes can be specific for motifs within the leader peptide, but promiscuous on the core once the enzyme-substrate complex is formed). Further work, including the reconstitution of a stand-alone YcaO, will be required to test these hypotheses.
Figure 5.9 | An updated model for azoline formation by TOMM cyclodehydratases. Based on the data presented in this report, two models for azoline formation are provided. (a) In TOMM cyclodehydratase complexes with a full length C protein (similar to Balh cluster 1) the TOMM E1 homolog (C protein) allosterically activates the YcaO domain (D protein) and binds the leader peptide (LP). Leader peptide binding by the C protein places the core peptide in proximity of the YcaO active site to facilitate the cyclodehydration reaction. (b) In TOMM cyclodehydratase complexes with a truncated C protein (similar to Balh cluster 2), the canonical C protein is not responsible for leader peptide recognition. Instead the ThiF-like protein binds the leader peptide.
5.7 Experimental

5.7.1 General methods. Unless otherwise specified, all chemicals were purchased from Sigma or Fisher Scientific. DNA sequencing was performed by the Roy J. Carver Biotechnology Center (UIUC). Restriction enzymes were purchased from New England Biolabs (NEB), Pfu Turbo was purchased from Agilent, and dNTPs were purchased from either NEB. Oligonucleotide primers were synthesized by Integrated DNA Technologies (IDT). Fluorescein labeled BalhA1 leader peptide was purchased from GenScript as an N-terminal FITC-Ahx (fluorescein isothiocynate, aminohexyl linker) conjugate with a single glycine spacer.

5.7.2 Balh cluster 2 cloning. Balh2A, 2CD, 2B and 2F were amplified by PCR from Bacillus sp. Al Hakam cells using the forward and reverse primers listed in table A.4. Polymerase reactions were carried out with PFU Turbo and the amplified product was digested with the indicated restriction enzymes following a gel extraction. The digested gene was PCR-purified and ligated into an appropriately digested, modified pET28 vector. For Balh2A the vector contained a tobacco etch virus (TEV) protease cleavable, N-terminal MBP-tag. For Balh2CD, 2B and 2F the vector contained a thrombin protease cleavable N-terminal His\(_6\)-tag.

5.7.3 Protein overexpression and purification. All MBP-tagged and His\(_6\)-tagged proteins were overexpressed and purified as previously described (see Chapters 2 and 3).\(^1\,3\)

5.7.4 \(^{18}\text{O}\) labeling of BalhD-treated BalhA1 reactions. 25 μM BalhD and 50 μM BalhA1 (both fused to MBP) were incubated with 0.2 μg/mL TEV protease in synthetase buffer for 18 h at 25 °C. Following treatment, BalhD, MBP and TEV were precipitated from the sample by the addition of acetonitrile to a final concentration of 50\% (v/v). The precipitated proteins were removed from the sample by centrifugation at 15,000 \(\times\) g and the supernatant was dried on a SpeedVac (Savant, Thermo Scientific). The resultant solid was resuspended in \(^{18}\text{O}\)-H\(_2\)O and
isotope labeling, trypsin digestion, and FT-MS/MS sequencing was carried out as described in Chapter 3.

5.7.5 BalhC leader peptide recognition. Equilibrium BalhC/BalhD fluorescence polarization binding assays were performed at 25 °C in non-binding surface, 384 black well, polystyrene microplates (Corning) and measured using a FilterMax F5 multi-mode microplate reader (Molecular Devices) with default settings. For each titration, protein was serially diluted into binding buffer [50 mM HEPES pH 7.5, 300 mM NaCl, 2.5% (v/v) glycerol, 0.5 mM TCEP], mixed with 10 nM fluorescein-labeled BalhA1 leader peptide (FP-BalhA1-LP), and equilibrated for 15 minutes with shaking prior to measurement. Data from three independent titrations were background subtracted and fitted using a non-linear dose response curve in OriginPro9 (OriginLab).

5.7.6 BalhD-A₄LEH₆ activity. Heterocycle formation assays were carried out with 50 µM MBP-BalhA1 and either 2 µM MBP-BalhC/D (CD activity) or 25 µM MBP-BalhD (D-only activity) in synthetase buffer [50 mM Tris (pH 7.5), 125 mM NaCl, 10 mM DTT, 20 mM MgCl₂ and 3 mM ATP]. The MBP tags were removed using 0.05 mg/mL of TEV protease and reactions were carried out for 18 h at 25 °C. Samples were desalted via C₁₈ ZipTip (Millipore) according to the manufacturer’s instructions and analyzed on a Bruker Daltonics UltrafleXtreme MALDI-TOF spectrometer. Spectra were obtained in positive reflector mode using α-cyano-hydroxycinnamic acid as the matrix.

Kinetic studies were performed using a previously described purine nucleoside phosphorylase (PNP)-coupled assay.³²⁸ The rate of ATP hydrolysis at variable ATP concentration (0.1-5 mM) was measured for 3 µM MBP-tagged BalhD-A₄LEH₆ with and without 3 µM MBP-tagged BalhC. Potentiation experiments were carried with 1 µM of the
indicated MBP-tagged enzymes and 100 µM BalhA1 (where applicable). Reactions were carried out in triplicate. Regression analyses to obtain the kinetic parameters for both substrates were carried out with IGOR Pro version 6.12 (WaveMetrics).

5.7.7 Balh cluster 2 overnight heterocyclization assays. Heterocycle formation assays were carried out with 50 µM MBP-Balh2A and 2 µM His6-Balh2CD/2F in synthetase buffer [50 mM Tris (pH 7.5), 125 mM NaCl, 10 mM TCEP, 20 mM MgCl₂ and 3 mM ATP]. The MBP tag was removed using 0.05 mg/mL of TEV protease and reactions were carried out for 18 h at 25 °C. Samples were desalted via C₁₈ ZipTip (Millipore) according to the manufacturer’s instructions, eluted in a saturated solution of sinapinic acid in 70% ACN, and analyzed on a Bruker Daltonics UltrafleXtreme MALDI-TOF spectrometer. Spectra were obtained in linear positive mode.

5.7.8 Size-exclusion chromatography. A 200 µL sample containing 25 µM of the indicated proteins (MBP-Balh2A LP, His6-Balh2CD, His6-Balh2F) was prepared in SEC running buffer [20 mM Tris (pH 7.5), 500 mM NaCl, 10 % glycerol (v/v), 0.5 mM TCEP]. Samples were incubated at 4 °C for 30 min prior to analysis via analytical SEC. The oligomeric state of the proteins and size of the pertinent protein complexes were assessed on a Flexar HPLC (Perkin Elmer) equipped with an analytical Yarra SEC-3000 (300 x 4.6 mm, Phenomenex) equilibrated with running buffer. Peaks of interest were collected and their composition was determined via a Coomassie stained 12% SDS-PAGE gel. The approximate molecular weights were determined by generating a standard curve with a 12-200 kDa molecular weight standard kit (Sigma). Chromatograms were analyzed using Flexar Manager (Perkin Elmer).

5.8 References


## APPENDIX A: LIST OF PRIMERS

### Table A.1 | Primers for chapter 3 constructs.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Cut Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>BalhA1 Core F</td>
<td>AGGATCTCTTGTTGTCAGGTATGATG</td>
<td>5'- BamHI</td>
</tr>
<tr>
<td>BalhA1 Core R</td>
<td>TTTGCGGCCGCCTTTTTTAATCCTGAGCATTATTTTTGCAAACCC</td>
<td>3'- Ndel</td>
</tr>
<tr>
<td>BalhA1 EPL F</td>
<td>GCAGCCATATGGAAATCGAGGCGATG</td>
<td>5'- Ndel</td>
</tr>
<tr>
<td>BalhA1 EPL R</td>
<td>ATATATCCCGGCGCCTGCG</td>
<td>3'- Smal</td>
</tr>
<tr>
<td>McbC (K201A) F</td>
<td>CATTATTTTTGAAAAAGCTTTTGCCATCAGCAGGATATCGGATG</td>
<td>NA</td>
</tr>
<tr>
<td>McbC (K201A) R</td>
<td>CATCCGATATCTCTCTGTGAGCATATGCGAACAAAGCTTTTTGCAAAAATATG</td>
<td>NA</td>
</tr>
<tr>
<td>McbC (Y202A) F</td>
<td>TCTATGCGATTTTTGAAAAAGCTTTTGCCATCAGCAGGATATCGGATG</td>
<td>NA</td>
</tr>
<tr>
<td>McbC (Y202A) R</td>
<td>CGATATCCTCTTGAGGCGGCTTTTGCAAAAGCTTTTTGCAAAAATATGAGAAGC</td>
<td>NA</td>
</tr>
</tbody>
</table>

Note: BalhA1 EPL F/R were used to generate an expressed protein ligation vector for the semi-synthetic preparation of an 18O-labeled BalhA1 substrate. The yields from the EPL synthetic strategy were not high enough to be useful for mechanistic investigations of the TOMM cyclodehydratase.

### Table A.2 | Primers for chapter 4 constructs. *, denote stop codons.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Cut Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>BalhD F</td>
<td>CCGGATCCATGGGTATACAGAATG</td>
<td>5'- BamHI</td>
</tr>
<tr>
<td>BalhD (S72A) F</td>
<td>CAGCCTAATAGCGAGGCGAGGAGAAATTCTTGAGCGGTTATTGC</td>
<td></td>
</tr>
<tr>
<td>BalhD (S72A) R</td>
<td>TCTCCGATATCTCTGTGAGCATATGCGAACAAAGCTTTTTGCAAAAATATG</td>
<td></td>
</tr>
<tr>
<td>BalhD (E76A) F</td>
<td>CTATATCAGGCGAGGAGAAATTCTTGAGCGGTTATTGC</td>
<td></td>
</tr>
<tr>
<td>BalhD (E76A) R</td>
<td>GCAATACGCTCAAGAATTCTGCGGACTCAGTTATG</td>
<td></td>
</tr>
<tr>
<td>BalhD (E79A) F</td>
<td>GGCAGTTTGGAAGAAATTCTTGCGGACTCAGTTATG</td>
<td></td>
</tr>
<tr>
<td>BalhD (E79A) R</td>
<td>GATAATCGGCAAAATAACCGCGGCAAAATAACCGCGTGACTCAGATG</td>
<td></td>
</tr>
<tr>
<td>BalhD (R80A) F</td>
<td>CGGCGATTTGGAGAAATTCTTGAGGCGGTTATTGC</td>
<td></td>
</tr>
<tr>
<td>BalhD (R80A) R</td>
<td>TATTTAGATAAATAAGCAGTAAGCAATAGCTCAAGAATTCTTGCGGCTTATG</td>
<td></td>
</tr>
<tr>
<td>BalhD (Q186A) F</td>
<td>CTACAAGGTGGCAGCAATAGAAAAACGCGGCAACTAGAAATG</td>
<td></td>
</tr>
<tr>
<td>BalhD (Q186A) R</td>
<td>GTTTTCTATTGCTGCCAAGCATTGGAACCTGCTCGGTAACCCCAG</td>
<td></td>
</tr>
<tr>
<td>BalhD (N190A) F</td>
<td>GATTTCTACAAAGGTGCGAAGCAGATAAGGCGGCAACTAGAAATG</td>
<td></td>
</tr>
<tr>
<td>BalhD (N190A) R</td>
<td>ATTCTAGTGCGGCGGCTTTCTTATGCGCTAGGAAACCTGCTGAGGACCC</td>
<td></td>
</tr>
<tr>
<td>BalhD (E194A) F</td>
<td>TAGAAAACGCGGACACTAGCTAGTATAGAAAGAGAAGACGGAGG</td>
<td></td>
</tr>
<tr>
<td>BalhD (E194A) R</td>
<td>GCGTCTCCTTCTATCATACTAGTGATGCGCGGCTTTCTATA</td>
<td></td>
</tr>
<tr>
<td>BalhD (E197A) F</td>
<td>CGGCGGACAAGGAATGGAAGCAGAAAGAGAAGACGGAGG</td>
<td></td>
</tr>
<tr>
<td>BalhD (E197A) R</td>
<td>ATAATCGGCTCTCTTGCTATACGTATGTGCGCGG</td>
<td></td>
</tr>
<tr>
<td>BalhD (R198A) F</td>
<td>AATTGATAGAAGCAGGAGGCGGAGGATGATGACATGCAAGTATTGAAATG</td>
<td></td>
</tr>
<tr>
<td>BalhD (R198A) R</td>
<td>ATAATCGGCTCTCTTCTATACGTATGTGCGCGG</td>
<td></td>
</tr>
<tr>
<td>BalhD (K281A) F</td>
<td>CGATCCCTTATAGCGGCGGAGGCTTTTATGAGACGG</td>
<td></td>
</tr>
</tbody>
</table>
### Table A.2 Continued

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Cut Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>BalhD (K281A) R</td>
<td>CGTCTCCATTAAAGCTCCGCAATGCTATAGAGGATCG</td>
<td></td>
</tr>
<tr>
<td>BalhD (E286A) F</td>
<td>GAAGGGAGCTTTATGGAAGGGTGCCTGCCCTCCCTC</td>
<td></td>
</tr>
<tr>
<td>BalhD (E286A) R</td>
<td>TTGAGCTTGCAGCTGCCATTAAAGCTCCCTC</td>
<td></td>
</tr>
<tr>
<td>BalhD (P425*) F</td>
<td>CGCTTACGGAAATGGATGCTAATTCTTATTTAAAGCTTTTCGCTGACATCC</td>
<td>3' - NotI</td>
</tr>
<tr>
<td>BalhD (P425*) R</td>
<td>AAGGGAGCTTTATGGAAGGGTGCCTGCCCTCCTTCTCCGCTGACATCC</td>
<td>3' - NotI</td>
</tr>
<tr>
<td>BalhD (P427*) F</td>
<td>GAGCTTACGGGCGCTTACCCGAAATTGATGCTAATTCTTATTTAAAGCTTTTCGCTGACATCC</td>
<td>3' - NotI</td>
</tr>
<tr>
<td>BalhD (P427*) R</td>
<td>GAGCTTACGGGCGCTTACCCGAAATTGATGCTAATTCTTATTTAAAGCTTTTCGCTGACATCC</td>
<td>3' - NotI</td>
</tr>
<tr>
<td>BalhD (P429*) F</td>
<td>GAGCTTACGGGCGCTTACCCGAAATTGATGCTAATTCTTATTTAAAGCTTTTCGCTGACATCC</td>
<td>3' - NotI</td>
</tr>
<tr>
<td>BalhD (P429*) R</td>
<td>GAGCTTACGGGCGCTTACCCGAAATTGATGCTAATTCTTATTTAAAGCTTTTCGCTGACATCC</td>
<td>3' - NotI</td>
</tr>
<tr>
<td>BalhD (P429G) R</td>
<td>GAGCTTACGGGCGCTTACCCGAAATTGATGCTAATTCTTATTTAAAGCTTTTCGCTGACATCC</td>
<td>3' - NotI</td>
</tr>
<tr>
<td>BalhD (PxPxPG) R</td>
<td>GAGCTTACGGGCGCTTACCCGAAATTGATGCTAATTCTTATTTAAAGCTTTTCGCTGACATCC</td>
<td>3' - NotI</td>
</tr>
<tr>
<td>BalhD (∆A418/K419) F</td>
<td>GAGCTTACGGGCGCTTACCCGAAATTGATGCTAATTCTTATTTAAAGCTTTTCGCTGACATCC</td>
<td>3' - NotI</td>
</tr>
<tr>
<td>BalhD (∆A418/K419) R</td>
<td>GAGCTTACGGGCGCTTACCCGAAATTGATGCTAATTCTTATTTAAAGCTTTTCGCTGACATCC</td>
<td>3' - NotI</td>
</tr>
</tbody>
</table>

### Table A.3 | Primers for YcaO constructs cloned for crystallography screens.

All proteins are identified by locus tag. Primers required for cloning the construct into a vector different than the standard pET28-MBP vector have the vector listed in parentheses. Additional information for each protein can be found in Table C.1.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Cut Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>Azl_d00180 F</td>
<td>AAAGGATCCATGCGCCCTGACATCTGCGATGCTCAAGGC</td>
<td>5' - BamHI</td>
</tr>
<tr>
<td>Azl_d00180 R</td>
<td>AAAGGATCCATGCGCCCTGACATCTGCGATGCTCAAGGC</td>
<td>3' - Xhol</td>
</tr>
<tr>
<td>Azl_022710 F</td>
<td>AAAGGATCCATGCGCCCTGACATCTGCGATGCTCAAGGC</td>
<td>5' - BamHI</td>
</tr>
<tr>
<td>Azl_022710 R</td>
<td>AAAGGATCCATGCGCCCTGACATCTGCGATGCTCAAGGC</td>
<td>3' - HindIII</td>
</tr>
<tr>
<td>Azl_a09740 F</td>
<td>AAAGGATCCATGCGCCCTGACATCTGCGATGCTCAAGGC</td>
<td>5' - BamHI</td>
</tr>
<tr>
<td>Azl_a09740 R</td>
<td>AAAGGATCCATGCGCCCTGACATCTGCGATGCTCAAGGC</td>
<td>3' - NotI</td>
</tr>
<tr>
<td>BalhD 1-414 R</td>
<td>TTGAGCTTGCAGCTGCCATTAAAGCTCCCTC</td>
<td>3' - NotI</td>
</tr>
<tr>
<td>BalhD a57 F</td>
<td>AAGGATCCATGCGCCCTGACATCTGCGATGCTCAAGGC</td>
<td>5' - BglII</td>
</tr>
<tr>
<td>BpumD F</td>
<td>AAGGATCCATGCGCCCTGACATCTGCGATGCTCAAGGC</td>
<td>3' - Sall</td>
</tr>
</tbody>
</table>

188
Table A.3 Continued

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Cut Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>BpumD R</td>
<td>AATTGGGCCGCTTAAAGGAAATGGATGCGGAATTGGATTCAGTTCTTCC</td>
<td>3'-NotI</td>
</tr>
<tr>
<td>Ec-YcaO Δ59 F</td>
<td>AAGGATCCGCTAAGGCGCAACCAAGAAAGCGGC</td>
<td></td>
</tr>
<tr>
<td>Ec-YcaO 1-420 R</td>
<td>TGGCCGGCGCTTATTAGCTGCTGGTATCGCAAAATGGCTTACGTTATG</td>
<td>3'-NotI</td>
</tr>
<tr>
<td>M446_2287 F</td>
<td>AAGGATCCGCTAAGGCGCAACCAAGAAAGCGGC</td>
<td></td>
</tr>
<tr>
<td>M446_2287 R</td>
<td>TTTAGCTTTACGCGCCCTCCCG</td>
<td>3'-HindII</td>
</tr>
<tr>
<td>ERDMAN_1534 F</td>
<td>AAAGGATCCATGTTTGCGGCCCAGATTTGTGTC</td>
<td>5'-BamHI</td>
</tr>
<tr>
<td>ERDMAN_1534 R</td>
<td>TTTAGCTTTACGCGCCCTCCCG</td>
<td>3'-HindII</td>
</tr>
<tr>
<td>HMPREF9018_0535 F</td>
<td>AAAGGATCCATGCTGCTGTTTGAATTGCACATG</td>
<td>5'-BamHI</td>
</tr>
<tr>
<td>HMPREF9018_0535 R</td>
<td>TTTCGCGCGCTTTAGGGGAATGGATGTGGATATACATTTCG</td>
<td>3'-NotI</td>
</tr>
<tr>
<td>Scab_1701 F1</td>
<td>AAACCTGAGATGCACGCGACCGCTCGACCAGT</td>
<td>5'Xhol</td>
</tr>
<tr>
<td>Scab_1701 F2 (pALM)</td>
<td>AAAAGGCTTCCATGCGCGAAGCGGCCGTCACCTTG</td>
<td>5'-HindII</td>
</tr>
<tr>
<td>Scab_5667 F</td>
<td>TTTCGCGCGCTTACGCGACCGCTCGACCAGT</td>
<td>5'-BamHI</td>
</tr>
<tr>
<td>Scab_5667 R</td>
<td>TTTCGCGCGCTTACGCGACCGCTCGACCAGT</td>
<td>3'-NotI</td>
</tr>
<tr>
<td>Scab_56661 (∆100) F</td>
<td>AAAGGATCCATGCGCGAAGCGGCCGTCGAGCG</td>
<td>5'-BamHI</td>
</tr>
<tr>
<td>Scab_56661 F2</td>
<td>AAAGGATCCATGCGCGAAGCGGCCGTCGAGCG</td>
<td>5'-BamHI</td>
</tr>
<tr>
<td>Scab_56661 R</td>
<td>TTTCGCGCGCTTACGCGACCGCTCGACCAGT</td>
<td>3'-HindII</td>
</tr>
<tr>
<td>SO_0795 F</td>
<td>AAAGGATCCATGCGCGAAGCGGCCGTCGAGCG</td>
<td>5'-BamHI</td>
</tr>
<tr>
<td>SO_0795 R</td>
<td>TTTAGCTTTAACGACGATTTTCGACAGTGTGCGGCGGTGAGCG</td>
<td>3'-HindII</td>
</tr>
</tbody>
</table>

Table A.4 | Primers for chapter 5 constructs. *, denote stop codons. Note, BAT_3535 and bcer0022_47460 are Balh2B homologs from *B. pumilus* ATCC 7061 and *B. cereus* Rock3-44, respectively.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Cut Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balh2A F</td>
<td>AAGTCGACAAACGATGAAATCAGTTTCAAAACAGAATCGATCAATCATTAAACC</td>
<td>5'-SalI</td>
</tr>
<tr>
<td>Balh2A R</td>
<td>TTTCGCGCGCTAATATGATCACAGTTTCAAAACAGAATCGATCAATCATTAAACC</td>
<td></td>
</tr>
<tr>
<td>Balh2A LP F</td>
<td>CGCGATCCGCGCGGCGCGCGACTACGTCGAGCTTGATGGATATAGTAGTA</td>
<td>3'-NotI</td>
</tr>
<tr>
<td>Balh2A LP R</td>
<td>TACTACTATATTTCAAGACGATGCACTAGTGTGGCGGCGGTGAGCG</td>
<td></td>
</tr>
<tr>
<td>Balh2B F</td>
<td>AAGGATCCATGCGCGAAGCGGCCGTCGAGGCATTTACGCGAATCGATCATTAAACC</td>
<td>5'-BamHI</td>
</tr>
<tr>
<td>Balh2B R</td>
<td>TTTCGCGCGCTAATATGATCACAGTTTCAAAACAGAATCGATCAATCATTAAACC</td>
<td>3'-NotI</td>
</tr>
<tr>
<td>Balh2CD F</td>
<td>AAGGATCCATGCGCGAAGCGGCCGTCGAGGCATTTACGCGAATCGATCATTAAACC</td>
<td>5'-BamHI</td>
</tr>
<tr>
<td>Balh2CD R</td>
<td>TTTCGCGCGCTAATATGATCACAGTTTCAAAACAGAATCGATCATTAAACC</td>
<td>3'-NotI</td>
</tr>
<tr>
<td>Balh2F F</td>
<td>AAGGATCCATGCGCGAAGCGGCCGTCGAGGCATTTACGCGAATCGATCATTAAACC</td>
<td>5'-BamHI</td>
</tr>
<tr>
<td>Balh2F R</td>
<td>AAAGGATCCATGCGCGAAGCGGCCGTCGAGGCATTTACGCGAATCGATCATTAAACC</td>
<td>3'-NotI</td>
</tr>
<tr>
<td>BAT_3535 F</td>
<td>AAGGATCCATGCGCGAAGCGGCCGTCGAGGCATTTACGCGAATCGATCATTAAACC</td>
<td>5'-BamHI</td>
</tr>
<tr>
<td>BAT_3535 R</td>
<td>TTTCGCGCGCTAATATGATCACAGTTTCAAAACAGAATCGATCATTAAACC</td>
<td>3'-NotI</td>
</tr>
<tr>
<td>bcer0022_47460 F</td>
<td>AAAGGATCCATGCGCGAAGCGGCCGTCGAGGCATTTACGCGAATCGATCATTAAACC</td>
<td>5'-BamHI</td>
</tr>
<tr>
<td>bcer0022_47460 R</td>
<td>TTTCGCGCGCTAATATGATCACAGTTTCAAAACAGAATCGATCATTAAACC</td>
<td>3'-NotI</td>
</tr>
<tr>
<td>BalhD F</td>
<td>CGCGATCCGCGCGAAGCGGCCGTCGAGGCATTTACGCGAATCGATCATTAAACC</td>
<td>5'-BamHI</td>
</tr>
<tr>
<td>BalhD (*430A) R</td>
<td>GTTCGCCGCTGCGCGAAGCGGCCGTCGAGGCATTTACGCGAATCGATCATTAAACC</td>
<td>3'-NotI</td>
</tr>
</tbody>
</table>
### Table A.5 | Primers for unpublished cyclodehydratase mutants

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>BalhC D89A F</td>
<td>CAATTCATTACATAAATTTATATTTTTCTGTAAGCGTTTTGTCGCTGTAACATTTT</td>
</tr>
<tr>
<td>BalhC D89A R</td>
<td>ATTTAGAGACGAAAAGATTAATATATTATACGCTTTACAGAAAATATATTATTTATGGAAGAAAA</td>
</tr>
<tr>
<td>BalhC S109A F</td>
<td>TGCAAGATATGGAAGATTCATCGCAAATAATTGGACAGGTCTTTATCC</td>
</tr>
<tr>
<td>BalhC S109A R</td>
<td>GGATAGAGATGGAAGAAGATTCATCGCAAATAATTGGACAGGTCTTTATCC</td>
</tr>
<tr>
<td>BalhC E136A F</td>
<td>CTTTACATTTCCTGTAATTCGCATTTCCTTTCCATTAATAATATTC</td>
</tr>
<tr>
<td>BalhC E136A R</td>
<td>GATATATGGAAGTGAATTCATCGCAAATAATTGGACAGGTCTTTATCC</td>
</tr>
<tr>
<td>BalhC Q139A F</td>
<td>CACAGGCTTTCTCTACATTTCCTTCGTAATTTCCTTTCCATTAATAATATTC</td>
</tr>
<tr>
<td>BalhC Q139A R</td>
<td>ATTTAGAGATGGAAGATTCATCGCAAATAATTGGACAGGTCTTTATCC</td>
</tr>
<tr>
<td>BalhC T211A F</td>
<td>GTATAGGGCCCAATTTCTTCTAATAGCTAATTTGTCGCTGTAACATTTT</td>
</tr>
<tr>
<td>BalhC T211A R</td>
<td>CCAGATTAAGATGGAAGATTCATCGCAAATAATTGGACAGGTCTTTATCC</td>
</tr>
<tr>
<td>BalhC R220A F</td>
<td>AGATAAGAGTTTTTTCATAAGTATATTTTCCGACTTAAAAAGCGTTTATAAATACCC</td>
</tr>
<tr>
<td>BalhC R220A R</td>
<td>GAAAGGGTTATTTAATAGCTAATTTTCCGACTTAAAAAGCGTTTATAAATACCC</td>
</tr>
<tr>
<td>BalhC E226A F</td>
<td>CATATATTTATAGCTAAGATAAGAGTTTTTCATATTTTCCGACTTAAAAAGCGTTTATAAATACCC</td>
</tr>
<tr>
<td>BalhC E226A R</td>
<td>TCTTTCCTACATTTCCTTCGTAATTCGCATTTCCTTTCCATTAATAATATTC</td>
</tr>
<tr>
<td>BalhC K221A F</td>
<td>GGTTATTTATAGCTAAGATAAGAGTTTTTCATATTTTCCGACTTAAAAAGCGTTTATAAATACCC</td>
</tr>
<tr>
<td>BalhC K221A R</td>
<td>CACAGGCTTTCTCTACATTTCCTTCGTAATTTCCTTTCCATTAATAATATTC</td>
</tr>
<tr>
<td>BalhC Y232A F</td>
<td>ACTTTCACTATTATTTTCTTCTTCTATATTTGTCGCTGTAACATTTT</td>
</tr>
<tr>
<td>BalhC Y232A R</td>
<td>CTCTTTCCTACATTTCCTTCGTAATTCGCATTTCCTTTCCATTAATAATATTC</td>
</tr>
<tr>
<td>BalhC E241A F</td>
<td>CCCCATCAAGGCTTTTCACTTTCACTTTCTTCTATATTTTCCGACTTAAAAAGCGTTTATAAATACCC</td>
</tr>
<tr>
<td>BalhC E241A R</td>
<td>TCTTTCCTACATTTCCTTCGTAATTCGCATTTCCTTTCCATTAATAATATTC</td>
</tr>
<tr>
<td>BalhC E244A F</td>
<td>CTGAAATTTATGGAAGATTCATCGCAAATAATTGGACAGGTCTTTATCC</td>
</tr>
<tr>
<td>BalhC E244A R</td>
<td>GAAAGGGTTATTTAATAGCTAATTTTCCGACTTAAAAAGCGTTTATAAATACCC</td>
</tr>
<tr>
<td>BalhC E276A F</td>
<td>CTTCCTGATTGTTGATTGCAAATAATTGGGAAATTTTCACTTTCTTCTTTTCCGACTTAAAAAGCGTTTATAAATACCC</td>
</tr>
<tr>
<td>BalhC E276A R</td>
<td>AATTTAATGGAAGATTCATCGCAAATAATTGGGAAATTTTCACTTTCTTCTTTTCCGACTTAAAAAGCGTTTATAAATACCC</td>
</tr>
<tr>
<td>BalhC K298A F</td>
<td>TCTTTCCTACATTTCCTTCGTAATTCGCATTTCCTTTCCATTAATAATATTC</td>
</tr>
<tr>
<td>BalhC K298A R</td>
<td>CACAGGCTTTCTCTACATTTCCTTCGTAATTTCCTTTCCATTAATAATATTC</td>
</tr>
<tr>
<td>BalhC K298A R</td>
<td>AATTTAATGGAAGATTCATCGCAAATAATTGGGAAATTTTCACTTTCTTCTTTTCCGACTTAAAAAGCGTTTATAAATACCC</td>
</tr>
<tr>
<td>BalhD G58A F</td>
<td>GAAATATTAGTGCTTTTAGGATATGCTATTGCTAGATGAC</td>
</tr>
<tr>
<td>BalhD G58A R</td>
<td>CTATATCTAAAAGCAACTTATAATTTGCCGAAGGTTTATATAGCCTT</td>
</tr>
<tr>
<td>BalhD L59K F</td>
<td>ATTATAGTGTAAGGATATGCTATTGCTAGATGAC</td>
</tr>
<tr>
<td>BalhD L59K R</td>
<td>TAGCTATATCTTCTTTACATTATAATTTCCGCGGATTATATG</td>
</tr>
<tr>
<td>BalhD E60A F</td>
<td>ATAGTGTTTTAGATATTAGCTATTGCTAGATGAC</td>
</tr>
<tr>
<td>BalhD E60A R</td>
<td>TGCTATATCTTCTTTACATTATAATTTCCGCGGATTATATG</td>
</tr>
<tr>
<td>BalhD Y81A F</td>
<td>TTACCTGATTGTTGATTGCAAATAATTGGGAAATTTTCACTTTCTTCTTTTCCGACTTAAAAAGCGTTTATAAATACCC</td>
</tr>
<tr>
<td>BalhD Y81A R</td>
<td>AATTTAATGGAAGATTCATCGCAAATAATTGGGAAATTTTCACTTTCTTCTTTTCCGACTTAAAAAGCGTTTATAAATACCC</td>
</tr>
<tr>
<td>Primer</td>
<td>Sequence</td>
</tr>
<tr>
<td>----------</td>
<td>------------------------------------------------------------</td>
</tr>
<tr>
<td>BalhD Q121A F</td>
<td>ATTACGCAGCCTTTAAGAGAGCAATATGCAGAAACATATGGTATTAGCAAAG</td>
</tr>
<tr>
<td>BalhD Q121A R</td>
<td>CTTTGCAATTACGATGTTTTTGCTGATTTGCTCTTTAAAGCGCTGAAAT</td>
</tr>
<tr>
<td>BalhD S175A F</td>
<td>CCCACATAAAGAGATCTTTGATTTAGGCAAACAGG</td>
</tr>
<tr>
<td>BalhD S175A R</td>
<td>CCTGTTAACAAACCAGTAACTGCAATCTCTTTTATATGTTGGG</td>
</tr>
<tr>
<td>BalhD E189A F</td>
<td>TGCAAGCAATAGCAAAACGCGCAGCTAAGATGTATAG</td>
</tr>
<tr>
<td>BalhD E189A R</td>
<td>AGTGGGCGCTTTTGCTTTGCACCTTTGAG</td>
</tr>
<tr>
<td>BalhD D199A F</td>
<td>CACTAGAATGTATAGAAAGAGCCCGATTTGATGACATACG</td>
</tr>
<tr>
<td>BalhD D199A R</td>
<td>CCATGTCATACAATCGCGCCCTTTTCTATACTTATATGTP</td>
</tr>
<tr>
<td>BalhD W205A F</td>
<td>TTTAGTACACAGCGTTAAATGAAATTCGTTGGTCCATTAAATG</td>
</tr>
<tr>
<td>BalhD W205A R</td>
<td>AATTCATTTAACGCTGTAGCATACAATCGCGCTCTTTTCTATAC</td>
</tr>
<tr>
<td>BalhD K271A F</td>
<td>AAATAGGCAGCACAGACATTACAGATGCTCTTTATAGTAGG</td>
</tr>
<tr>
<td>BalhD K271A R</td>
<td>TCGTAATGTGCTGCTGCTCTTTTGTATATGAGATATTATATTAAAGATCC</td>
</tr>
<tr>
<td>BalhD D275A F</td>
<td>AAGCAACATTTCGTCCTTCTTTAGCATTGAAGGGAGCTTTAATG</td>
</tr>
<tr>
<td>BalhD D275A R</td>
<td>CGTTAAGAGAGGCATATGCTTTTGCTCTATTTGTATATGAGG</td>
</tr>
<tr>
<td>BalhD K281R F</td>
<td>TTATAGGCTCCTCTTCTTATAGCATTGAAAGGAGCTTTAATG</td>
</tr>
<tr>
<td>BalhD K281R R</td>
<td>ATTTACTCCTCCCTCAATGCTATAAGAGAGCTGTAATGCTTTTGC</td>
</tr>
<tr>
<td>BalhD G282R F</td>
<td>TAGGAACTAAAGAGGACCTTTAATGGGAGCTTTGCAAGTC</td>
</tr>
<tr>
<td>BalhD G282R R</td>
<td>TCCATTTAAAGCCTCTCTTCTATAGCATAAGAGGAGCTGTAATG</td>
</tr>
<tr>
<td>BalhD P389G F</td>
<td>CTTATATGTATATGAGATGTTAGGAGGAGGCTTTTGAAAAATAACT</td>
</tr>
<tr>
<td>BalhD P389G R</td>
<td>AGTTATTTCCAAAAAGCCAACCTCTCCCATCAACACTCGATATACATATAAG</td>
</tr>
<tr>
<td>BalhD D370A F</td>
<td>TCATATGCTGCTACTGAGATAAGATGTTCTTTAGG</td>
</tr>
<tr>
<td>BalhD D370A R</td>
<td>TCGAGTTAGAGCATATGTATAGAGGTTTTAAATCTGTATAGTT</td>
</tr>
<tr>
<td>BalhD D375A F</td>
<td>TACTACTGAAAGCTATAAGTTCTTTAGGCTTATATGTATATCGAG</td>
</tr>
<tr>
<td>BalhD D375A R</td>
<td>AAAGAAGTATTACCGCTCTAGAATCTGATATGAGGTTTTAAATCTC</td>
</tr>
<tr>
<td>BalhD R405A F</td>
<td>TAACTTGTAATGCGTTATTAGATGCTCCGAAACACATGGG</td>
</tr>
<tr>
<td>BalhD R405A R</td>
<td>GCATCTAATAACGCGATTACAAACTTAGCTAGGTAAGTTATTTT</td>
</tr>
</tbody>
</table>
APPENDIX B: ADDITIONAL BALH CYCLODEHYDRATASE MUTANTS

**B.1 Unpublished BalhC mutants.** Mutations to BalhC were made in an attempt to locate residues responsible for YcaO potentiation. As structural information for a TOMM C protein was not available when the mutant library was constructed, the residues to be mutated were selected via bioinformatics. The effect of these mutations on the rate of cyclodehydration is displayed in table B.1.

**Table B.1 | Mutant BalhC activity.** *, reaction carried out with 15 μM BalhA1 and 3 mM ATP. All kinetic experiments were performed on the plate reader.

<table>
<thead>
<tr>
<th>Mutation</th>
<th>*Vo μM min⁻¹</th>
<th>% WT</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>6.6 ± 0.5</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td>D89A</td>
<td>0.11 ± 0.07</td>
<td>1</td>
<td>Potentially disrupted complex formation (see pg. 547)</td>
</tr>
<tr>
<td>S109A</td>
<td>6.9 ± 0.6</td>
<td>104</td>
<td>-</td>
</tr>
<tr>
<td>E136A</td>
<td>6.2 ± 0.2</td>
<td>94</td>
<td>-</td>
</tr>
<tr>
<td>E137A</td>
<td>0.5 ± 0.2</td>
<td>8</td>
<td>Only 5 mg/L obtained. Structurally destabilized.</td>
</tr>
<tr>
<td>Q139A</td>
<td>5.7 ± 0.6</td>
<td>91</td>
<td>-</td>
</tr>
<tr>
<td>T211A</td>
<td>5.6 ± 0.4</td>
<td>84</td>
<td>-</td>
</tr>
<tr>
<td>K221A</td>
<td>5.0 ± 0.3</td>
<td>76</td>
<td>-</td>
</tr>
<tr>
<td>R220A</td>
<td>0.16 ± 0.06</td>
<td>3</td>
<td>No measureable heterocycle formation. Disrupted interaction with BalhD?</td>
</tr>
<tr>
<td>E226A</td>
<td>4.6 ± 0.5</td>
<td>76</td>
<td>Small change in BalhA1 Kₘ and kₘₐₓ</td>
</tr>
<tr>
<td>Y232A</td>
<td>3.6 ± 0.2</td>
<td>55</td>
<td>-</td>
</tr>
<tr>
<td>E241A</td>
<td>5.4 ±0.5</td>
<td>85</td>
<td>-</td>
</tr>
<tr>
<td>E244A</td>
<td>3.5 ± 0.3</td>
<td>56</td>
<td>BalhA1 Kₘ increased 2-fold, kₘₐₓ down 20%</td>
</tr>
<tr>
<td>E276A</td>
<td>5.1 ±0.5</td>
<td>78</td>
<td>-</td>
</tr>
<tr>
<td>E281A</td>
<td>1.1 ± 0.3</td>
<td>17</td>
<td>Only 3 mg/L obtained. Structurally destabilized.</td>
</tr>
<tr>
<td>K298A</td>
<td>5.9 ± 0.4</td>
<td>94</td>
<td>-</td>
</tr>
</tbody>
</table>
B.2 | Unpublished BalhD mutants. Following the discovery that BalhD can install azoline heterocycles in the absence of BalhC, a mutant library was generated to locate the ATP-binding site and the catalytic residues in BalhD. All mutated residues were selected based on bioinformatics. Mutations that were not discussed in Chapters 4 and 5 are listed in tables B.2-B.3.

Table B.2 | Mutant BalhD activity. *, reaction carried out with 15 µM BalhA1 and 3 mM ATP.

<table>
<thead>
<tr>
<th>Mutation</th>
<th>( *V_0 ) µM min(^{-1} )</th>
<th>% WT</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>6.2 ± 0.5</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>G58A</td>
<td>2.55 ± 0.13</td>
<td>41</td>
<td>No background rate potentiation</td>
</tr>
<tr>
<td>L59K</td>
<td>0.59 ± 0.05</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>G60A</td>
<td>0.56 ± 0.05</td>
<td>9</td>
<td>No background rate potentiation</td>
</tr>
<tr>
<td>Y81A</td>
<td>0.13 ± 0.16</td>
<td>2</td>
<td>No background rate potentiation; Protein expresses very well</td>
</tr>
<tr>
<td>S175A</td>
<td>0.43 ± 0.07</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>E189A</td>
<td>-</td>
<td>-</td>
<td>Protein is structurally destabilized</td>
</tr>
<tr>
<td>D199A</td>
<td>-</td>
<td>-</td>
<td>Protein is structurally destabilized</td>
</tr>
<tr>
<td>W205A</td>
<td>-</td>
<td>-</td>
<td>Protein is structurally destabilized</td>
</tr>
<tr>
<td>K271A</td>
<td>5.1 ± 0.4</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td>D275A</td>
<td>4.3 ± 0.2</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>K281R</td>
<td>3 ± 0.2</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>D370A</td>
<td>0.86 ± 0.04</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>D375A</td>
<td>5.6 ± 0.8</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>R385A</td>
<td>0.16 ± 0.03</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>P389G</td>
<td>1.3 ± 0.1</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>R405A</td>
<td>0.28 ± 0.01</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>N424*</td>
<td>0.31 ± 0.14</td>
<td>5</td>
<td>No heterocycle formation</td>
</tr>
</tbody>
</table>
Table B.3 | Mutant BalhD BalhA1 and ATP Kinetics. *, reaction carried out with 3 mM ATP; **, reaction carried out with 80 µM BalhA1; -, kinetics not obtained; n.d., kinetics could not be determined.

<table>
<thead>
<tr>
<th>Mutation</th>
<th>BalhA1 Kinetics*</th>
<th>ATP Kinetics**</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k_{cat}$ min$^{-1}$</td>
<td>$K_M$ µM</td>
</tr>
<tr>
<td>WT</td>
<td>12.9 ± 0.4</td>
<td>16 ± 2</td>
</tr>
<tr>
<td>G58A</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>L59K</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>G60A</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Y81A</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>S175A</td>
<td>0.58 ± 0.03</td>
<td>18 ± 3</td>
</tr>
<tr>
<td>K271A</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D275A</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>K281R</td>
<td>6.3 ± 0.3</td>
<td>15 ± 2</td>
</tr>
<tr>
<td>D370A</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D375A</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>R385A</td>
<td>0.11 ± 0.01</td>
<td>17 ± 7</td>
</tr>
<tr>
<td>P389G</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>R405A</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
APPENDIX C: YCAO PROTEINS FOR CRYSTALLOGRAPHY SCREEN

C.1 YcaO proteins screened at UIUC

In an attempt to obtain an X-ray crystal structure of additional YcaO proteins, diverse proteins were cloned and expressed with an N-terminal fusion to MBP. Scab_1701, Scab_56671, and Scab_56661 were also expressed as an N-terminal His\textsubscript{6} fusion. All proteins were overexpressed at 22 °C for 16 h. A solubility screen was performed by treating each protein with TEV protease for 16 h at 25 °C and visually inspecting the sample for precipitate. Jonathan Chekan and Vinayak Agarwal (S. Nair Lab; UIUC) carried out all crystallography screens.

Table C.1 | YcaOs used for crystallography screens.

<table>
<thead>
<tr>
<th>Locus Tag</th>
<th>Organism</th>
<th>Subclass</th>
<th>Yield (mg/L)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Azl_d00180</td>
<td>Azospirillum sp. B510</td>
<td>TfuA Non-TOMM</td>
<td>60</td>
<td>Crystallization failed</td>
</tr>
<tr>
<td>Azl_022710</td>
<td>Azospirillum sp. B510</td>
<td>TfuA Non-TOMM</td>
<td>12</td>
<td>Precipitated during MBP cleavage</td>
</tr>
<tr>
<td>Azl_a09740</td>
<td>Azospirillum sp. B510</td>
<td>TfuA Non-TOMM</td>
<td>8</td>
<td>Highly truncated</td>
</tr>
<tr>
<td>BamD</td>
<td>Bacillus amyloliquefaciens</td>
<td>Plantazolicin</td>
<td>50</td>
<td>Crystallization failed</td>
</tr>
<tr>
<td>BpumD</td>
<td>Bacillus pumilus</td>
<td>Plantazolicin</td>
<td>17</td>
<td>Precipitated during MBP cleavage</td>
</tr>
<tr>
<td>ClosD</td>
<td>Clostridium botulinum</td>
<td>Cytolysin</td>
<td>20</td>
<td>Precipitated during MBP cleavage</td>
</tr>
<tr>
<td>HMPREF9018_0535</td>
<td>Prevotella amnii CRIS 21A-A</td>
<td>Uncharacterized TOMM</td>
<td>53</td>
<td>Crystallization failed</td>
</tr>
<tr>
<td>ERDMAN_1534</td>
<td>Mycobacterium tuberculosis str. Erdman</td>
<td>TfuA Non-TOMM</td>
<td>9</td>
<td>Precipitated during MBP cleavage</td>
</tr>
<tr>
<td>McbD</td>
<td>Escherichia Coli</td>
<td>Microcin B17</td>
<td>25</td>
<td>Precipitated during MBP cleavage</td>
</tr>
<tr>
<td>SagD</td>
<td>Streptococcus pyogenes</td>
<td>Cytolysin</td>
<td>15</td>
<td>Precipitated during MBP cleavage</td>
</tr>
<tr>
<td>Scab_1701</td>
<td>Streptomyces scabies 87.22</td>
<td>Uncharacterized TOMM</td>
<td>1</td>
<td>Highly truncated</td>
</tr>
<tr>
<td>Scab_56671</td>
<td>Streptomyces scabies 87.22</td>
<td>Bottromycin</td>
<td>2</td>
<td>Highly truncated</td>
</tr>
<tr>
<td>Scab_56661</td>
<td>Streptomyces scabies 87.22</td>
<td>Bottromycin</td>
<td>2</td>
<td>Highly truncated</td>
</tr>
<tr>
<td>SO_0795</td>
<td>Shewanella oneidensis MR-1</td>
<td>Non-TOMM</td>
<td>70</td>
<td>Crystallization failed</td>
</tr>
</tbody>
</table>
C.2 YcaO proteins screened by Steve Almo (Albert Einstein College of Medicine)

In collaboration with Steve Almo, an additional 38 TOMM and non-TOMM YcaO proteins were screened in an attempt to obtain additional structural information (table C.2). All proteins expressed as an N-terminal His₆ fusion. The proteins were expressed on small-scale cultures and solubility was determined via Western blot with an Anti-His₆ antibody on the whole cell lysate. All proteins with a high level of solubility (2 or 3) were taken forward to large-scale production and subjected to a crystallography screen. Unfortunately, all proteins proved refractory to crystallization.

Table C.2 | YcaOs screened by Steve Almo.

<table>
<thead>
<tr>
<th>Accession</th>
<th>Organism</th>
<th>Type</th>
<th>AA count</th>
<th>MW</th>
<th>Solubility</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAA09633.1</td>
<td><em>Streptomyces violaceoruber</em></td>
<td>CD Fusion</td>
<td>774</td>
<td>82.92</td>
<td>Cloning Failed</td>
</tr>
<tr>
<td>NP_250616.1</td>
<td><em>Pseudomonas aeruginosa</em></td>
<td>Non-TOMM</td>
<td>638</td>
<td>70.57</td>
<td>Cloning Failed</td>
</tr>
<tr>
<td>AAM03618.1</td>
<td><em>Methanosarcina acetivorans</em> C2A</td>
<td>methanogen</td>
<td>446</td>
<td>49.67</td>
<td>0</td>
</tr>
<tr>
<td>NP_633482.1</td>
<td><em>Methanosarcina mazei</em> Go1</td>
<td>methanogen</td>
<td>446</td>
<td>49.67</td>
<td>0</td>
</tr>
<tr>
<td>AAN69820.1</td>
<td><em>Pseudomonas putida</em> KT2440</td>
<td>Mcb 17</td>
<td>422</td>
<td>48.12</td>
<td>0</td>
</tr>
<tr>
<td>BAC89240.1</td>
<td><em>Gloeobacter violaceus</em> PCC 7421</td>
<td>TOMM</td>
<td>433</td>
<td>47.68</td>
<td>0</td>
</tr>
<tr>
<td>CAE50519.1</td>
<td><em>Corynebacterium diptheriae</em></td>
<td>Non-TOMM</td>
<td>348</td>
<td>38.06</td>
<td>0</td>
</tr>
<tr>
<td>AAS95701.1</td>
<td><em>Desulfovibrio vulgaris</em> str. Hildenborough</td>
<td>Non-TOMM</td>
<td>447</td>
<td>48.82</td>
<td>0</td>
</tr>
<tr>
<td>YP_078247.1</td>
<td><em>Bacillus licheniformis</em> DSM 13</td>
<td>Bcer CD</td>
<td>666</td>
<td>75.53</td>
<td>0</td>
</tr>
<tr>
<td>YP_135672.1</td>
<td><em>Haloarcula marismortui</em> ATCC 43049</td>
<td>Non-TOMM</td>
<td>717</td>
<td>78.49</td>
<td>0</td>
</tr>
<tr>
<td>CAI49392.1</td>
<td><em>Natronomonas pharaonis</em> DSM 2160</td>
<td>TOMM</td>
<td>585</td>
<td>60.93</td>
<td>Cloning Failed</td>
</tr>
<tr>
<td>ABA48314.1</td>
<td><em>Burkholderia pseudomallei</em> 1710b</td>
<td>NHLP-Burk</td>
<td>866</td>
<td>93.12</td>
<td>Cloning Failed</td>
</tr>
<tr>
<td>YP_388840.1</td>
<td><em>Desulfovibrio alaskensis</em> G20</td>
<td>Non-TOMM</td>
<td>604</td>
<td>66.3</td>
<td>1</td>
</tr>
<tr>
<td>ABC34570.1</td>
<td><em>Burkholderia thailandensis</em> E264</td>
<td>TfuA-YcaO</td>
<td>905</td>
<td>99.04</td>
<td>0</td>
</tr>
<tr>
<td>YP_441074.1</td>
<td><em>Burkholderia thailandensis</em> E264</td>
<td>NHLP Burk</td>
<td>778</td>
<td>84.02</td>
<td>0</td>
</tr>
<tr>
<td>NP_415425.4</td>
<td><em>Escherichia coli</em> str. K-12 substr. MG1655</td>
<td>Non-TOMM</td>
<td>608</td>
<td>68.17</td>
<td>3</td>
</tr>
<tr>
<td>YP_648520.1</td>
<td><em>Yersinia pestis</em> Nepal516</td>
<td>Non-TOMM</td>
<td>610</td>
<td>68.23</td>
<td>3</td>
</tr>
<tr>
<td>ABI90541.1</td>
<td><em>Burkholderia ambifaria</em> AMMD</td>
<td>NHLP Burk</td>
<td>767</td>
<td>83.45</td>
<td>0</td>
</tr>
<tr>
<td>ABJ57813.1</td>
<td><em>Lactobacillus delbrueckii</em></td>
<td>TOMM</td>
<td>355</td>
<td>40.15</td>
<td>0</td>
</tr>
<tr>
<td>YP_949181.1</td>
<td><em>Arthrobacter aurescens</em> TC1</td>
<td>TOMM</td>
<td>435</td>
<td>46.12</td>
<td>3</td>
</tr>
<tr>
<td>ABN06765.1</td>
<td><em>Methanocorpusculum labreanum</em> Z</td>
<td>TfuA Non-TOMM</td>
<td>425</td>
<td>47.16</td>
<td>3</td>
</tr>
<tr>
<td>ABN58160.1</td>
<td><em>Methanoculles marisnigri</em> JR1</td>
<td>Non-TOMM</td>
<td>419</td>
<td>46.52</td>
<td>2</td>
</tr>
<tr>
<td>YP_001045574.1</td>
<td><em>Rhodobacter sphaeroides</em> ATCC 17029</td>
<td>TOMM</td>
<td>460</td>
<td>49.66</td>
<td>Cloning Failed</td>
</tr>
<tr>
<td>ABO24899.1</td>
<td><em>Shewanella loihica</em> PV-4</td>
<td>Non-TOMM</td>
<td>750</td>
<td>84.27</td>
<td>3</td>
</tr>
<tr>
<td>Accession</td>
<td>Organism</td>
<td>Type</td>
<td>AA count</td>
<td>MW</td>
<td>Solubility</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------------------------</td>
<td>--------------</td>
<td>----------</td>
<td>--------</td>
<td>--------------------</td>
</tr>
<tr>
<td>ABP54882.1</td>
<td><em>Salinispora tropica</em> CNB-440</td>
<td>TOMM</td>
<td>652</td>
<td>71.19</td>
<td>0</td>
</tr>
<tr>
<td>ABO73127.1</td>
<td><em>Mycobacterium tuberculosis</em> H37Ra</td>
<td>TfuA Non-TOMM</td>
<td>461</td>
<td>49.74</td>
<td>0</td>
</tr>
<tr>
<td>YP_001363155.1</td>
<td><em>Kineococcus radiotolerans</em> SRS30216</td>
<td>TOMM</td>
<td>546</td>
<td>59.18</td>
<td>Cloning Failed</td>
</tr>
<tr>
<td>ABX43929.1</td>
<td><em>Clostridium phytofermentans</em> ISDg</td>
<td>TOMM</td>
<td>640</td>
<td>74.44</td>
<td>1</td>
</tr>
<tr>
<td>ACB53447.1</td>
<td><em>Cyanothecae sp.</em> ATCC 51492</td>
<td>Cyanobactin</td>
<td>786</td>
<td>90.5</td>
<td>0</td>
</tr>
<tr>
<td>ACK78345.1</td>
<td><em>Acidithiobacillus ferrooxidans</em> ATCC 23270</td>
<td>TOMM</td>
<td>456</td>
<td>51.41</td>
<td>0</td>
</tr>
<tr>
<td>ACL64263.1</td>
<td><em>Anaeromyxobacter dehalogenans</em> 2CP-1</td>
<td>TOMM</td>
<td>401</td>
<td>40.77</td>
<td>0</td>
</tr>
<tr>
<td>ACS40629.1</td>
<td><em>Methylobacterium extorquens</em> AM1</td>
<td>TfuA Non-TOMM</td>
<td>561</td>
<td>61.06</td>
<td>1</td>
</tr>
<tr>
<td>ACV48332.1</td>
<td><em>Halomonas mukohataei</em> DSM 12286</td>
<td>TOMM</td>
<td>599</td>
<td>62.81</td>
<td>1</td>
</tr>
<tr>
<td>ACZ11500.1</td>
<td><em>Sulfurospirillum deleyianum</em> DSM 6946</td>
<td>Non-TOMM</td>
<td>562</td>
<td>64.33</td>
<td>3</td>
</tr>
<tr>
<td>YP_003300351.1</td>
<td><em>Thermomonospora curvata</em> DSM 43183</td>
<td>TOMM</td>
<td>779</td>
<td>86.06</td>
<td>Cloning Failed</td>
</tr>
<tr>
<td>ACZ89043.1</td>
<td><em>Streptosporangium roseum</em> DSM 43021</td>
<td>Non-TOMM</td>
<td>443</td>
<td>48.35</td>
<td>Cloning Failed</td>
</tr>
<tr>
<td>AD35913.1</td>
<td><em>Methanococcus voltae</em> A3</td>
<td>TfuA Non-TOMM</td>
<td>465</td>
<td>52.45</td>
<td>3</td>
</tr>
<tr>
<td>EGE61713.1</td>
<td><em>Rhizobium etli</em> CNPAF512</td>
<td>TfuA Non-TOMM</td>
<td>443</td>
<td>47.7</td>
<td>Cloning Failed</td>
</tr>
</tbody>
</table>