UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

tRNA-Deacylase Directed Discovery of Biosynthetic Pathways

Permalink

https://escholarship.org/uc/item/5qs1r3pg

Author

Millar, Douglas

Publication Date

2024

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

tRNA-Deacylase Directed Discovery of Biosynthetic Pathways

By

Douglas C. Millar

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Chemical Engineering
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:
Professor Michelle C. Y. Chang, Chair
Professor Wenjun Zhang
Professor Jay D. Keasling
Professor Alanna Schepartz

Spring 2024

tRNA-Deacylase Directed Discovery of Biosynthetic Pathways

© 2024

by Douglas C. Millar

Abstract

tRNA-Deacylase Directed Discovery of Biosynthetic Pathways

by

Douglas C. Millar

Doctor of Philosophy in Chemical Engineering

University of California, Berkeley

Professor Michelle C. Y. Chang, Chair

Amino acids are one of nature's privileged scaffolds for chemical and biomolecular diversification. Beyond the standard twenty amino acids used in protein synthesis, hundreds more noncanonical amino acids (ncAAs) have been found in natural products with useful chemical features. Here we present a resistance-guided approach to discovering ncAA natural products with an emerging class of enzymes: tRNA-deacylases (tRDs). Beginning with the discovery and characterization of a tRD with specificity towards a ncAA, we performed family-wide analyses of tRDs in biosynthetic gene clusters (BGCs) and established tRDs as a widely present genomic marker of ncAA metabolism. These efforts enabled us to discover and elucidate biosynthetic pathways for several new ncAAs and natural products. Through strategies like protein engineering and cascade design, we employed several of these enzymes as biocatalysts to achieve transformations that have been challenging to access otherwise. We anticipate the advances in genome mining and experimental characterization developed in this work will have broad applications for discovering new natural products with ncAAs and useful biocatalysts.

Table of Contents

Table of Contents i
List of Figures, Schemes, and Tables ii
List of Abbreviations vii
Acknowledgements viii

Chapter 1: Introduction
1.1 Introduction 1
1.2 References 6

Chapter 2: tRNA-Deacylase Directed Discovery of Biosynthetic Pathways
2.1 Introduction 11
2.2 Materials and Methods 13
2.3 Results and Discussion 29
2.4 Conclusions 73
2.5 References 74

Chapter 3: Discovery of a Biocatalytic Cascade for Preparation of α,β-Substituted Amino/Hydroxy Acids via C-H Functionalization
3.1 Introduction 80
3.2 Materials and Methods 82
3.3 Results and Discussion 92
3.4 Conclusions 115
3.5 References 116

Chapter 4: Efforts towards Elucidating Unknown Functions of Biosynthetic Genes using tRDs and Engineering Site-Selective Incorporation
4.1 Introduction 122
4.2 Materials and Methods 123
4.3 Results and Discussion 132
4.4 Conclusions 176
4.5 References 177

Appendices
Appendix I: Plasmids, oligonucleotides, and gene sequences 182
Appendix II: Supplementary Tables 203
Appendix III: Supplementary Figures 234

List of Figures, Schemes, and Tables

Chapter 1

Figure 1.1	*Benefits of a chemoenzymatic approach*	2
Figure 1.2	*Noncanonical amino acids (ncAAs) in primary and secondary metabolism*	4

Chapter 2

TOC Figure	*tRNA-Deacylase Directed Discovery of Biosynthetic Pathways*	10
Figure 2.1	*tRNA-Deacylases (tRDs) are resistance proteins for noncanonical amino acid (ncAA) biosynthesis.*	12
Figure 2.2	*βes is a mimic of threonine.*	29
Figure 2.3	*The biosynthetic pathway for β-ethynylserine (βes)*	30
Figure 2.4	*Identification of a putative ncAA tRNA deacylase BesG*	30
Figure 2.5	*Chemical synthesis of β-ethynylserine (βes)*	31
Figure 2.6	*BesG shows specificity for βes-tRNA^{Thr}*	32
Figure 2.7	*Comparison of tRNA^{Thr}s from S. cattleya and E. coli provides insight into basis for BesG's tRNA preference.*	32
Table 2.1	*Proteomic analysis of BesG knockout strain shows bes incorporation at Thr sites. Proteomic analysis of BesG knockout strain shows βes incorporation at Thr sites.*	33
Figure 2.8	*MS/MS spectra showing bes incorporation into peptides at Thr sites*	35
Figure 2.9	*Sequence and structural comparison of BesG to canonical AlaX enzymes.*	36
Figure 2.10	*Substrate docking of βes in BesG.*	37
Figure 2.11	*tR3D analysis overview and SSN generation.*	39
Figure 2.12	*Annotated sequence similarity network (SSN) of the AlaX family.*	40
Figure 2.13	*Data cleaning for SSN of AlaX Family*	41
Figure 2.14	*Filtering of AlaX SSN to remove primary metabolism AlaXs*	42
Figure 2.15	*Amino acid biosynthetic genes co-occurring with AlaXs*	43
Table 2.2	*Primary metabolism AlaX entries in SSN*	43
Figure 2.16	*Representative gene neighborhoods for AlaXs in primary metabolism and class-defined secondary metabolism*	44
Figure 2.17	*Co-occurrence of class-defined secondary metabolism in AlaX SSN*	45
Figure 2.18	*tR3D analysis of AlaX tRNA-deacylases*	46
Figure 2.19	*Secondary metabolism genes that co-occur with AlaX tRNA-deacylases*	48
Figure 2.20	*Putative Biosynthetic Gene Clusters (BGCs) with AlaX tRNA-Deacylases*	49
Figure 2.21	*Discovery of a methionine hydroxylase AlfC*	51
Table 2.3	*Summary of NMR characterization of AlfC product*	51
Figure 2.22	*NMR characterization of (2S,3R)-hydroxymethionine*	52
Figure 2.23	*AlfB is a hydroxymethionine-tRNA^{Met} deacylase*	54
Figure 2.24	*A novel BGC containing the requisite genes for hydrazine formation*	55
Figure 2.25	*Discovery of a unique arginine hydroxylase*	56
Figure 2.26	*2D NMR analysis of (2S,3R)-hydroxyarginine*	57

Figure 2.27 *1D NMR analysis of (2S,3R)-hydroxyarginine* 58
Figure 2.28 *Protein structure comparison of FuzA and VioC* 59
Figure 2.29 *In vitro characterization of FuzC and FuzD provide insights into the natural product encoded by the Fuz BGC* 60
Figure 2.30 *In vivo detection of hydrazine-containing metabolites from the Fuscazine BGC* 61
Figure 2.31 *Comparative metabolomics reveals products of the Fuz BGC* 62
Figure 2.32 *Isotopic-labeling study to identify the components of the fuscazines* 63
Figure 2.33 *Proposed route for pipecolic acid biosynthesis via combination of Fuz cluster and native E. coli genes* 64
Figure 2.34 *Full biosynthetic pathway for Fuscazine A* 64
Figure 2.35 *Structural analysis of Fuscazines A and B* 65
Figure 2.36 *NMR analysis of Fuscazine A* 66
Figure 2.37 *In vitro reactivities of the ATPG-grasp enzymes from the Fuz cluster* 67
Figure 2.38 *Sequence similarity network of hydrazine synthetases* 68
Figure 2.39 *Docking analysis of Hydrazine Synthetase FuzG* 70
Figure 2.40 *tR3D Analysis of Hydrazine Synthetases* 71

Chapter 3

TOC Figure *Discovery of a Biocatalytic Cascade for Preparation of α,β-Substituted Amino/Hydroxy Acids via C-H Functionalization* 79
Figure 3.1 *Applications of α,β-functionalized amino acids.* 80
Figure 3.2 *Sequence similarity network analysis of AlaX tRD co-occurrence with TauD-like proteins* 92
Figure 3.3 *Phylogenetic analysis of TauD-like proteins* 93
Figure 3.4 *Bioinformatic discovery of C-H functionalizing biosynthetic pathway.* 94
Figure 3.5 *Genome mining to identify Pfrm biosynthetic gene cluster in Pseudomonas fluorescens R124* 95
Figure 3.6 *In vitro screen of Pfrm BGC enzymes with proteinogenic amino acids* 96
Figure 3.7 *Metabolomics reveals products of the Pfrm BGC* 97
Figure 3.8 *Experimental elucidation of the Pfrm biosynthetic pathway* 98
Figure 3.9 *In vitro reconstitution of β-phenylisoserine biosynthesis* 99
Figure 3.10 *PfrmB converts (S)-β-phenylalanine to (2R,3R)-β-phenylisoserine* 100
Figure 3.11 *NMR characterization of PfrmB product (2R,3R)- β-phenylisoserine* 101
Figure 3.12 *In vitro reconstitution of PfrmG's activity to produce (3R)-α-oxo- β-phenylalanine* 102
Figure 3.13 *In vitro reconstitution of α,β-diaminophenylalanine biosynthesis* 103
Figure 3.14 *PfrmD shows L-amino acid transaminase activity with (3R)- α -oxo-β-phenylalanine* 103
Figure 3.15 *PfrmD shows L-amino acid transaminase activity with phenylpyruvic acid* 104
Figure 3.16 *PfrmH ligates α,β-diaminophenylalanine to L-α-valine* 104
Figure 3.17 *Determining the connectivity of the diaminophenylalanine-valine dipeptide* 105
Figure 3.18 *Biocatalytic cascade for aromatic-substituted phenylalanine substrates* 106
Figure 3.19 *PfrmG forms D-phenylglycine from (2R,3R)-β-phenylisoserine* 107

Figure 3.20 *Established approaches for the synthesis of D-phenylglycine* 108
Figure 3.21 *Expanding PfrmB's reaction scope to access useful alternate diastereomers of β-phenylisoserine* 109
Figure 3.22 *Kinetics of WT PfrmB for (S)-b-phenylalanine* 110
Figure 3.23 *Docking study of PfrmB with (R)- and (S)-β-phenylalanine* 110
Figure 3.24 *Rational engineering of PfrmB to accept (R)-β-phenylalanine* 111
Figure 3.25 *Compiled results from PfrmB rational engineering* 112
Figure 3.26 *C-H functionalization of hydrocinnamic acid with PfrmB* 113
Figure 3.27 *Biocatalytic cascade to α-hydroxy and α,β-amino acids from diverse carboxylic acids* 114
Table 3.1 *Summary of PfrmB promiscuity towards carboxylic acid substrates* 114

Chapter 4

Figure 4.1 *Sequence Similarity Network (SSN) of AlaX tRDs in secondary metabolism with Fe/αKG oxygenases* 132
Figure 4.2 *Biosynthetic gene clusters (BGCs) containing both Fe/αKG-dependent oxygenases and AlaX tRNA-Deacylases* 133
Table 4.1 *Representative Accessions for AlaX tRNA-Deacylases with neighboring Fe/aKG-dependent oxygenases* 134
Figure 4.3 *Identification of an oxygenase-containing biosynthetic gene cluster in Photobacterium halotolerans* 135
Figure 4.4 *Identification of PhalA hydroxylation of leucine (Leu) and isoleucine (Ile)* 136
Figure 4.5 *Possible aziridination of Ile by PhalA* 137
Figure 4.6 *Small dipeptides produced by PhalF* 137
Figure 4.7 *Relative abundance of PhalF dipeptide products* 138
Figure 4.8 *Continuous assay of PhalF shows preference for valine* 138
Figure 4.9 *Experimental design for comparative metabolomics of the Phal BGC* 140
Table 4.2 *Summary of results from metabolomics experiment* 140
Figure 4.10 *PhalF and PhalD form hydroxylated dipeptides with proline in vivo* 141
Figure 4.11 *PhalF and PhalD form hydroxylated dipeptides with valine and valine/isoleucine in vivo* 142
Figure 4.12 *PhalA hydroxylates leucine and isoleucine in vivo* 143
Figure 4.13 *In vitro reconstitution of Val-Val and Val-Ile hydroxylated peptides* 144
Figure 4.14 *PhalD shows preference for valine in vitro* 146
Figure 4.15 *PhalD can produce both hydroxylated and nitro valine* 147
Figure 4.16 *Sequence-Similarity Network of Flavin-Dependent N-Monooxygenases* 148
Figure 4.17 *Valine is a building block for diazeniumdiolate-containing natural products* 149
Figure 4.18 *Biocatalytic approaches toward diazeniumdiolate-containing compounds* 150
Figure 4.19 *Diazeniumdiolate formation with PhalD and diethylamine NONOate* 151
Figure 4.20 *N-monooxygenation of valinol by PhalD* 152
Figure 4.21 *Structural models of putative nitrite reductases* 153
Figure 4.22 *Discovery of N-monooxygenases for L-alanosine synthesis* 154
Figure 4.23 *Sequence similarity network (SSN) of YbaK/INS family* 155

Table 4.3 *Characterized members of the YbaK/INS family* 155
Table 4.4 *Statistics for YbaK/INS family SSN generation* 156
Table 4.5 *Filtering SSN of short/poor genomic reads* 156
Table 4.6 *Filtering SSN of fragments* 156
Table 4.7 *Common primary metabolism genes that co-occur with YbaK/INS tRDs* 157
Table 4.8 *Remaining YbaK/INS entries after filtering steps* 157
Figure 4.24 *Overall analysis of secondary metabolism in filtered YbaK/INS SSN* 158
Figure 4.25 *Filtered SSN enriched for secondary metabolism* 159
Figure 4.26 *Overview of RiPP biosynthesis* 160
Figure 4.27 *Identification of a RiPP-like cluster in Pantoea sp. GM01* 160
Figure 4.28 *Highly conserved RiPP-like BGC* 161
Figure 4.29 *Structural model of PantB* 162
Figure 4.30 *In vitro screening of PantC and PantE with proteinogenic amino acids* 163
Figure 4.31 *Combining PantC and PantE yields unique products* 164
Figure 4.32 *PantE hydroxylates a Glu-Asn dipeptide* 165
Table 4.9 *Summary of results from metabolomics experiment of Pant BGC* 165
Figure 4.33 *Most abundant new metabolite from Pant BGC metabolomics* 166
Figure 4.34 *Additional hit from metabolomics of Pant BGC* 167
Figure 4.35 *Proposed biosynthetic pathway of the Pant BGC* 168
Figure 4.36 *BGCs with YbaK/INS tRDs* 169
Figure 4.37 *Genome mining to identify a DAP C_β hydroxylase* 169
Figure 4.38 *Overview of CuAAC reaction with terminal alkyne of Pra* 170
Figure 4.39 *Chimeric design of orthogonal aaRSs* 170
Figure 4.40 *Binding pocket for Ala-adenylate in EcAlaRS* 171
Figure 4.41 *In vitro promiscuity of chAlaRS for “clickable” ncAAs* 171
Figure 4.42 *Overview of protein display method for FACS-based engineering* 172
Figure 4.43 *Results of FACS with WT chAlaRS* 172
Figure 4.44 *Engineering approach for chAlaRS* 174
Figure 4.45 *Results from engineering of chAlaRS* 174
Figure 4.46 *chAlaRS variants sorted with FACS* 175

Appendix I

Table A1.1 *Strains, plasmids, oligonucleotides, and gene sequences for Chapter 2* 183
Table A1.2 *Strains, plasmids, oligonucleotides, and gene sequences for Chapter 3* 191
Table A1.3 *Strains, plasmids, oligonucleotides, and gene sequences for Chapter 4* 195

Appendix II

Table A2.1 *Minimal media amino acid supplement for comparative metabolomics experiments* 204
Table A2.2 *Proteomic analysis of BesG knockout strain.* 205

Table A2.3 *Bioinformatic analysis of tRD co-occurrence with TauD-like oxygenases* 210
Table A2.4 *Proteins discovered through genome mining from Chapter 4* 211
Table A2.5 *Enriched metabolites from fuscazine BGC metabolomics* 212
Table A2.6 *Enriched metabolites from Pfrm BGC metabolomics* 216
Table A2.7 *Enriched metabolites from Phal BGC metabolomics* 219
Table A2.8 *Enriched metabolites from Pant BGC metabolomics* 224

Appendix III

Figure A3.1 *MS/MS spectra of bes+2 m/z incorporation into peptides* 213
Figure A3.2 *Putative biosynthetic gene clusters with AlaX tRNA-Deacylases.* 214
Figure A3.3 *Purification results of BesG, ThrRSs, alf cluster, Af MetRS and fuz cluster* 216
Figure A3.4 *Genome mining to identify Pfrm biosynthetic gene cluster in Pseudomonas fluorescens R124* 217
Figure A3.5 *Purification of genes from Pfrm BGC* 218
Figure A3.6 *Purification of genes from Phal and Pant BGCs* 219
Figure A3.7 *Engineering of BesB* 220

List of Abbreviations

tRNA	transfer RNA
aaRS	aminoacyl-tRNA synthetase
BGC	biosynthetic gene cluster
tRD	tRNA-deacylase
ncAA	noncanonical amino acid
Pra	propargylglycine
βes	β-ethynylserine
HydS	hydrazine synthetase
SSN	sequence similarity network
GND	genomic neighborhood diagrams
EFI	Enzyme Function Instititute
NRPS	nonribosomal peptide synthetase
PKS	polyketide synthase
RiPP	ribosomally-synthesized and post-translationally modified peptide
chAlaRS	chimeric alanine-tRNA synthetase
αKG	α-ketoglutarate
ATP	adenosine-5′-phosphate
CuAAC	copper-catalyzed azide-alkyne cycloaddition
d	day
h	hour
s	seconds
DTT	DL-dithiothreitol
Da	dalton
HPLC	high pressure liquid chromatography
LC/MS	liquid chromatography mass spectrometry
LC/MS-MS	liquid chromatography tandem mass spectrometry
MWCO	molecular weight cutoff
QQQ	triple quadrupole mass spectrometer
QTOF	quadrupole-time of flight mass spectrometer

Acknowledgements

"For such selves in such a world, institutions…are places where one goes to perform, not to be formed" – Carl R. Trueman, *The Rise and Triumph of the Modern Self*

To be formed. Reading this quote early in the course of my PhD, I was drawn towards this idea of how schooling ought to be focused primarily on our formation, not solely on evaluation as it often is. I am grateful for how my PhD has formed me in so many ways.

I want to begin by thanking Michelle. A patient and trustworthy mentor. Creative, flexible, and persistent. I am thankful for her patience in guiding me through the PhD as I developed the various essential attributes of what I believe makes a good scientist – a keen eye for the details, unwavering perseverance, effective communication, and not caring too much what others think of you. Being able to change my project halfway through my PhD, apply for and acquire grant funding, and go where the science took me were all challenges but also privileges thanks to Michelle's consistent support. All the food we got along the way definitely helped too.

I am very thankful for a fantastic group of colleagues, who are incredibly talented yet do not take themselves too seriously. A thank you to my fellow six years, Eli Kissman and Max Sosa, for teaching me so much and modeling for me how to communicate science effectively. I would not be where I am without Jorge Marchand and Monica Neugebauer, teaching me when I joined the lab many of the essential skills I rely on to this day. Conversations with Dr. Kersh Theva and Eli during the pandemic helped me through those challenging times. I am thankful to Elizabeth Stone for sharing with me her fascination of stereochemistry. A thank you to Jason Fang for his help with synthetic chemistry and Andrew Whitten for his contagious work ethic. I am thankful to everyone for how they have contributed to an incredible lab culture: Dr. Edward Koleski, Veronica Stafford, Gabby Dolgonos, Dr. Yanni Kipouros, Dr. Edward Badding, Helena Roberts-Mataric, Nicole Wong, and the many amazing undergrads. A special shoutout to team tR3D: Claire Li, Tavis Chen, and Calvin Hu. Everyone's presence made coming to work way more enjoyable.

Finally, I want to thank my computational collaborator…and wife, Shannon. Her support through the challenges and successes of the PhD has kept me fueled. Her future-focus helps me dream big and strive for my best. Her care for others helps me see outside of myself to attend to the needs of others. I also want to thank my family (on both sides) for supporting me throughout my entire journey to achieve this milestone.

Chapter 1: Introduction

1.1 Introduction

Natural products have long been valued for their medicinal properties. As important molecules for defense, signaling, and nutrient acquisition, organisms such as fungi, plants, and bacteria rely on these compounds to stay alive and outcompete rivals [*1*]. As early as 2900 BCE, humans began repurposing these products to treat illness and pain, with records showing hundreds of plant extracts being formulated into pills and ointments [*2*]. Our understanding of the value of willow tree bark far preceded our understanding of the structure of salicin, the natural product from which aspirin is derived [*3*]. Only within the last two hundred years has the technology developed to where the chemical structures of these compounds could be elucidated, and with that, an understanding of the relationships between molecular structure and function [*4*].

In contrast, many of the pharmaceuticals prescribed today do not resemble the natural products from our early history. Of the 1,394 small molecule FDA-approved drugs developed since 1980, 32% are natural products or natural-product based [*5*]. While the reasons for this are varied and complex, a preference for a subset of organic reactions and a common panel of building blocks has led to an overrepresentation of certain molecular shapes [*6–8*]. The very factors that make natural product synthesis often more challenging than synthetic drugs (1.7-fold higher in molecular weight, 1.5-fold greater number of sp^3-hybridized carbon centers, 2-6-fold more stereocenters) are also essential for their efficacy [*9–12*]. As such, there is an emerging focus within medicinal chemistry to incorporate more natural product-like structural characteristics into the design

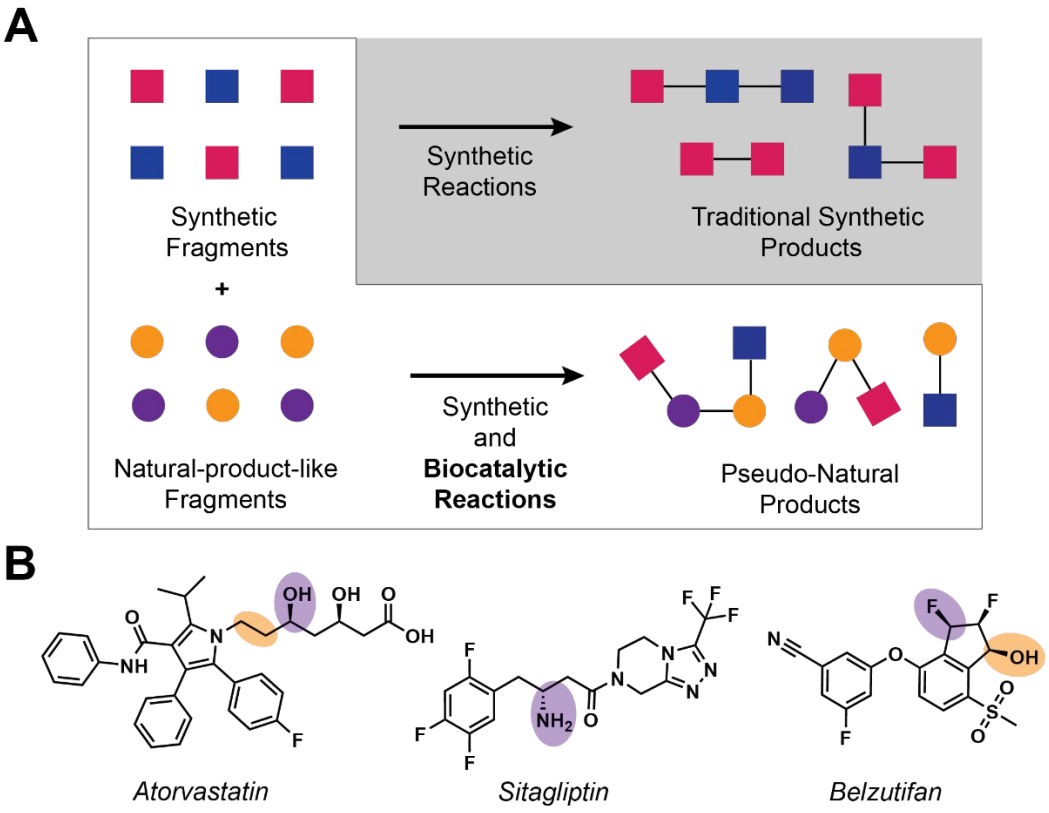


Figure 1.1. Benefits of a chemoenzymatic approach. (A) Utilizing biocatalysis in combination with synthetic chemistry enables access to new molecular features that can improve properties. (B) Outstanding examples of pharmaceuticals with components synthesized through biocatalysis (shaded).

of pharmaceuticals as an approach to increase the discovery rate [*13–15*]. One avenue that is achieving remarkable success is utilizing enzymes as biocatalysts in chemoenzymatic syntheses of moieties that are challenging to access synthetically (**Fig. 1.1A**).

To complement synthetic methods, enzymes offer new approaches to make complex structures using a relatively small number of steps from common functional groups, such as alcohols, aldehydes, ketones, amines, imines and alkenes [*16*]. In comparison, these structures can be difficult to prepare by total synthesis due to the high degrees of freedom involved in orienting reacting groups or the strain introduced upon ring formation. As such, biocatalysis offers potential for diversity-oriented synthesis (DOS) and fragment-based drug design (FBD) that can overcome some of the limitations of purely synthetic libraries [*13,15,17*]. Some outstanding examples include the use of a ketoreductase to install a stereocenter within atorvastatin, followed by C-C bond formation with a halohydrin dehalogenase (**Fig. 1.1B**) [*18*]. Perhaps one of the most famous examples of biocatalysis is in the synthesis of sitagliptin, where a transaminase installed a stereocenter at high yield and simplified the overall route by four steps [*19*]. Most recently, an oxygenase was employed to install a hydroxyl group for deoxyfluorination, followed by a ketoreductase-catalyzed step to install an additional stereocenter in the synthesis of belzutifan [*20*].

Enzymes, when leveraged effectively in synthetic schemes, can improve molecular complexity and route efficiency. A challenge in enabling this type of technology is identifying enzymes for the desired transformations. Fortunately, with advances in DNA sequencing technology, the number of available protein sequences is currently around 250 M and is increasing exponentially [*21*]. To expand the current toolbox of available biocatalysts, one needs to be able to predict, elucidate, and utilize the functions of new enzymes amidst the vast amounts of genomic data available.

To address these challenges, genome mining aims to identify whole groups of biosynthetic genes and connect their functions to natural product structure. With advances in computational and experimental approaches, genome mining is becoming an increasingly sophisticated field [*22*]. The genes that encode biosynthesis of a natural product are often grouped together in a biosynthetic gene cluster (BGC) in microbes, which aids in discovery [*23*]. BGCs can be identified with the help of class-defining enzymes that are essential components for specific types of natural products. Most often, these class-defining enzymes are used to identify polyketides, nonribosomal peptides, and ribosomally-synthesized and post-translationally modified peptides [*24–27*]. Additional tools have been developed that rely on machine learning, resistance enzymes, or phylogenetic outliers, but these tools have yet to gain widespread adoption [*28–31*]. Advances in protein structural prediction have also advanced our ability to discover new enzyme function [*32,33*]. Taken together, whole pathways can be discovered with many enzymes given preliminary function entirely *in silico*. There remains much to be discovered however, fueling interest in alternative approaches to discover new natural products and biocatalysts [*34*].

A unique group of natural products that can be difficult to discover with current genome mining approaches is noncanonical amino acids (ncAAs). Beyond the typical carboxylate and amine, the natural product literature reveals ncAAs contain a wide range of interesting and unique chemical modifications [*35*]. These include rare hetereoatom-carbon bonds, such as F, As, and Se, as well as useful functional groups including diazo, nitrile, phosphonate, and nitro moieties (**Fig. 1.2A**) [*36–42*]. Identifying the enzymes that install these types of transformations is beneficial towards enabling these challenging transformations in synthetic schemes.

Despite their similarities to proteinogenic amino acids, ncAAs are typically prevented from being translated into proteins through the impressive selectivity of aminoacyl-tRNA synthetases (aaRSs) [*43,44*]. As the enzymes responsible for the important process of ligating amino acids to tRNA, aaRSs display higher selectivity than other enzymes that recognize similar substrates [*45,46*]. One way in which this selectivity is achieved is through additional proofreading domains, which are found as additional parts of the proteins or act *in trans* to edit misacylated tRNA. The *trans*-editing proteins, termed tRNA-deacylases (tRDs), ensure accurate translation of alanine, proline, tyrosine, and phenylalanine [*47,48*]. Despite these extra means to prevent mistranslation, several ncAAs have been found to surpass these editing mechanisms and incorporate into proteins in place of near cognate amino acid [*49*].

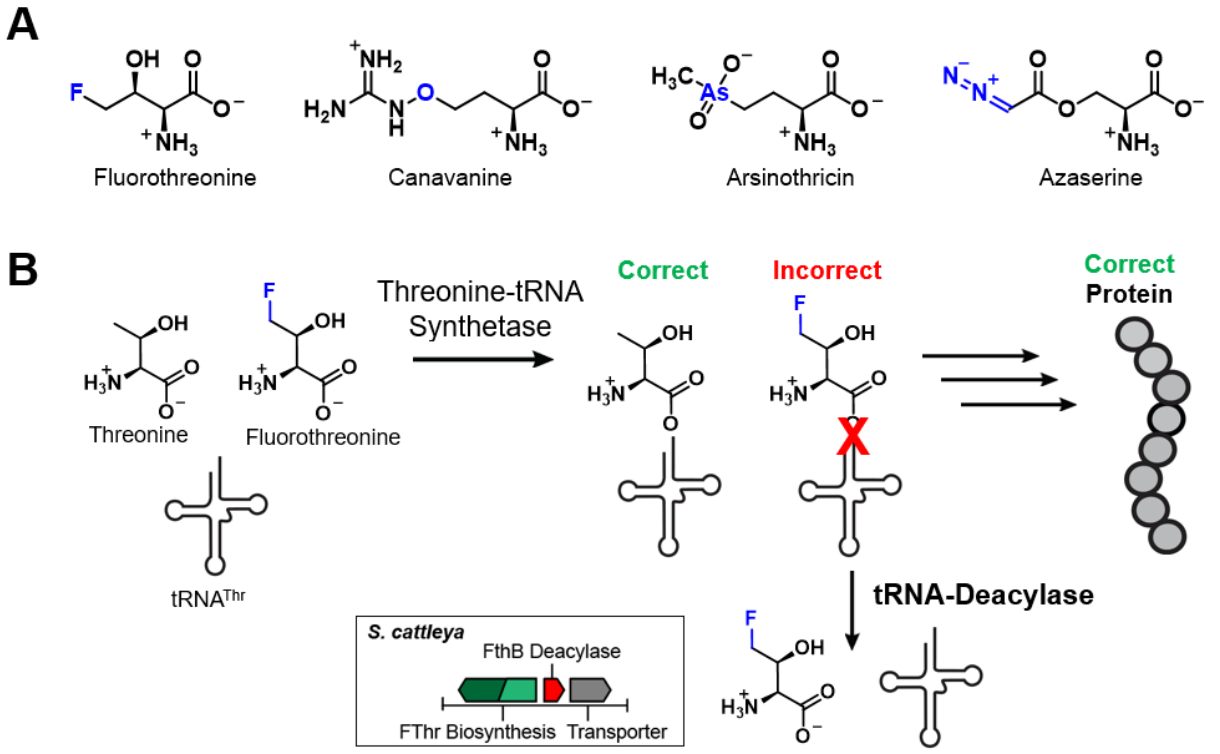


Figure 1.2. Noncanonical amino acids (ncAAs) in primary and secondary metabolism. (A) NcAAs contain an array of unique and useful functional groups, including fluorine, esters, rare metals, and diazo groups, as demonstrative examples. There are on the order of 500 ncAAs known to exist [*50*]. (B) Some ncAAs, such as FThr, are known to misincorporate for a similar proteinogenic amino acid. To prevent mistranslation, the native producer *S. cattleya* encodes a tRNA-deacylase (tRD) that is selective for FThr-tRNA^{Thr} and is genomically co-located with a FThr biosynthetic gene. In this work, we explore the extent to which tRDs are genomic markers of ncAA biosynthesis and utilize them for discovering novel ncAAs.

With the discovery of several ncAAs capable of entering the proteome, the question arises as to how translation is prevented in the producing host organism. Recently, several enzymes with homology to tRDs have been found in ncAA-producing organisms co-clustered with the metabolic genes for the ncAAs. For example, the ncAA 4-fluorothreonine (FThr) is a natural product of *Streptomyces cattleya* and can misincorporate for threonine in proteins, with antimicrobial effects. To prevent self-incorporation, *S. cattleya* encodes a tRD with selectivity for FThr that prevents this incorporation (**Fig. 1.2B**). Notably, this tRD is adjacent to a FThr biosynthetic gene, likely for co-

regulation [*51*]. A similar trend can be found for a tRD that prevents mistranslation of canavanine, an arginine mimic, which is genomically co-located with canavanine metabolic genes [*52*].

With these notable examples, we explored if tRDs could be used as genomic markers of ncAA biosynthesis. To do this, we developed a bioinformatic approach that explores whole enzyme families of tRDs for members involved in secondary metabolism. This approach was enabled with a data visualization tool we built to efficiently explore neighboring genes for enzymes likely involved in amino acid modifications, leading to the identification of new BGCs. Through experimental characterization, we discovered several new ncAAs and natural products that incorporate them. With this genome mining approach, we established tRDs as a novel class-defining enzyme for ncAA biosynthesis.

Building on our discoveries, we established several new biocatalysts capable of unique and challenging chemical reactions. This included an *in vitro* enzyme cascade we developed to install functional groups at the α and β positions of acid-containing substrates with stereo and regio-control using C-H functionalization. Through enzyme engineering and cascade design, we could efficiently synthesize several medicinally-relevant building blocks including phenylglycine and β-phenylisoserine. In addition, we discovered a new enzyme specificity within an underexplored class of N-monooxygenases. From this, we began efforts to synthesize a class of compounds, diazeniumdiolates, that have been challenging to access biocatalytically. Finally, we report on how tRDs can aid the discovery of entirely new enzyme functions, such as a possible new type of biosynthetic logic using a RiPP-like enzyme.

Altogether, with the goal of establishing biocatalytic approaches to desirable transformations and expanding the portfolio of fragments accessible to drug design, we developed a discovery approach for ncAA biosynthetic pathways and leveraged our discoveries to assemble valuable molecules. We anticipate these findings will have implications across the fields of natural product discovery, biocatalysis, and biochemistry, ultimately for the improvement of human health.

1.2 References

1. Croteau, R.; Kutchan, T. M.; Lewis, N. G. Natural Products (Secondary Metabolites). *Biochem. Mol. Biol. Plants* **2000**, *24*, 1250–1319.
2. Newman, D. J.; Cragg, G. M.; Snader, K. M. The Influence of Natural Products upon Drug Discovery. *Nat. Prod. Rep.* **2000**, *17* (3), 215–234. https://doi.org/10.1039/A902202C.
3. Mahdi, J. G.; Mahdi, A. J.; Mahdi, A. J.; Bowen, I. D. The Historical Analysis of Aspirin Discovery, Its Relation to the Willow Tree and Antiproliferative and Anticancer Potential. *Cell Prolif.* **2006**, *39* (2), 147–155. https://doi.org/10.1111/j.1365-2184.2006.00377.x.
4. Hoffmann, R. W. *Classical Methods in Structure Elucidation of Natural Products*; John Wiley & Sons, 2017.
5. Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **2020**, *83* (3), 770–803. https://doi.org/10.1021/acs.jnatprod.9b01285.
6. Ivanenkov, Y. A.; Zagribelnyy, B. A.; Aladinskiy, V. A. Are We Opening the Door to a New Era of Medicinal Chemistry or Being Collapsed to a Chemical Singularity? *J. Med. Chem.* **2019**, *62* (22), 10026–10043. https://doi.org/10.1021/acs.jmedchem.9b00004.
7. Brown, D. G.; Boström, J. Analysis of Past and Present Synthetic Methodologies on Medicinal Chemistry: Where Have All the New Reactions Gone? *J. Med. Chem.* **2016**, *59* (10), 4443–4458. https://doi.org/10.1021/acs.jmedchem.5b01409.
8. Lovering, F.; Bikker, J.; Humblet, C. Escape from Flatland: Increasing Saturation as an Approach to Improving Clinical Success. *J. Med. Chem.* **2009**, *52* (21), 6752–6756. https://doi.org/10.1021/jm901241e.
9. Atanasov, A. G.; Zotchev, S. B.; Dirsch, V. M.; Supuran, C. T. Natural Products in Drug Discovery: Advances and Opportunities. *Nat. Rev. Drug Discov.* **2021**, *20* (3), 200–216. https://doi.org/10.1038/s41573-020-00114-z.
10. Davison, E. K.; Brimble, M. A. Natural Product Derived Privileged Scaffolds in Drug Discovery. *Curr. Opin. Chem. Biol.* **2019**, *52*, 1–8. https://doi.org/10.1016/j.cbpa.2018.12.007.
11. Stone, S.; J. Newman, D.; L. Colletti, S.; S. Tan, D. Cheminformatic Analysis of Natural Product-Based Drugs and Chemical Probes. *Nat. Prod. Rep.* **2022**, *39* (1), 20–32. https://doi.org/10.1039/D1NP00039J.
12. Chen, Y.; Garcia de Lomana, M.; Friedrich, N.-O.; Kirchmair, J. Characterization of the Chemical Space of Known and Readily Obtainable Natural Products. *J. Chem. Inf. Model.* **2018**, *58* (8), 1518–1532. https://doi.org/10.1021/acs.jcim.8b00302.
13. Over, B.; Wetzel, S.; Grütter, C.; Nakai, Y.; Renner, S.; Rauh, D.; Waldmann, H. Natural-Product-Derived Fragments for Fragment-Based Ligand Discovery. *Nat. Chem.* **2013**, *5* (1), 21–28. https://doi.org/10.1038/nchem.1506.
14. Ertl, P.; Altmann, E.; McKenna, J. M. The Most Common Functional Groups in Bioactive Molecules and How Their Popularity Has Evolved over Time. *J. Med. Chem.* **2020**, *63* (15), 8408–8418. https://doi.org/10.1021/acs.jmedchem.0c00754.
15. Grigalunas, M.; Burhop, A.; Zinken, S.; Pahl, A.; Gally, J.-M.; Wild, N.; Mantel, Y.; Sievers, S.; Foley, D. J.; Scheel, R.; Strohmann, C.; Antonchick, A. P.; Waldmann, H. Natural Product

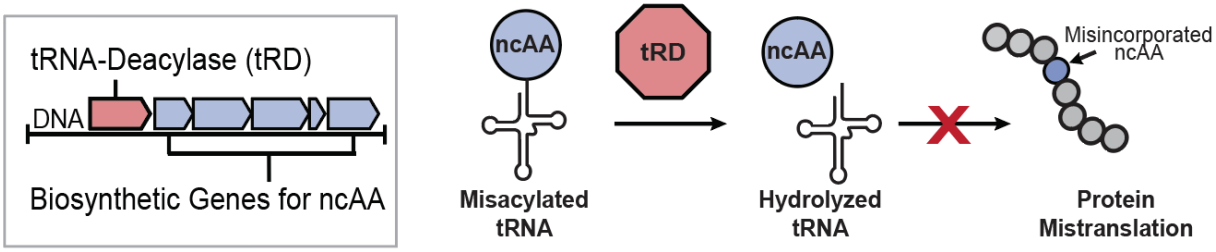
Fragment Combination to Performance-Diverse Pseudo-Natural Products. *Nat. Commun.* **2021**, *12* (1), 1883. https://doi.org/10.1038/s41467-021-22174-4.

16. Tang, M.-C.; Zou, Y.; Watanabe, K.; Walsh, C. T.; Tang, Y. Oxidative Cyclization in Natural Product Biosynthesis. *Chem. Rev.* **2017**, *117* (8), 5226–5333. https://doi.org/10.1021/acs.chemrev.6b00478.
17. Grigalunas, M.; Brakmann, S.; Waldmann, H. Chemical Evolution of Natural Product Structure. *J. Am. Chem. Soc.* **2022**, *144* (8), 3314–3329. https://doi.org/10.1021/jacs.1c11270.
18. Ma, S. K.; Gruber, J.; Davis, C.; Newman, L.; Gray, D.; Wang, A.; Grate, J.; Huisman, G. W.; Sheldon, R. A. A Green-by-Design Biocatalytic Process for Atorvastatin Intermediate. *Green Chem.* **2010**, *12* (1), 81–86. https://doi.org/10.1039/B919115C.
19. Savile, C. K.; Janey, J. M.; Mundorff, E. C.; Moore, J. C.; Tam, S.; Jarvis, W. R.; Colbeck, J. C.; Krebber, A.; Fleitz, F. J.; Brands, J.; Devine, P. N.; Huisman, G. W.; Hughes, G. J. Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture. *Science* **2010**, *329* (5989), 305–309. https://doi.org/10.1126/science.1188934.
20. DiRocco, D. A.; Zhong, Y.-L.; Le, D. N.; McCann, S. D.; Hethcox, J. C.; Kim, J.; Kolev, J. N.; Kosjek, B.; Dalby, S. M.; McMullen, J. P.; Gangam, R.; Morris, W. J. Evolution of a Green and Sustainable Manufacturing Process for Belzutifan: Part 1─Process History and Development Strategy. *Org. Process Res. Dev.* **2024**, *28* (2), 404–412. https://doi.org/10.1021/acs.oprd.3c00408.
21. *GenBank and WGS Statistics*. https://www.ncbi.nlm.nih.gov/genbank/statistics/ (accessed 2023-09-29).
22. Scherlach, K.; Hertweck, C. Mining and Unearthing Hidden Biosynthetic Potential. *Nat. Commun.* **2021**, *12* (1), 3864. https://doi.org/10.1038/s41467-021-24133-5.
23. Medema, M. H.; de Rond, T.; Moore, B. S. Mining Genomes to Illuminate the Specialized Chemistry of Life. *Nat. Rev. Genet.* **2021**, *22* (9), 553–571. https://doi.org/10.1038/s41576-021-00363-7.
24. Blin, K.; Shaw, S.; Kloosterman, A. M.; Charlop-Powers, Z.; van Wezel, G. P.; Medema, M. H.; Weber, T. antiSMASH 6.0: Improving Cluster Detection and Comparison Capabilities. *Nucleic Acids Res.* **2021**, *49* (W1), W29–W35. https://doi.org/10.1093/nar/gkab335.
25. Skinnider, M. A.; Johnston, C. W.; Gunabalasingam, M.; Merwin, N. J.; Kieliszek, A. M.; MacLellan, R. J.; Li, H.; Ranieri, M. R. M.; Webster, A. L. H.; Cao, M. P. T.; Pfeifle, A.; Spencer, N.; To, Q. H.; Wallace, D. P.; Dejong, C. A.; Magarvey, N. A. Comprehensive Prediction of Secondary Metabolite Structure and Biological Activity from Microbial Genome Sequences. *Nat. Commun.* **2020**, *11* (1), 6058. https://doi.org/10.1038/s41467-020-19986-1.
26. Montalbán-López, M.; Scott, T. A.; Ramesh, S.; Rahman, I. R.; van Heel, A. J.; Viel, J. H.; Bandarian, V.; Dittmann, E.; Genilloud, O.; Goto, Y.; Grande Burgos, M. J.; Hill, C.; Kim, S.; Koehnke, J.; Latham, J. A.; Link, A. J.; Martínez, B.; Nair, S. K.; Nicolet, Y.; Rebuffat, S.; Sahl, H.-G.; Sareen, D.; Schmidt, E. W.; Schmitt, L.; Severinov, K.; Süssmuth, R. D.; Truman, A. W.; Wang, H.; Weng, J.-K.; van Wezel, G. P.; Zhang, Q.; Zhong, J.; Piel, J.; Mitchell, D. A.; Kuipers, O. P.; van der Donk, W. A. New Developments in RiPP Discovery, Enzymology and Engineering. *Nat. Prod. Rep.* **2021**, *38* (1), 130–239. https://doi.org/10.1039/D0NP00027B.
27. Tietz, J. I.; Schwalen, C. J.; Patel, P. S.; Maxson, T.; Blair, P. M.; Tai, H.-C.; Zakai, U. I.; Mitchell, D. A. A New Genome-Mining Tool Redefines the Lasso Peptide Biosynthetic Landscape. *Nat. Chem. Biol.* **2017**, *13* (5), 470–478. https://doi.org/10.1038/nchembio.2319.

28. Kjærbølling, I.; Vesth, T.; Andersen, M. R. Resistance Gene-Directed Genome Mining of 50 Aspergillus Species. *mSystems 4* (4), e00085-19. https://doi.org/10.1128/mSystems.00085-19.
29. Hannigan, G. D.; Prihoda, D.; Palicka, A.; Soukup, J.; Klempir, O.; Rampula, L.; Durcak, J.; Wurst, M.; Kotowski, J.; Chang, D.; Wang, R.; Piizzi, G.; Temesi, G.; Hazuda, D. J.; Woelk, C. H.; Bitton, D. A. A Deep Learning Genome-Mining Strategy for Biosynthetic Gene Cluster Prediction. *Nucleic Acids Res.* **2019**, *47* (18), e110. https://doi.org/10.1093/nar/gkz654.
30. Sélem-Mojica, N.; Aguilar, C.; Gutiérrez-García, K.; Martínez-Guerrero, C. E.; Barona-Gómez, F. EvoMining Reveals the Origin and Fate of Natural Product Biosynthetic Enzymes. *Microb. Genomics* **2019**, *5* (12). https://doi.org/10.1099/mgen.0.000260.
31. Mungan, M. D.; Blin, K.; Ziemert, N. ARTS-DB: A Database for Antibiotic Resistant Targets. *Nucleic Acids Res.* **2021**, *50* (D1), D736–D740. https://doi.org/10.1093/nar/gkab940.
32. The UniProt Consortium. UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **2023**, *51* (D1), D523–D531. https://doi.org/10.1093/nar/gkac1052.
33. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, 1–11. https://doi.org/10.1038/s41586-021-03819-2.
34. Pavlopoulos, G. A.; Baltoumas, F. A.; Liu, S.; Selvitopi, O.; Camargo, A. P.; Nayfach, S.; Azad, A.; Roux, S.; Call, L.; Ivanova, N. N.; Chen, I. M.; Paez-Espino, D.; Karatzas, E.; Iliopoulos, I.; Konstantinidis, K.; Tiedje, J. M.; Pett-Ridge, J.; Baker, D.; Visel, A.; Ouzounis, C. A.; Ovchinnikov, S.; Buluç, A.; Kyrpides, N. C. Unraveling the Functional Dark Matter through Global Metagenomics. *Nature* **2023**, *622* (7983), 594–602. https://doi.org/10.1038/s41586-023-06583-7.
35. Hedges, J. B.; Ryan, K. S. Biosynthetic Pathways to Nonproteinogenic α-Amino Acids. *Chem. Rev.* **2020**, *120* (6), 3161–3209. https://doi.org/10.1021/acs.chemrev.9b00408.
36. O'Hagan, D.; Deng, H. Enzymatic Fluorination and Biotechnological Developments of the Fluorinase. *Chem. Rev.* **2015**, *115* (2), 634–649. https://doi.org/10.1021/cr500209t.
37. Galván, A. E.; Paul, N. P.; Chen, J.; Yoshinaga-Sakurai, K.; Utturkar, S. M.; Rosen, B. P.; Yoshinaga, M. Identification of the Biosynthetic Gene Cluster for the Organoarsenical Antibiotic Arsinothricin. *Microbiol. Spectr.* **2021**, *9* (1), e00502-21. https://doi.org/10.1128/Spectrum.00502-21.
38. Kayrouz, C. M.; Huang, J.; Hauser, N.; Seyedsayamdost, M. R. Biosynthesis of Selenium-Containing Small Molecules in Diverse Microorganisms. *Nature* **2022**, *610* (7930), 199–204. https://doi.org/10.1038/s41586-022-05174-2.
39. Woodyer, R. D.; Shao, Z.; Thomas, P. M.; Kelleher, N. L.; Blodgett, J. A. V.; Metcalf, W. W.; van der Donk, W. A.; Zhao, H. Heterologous Production of Fosfomycin and Identification of the Minimal Biosynthetic Gene Cluster. *Chem. Biol.* **2006**, *13* (11), 1171–1182. https://doi.org/10.1016/j.chembiol.2006.09.007.
40. Lee, J.; Simurdiak, M.; Zhao, H. Reconstitution and Characterization of Aminopyrrolnitrin Oxygenase, a Rieske N-Oxygenase That Catalyzes Unusual Arylamine Oxidation. *J. Biol. Chem.* **2005**, *280* (44), 36719–36728. https://doi.org/10.1074/jbc.M505334200.

41. Nelp, M. T.; Bandarian, V. A Single Enzyme Transforms a Carboxylic Acid into a Nitrile through an Amide Intermediate. *Angew. Chem. Int. Ed.* **2015**, *54* (36), 10627–10629. https://doi.org/10.1002/anie.201504505.
42. Sugai, Y.; Katsuyama, Y.; Ohnishi, Y. A Nitrous Acid Biosynthetic Pathway for Diazo Group Formation in Bacteria. *Nat. Chem. Biol.* **2016**, *12* (2), 73–75. https://doi.org/10.1038/nchembio.1991.
43. Yadavalli, S. S.; Ibba, M. Chapter 1 - Quality Control in Aminoacyl-tRNA Synthesis: Its Role in Translational Fidelity. In *Advances in Protein Chemistry and Structural Biology*; Marintchev, A., Ed.; Fidelity and Quality Control in Gene Expression; Academic Press, 2012; Vol. 86, pp 1–43. https://doi.org/10.1016/B978-0-12-386497-0.00001-3.
44. Mohler, K.; Ibba, M. Translational Fidelity and Mistranslation in the Cellular Response to Stress. *Nat. Microbiol.* **2017**, *2* (9), 1–9. https://doi.org/10.1038/nmicrobiol.2017.117.
45. Kaiser, F.; Krautwurst, S.; Salentin, S.; Haupt, V. J.; Leberecht, C.; Bittrich, S.; Labudde, D.; Schroeder, M. The Structural Basis of the Genetic Code: Amino Acid Recognition by Aminoacyl-tRNA Synthetases. *Sci. Rep.* **2020**, *10* (1), 12647. https://doi.org/10.1038/s41598-020-69100-0.
46. Tawfik, D. S.; Gruic-Sovulj, I. How Evolution Shapes Enzyme Selectivity – Lessons from Aminoacyl-tRNA Synthetases and Other Amino Acid Utilizing Enzymes. *FEBS J.* **2020**, *287* (7), 1284–1305. https://doi.org/10.1111/febs.15199.
47. Perona, J. J.; Gruic-Sovulj, I. Synthetic and Editing Mechanisms of Aminoacyl-tRNA Synthetases. In *Aminoacyl-tRNA Synthetases in Biology and Medicine*; Kim, S., Ed.; Topics in Current Chemistry; Springer Netherlands: Dordrecht, 2014; pp 1–41. https://doi.org/10.1007/128_2013_456.
48. Kuzmishin Nagy, A. B.; Bakhtina, M.; Musier-Forsyth, K. Chapter Four - Trans-Editing by Aminoacyl-tRNA Synthetase-like Editing Domains. In *The Enzymes*; Ribas de Pouplana, L., Kaguni, L. S., Eds.; Biology of Aminoacyl-tRNA Synthetases; Academic Press, 2020; Vol. 48, pp 69–115. https://doi.org/10.1016/bs.enz.2020.07.002.
49. Rodgers, K. J.; Shiozawa, N. Misincorporation of Amino Acid Analogues into Proteins by Biosynthesis. *Int. J. Biochem. Cell Biol.* **2008**, *40* (8), 1452–1466. https://doi.org/10.1016/j.biocel.2008.01.009.
50. Walsh, C. T.; O'Brien, R. V.; Khosla, C. Nonproteinogenic Amino Acid Building Blocks for Nonribosomal Peptide and Hybrid Polyketide Scaffolds. *Angew. Chem. Int. Ed.* **2013**, *52* (28), 7098–7124. https://doi.org/10.1002/anie.201208344.
51. McMurry, J. L.; Chang, M. C. Y. Fluorothreonyl-tRNA Deacylase Prevents Mistranslation in the Organofluorine Producer *Streptomyces Cattleya*. *Proc. Natl. Acad. Sci.* **2017**, *114* (45), 11920–11925. https://doi.org/10.1073/pnas.1711482114.
52. Hauth, F.; Funck, D.; Hartig, J. S. A Standalone Editing Protein Deacylates Mischarged Canavanyl-tRNA^{Arg} to Prevent Canavanine Incorporation into Proteins. *Nucleic Acids Res.* **2023**, *51* (5), 2001–2010. https://doi.org/10.1093/nar/gkac1197.

Chapter 2: *tRNA-Deacylase Directed Discovery of Biosynthetic Pathways*



Portions of this work were performed with the following people:

Jorge Marchand contributed towards the discovery and initial characterization of BesG, and constructed the *S. cattleya* Δ*besG* knockout strain.

Calvin Hu contributed towards the elucidation of the *Alf* BGC through protein purification and *in vitro* reactions.

Tavis Chen contributed towards the elucidation of the *Fuz* BGC through protein purification, *in vitro* reactions, and metabolomics.

Shannon Millar contributed towards the development of tR3D, a data visualization tool.

2.1 Introduction

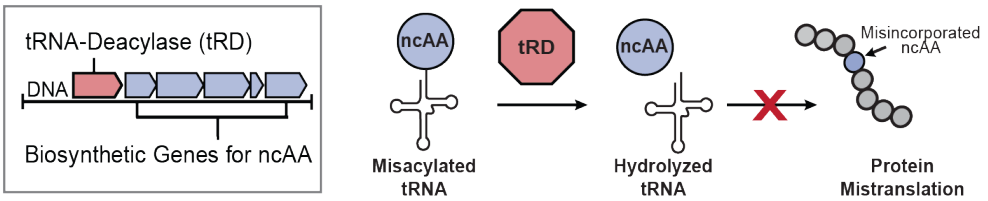
Amino acids are key chiral building blocks for the biosynthesis of proteins, peptides, and natural products. While the standard twenty amino acids are relatively limited in their structural scope, a large number of metabolic pathways have been discovered that yield noncanonical amino acids (ncAAs) with novel chemical, electronic, and structural properties [*1*]. Among notable examples, natural ncAAs have been observed to contain unusual side chains such as arsenic (arsinothricin), diazo (azaserine, thrazarine, duazomycin A), and terminal alkynes (β-ethynylserine, propargylglycine). [*2–6*] Beyond their observation as standalone α-amino acids, ncAAs have also been observed as components of larger metabolites that are key to their key to bioactivity - such as antibacterial β-lactams and neuromodulatory alkaloids [*7,8*].

There are likely many more ncAAs to be discovered, but a systematic approach of identifying biosynthetic gene clusters (BGCs) for ncAAs has yet to be established [*9*]. The diversity of known pathways for ncAAs extend beyond the class-defining features of other natural products, such as for polyketides, non-ribosomal peptides, and ribosomally-produced and post-translationally modified peptides (RiPPs) [*10–13*]. Genome mining for ncAAs typically relies on similarity-based searches for homologous enzymes, which provide helpful constraints of the types of molecules to be discovered but are usually limited to similar transformations [*14,15*]. Resistance enzymes are common features of natural product BGCs , and in contrast to class-defining enzymes, are agnostic to the biosynthetic pathways to achieve these transformations [*16,17*]. Borrowing inspiration from these approaches, we sought to develop a genome mining approach specific to discovering ncAAs and related natural products while remaining agnostic to the biosynthetic means.

In protein synthesis, editing domains and standalone editing proteins provide resistance against misincorporation of noncognate amino acids. For example, the alanine-tRNA synthetase requires an editing domain to compensate for challenges in distinguishing alanine from similar canonical amino acids serine and glycine, hydrolyzing Ser-tRNA^{Ala} and Gly-tRNA^{Ala} before product release [*18*]. As an additional quality control mechanism, some organisms encode standalone editing proteins termed tRNA-deacylases (tRDs) that act *in trans* on misacylated tRNA after product release, such as the AlaX tRD with misacylated tRNA^{Ala} [*19*]. Nine of the twenty aaRSs perform an editing function, and four known classes of tRDs are characterized to date [*20–23*]. While this highly-evolved system is effective for discriminating between the 20 standard amino acids, they are not sufficient to prevent the incorporation of some ncAAs, as demonstrated by many [*24–28*].

With this in view, recent findings have shown that ncAAs encoded by BGCs can compete with structurally similar canonical amino acids in protein synthesis, requiring specific tRDs in the native producer to prevent misincorporation (**Fig. 2.1A**). 4-fluorothreonine (F-Thr), a native metabolite of *Streptomyces cattleya*, can compete with threonine for the threonine-tRNA synthetase (ThrRS), yielding F-Thr-tRNA^{Thr} [*29*]. A tRD found co-localized with the BGC for F-Thr selectively edits mischarged F-Thr-tRNA^{Thr} to prevent downstream protein misincorporation. Canavanine, a metabolite of *R. leguminosarum*, has been known to misincorporate for arginine, and recently a selective tRD was found neighboring genes for canavanine metabolism [*30*]. Here, we describe our studies of the terminal alkyne amino acid β-ethynylserine (βes) and its corresponding resistance protein, the AlaX-like tRD BesG (**Fig. 2.1B**). Notably, each of these discoveries implicates a unique protein family of tRNA-editing proteins, and in each case the tRD co-localizes with genes relevant to the ncAA's biosynthesis or metabolism.

A



B

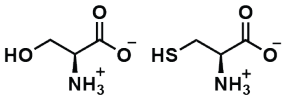
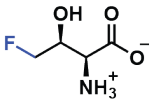
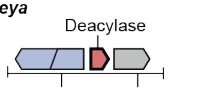
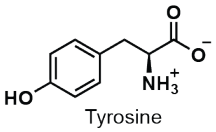
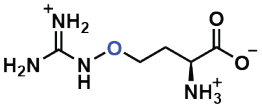
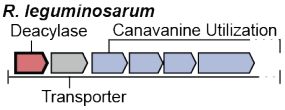
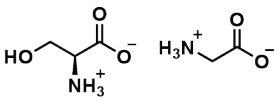
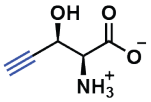
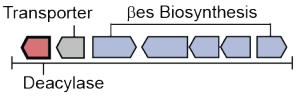
tRD Family (Cognate AA)	Canonical AA Editing Activity	Noncanonical AA Editing Activity	Biosynthetic Gene Cluster
INS (Proline)	 Serine, Cysteine	 4-fluorothreonine	*S. cattleya* 
B3/B4 (Phenylalanine)	 Tyrosine	 Canavanine	*R. leguminosarum* 
AlaX (Alanine)	 Serine, Glycine	 β-ethynylserine	*S. alboverticillatus* 

Figure 2.1. tRNA-Deacylases (tRDs) are resistance proteins for noncanonical amino acid (ncAA) biosynthesis. (A) tRNA-Deacylases (tRDs) with specificity towards ncAAs prevent misincorporation into proteins and are found co-localized with the biosynthetic genes for the ncAAs. (B) Known examples to date of tRDs gaining specificity for ncAAs compared to the canonical amino acids which are typically edited across three different tRD families.

In this report, we examine the entire family of AlaX-like tRDs for their co-occurrence in ncAA BGCs and leverage a data visualization tool we built called tR3D (tRNA-Deacylase Directed Discovery) to identify new BGCs, ncAAs, and a hydrazine-containing peptide. From this work, we establish deacylases as fruitful genomic markers for identifying ncAA-specific biosynthetic enzymes and anticipate this approach can be expanded to a diversity of enzyme classes for natural product discovery.

2.2 Materials and Methods

Commercial materials. Luria-Bertani (LB) Broth Miller, LB Agar Miller, Terrific Broth (TB), yeast extract, malt extract, and glycerol were purchased from EMD Biosciences (Darmstadt, Germany). Carbenicillin (Cb), kanamycin (Km), chloramphenicol (Cm), isopropyl-β-D-thiogalactopyranoside (IPTG), sodium chloride, dithiothreitol (DTT), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), magnesium chloride hexahydrate, acetonitrile, Nanosep 10k Omega filters, and dextrose were purchased from Fisher Scientific (Pittsburgh, PA). Adenosine triphosphate sodium salt (ATP), nicotinamide adenine dinucleotide reduced form dipotassium salt (NADH), flavin adenine dinucleotide disodium salt (FAD), β-nicotinamide adenine dinucleotide phosphate reduced form disodium salt (NADPH), lysozyme, magnesium sulfate heptahydrate, calcium chloride, ammonia chloride, β-mercaptoethanol (βME), sodium phosphate dibasic heptahydrate, sodium phosphate monobasic, apramycin sulfate salt (Am), dimethylsulfoxide (DMSO), acetonitrile (LC/MS-grade), ammonium formate (LC/MS-grade), copper sulfate, Tris(3-hydroxypropyltriazolylmethyl)amine (THPTA), α-ketoglutaric acid sodium salt, sodium ascorbate, maltose, L-propargylglycine (Pra), all 20 proteinogenic amino acids (L-alanine, L-arginine, L-asparagine, L-aspartic acid, L-cysteine (HCl), L-glutamic acid (potassium salt), L-glutamine, L-glycine, L-histidine, L-isoleucine, L-leucine, L-lysine, L-methionine, L-phenylalanine, L-proline, L-serine, L-serine, L-threonine, L-tryptophan, L-tyrosine, L-valine), 24-well plates and Aeraseal covers were purchased from Sigma-Aldrich (St. Louis, MO). PageRuler Plus Prestained Protein Ladder was purchased from ThermoFisherScientific (Waltham, MA). Tamra-Azide, 5-isomer was purchased from Lumiprobe (Hunt Valley, Maryland). BactoTM Agar was purchased from BD (Sparks, Maryland). Soy flour was purchased from Berkeley Bowl (Berkeley, CA). Formic acid was purchased from Acros Organics (Morris Plains, NJ). Restriction enzymes, Phusion DNA polymerase, PURExpress (Δaa, tRNA) IVTT kit, T7 HiScribe Quick High Yield RNA Synthesis Kit, T5 exonuclease, and Taq DNA ligase were purchased from New England Biolabs (Ipswich, MA). Deoxynucleotides (dNTPs), were purchased from Invitrogen (Carlsbad, CA). Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA), resuspended at a stock concentration of 100 μM in 10 mM Tris-HCl, pH 8.5, and stored at 4 °C. DNA purification kits, miRNeasy Mini Kit, and Ni-NTA agarose were purchased from Qiagen (Valencia, CA). SoluLyse reagent was purchased from Genlantis (San Diego, CA). Complete EDTA-free protease inhibitor was purchased from Roche Applied Science (Penzberg, Germany). PD-10 desalting columns were purchased from GE Healthcare Life, (Pittsburg, PA). Amicon Ultra 10,000 MWCO and 30,000 MWCO centrifugal concentrators were purchased from Millipore (Billerica, MA). Acrylamide/bis-acrylamide (30%, 37.5:1), electrophoresis grade sodium dodecyl sulfate (SDS) were purchased from Bio-Rad Laboratories (Hercules, CA). Isotopically-labelled amino acids were purchased from Cambridge Isotope Laboratories (Andover, MA). All chemicals were used as purchased without purification.

Bacterial strains. *Streptomyces cattleya* NRRL 8057 (ATCC 35852) was purchased from the American Tissue Type Collection (Manassas, VA). *E. coli* DH10B-T1^R was used for plasmid construction. *E. coli* BL21 Star (DE3) was used for heterologous protein production and metabolomics experiments. *E. coli* GM272 harboring the non-transmissible, oriT-mobilizing plasmid pUZ8002 was used for conjugative plasmid transfer into *S. cattleya*.

Construction of plasmids. Gibson ligation was used to carry out plasmid construction using *E. coli* DH10B-T1^R as the cloning host. All PCR amplifications were carried out with Phusion polymerase. Oligonucleotides used are listed in **Table A1.1B**. Unless otherwise noted, pET16b-

His_{10} was used for protein expression. To insert a gene in pET16b, the plasmid was digested at NdeI and BamHI. To insert into pSV272.1-His_6MBP, the plasmid was digested at SfoI and EcoRI. For pDuet plasmids (pETDuet, pACYCDuet, and pRSFDuet), the first gene was inserted following digestion at NcoI and HindIII sites, and the second gene was inserted following digestion at BglII or NdeI, and AvrII sites. The *Mf* FADR was inserted into pACYCDuet-Cupin-HydS following digestion at HindIII and NotI. For Peptide C, the whole plasmid was PCR amplified, digested with DpnI overnight at 37°C, and purified for Gibson ligation.

Transformation of chemically competent *E. coli*. A 100 μL aliquot of chemically competent *E. coli* is transformed with 1 μL of pure plasmid or full Gibson mixture, 20 μL of KCM buffer (500 mM KCl, 150 mM $CaCl_2$, 250 mM $MgCl_2$), and 160 μL MilliQ H_2O. The cells are incubated on ice for 30 min, followed by heat shock at 42 °C for 90 s. The cells are plated onto agar immediately (Cb) or after 1 h recovery in LB at 37 °C (Km, Cm).

Protein Expression Procedure. Single colonies of *E. coli* BL21 Star (DE3) transformed with plasmid were inoculated into TB (25 mL) containing the appropriate antibiotics and grown overnight at 37°C with shaking at 200 rpm. Overnight cultures were inoculated into 1L of TB with 50 μg/mL antibiotic in a 2.8 L-baffled shake flask to OD_{600} = 0.05 and grown at 37 °C and 200 rpm until the OD_{600} reached 0.6 - 0.8. Cultures were cooled on ice for 20 min before inducing with IPTG (0.25 mM) and then cultured for 18 h at 16°C and 200 rpm. Cells were pelleted and stored at -80 °C for storage until purification.

Protein Purification via Ni-NTA Chromatography. Adapted from [*31*]. Frozen cell pellets were thawed and resuspended in 45 ml of lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM imidazole, 20 mM βME and 20% (*v/v*) glycerol, pH 7.5) supplemented with EDTA-free protease inhibitor cocktail (Roche), lysozyme (0.3 mg/mL), and benzonase (5 U/mL). The cell paste was homogenized and then lysed by sonication with a Qsonica Sonicator 3000 (power= 50, 5 s on, 25 s off, 2min total process time, 0.5-inch tip). The lysate was then centrifuged at 13,000 × *g* for 10 min at 4 °C to separate the soluble and insoluble fractions. The soluble lysate was incubated with Ni-NTA (0.5ml of resin per gram of cell paste) for 45 min at 4 °C, resuspended and loaded onto a column by gravity flow. The column was washed with wash buffer (50 mM HEPES, 300 mM NaCl, 20 mM imidazole, 20 mM βME and 20% (*v/v*) glycerol, pH 7.5) for 15–20 column volumes. The column was then eluted with elution buffer (50 mM HEPES, 300 mM NaCl, 300 mM imidazole, 20 mM βME and 20% (*v/v*) glycerol, pH 7.5). Fractions containing the target protein were pooled according to absorbance at 280 nm and concentrated using an Amicon Ultra spin concentrator (10 or 30-kD MWCO; Millipore). Protein was then exchanged into storage buffer (50 mM HEPES, 100 mM sodium chloride, 10% (*v/v*) glycerol and 1 mM DTT, pH 7.5) using PD-10 desalting columns. Final protein concentrations before storage were estimated using the $\varepsilon_{280\ nm}$ calculated by ExPASy ProtParam [*32*] and measured using a Nanodrop spectrometer. They are as follows: *S. cattleya* ThrRS: 5.39 mg/mL ($\varepsilon_{280\ nm}$ = 75,770 $M^{-1}cm^{-1}$), *E. coli* ThrRS: 5.50 mg/mL ($\varepsilon_{280\ nm}$ = 96,260 $M^{-1}cm^{-1}$), *A. faecalis* MetRS: 12.4 mg/mL ($\varepsilon_{280\ nm}$ = 90,760 $M^{-1}cm^{-1}$), AlfB: 2.69 mg/mL ($\varepsilon_{280\ nm}$ = 19,940 $M^{-1}cm^{-1}$), AlfC: 1.81 mg/mL ($\varepsilon_{280\ nm}$ = 32,890 $M^{-1}cm^{-1}$), FuzA: 5.19 mg/mL ($\varepsilon_{280\ nm}$ = 53,860 $M^{-1}cm^{-1}$), FuzC: 24.9 mg/mL ($\varepsilon_{280\ nm}$ = 28,420 $M^{-1}cm^{-1}$), FuzD: 3.07 mg/mL ($\varepsilon_{280\ nm}$ = 57,870 $M^{-1}cm^{-1}$), FuzE: 12.3 mg/mL ($\varepsilon_{280\ nm}$ = 24,980 $M^{-1}cm^{-1}$), FuzF: 5.51 mg/mL ($\varepsilon_{280\ nm}$ = 41,940 $M^{-1}cm^{-1}$), FuzG: 6.54 mg/mL ($\varepsilon_{280\ nm}$ = 71,850 $M^{-1}cm^{-1}$), FuzH: 5.73 mg/mL ($\varepsilon_{280\ nm}$ = 23,380 $M^{-1}cm^{-1}$). All proteins were aliquoted, flash frozen in liquid nitrogen and stored at −80 °C.

Protein Purification via Amylose Chromatography. Frozen cell pellet was thawed and resuspended in 45 ml of lysis buffer (50 mM HEPES, 300 mM NaCl, 20 mM βME and 20% (*v/v*) glycerol, pH 7.5) supplemented with EDTA-free protease inhibitor cocktail (Roche), lysozyme (0.3 mg/mL), and benzonase (5 U/mL). The cell paste was homogenized and then lysed by sonication with a Qsonica Sonicator 3000 (power= 50, 5 s on, 25 s off, 2 min. total process time, 0.5-inch tip). The lysate was then centrifuged at 13,000 × *g* for 10 min at 4 °C to separate the soluble and insoluble fractions. The soluble lysate was diluted 5x in additional lysis buffer and loaded onto a column containing amylose resin (3 ml of resin per gram of cell paste) by gravity flow at 4 °C. The column was washed with an additional 5 column volumes of lysis. The column was then eluted with elution buffer (50 mM HEPES, 300 mM NaCl, 10 mM maltose, 20 mM βME and 20% (*v/v*) glycerol, pH 7.5). Fractions containing the target protein were pooled according to absorbance at 280 nm and concentrated using an Amicon Ultra spin concentrator (30-kD MWCO; Millipore). Protein was then exchanged into storage buffer (50 mM HEPES, 100 mM sodium chloride, 10% (*v/v*) glycerol and 1 mM DTT, pH 7.5) using PD-10 desalting columns. Final protein concentrations before storage were estimated using the $\varepsilon_{280\ nm}$ calculated by ExPASy ProtParam [*32*] and measured using a Nanodrop spectrometer. For MBP-BesG, protein concentration was 12.39 mg/mL ($\varepsilon_{280\ nm}$ = 94,770 $M^{-1}cm^{-1}$). Protein was aliquoted and flash frozen in liquid nitrogen and stored at −80 °C.

General procedure for high-resolution HPLC/MS analysis of polar metabolites with hydrophilic interaction chromatography (HILIC). Adapted from [*31*]. Samples containing polar metabolites were analyzed using an Agilent 1290 UPLC on a SeQuant ZIC-pHILIC (5 μm, 2.1×100 mm; EMD-Millipore) using the following buffers: buffer A (90% water, 10% acetonitrile and 10 mM ammonium formate) and buffer B (90% acetonitrile, 10% water and 10 mM ammonium formate). After 2 min at 100% buffer B, a linear gradient from 0% to 40% buffer A was applied over 15 min followed by a linear gradient from 40% to 60% buffer A over 5 min and a linear gradient from 60% to 70% buffer A over 2 min at a flow rate of 0.2 mL/min. Mass spectra were acquired in positive-ionization or negative-ionization mode using an Agilent 6530 QTOF (Agilent) with the following source and acquisition parameters: Gas temperature = 300 ºC; drying gas = 8 L/min; nebulizer = 35 psig; sheath gas temperature = 275 °C; capillary voltage = 3500 V; fragmentor 175 V; skimmer 65 V; oct 1 RF vpp = 750 V; acquisition rate = 1 spectra/s; acquisition time = 1000 ms/spectrum; min range = 50 *m/z*; max range = 500 *m/z*.

General procedure for rapid HPLC/MS analysis of polar metabolites with hydrophilic interaction chromatography (HILIC). HPLC fractions containing polar metabolites were screened using an Agilent 1290 UPLC on a SeQuant ZIC-HILIC (5 μm, 2.1×50 mm; Sigma-Aldrich) using the following buffers: buffer A (90% water, 10% acetonitrile and 10 mM ammonium formate) and buffer B (90% acetonitrile, 10% water and 10 mM ammonium formate). After 0.2 min at 100% buffer B, a linear gradient from 0% to 50% buffer A was applied over 1.80 min followed by a linear gradient from 50% to 100% buffer A over 1 min followed by a 1 min hold at 100% A at a flow rate of 0.6 mL/min. Mass spectra were acquired in positive-ionization mode using an Agilent 6530 QTOF (Agilent) with the following source and acquisition parameters: Gas temperature = 300 ºC; drying gas = 8 L/min; nebulizer = 35 psig; sheath gas temperature = 275 °C; capillary voltage = 3500 V; fragmentor 175 V; skimmer 65 V; oct 1 RF vpp = 750 V; acquisition rate = 1 spectra/s; acquisition time = 1000 ms/spectrum; min range = 100 *m/z*; max range = 1700 *m/z*.

General procedure for high-resolution HPLC/MS analysis of nonpolar metabolites with reverse-phase chromatography. Samples containing nonpolar metabolites were analyzed using an Agilent 1290 UPLC on a Poroshell 120 EC-C18 (5 μm, 2.1 × 30 mm; Agilent) using the following buffers: Buffer A (100% water, 0.1% *v/v* formic acid) and Buffer B (100% acetonitrile (LC/MS grade)). After 1 min at 100% buffer A, a linear gradient from 0% to 100% Buffer B was applied over 5 min was followed by 30 s at 100% Buffer B at a flow rate of 0.6 mL/min. Mass spectra were acquired on either positive or negative ionization modes using an Agilent 6530 QTOF (Agilent) with the following source and acquisition parameters: Gas temperature = 300 ºC; drying gas = 8 L/min; nebulizer = 35 psig; sheath gas temperature = 275 °C; capillary voltage = 3500 V; fragmentor 175 V; skimmer 65 V; oct 1 RF vpp = 750 V; acquisition rate = 1 spectra/s; acquisition time = 1000 ms/spectrum; min range = 50 *m/z*; max range = 1700 *m/z*.

General procedure for high-resolution HPLC/MS-MS analysis of aminoacyl-tRNA with reverse-phase chromatography. In an approach adapted from [*29*], adenylate samples were analyzed using an Agilent 1290 UPLC on a Poroshell 120 EC-C18 (5 μm, 2.1 × 30 mm; Agilent) using the following buffers: Buffer A (100% water, 10 mM ammonium formate) and Buffer B (100% acetonitrile (LC/MS grade)). after 50 s at 95% buffer A, a linear gradient from 5% to 30% Buffer B was applied over 40 s followed by a linear gradient to 100% Buffer B over 90 s at a flow rate of 0.7 mL/min. Mass spectra were acquired on either positive ionization modes using an Agilent 6460 QQQ (Agilent) with the following source and acquisition parameters: Gas temperature = 250 ºC; drying gas = 5 L/min; nebulizer = 45 psig; sheath gas temperature = 250 °C; capillary voltage = 3500 V; fragmentor = 60 V; collision energy = 0 V; cell accelerator voltage = 7. Ions were monitored by MRM using the following transitions: βes-adenylate (*m/z* = 379.1→136.1), Thr-adenylate (*m/z* = 369.1→136.1), 3-OH-Met-adenylate (*m/z* = 415.1→136.1), Met-adenylate (*m/z* = 399.1→136.1).

General procedure for high-resolution HPLC/MS-MS analysis of aminoacyl-tRNA with hydrophilic interaction chromatography (HILIC). In an approach adapted from [*29*], samples containing polar metabolites were analyzed using an Agilent 1290 UPLC on a SeQuant ZIC-pHILIC (5 μm, 2.1×100 mm; EMD-Millipore) using the following buffers: buffer A (90% water, 10% acetonitrile and 10 mM ammonium formate) and buffer B (90% acetonitrile, 10% water and 10 mM ammonium formate). After 2 min at 100% buffer B, a linear gradient from 0% to 33% buffer A was applied over 15 min followed by a linear gradient from 33% to 60% buffer A over 3 min followed by 2 min at 60% buffer A at a flow rate of 0.2 mL/min. Mass spectra were acquired in positive-ionization mode using an Agilent 6460 QQQ (Agilent) with the following source and acquisition parameters: Gas temperature = 300 ºC; drying gas = 5 L/min; nebulizer = 45 psig; sheath gas temperature = 250 °C; capillary voltage = 3500 V; fragmentor 60 V; collision energy = 0 V; cell accelerator voltage = 7. Ions were monitored by MRM using the following transitions: βes (*m/z* = 130.1→84.1), Thr (*m/z* = 120.1→74.1).

General procedures for the synthesis of βes. All chemicals were purchased from Sigma-Aldrich and used without further purification. NMR characterization was performed on a Bruker AV-600 (600 MHz) instrument equipped with either a 5 mm broadband observe Z-gradient probe or a 5 mm triple inverse broadband Z-gradient probe. Spectra were acquired at 298 K and chemical shifts were expressed relative to tetramethylsilane. Commercially available N-Boc-D-serine methyl ester was converted to βes-Cpd4 by known methods [*33–37*]. Next, the secondary alcohol was protected as acetyl ester βes-Cpd5. Conditions for N,O-acetonide deprotection were screened, and the optimal condition of heating βes-Cpd5 in 3:1 acetic acid/water at 50 ºC was used to

generate βes-Cpd6. Our choices of the acetyl ester and deprotection conditions were crucial for minimizing ester migration to the primary alcohol, and for leaving the Boc group intact, providing a new approach to these two challenges previously observed by Chan and O'Hagan [*33*]. Finally, Jones oxidation followed by basic and acidic deprotections according to [*33*] gave β-ethynylserine hydrochloride, in 5 steps and 34% yield from the literature intermediate βes-Cpd4.

Synthesis of (*R*)-2,2-dimethyloxazolidine-3,4-dicarboxylic acid 3-tert-butyl ester 4-methyl ester (βes-Cpd2, D-Garner's ester). Commercially available N-Boc-D-serine methyl ester was converted to βes-Cpd4 by known methods [*33–37*]. To *N*-(*tert*-butoxycarbonyl)-D-serine methyl ester (10 g, 45.6 mmol) dissolved in acetone (140 mL) was added p-toluenesulfonic acid (100 mg) and 2,2-dimethoxypropane (56 mL, 0.46 mol). The mixture was heated to reflux and stirred for 4 h. Additional *p*-toluenesulfonic acid (100 mg) and 2,2-dimethoxypropane (28 mL, 0.23 mol) was added, following which the reflux was continued for 4 h and the resulting mixture was concentrated. The residue was dissolved in Et_2O (80 mL) and washed with sat. $NaHCO_3$ (3 × 50 mL). The combined aqueous layers were back-extracted with Et_2O (3 × 40 mL). The combined organic layers were then washed with brine (50 mL), dried over Na_2SO_4, and concentrated. Purification by silica gel chromatography (1:1 hexanes/Et_2O) provided the product as a yellow oil (9.86 g, 83%). ^{1}H NMR (600 MHz, $CDCl_3$) δ 4.49 (dd, *J* = 6.8, 2.6 Hz, 0.4H), 4.38 (dd, *J* = 7.1, 3.1 Hz, 0.6H), 4.17-4.10 (m, 1H), 4.07-4.01 (m, 1H), 3.76 (s, 3H), 1.67 (s, 1.7H), 1.64 (s, 1.3H), 1.53 (s, 1.8H), 1.49 (s, 5H), 1.41 (s, 5.2H). ^{13}C NMR (151 MHz, $CDCl_3$) δ 171.6, 171.2, 152.0, 151.1, 95.0, 94.3, 80.8, 80.2, 66.2, 65.9, 59.2, 59.1, 52.3, 52.2, 28.3, 28.2, 26.0, 25.1, 24.9, 24.3. Most peaks are doubled due to carbamate rotamers.

Synthesis of (*R*)-4-formyl-2,2-dimethyloxazolidine-3-carboxylic acid *tert*-butyl ester (βes-Cpd3, *D*-Garner's aldehyde). In an oversized flask under nitrogen atmosphere, D-Garner's ester (5.19 g, 20 mmol) was dissolved in anhydrous toluene (40 mL). The contents were cooled to -78 ºC, then a solution of DIBAL-H (1 M in toluene, 35 mL) was added dropwise via syringe over 15 min, ensuring the reagent was trickled down the wall of the cold flask. The reaction was stirred at -78 ºC for 2 h then slowly quenched with MeOH (10 mL). The mixture was poured into 1 M Rochelle's salt (120 mL) and stirred vigorously for 1 h to obtain clear phase separation. The organic layer was set aside, and the aqueous layer was extracted with Et_2O (3 × 50 mL). The combined organic layers were washed with brine (50 mL), dried over Na_2SO_4, and concentrated. Purification by silica gel chromatography (1:1 hexanes/Et_2O) provided the product as a light yellow oil (3.66 g, 80%). ^{1}H NMR (600 MHz, $CDCl_3$) δ 9.61 (s, 0.4H), 9.55 (d, *J* = 2.6 Hz, 0.6H), 4.36-4.32 (m, 0.4H), 4.21-4.17 (m, 0.6H), 4.13-4.03 (m, 2H), 1.65 (s, 2H), 1.60 (s, 1.6H), 1.56 (s, 2.4H), 1.51 (s, 3H), 143 (s, 6H). ^{13}C NMR (151 MHz, $CDCl_3$) δ 199.4, 199.3, 152.6, 152.3, 95.0, 94.3, 81.3, 81.0, 64.7, 64.6, 63.9, 63.4, 28.25, 28.20, 26.6, 25.7, 24.6, 23.7. Most peaks are doubled due to carbamate rotamers.

Synthesis of (*R*)-4-((*R*)-1-hydroxy-3-(trimethylsilyl) prop-2-yn-1-yl) - 2,2-dimethyloxazolidine 3-carboxylate *tert*-butyl ester (βes-Cpd4). A flask fitted with a reflux condenser was placed under nitrogen atmosphere, charged with anhydrous THF (30 mL), and cooled to 0 ºC. Ethynyltrimethylsilane (2.2 mL, 16 mmol) was added, followed by ethylmagnesium bromide (1 M in THF, 15 mL) dropwise. The mixture was heated to reflux with stirring for 2 h and then cooled to room temperature. Meanwhile, in a second flask under nitrogen, copper(I) iodide (4.19 g, 22 mmol) was dissolved in anhydrous THF (30 mL) and dimethyl sulfide (12 mL). This solution was stirred at room temperature for 1.5 h, then cooled to -78 ºC. The Grignard solution was cannulated into the copper solution dropwise, then the mixture was warmed

to -30 ºC (acetonitrile/dry ice) and stirred for 30 min. After cooling the organocopper reagent to -78 ºC, D-Garner's aldehyde (2.29 g, 10 mmol) was added dropwise as a solution in THF (5 mL). The mixture was stirred for 15 min at -78 ºC, then warmed to room temperature and stirred for 16 h overnight. The reaction was quenched with sat. NH_4Cl (100 mL) and stirred for 30 min to achieve complexation of copper salts, then the product was extracted with Et_2O (3 × 80 mL). The combined organic layers were washed with 0.5 M HCl (2 × 50 mL, precipitates residual copper salts) and brine (50 mL), then dried over Na_2SO_4 and concentrated to ~20 mL. Precipitates were removed by filtration over Celite and rinsing the solids with Et_2O (3 × 20 mL). After further concentration to ~5 mL, the final precipitates were removed by filtration through a glass pipette with a plug of filter paper. Purification by silica gel chromatography (1:1 hexanes/Et_2O) provided the product as a very viscous yellow oil (2.19 g, 67%). ^{1}H NMR (600 MHz, DMSO-d_6) δ 5.75 (d, J = 5.5 Hz, 0.5H), 5.72 (d, J = 5.5 Hz, 0.5H), 4.69 (t, J = 5.1 Hz, 0.5H), 4.64 (t, J = 5.2 Hz, 0.5H), 4.01 (dd, J = 9.3, 3.4 Hz, 1H), 3.94 (q, J = 7.9 Hz, 1H), 3.82 (m, 0.5H), 3.78 (m, 0.5H), 1.51 (s, 3.5H), 1.41-1.39 (2 s, 11.5 H), 0.10 (s, 9H). ^{13}C NMR (151 MHz, DMSO-d_6) δ 152.3, 151.6, 106.7, 106.5, 94.3, 94.0, 89.3, 89.1, 80.0, 79.6, 64.5, 64.1, 62.0, 61.1, 60.5, 28.45, 28.39, 26.8, 25.9, 25.2, 23.9, 0.2. Most peaks are doubled due to carbamate rotamers.

Synthesis of (*R*)-4-((R)-1-acetoxy-3- (trimethylsilyl) prop-2-yn-1-yl) - 2,2-dimethyloxazolidine-3-carboxylate *tert*-butyl ester (βes-Cpd5). βes-Cpd4 (1.15 g, 3.5 mmol) was dissolved in pyridine (10 mL), then acetic anhydride (1.0 mL, 10.5 mmol) and DMAP (21 mg, 0.18 mmol) were added. The mixture was stirred at room temperature for 1.5 h, then was diluted with Et_2O (30 mL) and washed with water (3 × 20 mL). The combined aqueous layers were back-extracted with Et_2O (3 × 20 mL). The combined organic layers were then washed with brine (40 mL), dried over Na_2SO_4, and concentrated. Purification by silica gel chromatography (4:1 hexanes/Et_2O) provided the product as a light yellow oil that crystallized to a white solid upon cooling (1.10 g, 85%). ^{1}H NMR (600 MHz, $CDCl_3$) δ 5.88 (d, J = 4.9 Hz, 0.5H), 5.80 (d, J = 6.1 Hz, 0.5H), 4.21-4.11 (m, 1.5H), 4.03-3.97 (m, 1H), 3.96-3.91 (m, 0.5H), 2.08-2.06 (2 s, 3H), 1.64-1.48 (5 s, 15 H), 0.16 (s, 9H). ^{13}C NMR (151 MHz, $CDCl_3$) δ 169.1, 169.0, 152.4, 151.7, 100.0, 99.5, 95.0, 94.4, 92.4, 91.9, 80.6, 64.7, 64.6, 64.2, 64.1, 59.3, 58.4, 28.34, 28.28, 26.8, 25.9, 24.8, 23.6, 20.9, -0.4. Most peaks are doubled due to carbamate rotamers.

Synthesis of (3*R*,4*R*)-4-((*tert*-butoxycarbonyl)amino)-5-hydroxy-1-(trimethylsilyl)pent-1-yn-3-yl acetate (βes-Cpd6). βes-Cpd5 (369.5 mg, 1 mmol) was dissolved in 3:1 AcOH/H_2O (20 mL) and stirred with heating at 50 ºC for 16 h overnight. The solvent was evaporated, then the residue was dissolved in EtOAc (30 mL) and washed with sat. $NaHCO_3$ (2 × 20 mL). The combined aqueous layers were back-extracted with EtOAc (3 × 20 mL). The combined organic layers were then washed with brine (40 mL), dried over Na_2SO_4, and concentrated. Purification by silica gel chromatography (2:1 hexanes/EtOAc) provided the product as a white solid (Rf 0.17, 161.4 mg, 49%). ^{1}H NMR (600 MHz, $CDCl_3$) δ 5.59 (d, J = 5.5 Hz, 1H), 4.98 (d, J = 9.1 Hz, 1H), 3.99-3.92 (m, 1H), 3.90-3.84 (m, 1H), 3.69-3.63 (m, 1H), 2.24-2.18 (m, 1H), 2.11 (s, 3H), 1.45 (s, 9H), 0.17 (s, 9H). ^{13}C NMR (151 MHz, $CDCl_3$) δ 169.6, 155.6, 99.9, 92.5, 79.9, 63.5, 61.9, 54.7, 28.3, -0.4. The minor product in which the acetyl group migrated to the primary position can be recovered as a yellow oil (Rf 0.30, 40.3 mg, 12%). ^{1}H NMR (600 MHz, $CDCl_3$) δ 5.01 (d, J = 7.9 Hz, 1H), 4.51 (t, J = 5.4 Hz, 1H), 4.34-4.28 (m, 1H), 4.28-4.21 (m, 1H), 4.06-3.99 (m, 1H), 2.92 (d, J = 5.1 Hz, 1H), 2.11 (s, 3H), 1.48 (s, 9H), 0.19 (s, 9H). ^{13}C NMR (151 MHz, $CDCl_3$) δ 171.1, 156.0, 103.1, 91.2, 80.0, 63.0, 62.5, 53.8, 28.3, 20.9, -0.3.

Synthesis of (2*S*,3*R*)-3-acetoxy-2-((tert-butoxycarbonyl)amino)-5-(trimethylsilyl)pent-4-ynoic acid (βes-Cpd7). Following a modified preparation for βes-Cpd7 from Chan and O'Hagan [*33*] including Jones oxidation followed by basic and acidic deprotections, βes-Cpd6 (108.7 mg, 0.33 mmol) was dissolved in acetone (10 mL). Jones reagent (1 M CrO_3 in 10% aqueous H_2SO_4, 2.0 mL) was added dropwise and the mixture stirred for 16 h overnight at room temperature. Isopropanol (2 mL) was added to quench the reaction and the majority of solvent was evaporated. The residue was dissolved in EtOAc (30 mL) and washed with 0.5 M HCl (2 × 20 mL). The combined aqueous layers were back-extracted with EtOAc (3 × 20 mL). The combined organic layers were then washed with brine (40 mL), dried over Na_2SO_4, and concentrated. Purification by silica gel chromatography (2:1 hexanes/EtOAc + 1% AcOH) followed by concentration from toluene to remove acetic acid provided the product as a white solid (98.0 mg, 86%). ^{1}H NMR (600 MHz, methanol-d_4) δ 5.79 (br, 1H), 4.94 (br, 1H), 2.08 (s, 3H), 1.50 (s, 9H), 0.17 (s, 9H). ^{13}C NMR (151 MHz, methanol-d_4) δ 169.4, 156.5, 99.3, 91.1, 79.4, 64.4, 27.4, 19.1, -1.7. Some carbon resonances are not visible due to deuterium-induced signal broadening.

Synthesis of (2*S*,3*R*)-2-((*tert*-butoxycarbonyl)amino)-3-hydroxypent-4-ynoic acid (βes-Cpd8) and (2*S*,3*R*)-2-Amino-3-hydroxypent-4-ynoic acid hydrochloride (βes-HCl). In an approach adapted from [*33*], βes-Cpd7 (51.5 mg, 0.15 mmol) was dissolved in 2 M methanolic KOH (2 mL) and stirred at room temperature for 2 h. The mixture was diluted with methanol (20 mL) and then pre-washed Dowex 50W was added until the solution was neutral to water-wetted pH paper. The resin was filtered off and rinsed with methanol (2 ×20 mL). The solvent was evaporated to provide crude βes-Cpd8 that was used directly in the next step. βes-Cpd8 was dissolved in 6 M HCl (2 mL) and stirred at room temperature for 2 h. The mixture was diluted with water (6 mL) and washed with EtOAc (3 × 6 mL). The aqueous layer was lyophilized to provide the product as a light brown powder (23.6 mg, 95% over 2 steps). ^{1}H NMR (600 MHz, D_2O) δ 5.00 (t, J = 2.6 Hz, 1H), 4.17 (d, J = 3.0 Hz, 1H), 2.98 (d, J = 2.2 Hz, 1H). ^{13}C NMR (151 MHz, D_2O) δ 168.9, 78.7, 77.6, 59.5, 57.6.

***In vitro* transcription and translation of Thr/βes-containing peptides.** The IVTT reactions were performed using the PURExpress Δ(aa, tRNA) kit (NEB), which separates the amino acids and tRNA from the other reaction components. Reactions contained the plasmid encoding the peptide (12.5 ng), all necessary amino acids (8 total) for the encoded peptide (METRSKNFL) (0.33 mM) except Thr, and the remaining PURExpress kit components according to their protocol. Thr or βes was added (0.40 mM) depending on the reaction condition. The reactions (12.5 µl) were incubated at 37 °C for 4 h. Protein components were removed from the reaction by passage through a Pall Nanosep 10-kD spin column, resuspended with 10 µL water, and passed through the filter again, yielding ~20 µL final sample volume. The flow-through was analyzed with LC/MS using the procedure for nonpolar metabolites.

***In vitro* transcription and purification of tRNA and aminoacyl-tRNA.** dsDNA oligos for *in vitro* tRNA transcription containing a T7 promoter and corresponding tRNA were synthesized by IDT. PCR using Phusion polymerase with forward and reverse primers (primer name reads acceptor stem 5′ to 3′) was used to amplify DNA from the synthetic oligos to obtain high-concentration (150-200 ng/uL) DNA used for *in vitro* transcription. The primers for *Af* Met-tRNA primers contained a methylguanosine as the second to last base on the 3′ end to minimize RNA polymerase run-off, and the extended primer length eliminated the need for a dsDNA template [*38*]. The HiScribe T7 High Yield RNA Synthesis Kit (NEB) was used for *in vitro* transcription, with the recommended conditions for short RNAs. Transcription reactions contained 1 µg of DNA

template in 20 μL and were ran overnight at 37 °C. For aminoacylation reactions with βes and Thr, the reaction contained ThrRS (8 μM), ATP (5 mM), $MgCl_2$ (10 mM), $ZnCl_2$ (0.5mM), amino acid (βes or Thr, 4 mM), HEPES (40 mM, pH 7.5) with nuclease-free H_2O to reach 15 μL (35 μL total). For *S. cattleya* tRNA sequences, the *S. cattleya* ThrRS was used, and for the *E. coli* tRNA sequences, the *E. coli* ThrRS was used. Prior to beginning the reaction, tRNA was first refolded according to a literature protocol [*39*]. Briefly, HEPES buffer and water were added to the transcription reaction, the sample was heated to 98°C on a thermocycler for 2 min, and slowly cooled to room temperature over 5 min. $MgCl_2$ was added, and the sample was incubated at room temperature for 5 min, followed by 3 min on ice. The rest of the components were added, with enzyme as the last component to begin the acylation reaction. The reaction was allowed to incubate for 4 h at 37 °C. tRNA or aminoacylated-tRNA was purified from these reactions directly using miRNeasy purification kit (QIAGEN) following manufacturer protocols.

Quantification of aminoacyl-tRNA. In an approach adapted from [*29*], aminoacyl-tRNA was quantified after purification by hydrolysis of a subset of βes- or Thr-tRNA followed by quantification using LC/MS-MS. To hydrolyze βes- or Thr-tRNA, 5 μL of purified tRNA was mixed with 5 μL of 0.1 M NaOH for 1 h at 37 °C. Reactions were neutralized by adding 10 μL of 0.1 M $NaHPO_4$, then mixed 1:1 (*v/v*) with MeOH for analysis on polar phase LC method. External standards of βes or Thr (0 – 100 μM) were run alongside hydrolyzed samples using the general procedure for high resolution HPLC/MS-MS analysis of aminoacyl-tRNA using HILIC chromatography and fit using a linear standard curve (R^2 = 0.97 for βes and R^2 = 0.99 for Thr). The MRM transitions βes (*m/z* = 130.1→112.1) and Thr (*m/z* =120.1→102.1) were chosen for generating standard curve.

***In vitro* deacylation timecourse of various $tRNA^{Thr}$ by BesG.** Deacylation studies of $tRNA^{Thr}$ from *S. cattleya* and *E. coli* were performed by using $tRNA^{Thr}$ acylated with respective ThrRS (either from *S. cattleya* or *E. coli*, respectively). A high concentration of tRNA was prepared by IVT (50 – 60 μM). A portion of aminoacyl-tRNA was hydrolyzed for quantification as described above. Reactions for the deacylation timecourse contained $MgCl_2$ (10 mM), $ZnCl_2$ (0.5 mM), BesG (16 μM), HEPES (50 mM, pH 7.5) in nuclease-free water. Prior to beginning the reaction, BesG was exchanged from a 10% (*v/v*) glycerol storage buffer to a glycerol-free storage buffer (50 mM HEPES, 100 mM NaCl) after 2 rounds of concentrating and diluting 10-fold at 4°C with a final round of concentration at 4°C, at which point the concentration of the protein was measured with a Nanodrop spectrometer (MBP-BesG: 12.39 mg/mL ($\varepsilon_{280\ nm}$ = 94,770 $M^{-1}cm^{-1}$). Reactions were initiated by addition of aminoacyl-tRNA (5-8 μM) and immediately incubated at 37°C. Samples (1.5 μL) were removed at 0, 0.5, 1, 2.5, and 4 min and added to a solution (3 μL) of RNaseA (1 U/μL) in ammonium acetate (200 mM). Protein was removed by precipitation with 5% TCA (*v/v*) and centrifugation at 14,000 × *g* for 10 min. Samples were then analyzed by LC/MS-MS on an Agilent 1290 UPLC-6530 QQQ using the protocol for aminoacyl-tRNA analysis with reverse-phase chromatography.

Construction of plasmids for *besG* gene disruption and complementation in *S. cattleya*. Primers for amplifying homology arms for *besG* gene disruption were designed to produce amplicons of either 2 kB homology upstream of start codon and downstream of stop codon of *besG* (oligonucleotides pIJ773-BesG-A/B-Fwd/Rev) (**Table A1.1B**). *S. cattleya* genomic DNA was used as a template for amplification with primer sets pIJ773-BesG-A-Fwd/Rev and pIJ773-BesG-B-Fwd/Rev. Amplified DNA for pIJ773-BesG-A and pIJ773-BesG-B were cloned into pIJ773 sequentially using restriction-ligation with NotI/SpeI and HindIII/KpnI cut sites. Standard

molecular biology techniques were used to carry out plasmid construction using *E. coli* DH10B-T1^R as the cloning host. Following plasmid construction, all cloned inserts were sequenced at Genewiz (San Francisco, CA).

Conjugative transfer of plasmids from *E. coli* to *S. cattleya*. Plasmids pIJ773-BesG for gene disruption was transferred into *S. cattleya* by conjugation with *E. coli* GM272 harboring the *oriT*-mobilizing plasmid pUZ8002 using the literature protocol [*40*]. Gene disruption was verified by PCR amplification of the target locus and neighboring loci using primers SeqConfirm-BesG-KO-Fwd/Rev, followed by sequencing to validate correct disruption, generating *S. cattleya* Δ*besG*::Am^R.

***S. cattleya* growth to produce spent media.** Spores of *S. cattleya* were heat shocked at 55°C for 5 min before inoculation into 50 mL seed cultures of GYM media (yeast extract, 4 g/L; malt extract, 10 g/L; glucose, 4 g/L; presterile pH 7.2) in 250 mL baffled flasks with glass beads. After 14 h of growth, cultures were re-inoculated to OD 0.05 in 200 mL GYM media in 500 mL UltraYield flasks with glass beads. Spent media was isolated from biomass 10 d subsequent to inoculation through centrifugation at 18,000 × *g* at room temperature to pellet the cells and sterile filtration at 0.2 μm to remove spores.

***S. cattleya* growth with spent media.** Spores of *S. cattleya* were heat shocked at 55°C for 5 min before inoculation into 50 mL seed cultures of GYM media (yeast extract, 4 g/L; malt extract, 10 g/L; glucose, 4 g/L; presterile pH 7.2) in 250 mL baffled flasks with glass beads. After 14 h of growth, cultures were re-inoculated to $OD_{600\ nm}$ = 0.05 in a total of 50 mL media containing spent media in 250 mL baffled flasks with glass beads. The conditions included 100% fresh GYM media, 50:50 fresh:spent media,, 25:75 fresh:spent media, and 100% spent media. Biomass was isolated 7 d subsequent to inoculation through centrifugation at 9,800 × *g* for 7 min at 4ºC to pellet the cells.

Derivatization and visualization of alkyne-labeled proteins with Tamra-azide. Cell cultures (1 mL) were collected and centrifuged at 9,800 × *g* for 7 min at 4ºC. The resulting cell pellet was washed once with 1 mL M9 salts and resuspended in SoluLyse (150 μL) and incubated for 1 h at room temperature. Samples were centrifuged at 20,000 × *g* for 10 min and the supernatant (50 μL) was mixed with an assay solution (50 μL) containing Tamra-azide (40 μM), THPTA (100 μM), $CuSO_4$ (50 μM) and sodium ascorbate (10 mM) in water. The labeling reaction was incubated for 1 h at room temperature. Laemmli buffer was added and samples were heated at 95ºC for 5 min before analyzing by SDS-PAGE (MiniProtean 8-16% TGX gels; BioRad) using a constant voltage (150 V) for 1h. Gels were imaged for fluorescence using a ChemiDoc MP Imager (Biorad; λ_{ex} = 520-545 nm, λ_{em} = 577-613 nm, exposure = 150 s).

Proteomics of βes incorporation of ΔBesG Knockout Strain. Cell pellet from 1 mL of cell culture was sent to the UC-Davis Proteomic Facility for analysis. Sample was lysed in 500 μL of RIPA lysis buffer (Pierce) with a MagNA Lyser (Roche) with three 10 s iterations and centrifuged for 10 min at 12,000 × *g* to remove precipitate. Protein was precipitated with Proteoextract reagent (CalBiochem) and incubated for 1 h at -20°C. After centrifuging for 10 min at 12,000 × *g*, supernatant was removed, and this was repeated another two times with cold ethanol washes. Protein concentration was measured by BCA assay (Thermo Scientific, USA). All protein samples were subjected to clean-up / reduction / alkylation / tryptic proteolysis by using suspension-trap (ProtiFi) devices. Proteins were resuspended in 50 μL SDS solubilization buffer consisting of 5% (*w/v*) SDS, 50 mM TEAB, pH 7.55). Disulfide bonds were reduced with dithiothreitol and

alkylated in the dark with iodoacetamide in 50 mM TEAB buffer, pH 8.5. Digestion constituted of a first addition of trypsin 1:100 enzyme: protein (*w/w*) for 4 h at 37 °C, followed by a boost addition of trypsin (1:100 *w/w*) for overnight digestion at 37 °C. To stop the digestion, the reaction mixture was acidified with 1% (*v/v*) trifluoroacetic acid (TFA). The eluted tryptic peptides were dried in a vacuum centrifuge and re-constituted in water with 2% acetonitrile (ACN) and 2 µg was loaded onto a ziptip (Millipore), eluted, and analyzed as described below.

Procedure for high-resolution UHPLC/MS-MS analysis of trypsinized peptides. Liquid chromatography was performed on an ultra-high pressure liquid chromatography system, nanoElute (Bruker Daltonics), at 40 °C and with a constant flow of 400 nL/min on a PepSep 150umx25cm C18 column (PepSep, Denmark) with 1.5 µm particle size (100 Å pores) (Bruker Daltronics), and a ZDV spray emitter (Bruker Daltronics). Mobile phases A and B were water with 0.1% formic acid (*v/v*) and 80/20/0.1% ACN/water/formic acid (*v/v*), respectively. Peptides were separated using a 30 min gradient. Mass spectrometry was performed on a hybrid trapped ion mobility spectrometry-quadrupole time of flight mass spectrometer (timsTOF Pro, (Bruker Daltonics, Bremen, Germany) with a modified nano-electrospray ion source (CaptiveSpray, Bruker Daltonics) operated in PASEF mode under Data Dependent Acquisition (DDA). Desolvated ions entered the vacuum region through the glass capillary and deflected into the TIMS tunnel which is electrically separated into two parts (dual TIMS). Here, the first region is operated as an ion accumulation trap that primarily stores all ions entering the mass spectrometer, while the second part performs trapped ion mobility analysis. A single tims-MS cycle is 0.57 ms and is composed of many individual TOF scans. All experiments were acquired with ramp and accumulation times of 75 ms and 6 PASEF MS/MS scans per topN acquisition cycle. MS and MS/MS spectra were recorded from 100 to 1700 Da. A polygon filter was applied to the *m/z* and ion mobility plane to select features most likely representing peptide precursors rather than singly charged background ions. The quadrupole isolation width was set to 2 Th for *m/z* under 700 and 3 Th for *m/z* larger than 800, and the collision energy was ramped linearly as a function of the mobility from 59 eV at $1/K_0$=1.4 to 20 eV at $1/K_0$=0.6.

Proteomic analysis. Proteomic data were analyzed with Fragpipe (v 21.1) [*41*] with a default offset search using MSFragger (v 3.8) with the following parameters. The database search was performed against the *S. cattleya* proteome (UP000007842) with reversed protein sequences as decoys. Precursor and fragment mass tolerance was set to 10 ppm. A variable modification of Thr+9.9844 (βes) was specified with 2 maximum occurrences per peptide, and the mass offset of 9.9844 and 12 was localized to T (Thr). Additional allowed variable modifications were oxidation of methionine (+15.995 Da) and N-terminal acetylation (+42.01 Da) with a fixed carbamidomethyl modification on cysteine (+57.0214 Da). The delta mass exclude range was constrained to -0.5 to 0.5 and the mass shift was localized to T (Thr). MSBooster was enabled, contaminants were removed, and both peptide- and protein-level summaries were produced. MS1 Quantification was performed with IonQuant (v 1.9.8), and labeling was specified as T0 (light) and T9.9844, T12 (heavy). MS/MS spectra were generated using FragPipe-PDV viewer and prepared for publication with Microsoft Illustrator.

***In vitro* screening of non-heme iron enzymes AlfC and FuzA.** Frozen enzyme aliquots were thawed on ice and exchanged to an ammonium bicarbonate buffer (50 mM Ammonium Bicarbonate, 100 mM NaCl, pH 7.5) using BioSpin6® columns (BioRad). Reactions (25 µL) contained all 20 canonical amino acids (L-alanine, L-arginine, L-asparagine, L-aspartate, L-cysteine, L-glutamine, L-glutamate, L-glycine, L-histidine, L-isoleucine, L-leucine, L-lysine, L-

methionine, L-phenylalanine, L-proline, L-serine, L-threonine, L-tryptophan, L-tyrosine, L-valine) (0.5 mM each), as well as sodium αKG (5 mM), sodium ascorbate (5 mM), $(NH_4)_2Fe(SO_4)_2 \cdot 6$ H2O (1 mM) and sodium chloride (100 mM) in 50 mM ammonium bicarbonate buffer (pH 7.5). Reactions were initiated by addition of non-heme iron enzyme (15 μM final concentration) and allowed to proceed at room temperature before quenching 10 μL at 2 h and 18 h in 1 volume of methanol with 1% (*v/v*) formic acid. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C to pellet protein, and the supernatant was analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis. Following identification of substrates from the amino acid pool, the reactions were repeated with individual amino acids for confirmation.

***In vitro* scale-up of non-heme enzymatic reactions for NMR characterization.** Frozen enzyme aliquots were thawed on ice and exchanged to an ammonium bicarbonate buffer (50 mM ammonium bicarbonate, 100 mM NaCl, pH 7.5) using BioSpin6® columns (BioRad). Reactions (500 μL) contained amino acid substrate (0.5-4 mM), sodium αKG (10 mM), sodium ascorbate (5 mM), $(NH_4)_2Fe(SO_4)_2 \cdot 6$ H_2O (1 mM) and sodium chloride (100 mM) in 50 mM ammonium bicarbonate buffer (pH 7.5). Reactions were initiated by addition of non-heme iron enzyme (30 μM final concentration) and allowed to proceed at room temperature for 18 h before quenching in 2 volumes of methanol with 1% (*v/v*) formic acid or passage through a Pall Nanosep 10-kD spin column. Precipitated protein was removed by centrifuging at 20,000 × *g* for 10 min at 4 °C and the supernatant was collected for further purification as described.

Ion exchange (IEX) purification of (2*S*,3*R*)-hydroxyarginine. A HiTrap Q HP 5 mL IEX column (Cytiva) was used for separation of 3-hydroxy-arginine from the negatively charged cofactors. Following removal of protein components from the reaction by passage through a Pall Nanosep 10-kD spin column, the column was flushed by 25 mL of 1 M NH_4HCO_2 followed by 25 mL of MilliQ H_2O. Sample (500 mL) was added and eluted with 10 mL MilliQ H_2O. The eluent was dried using N_2 and resuspended in D_2O for NMR analysis.

HPLC purification of (2*S*,3*R*)-hydroxyarginine, (2*S*,3*R*)-hydroxyarginine, and fuscazine A. Samples (1.5mL) were purified with an Agilent 1290 UPLC using a SeQuant ZIC-HILIC column (5 μm, 200 Å, 150 × 21.2 mm; Merck) using the following buffers: buffer A (90% water, 10% acetonitrile and 10 mM ammonium formate) and buffer B (90% acetonitrile, 10% water and 10 mM ammonium formate). After 10 min at 100% buffer B, a linear gradient from 0 to 100% buffer A was applied over 80 min followed by 10 min at 100% buffer A at a flowrate of 5 mL/min. Fractions were collected in 2 mL fractions and checked by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the rapid protocol for polar metabolite analysis. Fractions containing the product were collected, pooled, dried to a solid with N_2, and resuspended in D_2O for NMR analysis.

tRNA competitive acylation with *Af* MetRS and AlfB. (2*S*,3*R*)-hydroxymethionine was generated biocatalytically in a 50 μL reaction containing sodium αKG (5 mM), sodium ascorbate (5 mM), $(NH_4)_2Fe(SO_4)_2 \cdot 6$ H2O (1 mM), sodium chloride (100 mM), methionine (5 mM) in 50 mM HEPES buffer (pH 7.5). Reaction was initiated with AlfC (25 μM) and allowed to proceed 4 h at 37 °C to generate (2*S*,3*R*)-hydroxymethionine for subsequent reactions without purification. tRNA was prepared by *in vitro* transcription and purified using miRNeasy purification kit (QIAGEN) following manufacturer protocols. Competitive acylation/deacylation reactions contained HEPES buffer (pH 7.5) (40 mM), $MgCl_2$ (15 mM), ATP (5 mM), methionine (2.5 mM) or 3-hydroxymethionine (*in situ*), *Af* MetRS (5 μM), *Af* MetX (0, 5, and 25 μM), and purified tRNA (4 μL). Prior to beginning the reaction, tRNA was first refolded according to a literature

protocol [*39*]. Briefly, HEPES buffer and water were added to the transcription reaction, the sample was heated to 98°C on a thermocycler for 2 min, and slowly cooled to room temperature over 5 min. $MgCl_2$ was added, and the sample was incubated at room temperature for 5 min, followed by 3 min on ice. The rest of the components were added and initiated by addition of enzyme. Samples (1.5 µL) were removed at 0 and 45 min and added to a solution (3.5 µL) of RNaseA (1 U/µL) in ammonium acetate (200 mM). Protein was removed by precipitation with 5% TCA (*v/v*) followed by centrifugation at 14,000 × *g* for 10 min. Samples were then analyzed by LC/MS-MS on an Agilent 1290 UPLC-6530 QQQ using the protocol for aminoacyl-tRNA analysis with reverse-phase chromatography.

***In vitro* reconstitution of FuzA activity.** The reaction (25 µL) contained L-arginine (0.5mM), sodium αKG (5 mM), sodium ascorbate (5 mM), $(NH_4)_2Fe(SO_4)_2 \cdot 6\ H_2O$ (1 mM) and sodium chloride (100 mM) in 50 mM HEPES buffer (pH 7.5). The reaction was initiated by addition of FuzA (15 µM final concentration) and allowed to proceed at room temperature for 18 h before quenching in 1 volume of acetonitrile. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C to pellet protein and 20 µL of supernatant was derivatized with 5 µL Fmoc-Cl (10 mM stock solution in DCM) and 6.25 µL Borate buffer (200 mM sodium tetraborate decahydrate, pH 8.2) for 15 min at room temperature to reach a final pH of 8 before quenching in 1 volume of methanol. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C to remove any precipitate, and the supernatant was analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for nonpolar metabolite analysis.

***In vitro* reactions with ATP-grasp enzymes FuzC and FuzH.** Reactions (25 µL) contained ATP (5 mM), $MgCl_2$ (15 mM), NaCl (100 mM) in 100 mM HEPES buffer (pH 7.5). In addition, a selection from or all of the 20 canonical amino acids (L-alanine, L-arginine, L-asparagine, L-aspartate, L-cysteine, L-glutamine, L-glutamate, L-glycine, L-histidine, L-isoleucine, L-leucine, L-lysine, L-methionine, L-phenylalanine, L-proline, L-serine, L-threonine, L-tryptophan, L-tyrosine, L-valine) were included along with L-pipecolic acid (0.5 mM each) when FuzH is present. Reactions were initiated by addition of FuzC and/or FuzH (25 µM final concentration) and allowed to proceed at room temperature before quenching at 18 h in either 1 volume of methanol with 1% (*v/v*) formic acid for polar metabolite analysis or 1 volume acetonitrile for nonpolar metabolite analysis. Samples for polar metabolite analysis were centrifuged at 20,000 × *g* for 10 min at 4 °C to pellet protein, and the supernatant was analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis. Samples for nonpolar metabolite analysis were centrifuged at 20,000 × *g* for 10 min at 4 °C to pellet protein, and 20 µL of supernatant was derivatized with 5 µL Fmoc-Cl (10 mM stock solution in DCM) and 6.25 µL Borate buffer (200 mM sodium tetraborate decahydrate, pH 8.2) for 15 min at room temperature to reach a final pH of 8 before quenching in 1 volume of methanol. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C to remove excess Fmoc, and the supernatant was analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for nonpolar metabolite analysis.

***In vitro* reconstitution of lysine N^6-monooxygenase FuzD.** Reactions (25 µL) contained either L-lysine or L-ornithine (0.5 mM), NADPH (2 mM), FAD (200 µM), NaCl (100 mM) in 50 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of enzyme (15 µM final concentration) and allowed to proceed at room temperature before quenching at 2 h and 18 h in 1 volume of acetonitrile. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C to pellet protein and 20 µL of supernatant was derivatized with 5 µL Fmoc-Cl (10 mM stock solution in DCM) and 6.25 µL Borate buffer (200 mM sodium tetraborate decahydrate, pH 8.2) for 15 min at room

temperature to reach a final pH of 8 before quenching in 1 volume of methanol. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C to remove excess Fmoc, and the supernatant was analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for nonpolar metabolite analysis.

***In vitro* reaction for formation of hydrazine-containing dipeptide.** Reactions (25 μL) contained L-lysine (0.5 mM), L-arginine (0.5mM), NADPH (2 mM), FAD (200 μM), ATP (5 mM), $MgCl_2$ (10 mM), $ZnCl_2$ (500 μM), NaCl (100 mM) in 50 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of enzyme (10 μM final concentration) and allowed to proceed at room temperature before quenching at 2 h in 1 volume of acetonitrile. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C to pellet protein and 20 μL of supernatant was derivatized with 5 μL Fmoc-Cl (10 mM stock solution in DCM) and 6.25 μL Borate buffer (200 mM sodium tetraborate decahydrate, pH 8.2) for 15 min at room temperature to reach a final pH of 8 before quenching in 1 volume of methanol. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C to remove excess Fmoc, and the supernatant was analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for nonpolar metabolite analysis.

M9 minimal media for metabolomics. M9 minimal media made according to the University of Wisconsin-Madison's *E. coli* Genome Project [*42*] was utilized for metabolomics experiments, containing Na_2HPO_4 (47.7 mM), KH_2PO_4 (22.0 mM), NaCl (8.55 mM), NH_4Cl (18.7 mM), glucose (0.4% *w/v*), $MgSO_4$ (2 mM), $CaCl_2$ (100 μM), thiamine (1 mg/mL), and an amino acid mixture diluted from a 5X solution (**Table A2.1**). For isotopically labelled amino acid experiments, an equivalent amount of the labelled version was supplied exogenously to the complete medium.

Metabolomics for identifying nitrogen-nitrogen-containing natural product produced by Fuz Cluster. Single colonies of *E. coli* BL21 Star (DE3) transformed with plasmid were inoculated in triplicate per condition into LB (5 mL) in a 24-well plate containing the appropriate antibiotics (50 μg/mL) and grown overnight at 37°C with shaking at 200 rpm. Overnight cultures were inoculated into 5 mL of LB with appropriate antibiotics (50 μg/mL) in a 24-well plate to $OD_{600\ nm}$ = 0.05 and grown at 37 °C and 200 rpm until the $OD_{600\ nm}$ reached 0.4-0.6. The plate was cooled on ice for 20 min before spinning down at 4,000 × *g* for 10 min at 4 °C. Media was removed, and cell pellets were resuspended in M9 minimal media for metabolomics. Additional components were added, including appropriate antibiotics (50 μg/mL), IPTG (0.25 mM), and isotopically labelled amino acids $^{13}C_6$,$^{15}N_4$-L-arginine, $^{13}C_6$,$^{15}N_2$-L-lysine, or $^{13}C_3$,^{15}N-L-alanine (equal concentration to amount supplied in minimal media). After the plate was cultured for 48 h at 30°C and 200 rpm, 1 mL aliquots were taken in duplicate from each condition and centrifuged at 10,000 × *g* for 5 min at RT. After decanting the supernatant, 1 set of pellets was resuspended in 50 μL of methanol with 1% (*v/v*) formic acid, lysed by vortexing for 10 s, and centrifuged at 20,000 × *g* at 4°C for 10 min. Supernatant was analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis. 1 set of pellets was resuspended in 50 μL of acetonitrile, lysed by vortexing for 10 s, and centrifuged at 20,000 × *g* at 4°C for 10 min. 40 μL of supernatant was derivatized with 10 μL Fmoc-Cl (10 mM stock solution in DCM) and 12.5 μL Borate buffer (200 mM sodium tetraborate decahydrate, pH 8.2) for 15 min at room temperature to reach a final pH of 8 before quenching in 1 volume of methanol. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C to remove excess Fmoc, and the supernatant was analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for nonpolar metabolite analysis.

Scale-up fermentation and extraction for isolation of fuscazine A. Single colonies of *E. coli* BL21 Star (DE3) transformed with plasmids pETDuet-TauD-LysOH, pACYCDuet-Cupin-FADR-HydS, and pRSFDuet-ATPGA-ATPGB were inoculated in 200 mL TB containing appropriate antibiotics (50 μg/mL) and grown overnight at 37°C with shaking at 200 rpm. Overnight cultures were inoculated into 6 × 1 L cultures of TB with 50 μg/mL antibiotic in 2.8 L-baffled shake flasks to $OD_{600\ nm}$ = 0.05 and grown at 37 °C and 200 rpm until the $OD_{600\ nm}$ reached 0.6 - 0.8. Cultures were cooled on ice for 20 min before inducing with IPTG (0.25 mM) and then cultured for 72 h at 30°C and 200 rpm. Cells were pelleted and stored at -80 °C for storage until extraction. Cell pellets were thawed in water at RT, and resuspended with 20 mL ice-cold 50:50 MeOH:H_2O. Cell pellets were lysed by vortexing and cell pellet was pelleted at 13,000 × *g* for 10 min at 4 °C. The liquid fraction was subjected to one freeze-thaw cycle, pelleted again at 13,000 × *g* for 10 min at 4°C, and the remaining liquid fraction was dried by an air stream. This process was repeated for each cell pellet with 100% ice-cold MeOH. Dried solid was resuspended in 2 mL 100% MeOH. These resuspensions were pelleted at 13,000 × *g* for 10 min at 4 °C, and the liquid fractions were removed and pooled in sets of 3. These were dried down using N_2 to final volumes of 1.5 mL for HPLC purification.

NMR spectroscopy. 2D experiments were recorded on a Bruker Avance II spectrometer operating at 900 MHz and at 298 K, while 1D experiments, as noted, were recorded on a Bruker Avance III spectrometer operating at 600 MHz and at 298 K or a Bruker Avance I spectrometer operating at 700 MHz and at 298K. The 900 MHz instrument was equipped with a CP TXI cryoprobe, the 600 MHz instrument was equipped with a 5 mm 1H/BB Prodigy CryoProbe, and the 700 MHz instrument was equipped with a 5 mm triple resonance $^{1}H/^{13}C/^{15}N$ TXI cryo-probe. NMR solvents, chloroform-*d* ($CDCl_3$; Cambridge Isotope Laboratories) and D_2O (99.9%; Sigma-Aldrich) were used without further purification. All instruments were controlled using Topspin (v 3.2) software. Data were processed using MestReNova 14.2.1 using automatic phasing and Bernstein Polynomial Fit baseline correction capabilities. Splitting was determined using the automatic multiplet analysis function with intervention as necessary. Spectral data are reported as follows: chemical shift (multiplicity [singlet (s), broad singlet (bs), doublet (d), triplet (t), quartet (q), pentet (p), multiplet (m), doublet of doublets (dd), doublet of doublet of doublets (ddd), doublet of triplet of doublets (dtd), doublet of doublet of doublet of doublets (dddd), doublet of triplets (dt), triplet of doublets (td), complex (comp), etc.], coupling constant (Hz), integration). Chemical shifts are reported in ppm (δ), and coupling constants are reported in Hz. Resonances are referenced to solvent residual peaks for $CDCl_3$ (^{1}H = 7.26 ppm, ^{13}C = 77.2 ppm) or D_2O (^{1}H = 4.79 ppm).

Regarding 2D experiments, a constant-time ^{1}H-^{13}C HSQC experiment (Bruker pulse sequence hsqcctetgpsp) was recorded with carrier frequencies set to 5.48 ppm (^{1}H) and 40 ppm (^{13}C)) and the spectral widths set to 16 ppm (^{1}H) and 80 ppm (^{13}C). The carboxyl chemical shift was set to 176 ppm, and the ^{1}H-^{13}C coupling constant was set to 145 Hz. The recycle delay was set to 1.5 sec. The constant-time evolution period was set to 13.6 ms, to distinguish ^{1}H-^{13}C pairs with one or three carbon neighbors from those with two carbon neighbors. A 2D ^{1}H-^{13}C version of the 3D HCCH-COSY experiment (Bruker pulse sequence hcchcogp3d2) was used to identify neighboring ^{1}H and ^{13}C resonances. The carrier frequencies were set to 5.48 ppm (^{1}H) and 39 ppm (^{13}C) with spectral widths of 14 ppm (^{1}H) and 75 ppm (^{13}C). The recycle delay was set to 1.8 sec. A ^{1}H-^{13}C HMBC experiment (Bruker pulse sequence hmbcgplpndqf) was used to correlate protons with carbon spins two and three bonds removed via long-range proton-carbon couplings. The experiment was optimized for a long-range proton-carbon coupling of 10 Hz. Carrier frequencies

were set to 5 ppm (^{1}H) and 50 ppm (^{13}C), with spectral widths of 11 ppm (^{1}H) and 100 ppm (^{13}C). The recycle delay was set to 1.5 s.

Bioinformatic method for tRNA-Deacylase Directed Discovery (tR3D). For generating the sequence similarity network (SSN) of the AlaX family of tRNA-Deacylases, we utilized the EFI-EST tool for generating a SSN of IPR018163 (tRNA-editing domain) as a Uniref50 dataset containing 128,218 proteins (Sept. 2022) (including fragments) [*43*]. Uniref50 is a representative dataset where each node in the SSN represents proteins that are ≥50% sequence-similar to a representative protein. After generating the initial dataset, we set the sequence alignment score to an arbitrary value of 35%, which was equivalent to a sequence similarity of 30-40%. A maximum sequence length was specified as 480 amino acids to exclude any Alanine-tRNA Synthetases that have tRNA-editing domains, resulting in 23,260 proteins. The SSN was visualized with Cytoscape, and the characterized AlaXs were mapped onto the SSN. We found that the majority of these proteins were in the same cluster, so the sequence alignment score was increased until it reached 43%, which was the highest value before most characterized AlaXs split from each other. This SSN was submitted to the EFI-GNT tool to generate neighborhood data files. The neighborhood distance was set to 10 genes, and the co-occurrence cut-off to 0% to maximize the amount of genomic information available. After completion, we downloaded the Color SSN, SSN Cluster GNN Hub-Node file, and the Genomic Neighborhood Diagrams (GNDs) file.

We used the Color SSN for the next round of analysis. This SSN contained 23,260 proteins represented as 3,676 nodes. To clean the data for analysis, we removed entries with poor neighborhood information resulting from short reads by only keeping nodes with >6 neighbors. Additionally, we removed nodes where every represented sequence is a fragment. Empirically, these sequences often had reads starting in the middle of the editing domain AlaX, and likely belong to a full AlaRSs. This resulted in 1,879 nodes (51% of original) and 20,123 AlaX proteins (87% of original).

For the next stage of the analysis, we uploaded the SSN Cluster GNN Hub-Node file to Cytoscape and downloaded the table as a CSV. This is the input file for the tR3D data visualization tool. Using tR3D, we identified the most common protein families to co-occur with the deacylases for each cluster of the SSN. We also used the EFI-GND tool for viewing neighborhood data directly. From this analysis, we generated a list of protein families that participate in primary metabolism, filtered for all AlaX proteins that co-occur with these genes, and removed them from further analysis. This yielded a much smaller SSN with 844 nodes (23% of original) and 2,675 genomes (11.5% of original) that likely contain mostly secondary metabolism-related BGCs. This SSN was resubmitted to the EFI-GNT tool using the same specifications as above.

The resulting SSN was analyzed for traditional secondary metabolism genes, including NRPSs, PKSs, and RiPPs. The SSN Cluster Hub-Node file was visualized with the tR3D tool and clusters were explored for BGC-like character.

Deploying the tR3D Visualization Tool. The tR3D visualization tool leverages data provided from the EFI-GNT tool. This data can be acquired through the Cytoscape platform. The raw data was attained by uploading the SSN Cluster GNN Hub-Node file into Cytoscape and then downloading the corresponding CSV file. The resulting table includes features such as Protein Family Identification, Number of Occurrences, Median Distance from the Query Gene, and Related Cluster. This file was uploaded to the tR3D app, which can be accessed at https://github.com/shanenak/tR3D-app and is deployed via Streamlit with VS code.

Bioinformatic analysis of hydrazine synthetases. For generating the sequence similarity network (SSN) of all hydrazine synthetases (HydSs), we utilized the EFI-EST tool for generating a SSN of PF09334 (Methionine-tRNA synthetase (MetRS) family) as a Uniref50 dataset containing 84,733 proteins (Oct. 2022) (including fragments) [*43*]. Uniref50 is a representative dataset where each node in the SSN represents proteins that are ≥50% sequence-similar to a representative protein. After generating the initial dataset, we set the sequence alignment score to an arbitrary value of 55%, which was equivalent to a sequence similarity of 30-40%. The SSN was visualized with Cytoscape, and the characterized HydSs were mapped onto the SSN. We found that all of these proteins were co-located in a very large cluster, so the sequence alignment score was increased until it reached 87% to separate HydSs from canonical MetRSs without splitting HydSs from each other. The resulting SSN was submitted to the EFI-GNT tool to generate neighborhood data files. The neighborhood distance was set to 10 genes, and the co-occurrence cut-off to 0% to maximize the amount of genomic information available. After completion, we downloaded the Color SSN, SSN Cluster Hub-Node file, and the Genomic Neighborhood Diagrams (GNDs) file.

Computational modeling and docking. These experiments were performed at the Molecular Graphics and Computation Facility (Berkeley, CA). To postulate the alkyne binding site in BesG, a model of the enzyme was built using AlphaFold2 [*44*]. This model was then aligned with the "super" command in PyMOL with the crystal structure of the *Pyrococcus horikoshii* AlaX-S (PDB: 1WNU) and AlaX-M (PDB: 2E1B), with RMSD values of 2.265 Å and 1.85 Å respectively. Based on the alignment, a putative amino acid binding site was identified, and βes was rigidly docked into the structure using GLIDE via Maestro 2022-2, with Ser in AlaX-S as a starting position. Viable ligand poses were selected based on comparison to the existing Ser-bound AlaX-S.

To better understand the structural basis behind the substrate preference of FuzG, a model of the enzyme was built using AlphaFold2 [*44*]. This model was then aligned using the "super" command in PyMOL with the crystal structure of the *Escherichia coli* methionyl-tRNA synthetase complexed with methionine (PDB: 1F4L), as well as the structure of the *Aquifex aeolicus* methionyl-tRNA synthetase complexed with a methionyl-adenylate analogue and tRNA(Met) (PDB: 2CT8), with RMSD values of 2.569 and 2.551 Å respectively. Based on the alignment, a putative amino acid binding site was identified, and (2*S*,3*R*)-hyroxyargininyl-adenylate was rigidly docked into the structure using GLIDE via Maestro 2022-2. Viable ligand poses were selected based on comparison to existing substrate-bound methionyl-tRNA synthetase structures.

2.3 Results and Discussion

Discovery of the βes-specific deacylase BesG. In previous work, we elucidated the biosynthesis of the terminal alkyne amino acids propargylglycine (Pra, **2**) and β-ethynylserine (βes, **3**) (**Fig. 2.2A**) [*6*]. Beginning with the canonical amino acid lysine (**1**), three biosynthetic steps yield Pra and an additional three enzymatic steps produce the hydroxylated version, βes. This pathway includes an array of unique enzymatic transformations, including halogenation, alkene formation, peptide ligation, and hydroxylation (**Fig. 2.3**) [*45,46*]. Following the pathway elucidation in *S. cattleya*, comparative genomics revealed a subset of unique proteins between organisms that only contained the necessary genes for Pra compared to the full BGC for βes [*6*]. We found in several of the βes-BGC-containing organisms a neighboring tRD belonging to the AlaX family (**Fig. 2.2A, Fig. 2.4**). Interestingly, a homologous AlaX can be found in *S. cattleya* but ~80 kb from βes biosynthesis. We annotated this enzyme "BesG" and hypothesized it serves as a resistance marker against βes self-incorporation, analogous to FthB's role in preventing self-incorporation of F-Thr [*29*]. The BGCs for Pra do not contain a tRNA-deacylase, indicating Pra does not naturally incorporate into proteins but instead requires an engineered aaRS [*47*]. With the structural similarity of βes to the proteinogenic amino acid threonine (Thr) along with the presence of the AlaX-like deacylase BesG, we hypothesized βes could misincorporate for Thr.

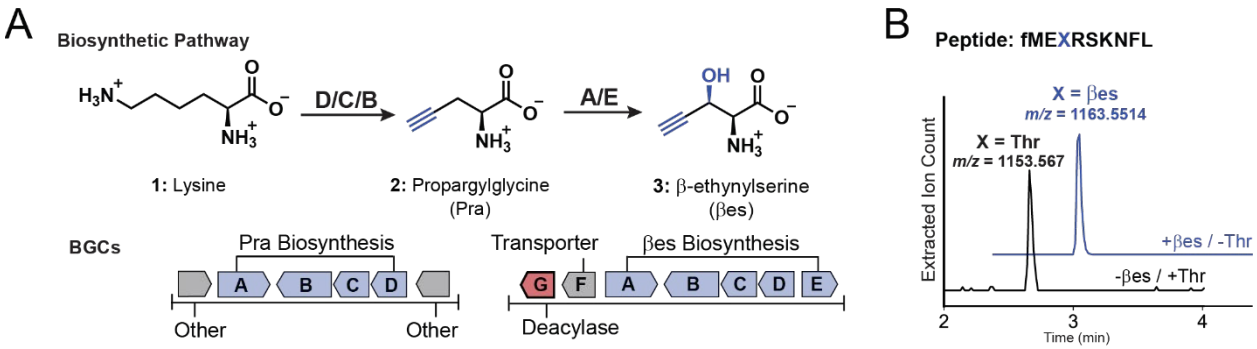
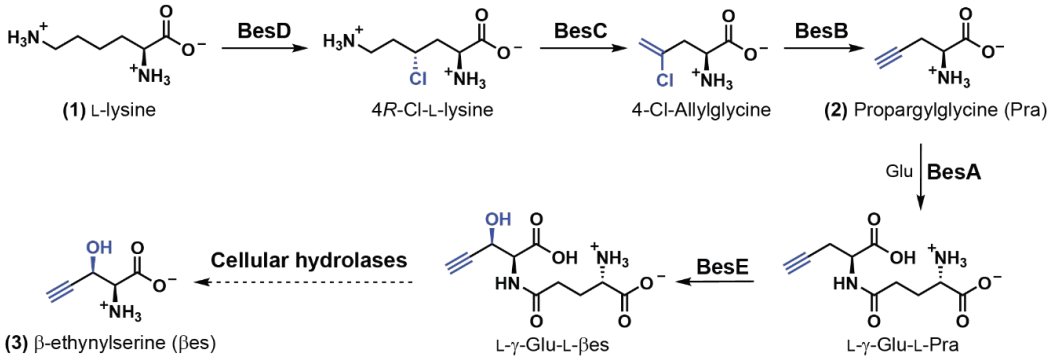


Figure 2.2. βes is a mimic of threonine. (A) Comparing the biosynthetic pathways for terminal alkyne amino acids Pra (**2**) and βes (**3**) reveal a deacylase in the context of βes biosynthesis, which we hypothesized to be a resistance mechanism against misincorporation (B) *In vitro* transcription/translation indicates βes can misincorporate for Thr, suggesting BesG may be necessary to prevent βes mistranslation.

To evaluate if βes is a Thr mimic, we performed *in vitro* transcription/translation reactions to determine whether *S. cattleya*'s threonine-tRNA synthetase (ThrRS) would mischarge $tRNA^{Thr}$ with βes (**Fig. 2.2B**). We designed a short peptide with a single Thr site, and when only the necessary canonical amino acids were supplied, we detected the mass correlating to Thr incorporation. As βes was not commercially available, a synthetic route was developed (**Fig. 2.5**). When Thr was replaced with βes in the reaction conditions, we detected the expected mass shift for βes incorporation. This data supports βes incorporation at Thr sites via the inability of the threonine-tRNA synthetase to effectively discriminate against this amino acid. This conclusion is further supported by the recent report of βes serving as a clickable probe for proteomics studies [*48*].

With the inability of the ThrRS to effectively discriminate between Thr and βes, we explored whether BesG specifically removes βes from mischarged $tRNA^{Thr}$. We expressed and purified both the *E. coli* and *S. cattleya* ThrRSs with N-terminal His_{10} tags from *E. coli*, and the *S. cattleya* BesG

A

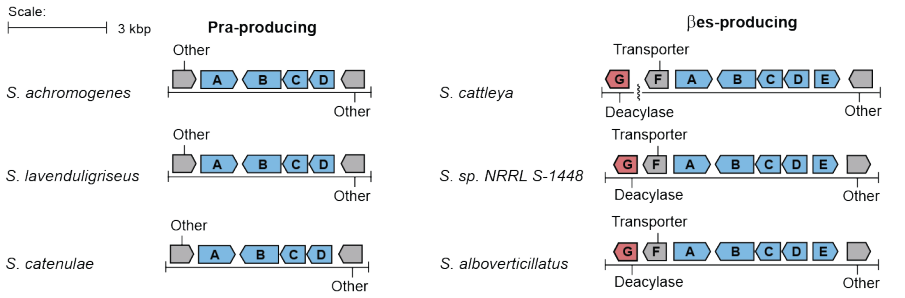


B

Enzyme Name	Uniprot Accession (*S. cattleya*)	Function	Enzyme Class
BesA	G8XHD8	L-γ-Glu-L-Pra ligase	ATP-grasp
BesB	G8XHD7	4-Cl-allylglycine desaturase	PLP-dependent transferase
BesC	F8JJ25	4-Cl-allylglycine synthase	Haem-oxygenase
BesD	G8XHD5	L-lysine 4-chlorinase	Non-heme iron halogenase
BesE	F8JJ27	L-γ-Glu-L-Pra hydroxylase	Non-heme iron hydroxylase
BesF	F8JJ22	Probable βes exporter	EamA transporter
BesG	F8JJ98	βes-tRNA hydrolase	AlaX tRNA-Deacylase

Figure 2.3. The biosynthetic pathway for β-ethynylserine (βes). (A) The biosynthesis of βes requires 5 enzymatic steps encoded in a biosynthetic gene cluster (BGC) to produce a Glu-βes dipeptide with promiscuous cellular transglutaminases yielding free βes [*6*]. (B) Enzymes from the elucidated pathway in *Streptomyces cattleya* are annotated by name, Uniprot ID, function, and enzyme family.

A



B

Enzyme Name	Organism	NCBI Accession	% ID/%S
BesG	*Streptomyces cattleya*	AEW98792.1	-
BesG	*Streptomyces sp. NRRL S-1448*	WP_051822990.1	60/70
BesG	*Streptomyces alboverticillatus*	WP_190062247.1	60/67

Figure 2.4. Identification of a putative ncAA tRNA deacylase BesG. (A) An enzyme with homology to AlaX tRNA-deacylases, BesG, is unique to βes-producing organisms (B) *S. cattleya* BesG shows high sequence similarity compared to BesG homologs that co-localize with the βes BGC in other organisms.

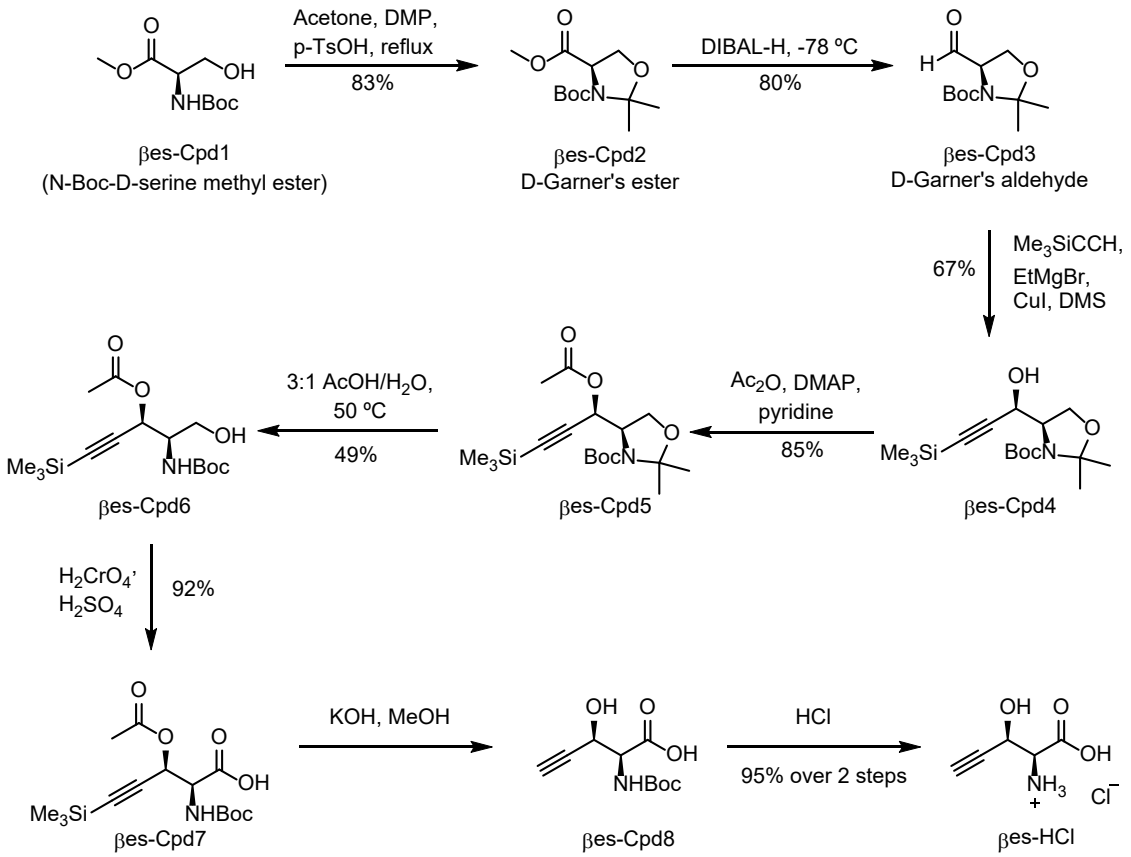


Figure 2.5. Chemical synthesis of β-ethynylserine (βes). Commercially available N-Boc-D-serine methyl ester was converted to βes-Cpd4 by known methods [*33*–*37*]. Next, the secondary alcohol was protected as acetyl ester βes-Cpd5. Conditions for N,O-acetonide deprotection were screened, and the optimal condition of heating βes-Cpd5 in 3:1 acetic acid/water at 50 ºC was used to generate βes-Cpd6. Our choices of the acetyl ester and deprotection conditions were crucial for minimizing ester migration to the primary alcohol, and for leaving the Boc group intact, providing a new approach to these two challenges previously observed by Chan and O'Hagan [*33*]. Finally, Jones oxidation followed by basic and acidic deprotections according to [*33*] gave β-ethynylserine hydrochloride, in 5 steps and 34% yield from the literature intermediate βes-Cpd4.

with an N-terminal His_6-MBP tag from *E. coli* (**Fig. A3.3**). In addition, we synthesized *E. coli* and *S. cattleya* $tRNA^{Thr}$ *in vitro*. With these components, we performed acylation reactions with combinations of *E. coli* or *S. cattleya* $tRNA^{Thr}$ with βes or Thr to prepare substrates for BesG, and subsequently performed a time course deacylation experiment of BesG with βes-$tRNA^{Thr}$ or Thr-$tRNA^{Thr}$ using LC-MS/MS (**Fig. 2.6A**). Our results show βes is preferentially deacylated compared to Thr, indicating BesG can discriminate between these two amino acids and specifically edit βes-$tRNA^{Thr}$. In addition, we found BesG was less active towards *E. coli* $tRNA^{Thr}$ than *S. cattleya* $tRNA^{Thr}$. We hypothesized the basis for this preference was slight differences in the tRNA acceptor stem [*49*]. Specifically, there is a uracil at the 4th nucleotide from the 3′ end of Streptomyces $tRNA^{Thr}$ compared to an adenine in *E. coli* $tRNA^{Thr}$. This nucleotide is important for ThrRS aminoacylation and is termed the discriminator base [*50*]. When this base was changed to uracil in *E. coli* $tRNA^{Thr}$, this substrate was accepted as well as the native S. cattleya $tRNA^{Thr}$ by BesG, despite the overall sequence similarity being 67%, indicating this is an important tRNA binding determinant for BesG activity (**Fig. 2.7**).

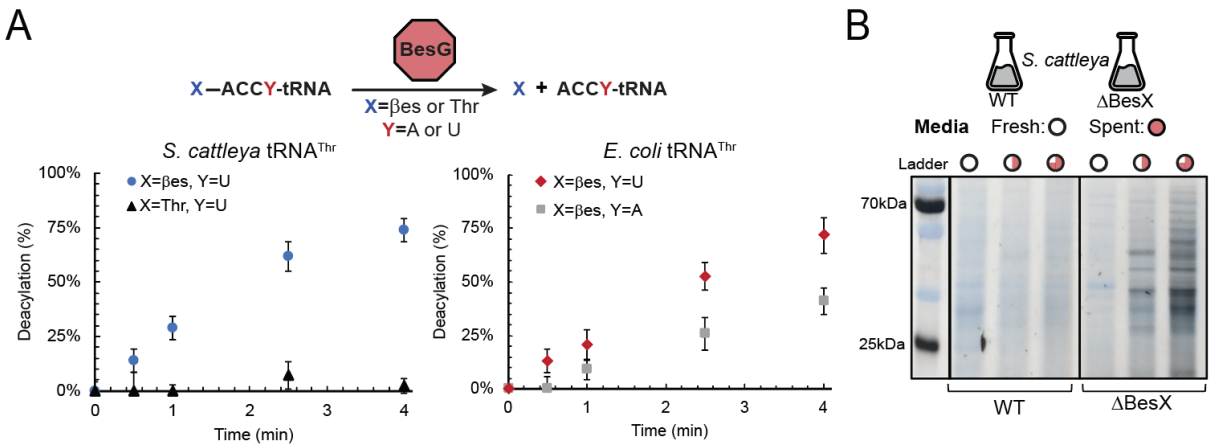


Figure 2.6. BesG shows specificity for βes-tRNA^{Thr}. (A) BesG shows preference for βes-tRNA^{Thr} over Thr-tRNA^{Thr} in an *in vitro* reaction where % deacylation is measured over time. In addition, BesG shows a preference for the tRNA acceptor stem of *S. cattleya* (Y=U) over *E. coli* (Y=A) as indicated by a higher level of deacylation over time. (D) We performed a genomic knockout of BesG in the native producer *S. cattleya* supplied with various levels of βes via spent media (0%, 50%, 75% spent media compared to fresh media). Analyzing the proteome via CuAAC-based fluorescence tagging of terminal alkynes shows higher incorporation of βes into the proteome compared to the WT strain, suggesting BesG can effectively edit misacylated βes-tRNA^{Thr} *in vivo*.

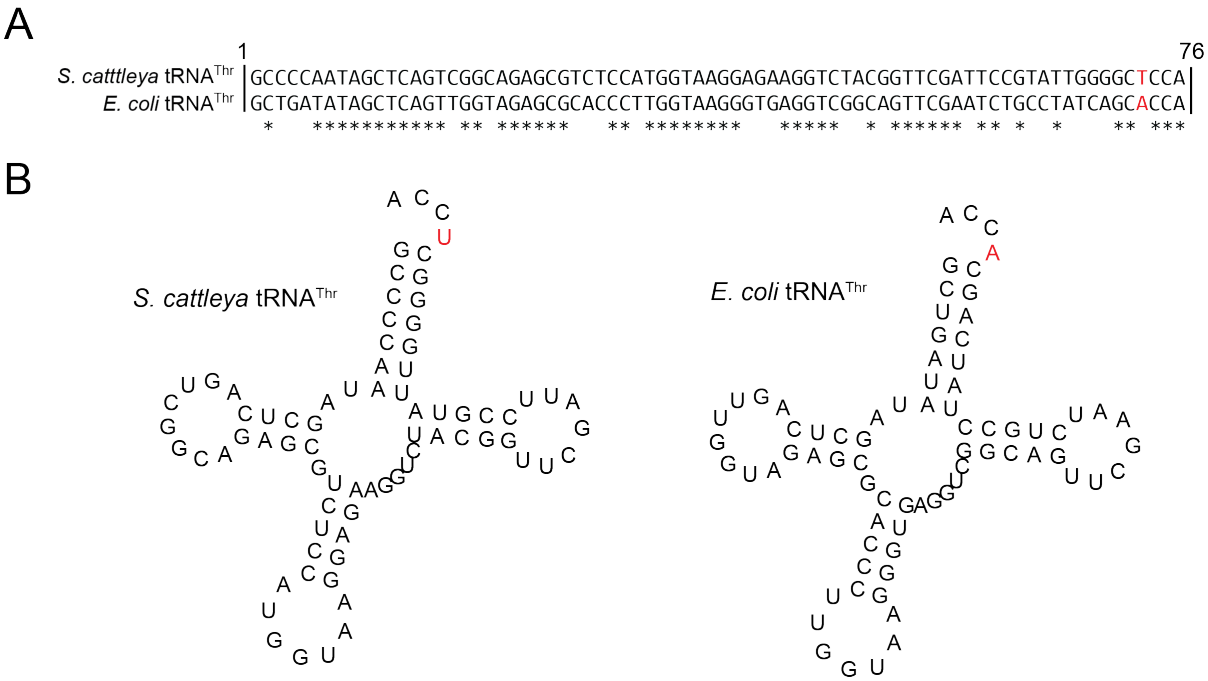


Figure 2.7. Comparison of tRNA^{Thr}s from *S. cattleya* and *E. coli* provides insight into basis for BesG's tRNA preference. (A) Threonine-encoding tRNAs (ACC codon) from *S. cattleya* and *E. coli* are 67% identical (51/76 bases) [*51*], with one notable difference being the 4th base from the 5′-CCA-3′ end termed the discriminator base (shown in red) [*50*]. From experimental data, when this base in *E. coli* tRNA^{Thr} was changed to T/U, this tRNA was a better substrate for BesG. (B) Secondary structure of *E. coli* tRNA^{Thr} and *S. cattleya* tRNA^{Thr} shows the discriminator base is not paired, as predicted by the genomic tRNA database and NAVIEW [*52,53*].

Table 2.1. Proteomic analysis of BesG knockout strain shows βes incorporation at Thr sites. Tryptic peptides and corresponding proteins detected by MSFragger to contain at least one βes incorporation at a Thr site in the BesG knockout strain through detection of the addition 10 Da of the alkyne at Thr sites.

Protein ID	Protein	Peptide Sequence
F8JUX7	3-demethylubiquinone-9 3-methyltransferase	YGEAGPGTPGTVMTVVFELAGQR
G8WUC8	3-hydroxy-acyl-CoA dehydrogenase	AATVPATGLAELAGAGLVIEAIAER
G8X087	AAA+ ATPase domain-containing protein	LGPLGTALLPAR
G8WUP1	ABC transporter protein	LSDTIMQNDVPAVPFAYDK
F8JZM5	Acetamidase/Formamidase	DYGLSELDAYQLVTQAVESPLANVCDANYTVVAK
G8X0C6	Adenosylhomocysteinase	DGTLLFPAINVNDSVTK
F8K2J3	Aminoglycoside phosphotransferase	TTNPSAELLDNLLTMAGQTTAVR
F8K2J3	Aminoglycoside phosphotransferase	VTFPDGTTAIFK
F8JYC4	Aminotransferase	AATNLQSHATSNVSNVAQAAALAAVSGGLEAVATMR
F8JLR2	Catalase-peroxidase	YLGPEVPSETLLWQDPLPEVTHPLVDAGDIATLK
F8JVH3	Chaperone protein DnaK	DVLLLDVTPLSLGIETK
F8JRT9	Chaperone protein DnaK	MARAVGIDLGTTNSVVSVLEGGEPTVITNTEGAR
G8WS40	Chaperonin GroEL	EQIASTASISAADTQIGELIAEAMDK
F8JZD2	Chaperonin GroEL	AMLGDMAVLTGATVVAEEVGLK
G8WS40	Chaperonin GroEL	EQIASTASISAADTQIGELIAEAMDK
F8JZD1	Co-chaperonin GroES	VVVQPLEAEQTTASGLVIPDTAK
F8K0X7	Cytochrome bc1 complex cytochrome c subunit	AIYNQAEIDQLAAYIASLGPGPATPTK
F8JNU5	Cytochrome P450	AGDVIQGLDVLPVTW
F8JSH4	D-3-phosphoglycerate dehydrogenase	AGVMVVNAPTSNIVTAAELACGLLVSTAR
F8JYJ2	Dihydropyrimidinase	AAAWGGTTTIVDFAIQSVGGSLR
F8JYD3	DNA-directed RNA polymerase subunit beta'	LVDIGEAVGIIAAQSIGEPGTQLTMR
F8JYD7	Elongation factor Tu	VNETVDIIGIK
F8JYD7	Elongation factor Tu	TTDVTGVVTLPEGTEMVMPGDNSSMTVSLIQPVAMEEGLR
F8JYD7	Elongation factor Tu	NMITGAAQMDGAILVVAATDGPMPQTK
F8JPG4	Fluorinase	RPIIAFMSDLGTTDDSVAQCK
F8JX40	Fluoroacetaldehyde dehydrogenase	FDGYDDAISIANDTLYGLGAGVWTR
F8JVI8	Fructose-bisphosphate aldolase	PIATPEIYNEMLDR
F8JP94	GatB/YqeY domain-containing protein	QLTDEELGALVDEAVAEAAASGTEGPK
F8JYI5	Glutathione hydrolase proenzyme	SATSGLFLEHGKPIPFDQAVTSGLSVGTPGTPATWQAALDAWGTR
F8JYZ8	glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)	VPVPTGSVTDLVVELER
G8XET2	histidine kinase	GEILALKTTINTMVDQLSSFAAEVTR
F8JX19	Homogentisate 1,2-dioxygenase	WSPPPAPEGPTDFVDGLVTVGGNGDVLQR
F8K4P3	Integral membrane protein	LALTQPAPGGGITDPVHGYLAPVIAYMTAVGMARR
F8JYD0	Large ribosomal subunit protein uL10	IAANEAGISSLDDLLNGSTALAFVTGDPVEAAK
G8WX50	Large ribosomal subunit protein uL3	TSDASEYTLGQEIAADVFEAGVK
F8JTW6	L-lysine 6-transaminase	TQLLETVVR
G8WXU0	Mannosyl-oligosaccharide alpha-1,2-mannosidase	TGAVSGSSVPLAEVGTNVMEFGLLTR
F8JTR7	methylmalonate-semialdehyde dehydrogenase (CoA acylating)	VTGQVDFASIAEVDQAVGAAVTAFR

F8K2V4	Methylthioribose-1-phosphate isomerase	LATAATGTALGAILELADR
F8JML4	Methyltransferase domain-containing protein	EPTPETIEILGDFADLATTVDR
F8JZV2	Mycothiol-dependent maleylpyruvate isomerase metal-binding domain-containing protein	LLLADLLESAGPDAPTLCEGWTTR
F8JU26	N5,N10-methylenetetrahydromethanopterin reductase-related protein	TLDPFVTLAAVAAVTER
G8WS69	Nucleoid-associated protein SCATT_29380	AVDPEDTETLADLVLAAVR
F8JW22	Oxidoreductase	YAAGGADVTGDIVDLLDPAATR
F8K1I6	Peptidase M20 dimerisation domain-containing protein	ALETSFDGALVEAMQTALR
F8JSE4	Peptidyl-prolyl cis-trans isomerase	HTIFGEVADDAGKR
F8K155	Phosphate-binding protein	QTAGAIGYFELSYATASNIPTVK
F8K0Q5	Phosphoserine aminotransferase	NQTYNTPALATLFLLDQQLK
F8K1S4	Prephenate dehydrogenase	TAPAAETPAAADKASDSAAEASAAEAPAR
F8K1Y0	Probable cytosol aminopeptidase	LGLPVNVTGWLALAENMPSGTATRPGDVLR
F8K1Y0	Probable cytosol aminopeptidase	ASEEKPDAIVDVATLTGAMVLALGNR
F8JW21	Putative aldolase	LANAAATLGLPLAAFTTDAGVDIVSFGGTK
G8WZJ6	Putative ATP-dependent Clp protease	VSITDAALVAAATLADR
F8JNE1	Putative beta-lactam synthetase	FAGVTAVPAGTAVTLAVAGFDTEPLLTR
F8JNE1	Putative beta-lactam synthetase	GADDTLLVATSAGALLPLLGTAPVLGEGDR
F8K4U4	Putative elongation factor	AETGDTLSAKDDPLLMEPWLMPDPLLPVAIEAHSK
G8XGW0	Putative polyketide oxygenase/hydroxylase	TVLTEWFGAPLDD
F8K0V3	Pyrophosphate--fructose 6-phosphate 1-phosphotransferase	TIDNDLSATDYTFGFDTAVNIATEAIDR
F8JJ36	Secreted protein	EAVTAIEGAVGAALVDYTSGMALGTIGGDK
F8JP36	Serine protease	IGGTSLAAPVISAIQALAQQAR
F8K2P9	Short chain dehydrogenase	GSIINTASFVAVMGAATSQISYTASK
F8JYG3	Small ribosomal subunit protein uS8	TMTDPIADMLTR
F8JYG3	Small ribosomal subunit protein uS8	VLGGLGVAIISTSQGLLTGQQAAK
F8K496	Stress protein	YDLTEDASTETAMVFGELYR
F8K173	Thymidine phosphorylase	DVTGTVEAIPLIASSIMSK
F8JTY4	Transcriptional regulator, winged helix family	DAMLGGHSTAELTDR
F8K2J4	Uncharacterized protein	SQLLNVDETTESAVEQLALDFIK
F8K2J4	Uncharacterized protein	TTIDTTAITVELPEAFDPR
F8JWY4	Uncharacterized protein	GIGVGEVTTAASAGQGWATVTTRWSGGAR

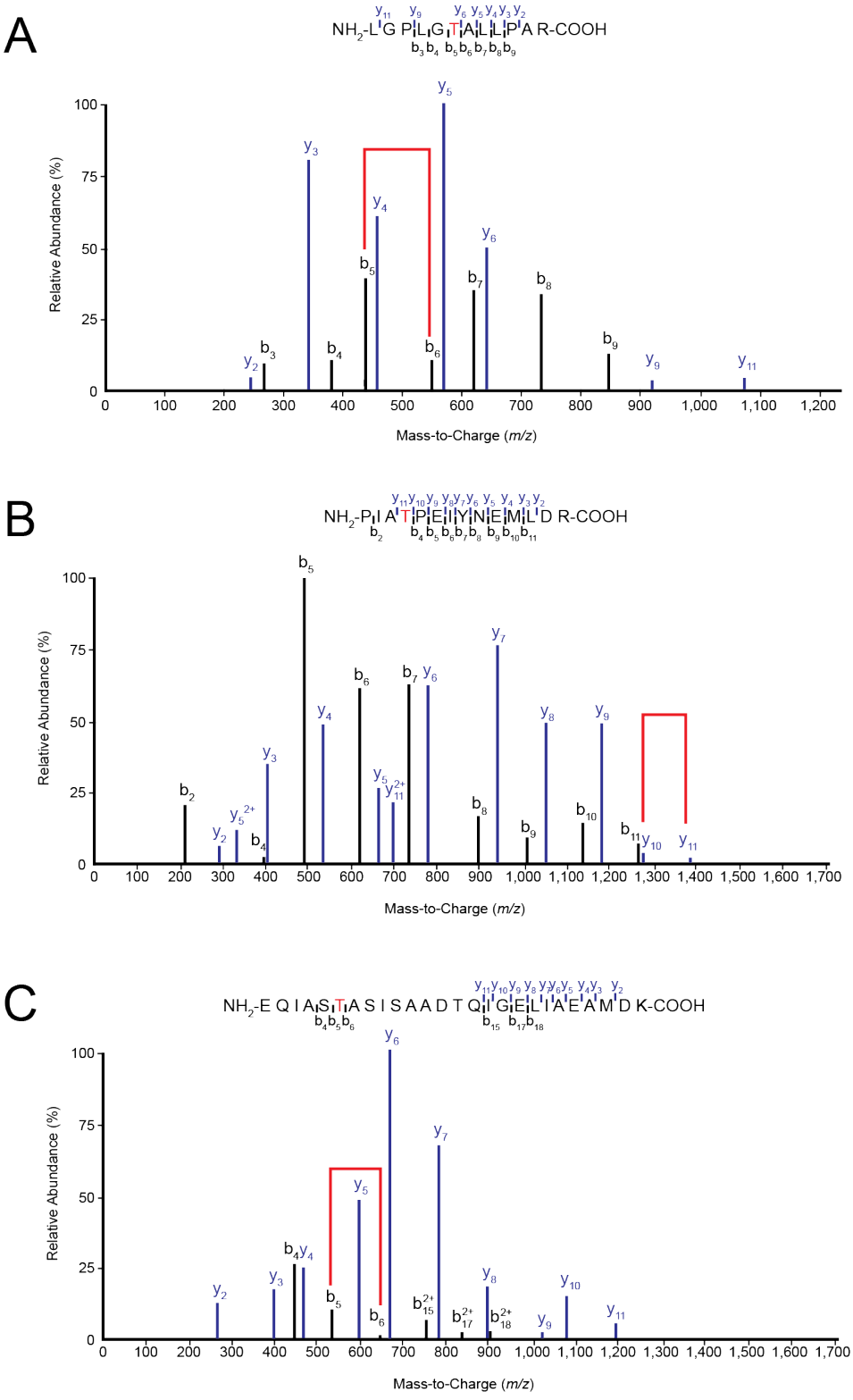


Figure 2.8. MS/MS spectra showing βes incorporation into peptides at Thr sites. High resolution MS/MS spectra showing βes incorporation at Thr positions in ΔBesG strain supplied with exogenous βes. (A) Peptide from the ATPase (G8X087). (B) Peptide from Fructose-bisphosphate aldolase (F8JVI8). (C) Peptide from Chaperonin GroEL (G8WS40).

A

```
AlaX-S |------------MYSIEVRT----------------------------------------|8
AlaX-M |MINMTRKLYYEDAYLKEAKGRVLEIRDNA----ILLDQTIFYPTGGGQPHDRGTINGVEV|56
BesG   |-MTQQRQIYLTDTYAYEASTRVVSAERGPDTVRIALADNIFHPQGGGQPDDRGWVDQVEV|59
                    *  *.  .

AlaX-S |HSALHVVKGAVVKVLGSEAKWTYSTYVKGNKGVLIVKFDRK----PSDEEIREIERLANE|64
AlaX-M |LDVYKDEEGNVWHVVKEP-----EKFKVGDEVELKIDWDYRYKLMRIHTGLHLLEHVLNE|111
BesG   |RPVRDTGPGLVHLTCPADAAWPREAVAVGTTVTSRIDPALR----RLHAALHTAGHLVDA|115
          .     * *  .         .    *      :.   .         :.   .: :

AlaX-S |KVKENAPIKI------------YELPREEAEK-MFGEDMYDLF--PVPEDVRIL------|103
AlaX-M |VLGEGNWQLVGSGMSVEKGR--YDIAYPENLN-KYKEQIISLFNKYVDEGGEVKIWWEGD|169
BesG   |LIDEWGYRHVGSNHFPGQARVEYDLGGLDCDKDALAEGLTERLGKPLAEALPVIPGERDG|175
          : *     :            *::   :  :    * : . :   : *   :

AlaX-S |-KVVVIEDWNVNACNKEHTKTTGEIGPI-KIRKVRFRKSKGLLEIHFELLELENPSLE|159
AlaX-M |RRYTQIRDFEVIPCGGTHVKDIKEIGHIKKLKRSSIGRGKQRLEMWLE----------|217
BesG   |RRTITIEGYATEFCAGTHVPDLSLLTGV-AIRSVKVKGGRLKVGYTAEHVPVDG----|228
          .   * .: .  *   *.     :  :  :.   .  ..  :    *
```

B

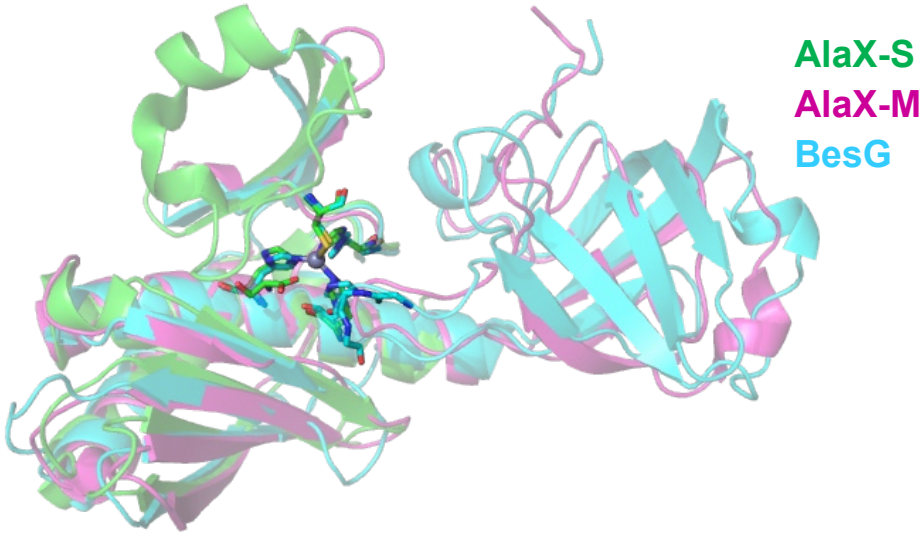


Figure 2.9. Sequence and structural comparison of BesG to canonical AlaX enzymes. (A) Sequence alignment of canonical AlaX-S (PDB: 1WNU) and AlaX-M (PDB: 2E1B) from *Pyrococcus horikoshii* to BesG (Uniprot ID: F8JJ98) [*51*]. Glycine-rich loop (green) and Zinc-binding motifs (**H/C**XXX**H**, red) are bolded and colored. (B) Overlay of solved crystal structures with BesG AlphaFold model show a similar overall fold.

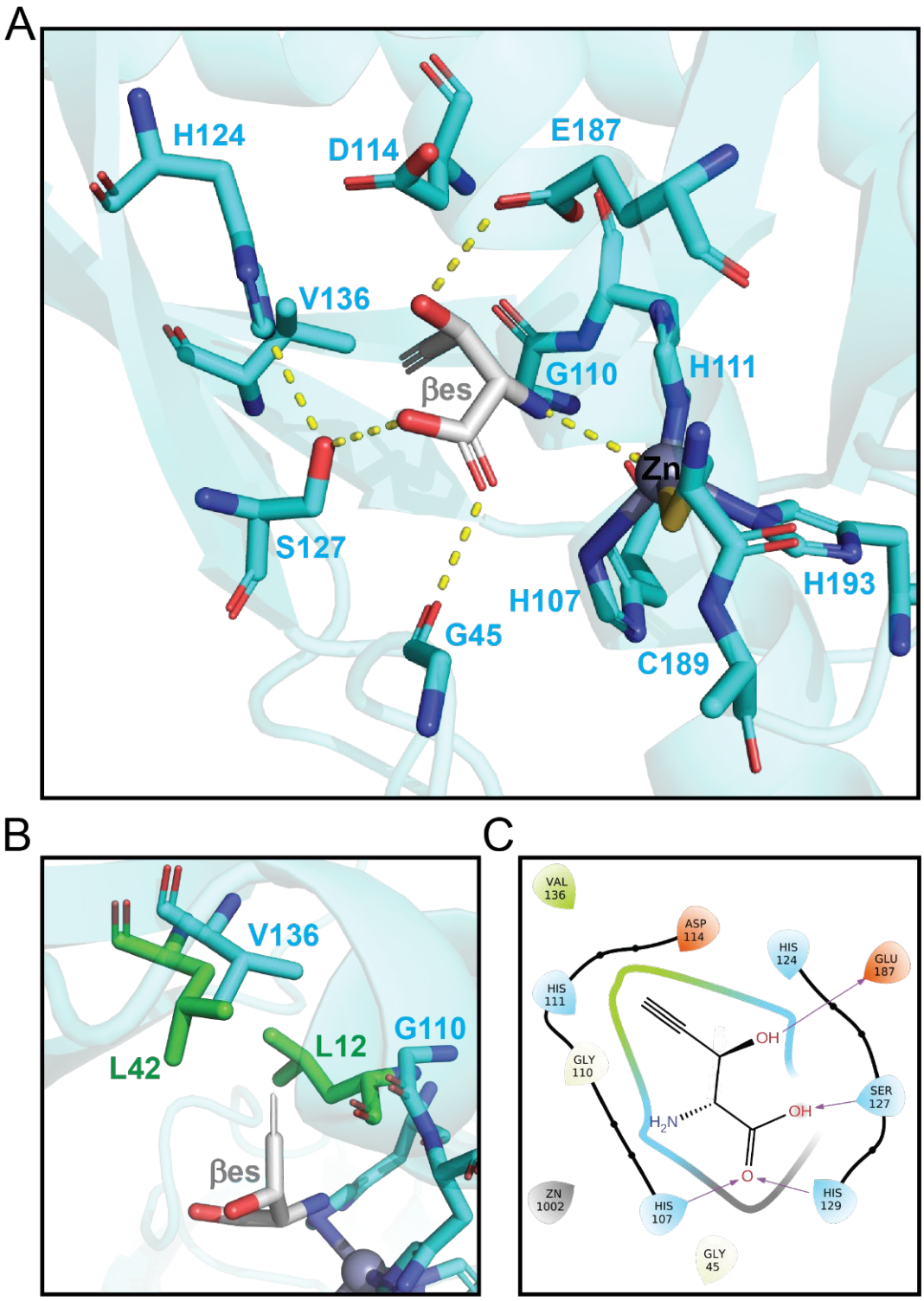


Figure 2.10. Substrate docking of βes in BesG. (A) The starting location of βes was estimated based on the location of serine bound in AlaX-S (PDB: 1WNU), and rigid docking was performed to optimize molecular interactions. A series of polar interactions support the substrate binding location. (B) Comparison of BesG active site with AlaX-S reveals several mutations to smaller residues in BesG, which we postulate could make room for the additional steric bulk of the alkyne compared to Ser. (C) 2D interaction map of βes in the active site of BesG following docking shows predominantly polar interactions.

Following these findings of BesG activity *in vitro*, we sought to evaluate the ability of BesG to prevent misincorporation in the native producer *S. cattleya*. We performed a genomic knockout of BesG in *S. cattleya* and supplied βes to both WT and ΔBesG strains through spent media. After a 7-day fermentation, the proteome was extracted and derivatized with copper-catalyzed azide-alkyne cycloaddition (CuAAC) to a TAMRA fluorophore, enabling detection of incorporation via fluorescence (**Fig. 2.6B**). We observed much higher fluorescence in the knockout strain compared to WT, demonstrating that BesG can mitigate self-incorporation of βes at physiological concentrations. Bottom-up proteomics confirmed βes incorporation at Thr sites in the ΔBesG strain **(Table 2.1, Fig. 2.8, Table A2.2, Figure A3.1**). Notably, BesG is the first AlaX-family deacylase reported to have specificity towards a naturally produced ncAA.

With BesG showing specificity for βes over Thr, we explored putative structural determinants for alkyne side chain recognition. We compared an AlphaFold2 model of BesG with solved crystal structures of AlaXs from primary metabolism and compared their sequences and structures (**Fig. 2.9**) [*44*]. The crystal structures show a Zn binding motif that binds the carbonyl of the bound substrate serine, along with a glycine-rich loop that coordinates glycine-tRNA^{Ala} unique to AlaX-M [22]. The AlphaFold2 structure of BesG aligns well with the AlaX-M of *P. horikoshii* (PDB: 2E1B, RMSD=1.85 Å) and contains the same metal-binding site and glycine-rich loop (**Fig. 2.9B**). We performed a docking experiment of βes in BesG using AlaX-S with bound serine (PDB: 1WNU) as a reference. The results suggest a similar binding pose for βes, where the alkyne side chain is accommodated by a pocket created by V136 and G110, which are larger leucines in AlaX-S (**Fig. 2.10**). Protein modeling provides a basis for understanding the specificity of BesG and serves as a starting point for characterizing other ncAA-selective AlaX-family tRNA-deacylases.

Genome mining of the AlaX family for ncAA biosynthesis. The discovery of BesG prompted us to explore the extent to which nature uses dedicated, co-clustered tRNA-deacylases as a general resistance mechanism in ncAA biosynthesis. We took an enzyme family-based approach, where all AlaX-like deacylases were included in the analysis and divided into clusters by sequence similarity, and neighboring genomic information was compared between clusters of tRDs. The tools available through the Enzyme Function Initiative (EFI) were used to generate sequence similarity networks (SSNs) and organize genomic neighborhood data by subclusters. In addition, we built a data visualization tool called tR3D (tRNA-Deacylase Directed Discovery) for quantifying the genomic neighborhood data for efficient analysis (**Fig. 2.11**). Since tR3D visualizes data directly output from EFI, we built this pipeline to be broadly applicable to any enzyme family of interest for discovering new biosynthetic pathways.

Since tRNA-deacylases could function in primary or secondary metabolism, we used our approach to first separate these two groups before proceeding to BGC identification. The SSN of the AlaX-like tRNA-editing family (IPR018163) contained 23,260 proteins, notably showing one very large cluster and many smaller clusters (**Fig. 2.12**). Of the few characterized members of this family, nine known to function in primary metabolism mapped to this large cluster (13,921 total AlaXs). In contrast, BesG mapped to one of the many smaller clusters. From this, we could loosely assign the members of the large cluster as participating in primary metabolism.

To more accurately sort the functions of tRDs as within primary or secondary metabolism, we identified the neighboring protein families upstream and downstream of each deacylase in our SSN using the EFI toolset and visualized this data with tR3D. tR3D quantifies the co-occurrence of neighboring genes for each cluster and enabled us to compare the most common protein families

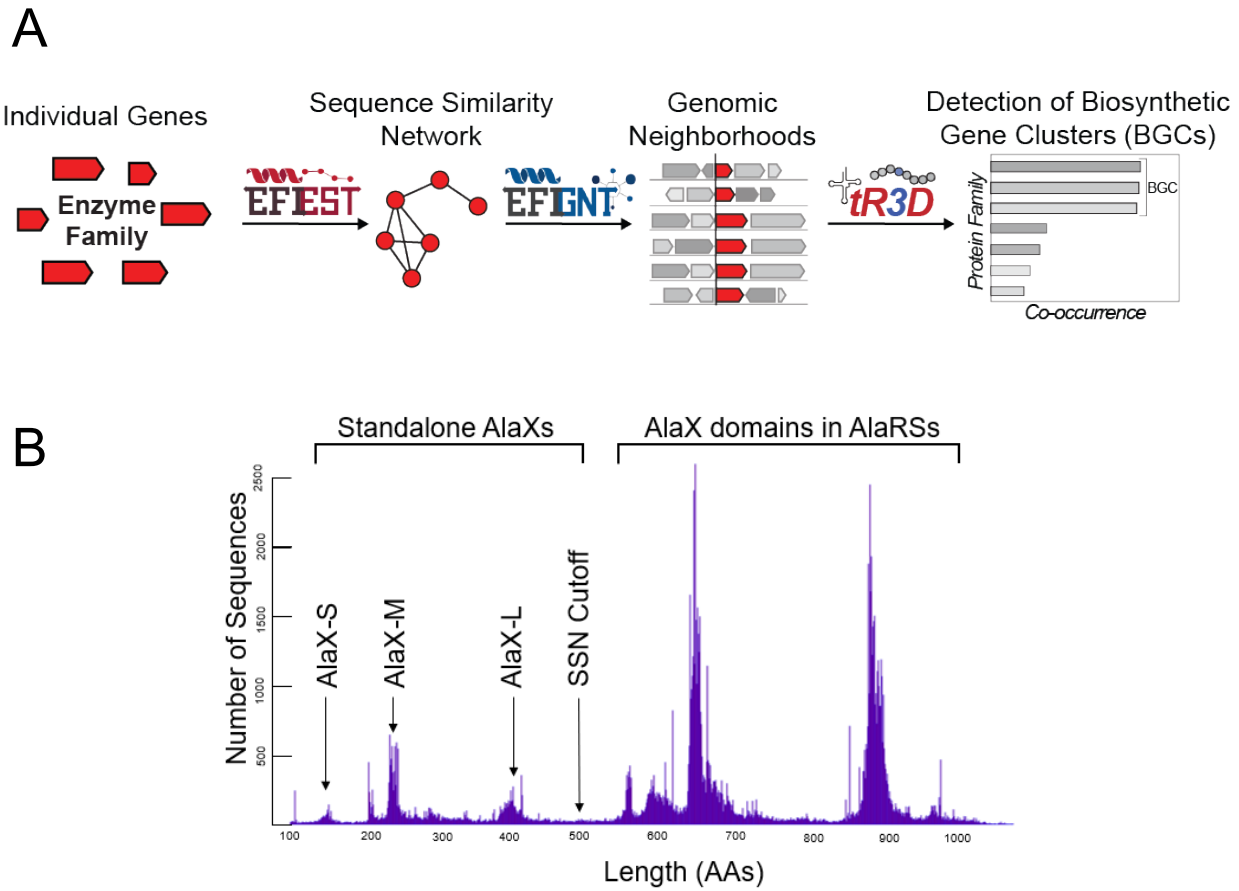


Figure 2.11. tR3D analysis overview and SSN generation. (A) Overall analysis approach, including visualizing enzyme family members with a sequence similarity network (SSN) [*43*], identifying their genomic contexts, and searching for BGCs with the visualization tool tR3D (tRNA-Deacylase Directed Discovery). (B) All sequences in the AlaX family (IPR018163) sorted by length with 3 major groups corresponding to the 3 main forms present in nature, AlaX-S (small), AlaX-M (medium, and AlaX-L (large). The initial cutoff was chosen to separate *trans*-editing deacylases from editing domains in AlaRSs which are involved in pnly primary metabolism [*19*].

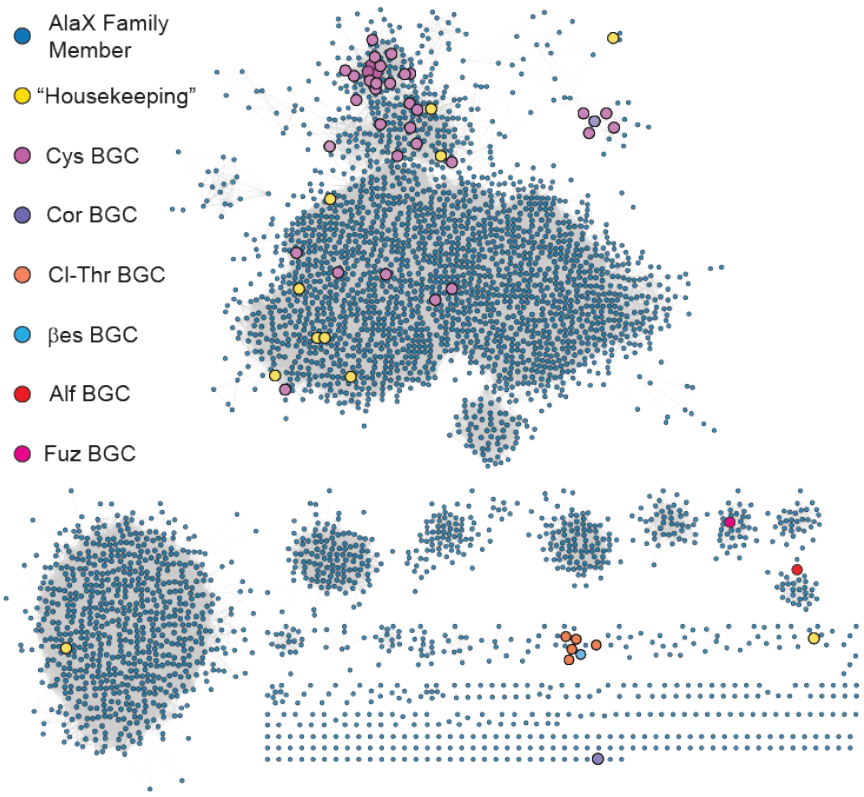


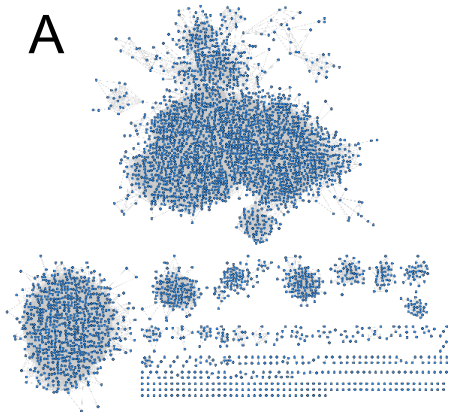
Figure 2.12: Annotated sequence similarity network (SSN) of the AlaX family. SSN of the AlaX family annotated with all AlaXs known to function in primary metabolism and newly discovered BGCs detected through the tR3D data visualization tool, which will be fully elucidated experimentally in this work.

between clusters. For tRDs participating in primary metabolism, we hypothesized the majority of neighboring genes would participate in primary metabolism, and tR3D would report low co-occurrence scores due to deacylase function not relating to neighboring activities. After several processing steps, including removing deacylases with short neighborhood reads and fragmented genes, we analyzed the genomic data for their co-occurrence with a list of primary metabolism genes we gathered from tR3D (**Fig. 2.13, Fig. 2.14**). If these were found in the genomic neighborhood, we assigned them as primary AlaXs and removed them from further analysis. Examples of commonly occurring primary metabolism genes included aminoacyl-tRNA synthetases, nucleotide-binding proteins, and ribosomal proteins.

Within primary metabolism, we were intrigued to find groups of tRDs in very high co-occurrence with genes in canonical amino acid biosynthesis. 11% of the deacylases in the large cluster (1,562 tRDs) co-located with cysteine biosynthetic genes (**Fig. 2.15, Table 2.2**), while others co-located with histidine (65) or threonine (42) biosynthesis (**Fig. 2.16A-C**). These findings highlight tRDs are general markers of amino acid metabolism. Using our co-occurrence-based approach, we estimate 89% of the AlaX tRDs included in this analysis function in primary metabolism, while the remaining could contain novel BGCs.

With the tRDs that passed the initial filtering (2,673), we utilized our neighborhood-based approach to identify secondary-metabolism BGCs. Class-defining protein families for non-ribosomal peptides, polyketides, and RiPPs are well-established, and we found that 17% of the tRDs (442) in this dataset contained genes for these pathways (**Fig. 2.16D-G, Fig. 2.17**). This

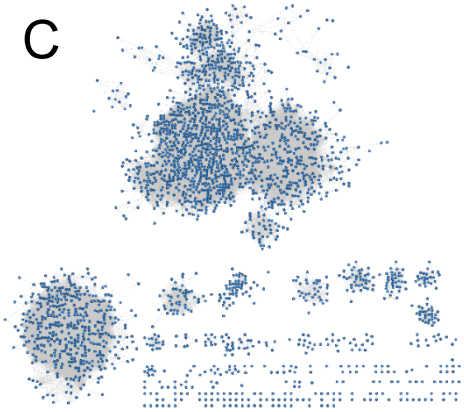
A



B

Description	No.
Total Nodes	3676
Total AlaX Sequences Represented	23260
Minimum Length Filter (AAs)	0
Maximum Length Filter (AAs)	480
Sequence Similarity Score Cutoff (%)	43
Neighborhood Size (upstream and downstream)	10
Co-occurrence Lower Limit (%)	0

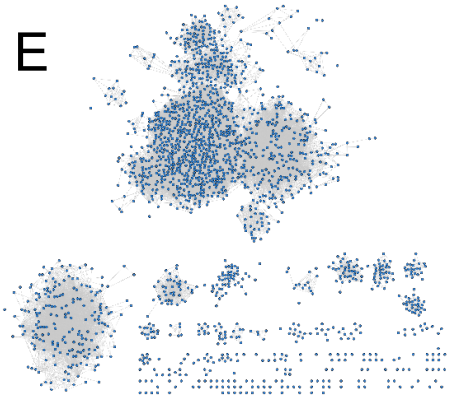
C



D

Description	Total Nodes (No.)	Represented Sequences (No.)
≤ 6 Neighboring Genes	1414	2021
>6 Neighboring Genes	2262	21239

E

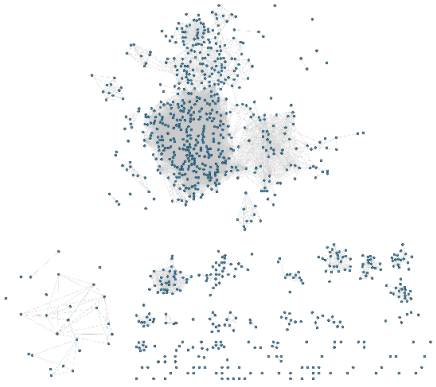


F

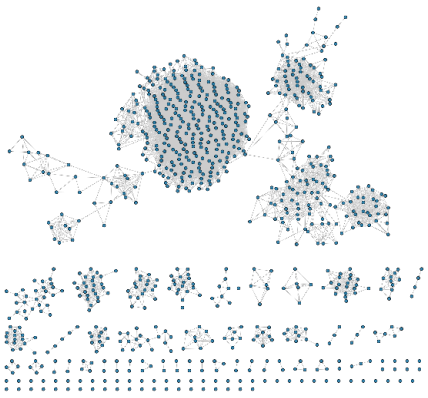
Description	Total Nodes (No.)	Represented Sequences (No.)
Every sequence is a fragment	383	1116
At least one sequence is complete	1879	20123

Figure 2.13. Data cleaning for SSN of AlaX Family. (A) Full SSN of Uniref50 AlaX Family IPR018163, where each node represents AlaXs with >50% similarity. (B) Parameters for SSN generation using the EFI tool [*43*]. (C) SSN was cleaned by removing any genomic reads with ≤6 reported neighboring protein families, eliminating poor genomic reads. (D) Table describing clusters removed and clusters remaining after this first cleaning step. (E) SSN was further cleaned by removing nodes where every included sequence contained a fragment (a start or stop codon was missing from the AlaX gene) (F) Number of sequences removed and number remaining after removing fragments.

A



B



C

Pfam	Description	Nodes Removed (No.)	Sequences Removed (No.)
PF01411	tRNA-synt_2c (Alanine-tRNA synthetase)	199	4261
PF02824	TGS (ThrRS Nucleotide-Binding Domain) (Neighbor and Attached)	66	2366
PF00587	tRNA-synt_2b (G/H/P/T/S)	179	1605
PF00293	NUDIX (pyrophosphohydrolases)	104	3973
PF00149	Calcineurin-like phosphoesterase-5'-nucleotidase	84	1442
PF13927	Immunoglobulin domain	15	655
PF00453	Ribosomal_L20	30	97
PF03129	HGTP_anticodon	22	79
PF00133	tRNA-synt_1 (I,L,M,V)	45	274
PF03652	Holliday Junction Resolvase	39	86
PF00448	Flagellar Biosynthesis protein	3	64
PF13393	Histidyl-tRNA synthetase	4	44
PF00749	tRNA-synt_1c (E,Q)	12	140
PF06144	DNA polymerase III, delta subunit	4	58
PF01084	Ribosomal protein S18	4	118
PF07686	Immunoglobulin V-set domain	11	605
PF13927	Immunoglobulin Domain	3	27
PF05879	Root hair defective 3 GTP-binding protein (RHD3) GTPase	1	17
PF00416	Ribosomal Protein S13	13	41
Q9BTE6	Neighbors of Human AlaX	56	924
O57734	Neighbors of P horikoshii AlaX-L	75	328
Q8GZ45	Neighbors of A thalianas AlaX	8	27
PF00291 AND PF00581	Cysteine Biosynthesis (Cysteine Synthase (PLP) AND Sulfurtransferase (Rhodanese))	7	100
PF00815 AND PF00977	Histidine Biosynthesis (Histidinol dehydrogenase and HisA isomerase)	1	1
IPR020568 AND PF00291	Threonine Biosynthesis (Kinase and PLP Synthase)	14	82
	Remaining Singletons	36	36
	Total	**1035**	**17450**

Figure 2.14. Filtering of AlaX SSN to remove primary metabolism AlaXs. (A) Resulting SSN after removing AlaX proteins with neighboring primary metabolism genes, resulting in 844 remaining nodes and 2,673 genomes. (B) Reformatted SSN of (A) for further analysis. (C) The neighboring protein families hypothesized to be specific for primary metabolism and used to remove AlaXs that are not for ncAA editing.

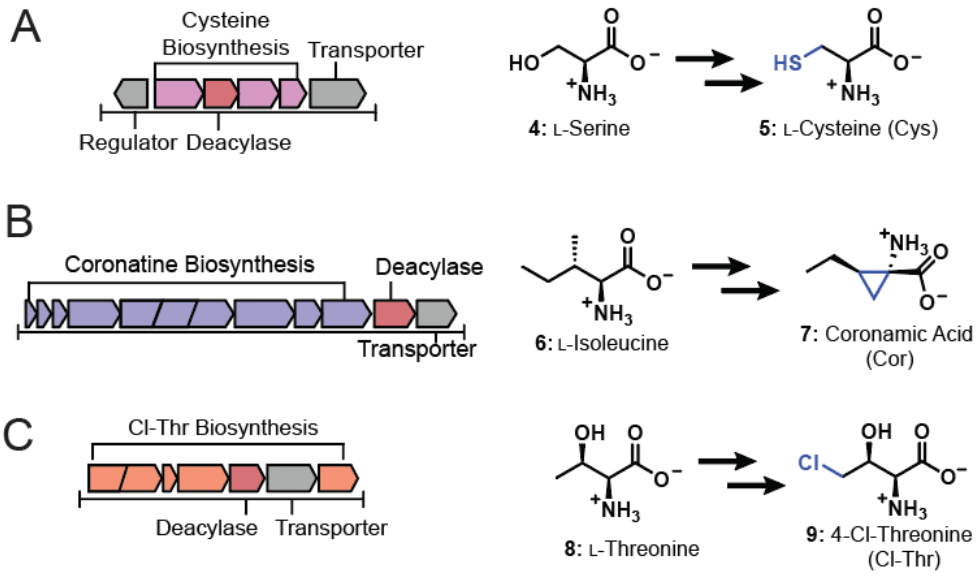


Figure 2.15. Amino acid biosynthetic genes co-occurring with AlaXs. (A) 11% of the AlaX enzymes in the large cluster are found in the context of cysteine biosynthesis (B) A deacylase was discovered within the biosynthetic gene cluster of the cyclopropyl amino acid-containing coronatine. (C) A deacylase was identified in the biosynthetic gene cluster of 4-Cl-Threonine. The pathways of B and C are known [*54,55*].

A

Gene IDs	SSN Cluster (No.)	Organism
Q9BTE6	1	*Homo sapiens*
Q6DEJ5	1	*Danio rerio*
Q3THG9	1	*Mus musculus*
Q7ZYJ9	1	*Xenopus laevis*
P53960	1	*Saccharomyces cerevisiae*
Q97ZX2	1	*Saccharolobus solfataricus*
Q8Q0A4	1	*Mathanosarcina mazei*
O57734	1	*Pyroccocus horikoshii* (AlaX-L)
O57848	1	*Pyroccocus horikoshii* (AlaX-M)
Q8GZ45	2	*Arabidopsis thaliana*
O58307	12	*Pyroccocus horikoshii* (AlaX-S)

B

Subcluster (No.)	Total AlaX Sequences (No.)	Fraction of Cluster 1 (%)
1	1321	9.5%
12	200	1.4%
39	41	0.3%
Total	**1562**	**11.2%**

Table 2.2. Primary metabolism AlaX entries in SSN. (A) Gene Uniprot IDs and SSN cluster location of characterized AlaXs that have been shown to function in primary metabolism. (B) Fraction of AlaX proteins in Cluster 1 that contain neighboring genes for cysteine biosynthesis (Cysteine Synthase PF00291 and Sulfurtransferase PF00581), which was calculated by splitting cluster 1 into smaller subclusters through increasing the alignment threshold (sequence alignment score = 90) and counting the total number of AlaX proteins in each relevant subcluster. We found that 11.2 % of the AlaXs in cluster 1 pertained to AlaXs in the context of Cys biosynthesis.

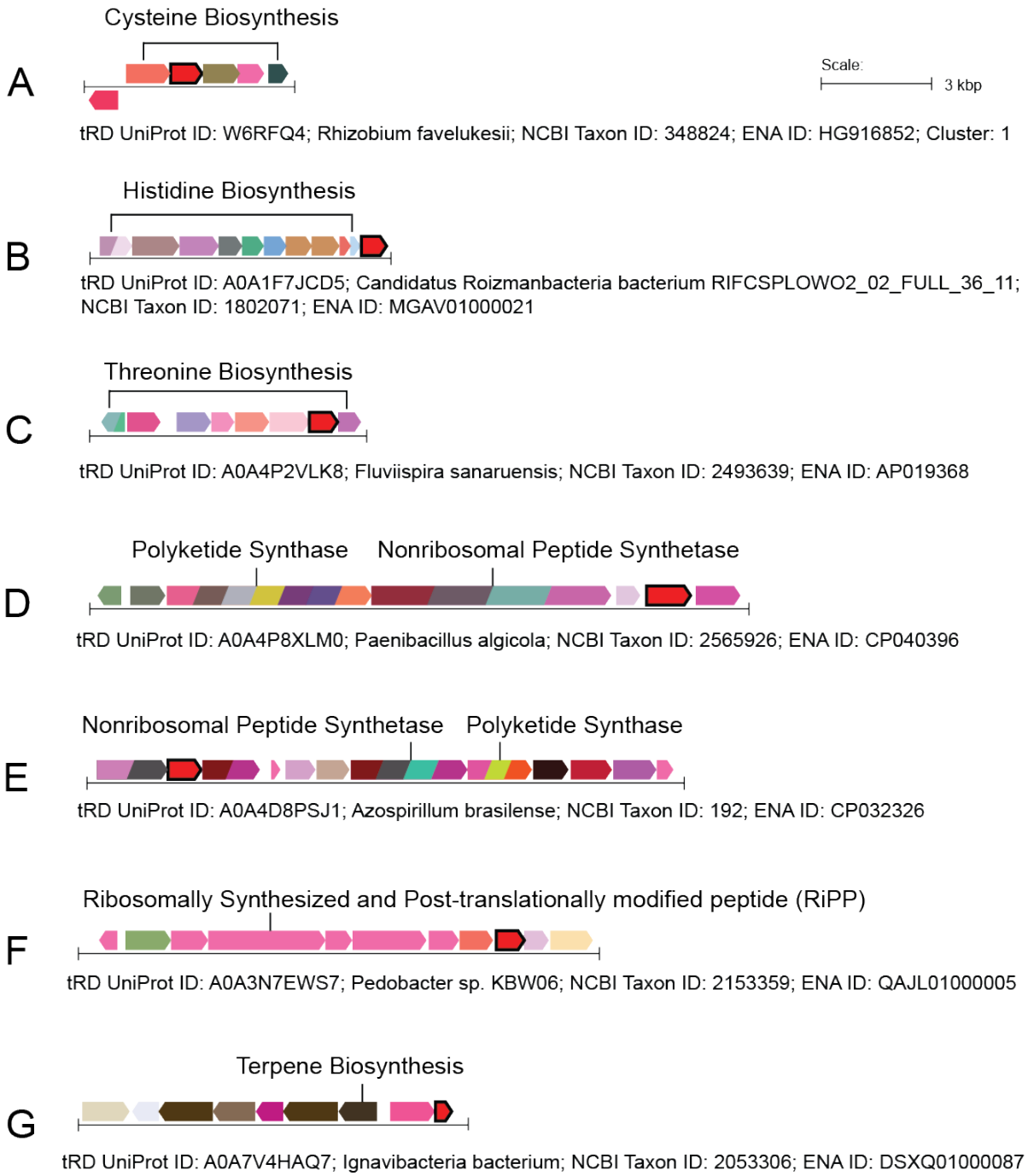
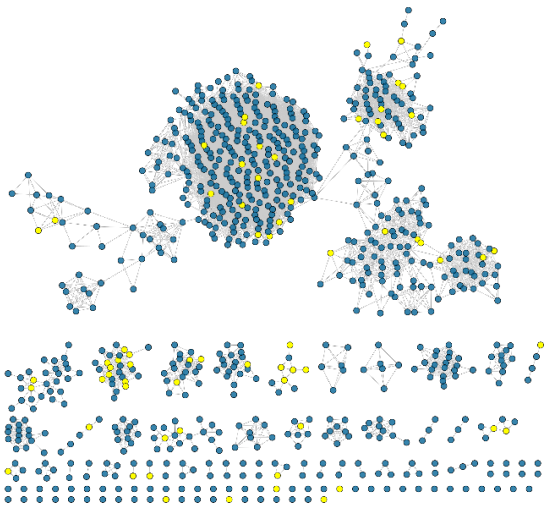
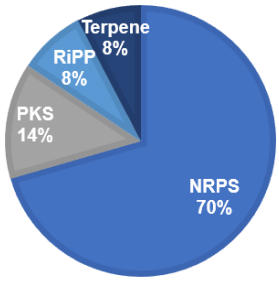


Figure 2.16. Representative gene neighborhoods for AlaXs in primary metabolism and class-defined secondary metabolism. (A) An AlaX in the context of Cys biosynthesis. (B) An AlaX in the context of His biosynthesis. (C) An AlaX in the context of Thr biosynthesis. (C) and (D) AlaXs co-located with Nonribosomal Peptide Synthetases (NRPSs) and Polyketide Synthases (PKSs). (E) An AlaX co-located with RiPP biosynthetic genes. (F) An AlaX co-located with terpene biosynthetic genes.

A



B



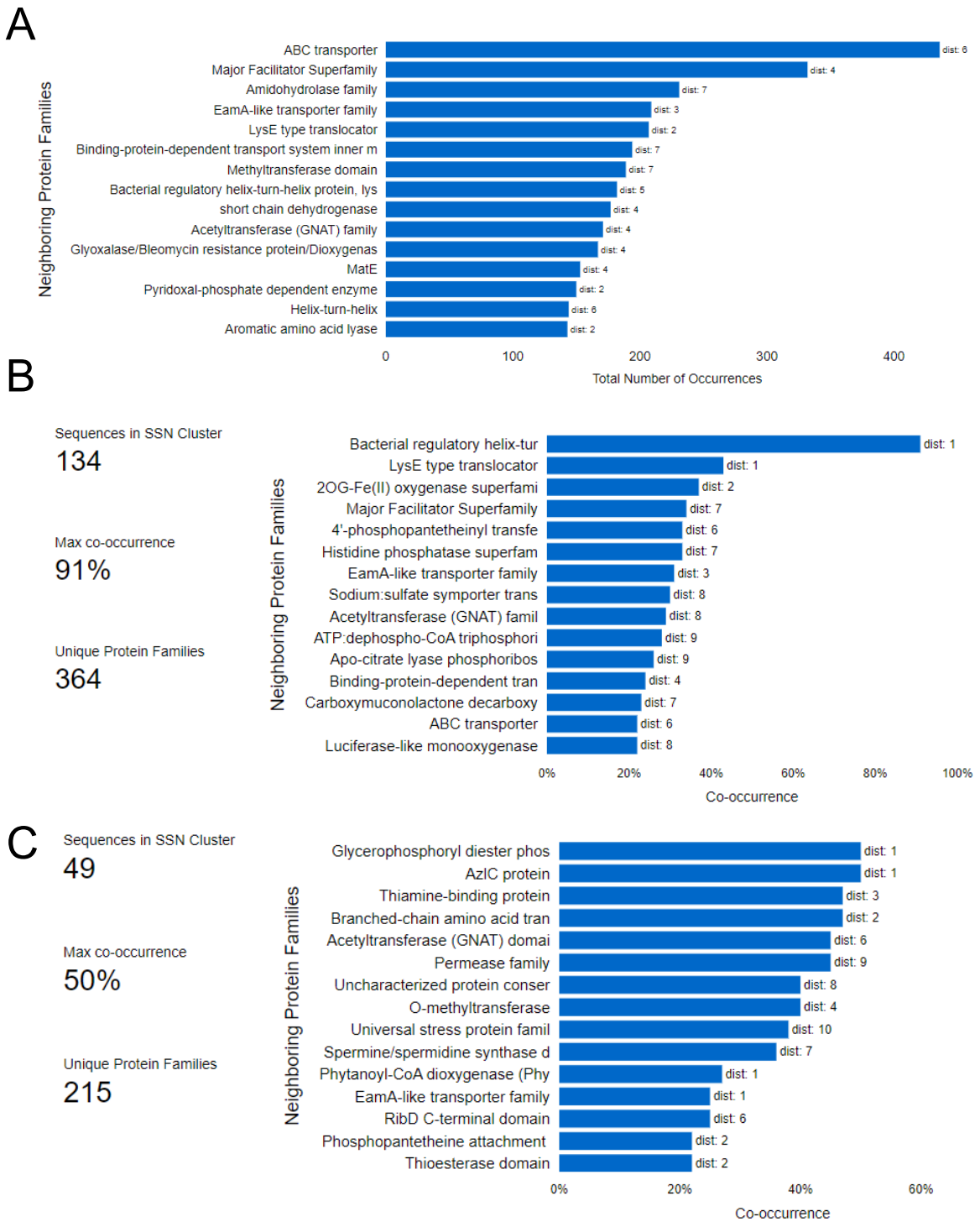
Class of Secondary Metabolism	Total Counts (No.)
NRPS	333
PKS	65
RiPP	38
Terpene	36

C

Pfam ID	Description	Selected Nodes (No.)	Represented Sequences(No.)
NRPS			
PF00668	Condensation Domain [NRPS]	10	80
PF00501	AMP-binding enzyme [NRPS]	52	333
PKS			
PF00109	Beta-ketoacyl synthase N-terminal [PKS]	18	65
PF02801	Beta-ketoacyl synthase C-terminal [PKS]	17	64
RiPP			
PF05147	Lanthionine synthetase C-like protein [lantipeptide/RiPP]	1	2
PF00733	Lasso Peptide Cyclase/Asparagine Synthase	4	29
PF13471	Lasso Peptide Protease/Transglutaminase	1	6
PF02624	YeaO Cyclodehydratase [RiPP]	1	1
Terpene			
PF00494	Squalene/phytoene synthase [Terpene]	1	33
PF01397	Terpene synthase, N-terminal [Terpene]	1	3
	Total (Unique):	**79**	**472**

Figure 2.17. Co-occurrence of class-defined secondary metabolism in AlaX SSN. (A) Revisualization of AlaX SSN after filtering out short reads, fragments, and AlaXs with co-occurring genes involved in primary metabolism. Class-defined secondary metabolism is highlighted in yellow (B) Table and pie chart describing distribution of class-defined secondary metabolism. (C) Protein families used for classification of BGCs [*56*].

Figure 2.18. tR3D analysis of AlaX tRNA-deacylases. (A) tR3D app output quantifying most abundant protein families in the gene neighborhoods of AlaXs after filtering (B) Genomic neighborhood analysis of cluster 5, which contains the 3-hydroxymethione-producing BGC. (C) Genomic neighborhood analysis of Cluster 11, which contains the BGC for coronamic acid. (D) Genomic neighborhood analysis of Cluster 13, which contains the BGC for 4-Cl-Threonine. (E) Genomic neighborhood analysis of Cluster 15, which contains the BGC for the fuscazines.

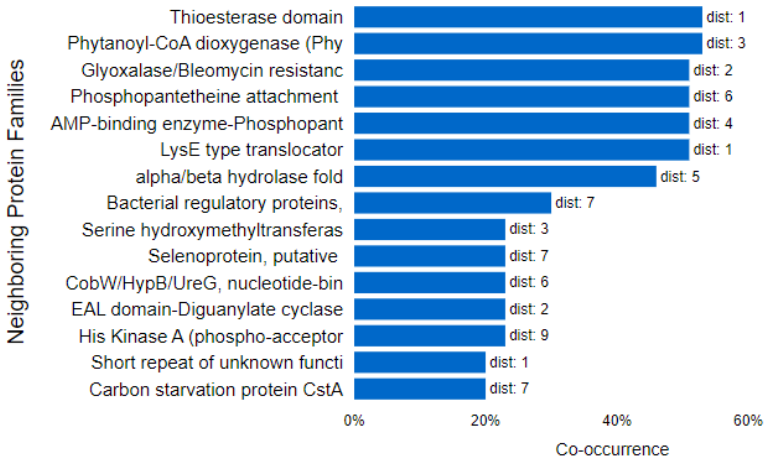


D

Sequences in SSN Cluster
42

Max co-occurrence
53%

Unique Protein Families
139



E

Sequences in SSN Cluster
34

Max co-occurrence
32%

Unique Protein Families
265

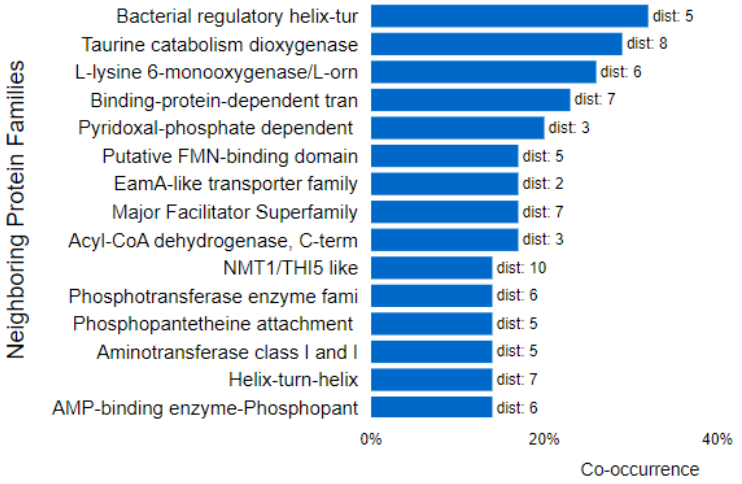
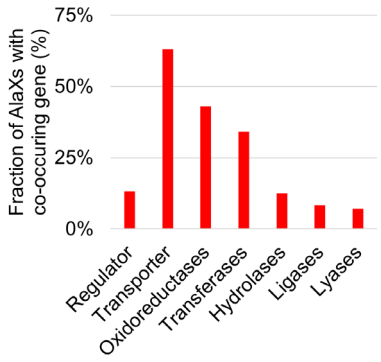


Figure 2.19. Secondary metabolism genes that co-occur with AlaX tRNA-deacylases. Table and summary figure generated from tR3D analysis of AlaX SSN, where members of the 5 most common enzyme types, as well as regulators and transporters, were identified and quantified as markers of ncAA biosynthesis.



Pfam ID	Description	Nodes (No.)
Regulator		
PF00126	HTH Regulator	58
PF00392	GntR Regulator	54
Transporter		
PF00005	ABC Transporter	249
PF01810	LysE Transporter	40
PF07690	MFS Transporter 1	132
PF00324	Amino Acid Permease	12
PF00892	EamA Transporter	86
PF03591	Branched Chain Transporter	14
Oxidoreductase		
PF13640	2OG	15
PF02668	TauD	17
PF00355	Rieske	18
PF00067	p450	30
PF05721	Phy CoA Dioxygenase	18
PF00743	FMO	3
PF13434	Lys/Orn Oxygenase	4
PF13454	CreE-like N-oxygenase	3
PF01070	FMN-dependent dehydrogenase	12
PF00296	Luciferase like monooxygenase	18
PF13460	NADP-binding protein	12
PF00248	Aldo/keto reductase	46
PF00106	Short chain dehydrogenase	49
PF04055	Radical SAM	116
PF13370	4Fe-4S	2
Ligase		
IPR011761/013865	ATP-Grasp	68
PF02655		2
Transferase		
PF00155	PLP	79
PF00291	PALP	99
PF00583	GNAT Acetyltransferase	83
PF01648	Phosphopantetheinyl transferase	8
PF10672	S-adenosylmethionine-dependent methyltransferase	12
PF02353	Mycolic Acid Cyclopropane Synthetase	7
Hydrolases		
PF01546	Peptidase M20/M25/M40	45
PF04389	Peptidase M28	14
PF04909	Amidohydrolase	16
PF12697	Alpha/beta hydrolase	30
Lyases		
PF00903	Glyoxylase/Bleomycin Resistance Protein	25
PF00205	Thiamine-pyrophosphate protein	6
IPR011118	Tannase/feruloyl esterase	4
PF00221	Aromatic Amino Acid lyase	12
PF00975	Thioesterase	13



Figure 2.20. Putative Biosynthetic Gene Clusters (BGCs) with AlaX tRNA-Deacylases. (A) Examples from the 140 unique biosynthetic gene clusters (BGCs) identified through tR3D to potentially produce secondary metabolites. Accessions for these tRDs are in **Figure A3.2**.

fraction was larger than we expected, as these biosynthetic strategies typically shuttle substrates from protein domain to protein domain, yet the presence of a deacylase suggests these ncAAs are free at some stage in these pathways. Notably, we found the biosynthetic pathways for several known ncAAs, coronamic acid and chlorothreonine, where the biosynthetic genes are characterized yet the neighboring deacylase is unexplored (**Fig. 2.15B,C**). These findings validated our approach and exemplify the unique chemical features of ncAAs.

The remaining tRDs in our dataset (2,201) were hypothesized to hold unique biosynthetic logic for producing ncAAs and were analyzed via tR3D for identifying unique BGCs (**Fig. 2.18, Fig. 2.19**). A majority of the gene neighborhoods contained at least one transporter or a regulator, which are common features of secondary metabolism. Additionally, nearly half of the neighborhoods contained an oxidoreductase, a third contained transferases, and one in ten contained a ligase. We estimate this dataset to contain at least 140 unique biosynthetic pathways with a diverse array of enzymes (**Fig. 2.20**). From the annotations of neighboring genes, we anticipate the discovery of new ncAAs incorporated into a diversity of larger structures including glycosides, nucleosides, and peptides. With these findings, we chose several pathways for experimental validation of ncAA and related natural product biosynthesis.

Discovery of a new amino acid modification, tRNA misacylation, and ncAA-tRNA editing. With the goal of discovering a new ncAA, we identified a BGC conserved across 216 genomes with tR3D containing a regulator, transporter, and an αKG-dependent non-heme iron/α-ketoglutarate-dependent dioxygenase co-clustered with a tRD (**Fig. 2.21**). We hypothesized this dioxygenase produces a ncAA, the tRD edits misacylated ncAA-tRNA, and the substrate for this dioxygenase is a standard amino acid, as this is a common feature of known ncAA biosynthetic pathways [*1*]. By limiting the likely substrate scope to the twenty standard amino acids, we had a straightforward approach for discovering new enzymatic activity. We selected the BGC from the gram-negative organism *Alacaligenes faecalis* and expressed and purified the dioxygenase (AlfC) and the tRD (AlfB) with His_{10} N-terminal tags from *E. coli* (**Fig. A3.3**). When we supplied the twenty amino acids with necessary cofactors to AlfC and analyzed this reaction by LC/MS-QTOF, we detected consumption of methionine (**10**) and the production of a new compound +16 *m/z* when enzyme was present (**11**) (**Fig. 2.21**). This mass shift indicated substrate hydroxylation, and while methionine can spontaneously oxidize to produce a nonenzymatic +16 *m/z* product, NMR spectroscopy confirmed methionine hydroxylation at C_β and stereochemistry was assigned as 3*R* (**Table 2.3, Fig. 2.22**). To our excitement, this result supported our hypothesis and is the first report of an enzymatic hydroxylation of methionine along its carbon chain to form a ncAA.

With this discovery of a new ncAA, we examined the specificity of the neighboring tRD, AlfB, towards (2*S*,3*R*)-hydroxy-methionine (**11**). We expressed and purified the methionine-tRNA synthetase of the native organism (*Af* MetRS) with an N-terminal His_{10} tag from *E. coli*, produced the corresponding $tRNA^{Met}$ *in vitro,* and performed a competitive acylation experiment containing the *Af* MetRS, $tRNA^{Met}$, and AlfB with either methionine or (2*S*,3*R*)-hydroxymethionine (**Fig. 2.23**). Without AlfB present, we found (2*S*,3*R*)-hydroxymethionine was acylated to $tRNA^{Met}$, indicating *Af* MetRS accepts this ncAA and cannot effectively discriminate against it. When AlfB is present, (2*S*,3*R*)-hydroxymethionine-$tRNA^{Met}$ levels decrease dramatically, compared to relatively little change with methionine. When AlfB is present 5:1 to *Af* MetRS, (2*S*,3*R*)-hydroxymethionine-$tRNA^{Met}$ is nearly undetectable. This finding supports our hypothesis, where

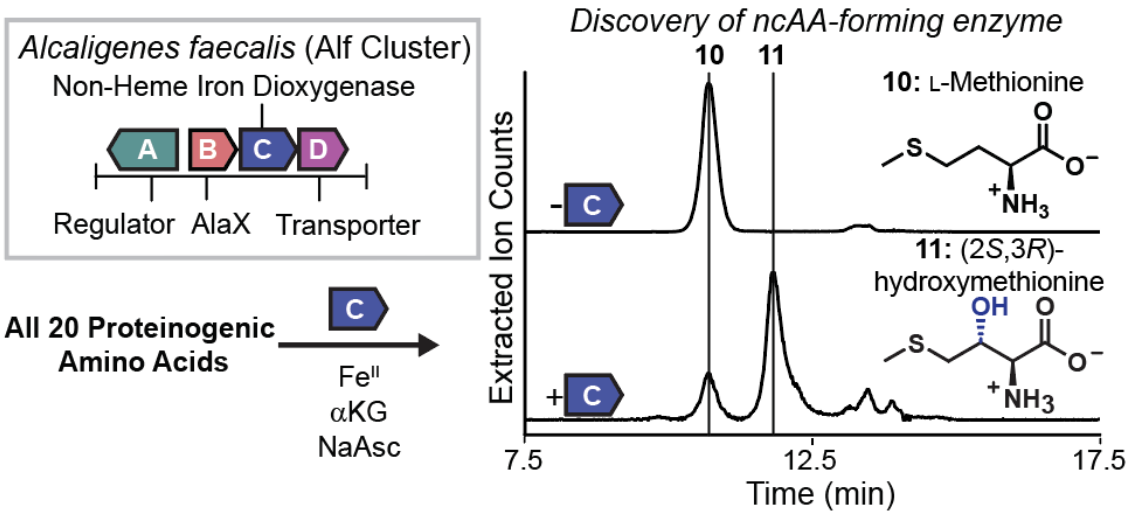
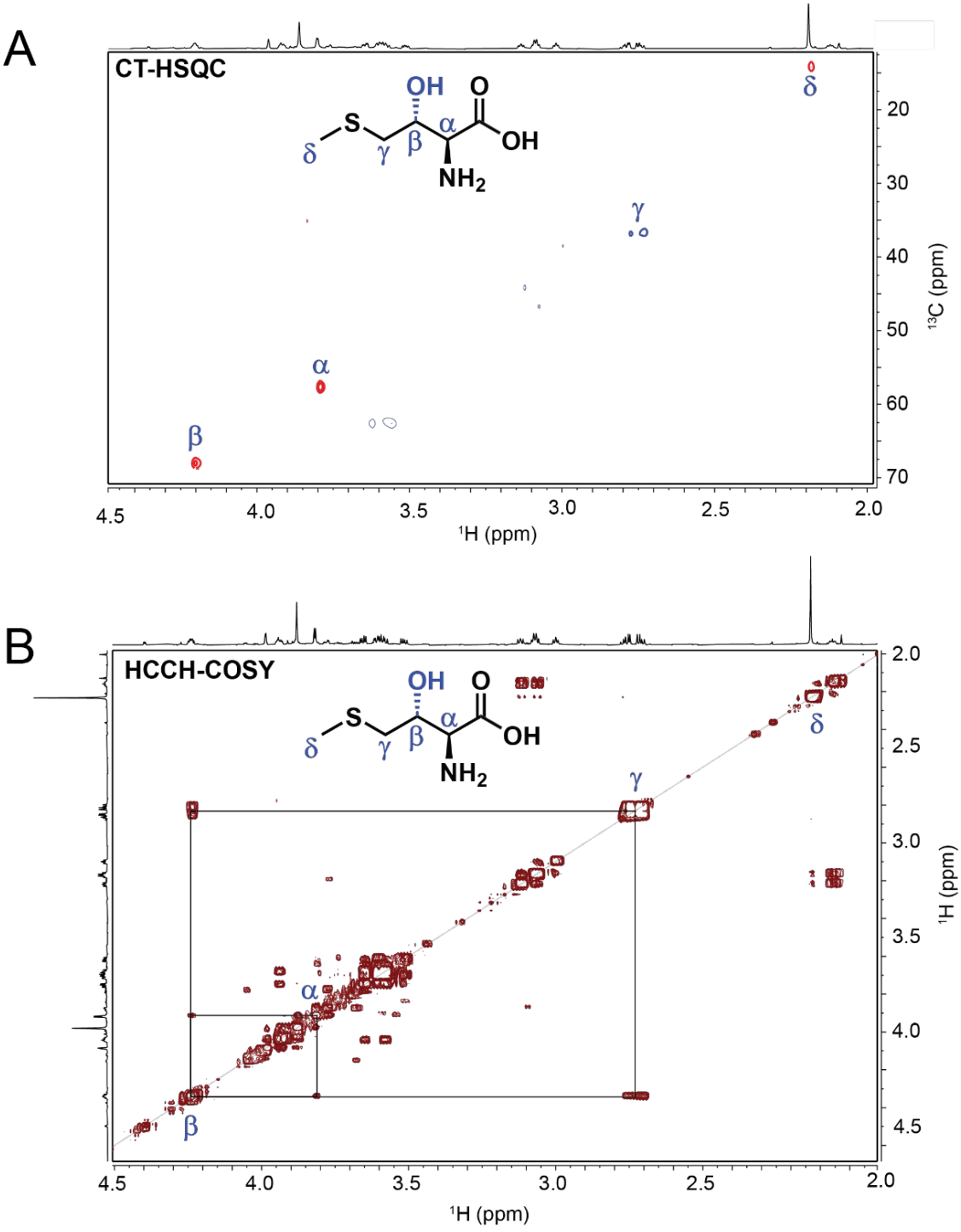


Figure 2.21. Discovery of a methionine hydroxylase AlfC. A biosynthetic gene cluster identified via tR3D contains a non-heme iron enzyme AlfC, which hydroxylates methionine (**10**) to produce (2*S*,3*R*)-hydroxymethionine (**11**), as detected by LC/MS.

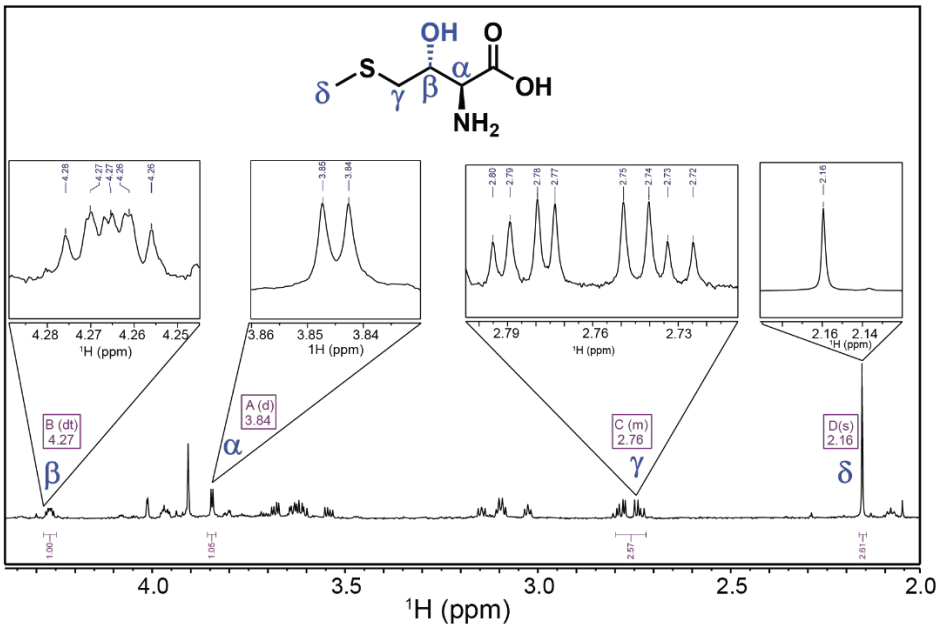
AlfC product	^{1}H (ppm)	^{13}C (ppm)
α	3.84	57.8
β	4.27	68.0
γ	2.74,2.78	36.8
δ	2.16	14.4

Table 2.3. Summary of NMR characterization of AlfC product. Through a combination of 1D and 2D experiments, the AlfC product was characterized as (2*S*,3*R*)-hydroxymethionine and not the nonenzymatic methionine sulfoxide product.

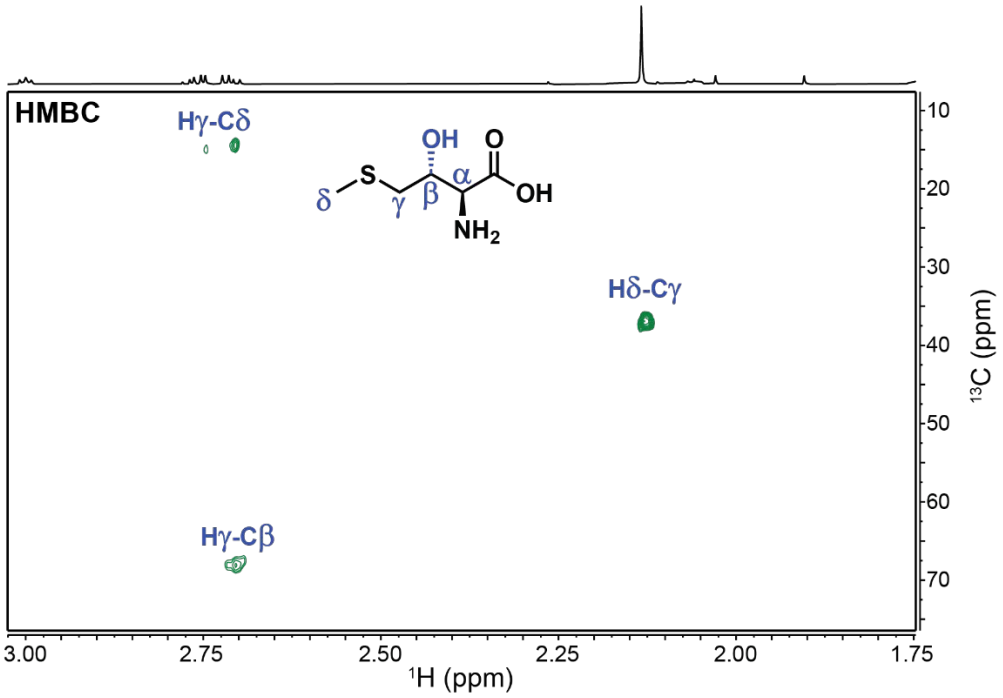
Figure 2.22. NMR characterization of (2*S*,3*R*)-hydroxymethionine. (A) 2D ^{1}H-^{13}C CT-HSQC was used to assign carbons α-δ, where phasing corresponds to CH and CH_3 (red) or CH_2 (blue) (B) 2D ^{1}H-^{13}C HCCH COSY was used to assign the site of the hydroxylation as C_β due to its relative downfield shift compared to C_α. (C) ^{1}H-NMR was used to calculate the *J*-values for each proton. (D) HMBC confirms the assignment of C_δ through three-bond correlations. These experiments were collected at 900 MHz. (E) Comparing the chemical shift of H_α and $J_{\alpha,\beta}$ to characterized β-hydroxy-α-amino acids (diastereomers of threonine) shows more similarity to the *anti* diastereomer, so we assigned the stereochemistry as (2*S*,3*R*)-hydroxymethionine [*57*].



C



D



E

Stereoisomer	H_α (ppm)	*J* (Hz)
AlfC product	3.84	$J_{\alpha,\beta}$ = 4.2
L-*allo*-threonine (*anti*)	3.83	$J_{\alpha,\beta}$ = 3.9-4
L-threonine (*syn*)	3.56	$J_{\alpha,\beta}$ = 4.9-5.0

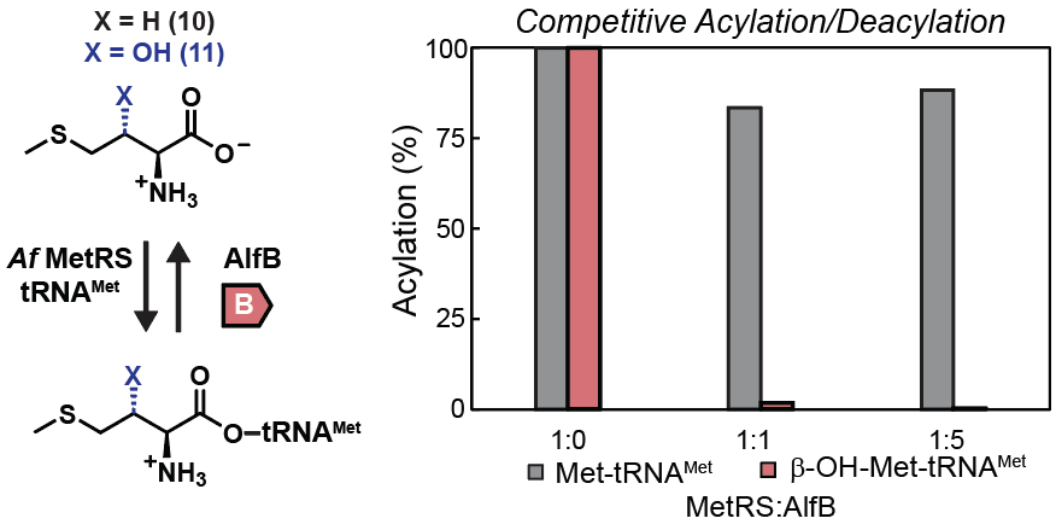


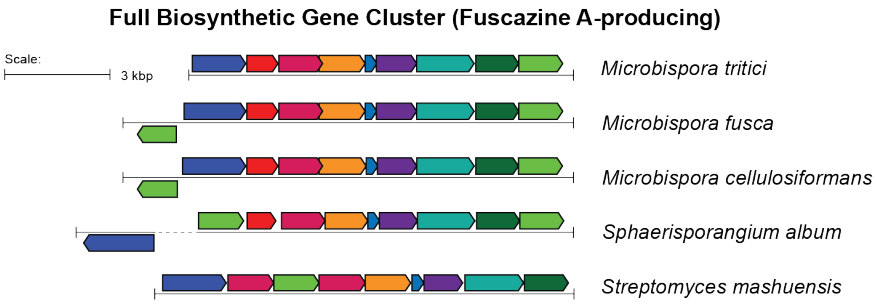
Figure 2.23. AlfB is a hydroxymethionine-tRNA^{Met} deacylase. With the methionyl-tRNA synthetase and tRNA^{Met} from the native organism *Alcaligenes faecalis*, both compound **10** and **11** could be acylated to tRNA, but **11** was preferentially hydrolyzed from tRNA^{Met} by AlfB. This finding supports our hypothesis of tRDs as resistance mechanisms to ncAA biosynthesis.

Alf B is a resistance gene to the product of AlfC. More importantly, with no prior insight into the function of AlfC beyond its classification as a dioxygenase, we could identify its function in producing a ncAA effectively through the co-localization of the tRD, demonstrating the utility of tR3D to discover new enzymes and natural products.

Identification of biosynthetic gene cluster for hydrazine-containing tripeptide. Following this discovery of a ncAA, we sought to expand this discovery approach towards a more complex ncAA-containing natural product. Using tR3D, we identified a conserved BGC containing a tRD with a non-heme iron/α-ketoglutarate-dependent dioxygenase along with a gene cassette for substituted hydrazine biosynthesis in 5 genomes (**Fig. 2.20E**, **Fig. 2.24**). This gene cassette consists of a flavin-dependent N-monooxygenase, a methionine-tRNA synthetase-like enzyme, and a cupin. This general biosynthetic strategy for hydrazine formation was first characterized fairly recently, where a hydroxylamine donor forms a nitrogen-nitrogen bond with the α-amine of a standard amino acid, with further enzymatic modification to form diverse products [*58,59*]. We were intrigued to see this cassette co-occuring with a tRD, as this suggested the BGC produces a hydrazine-containing ncAA. In addition to this gene cassette, this specific BGC contains two ATP-grasp enzymes, suggesting a peptide product with several functional groups. We chose the BGC from *Microbispora fusca* (Fuz BGC) for experimental characterization.

Taking a similar approach to the Alf cluster, we began exploring this cluster through individual *in vitro* reactions using purified enzymes with N-terminal His_6 tags heterologously expressed and purified from *E. coli* (**Fig A3.3**). The non-heme iron/α-ketoglutarate-dependent dioxygenase FuzA, when tested with all twenty standard amino acids and necessary cofactors, consumed arginine and produced a new compound +16 *m/z* larger, indicating hydroxylation (**Fig. 2.25A**). Using NMR spectroscopy, arginine hydroxylation was confirmed at C_β and we assigned the stereochemistry as (3*R*) (**Fig. 2.25B, 2.26, Fig. 2.27**). Notably, this is distinct to the well-characterized arginine (3*S*)-

A



Gene	Enzyme Family	Proposed Function
FuzA	αKG-dependent non-heme iron dioxygenase	Hydroxylase
FuzB	AlaX tRNA-deacylase	NcAA tRNA-deacylase
FuzC	ATP-grasp	NcAA/AA ligase
FuzD	Flavin-dependent N-monooxygenase	Lysine N^6-monooxygenase
FuzE	Cupin	Hydrazine-forming Protein
FuzF	Flavin Reductase	Hydrazine Reductase
FuzG	Methionine-tRNA Ligase	Hydrazine Synthetase
FuzH	ATP-grasp	NcAA/AA ligase
FuzI	MFS Transporter	Product Exporter



Streptomyces glebosus
Streptomyces qaidamensis
Streptomyces sp. 46
Streptomyces collinus
Streptomyces paradoxus
Streptomyces sp. W1SF4
Streptomyces sp. WAC07149
Streptomyces sp. WAC07061

B

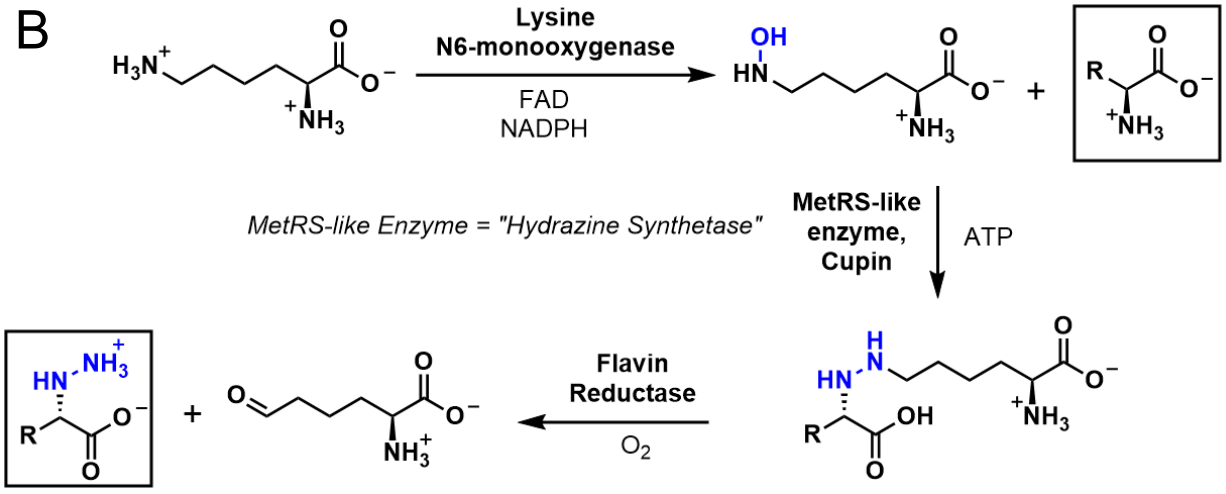


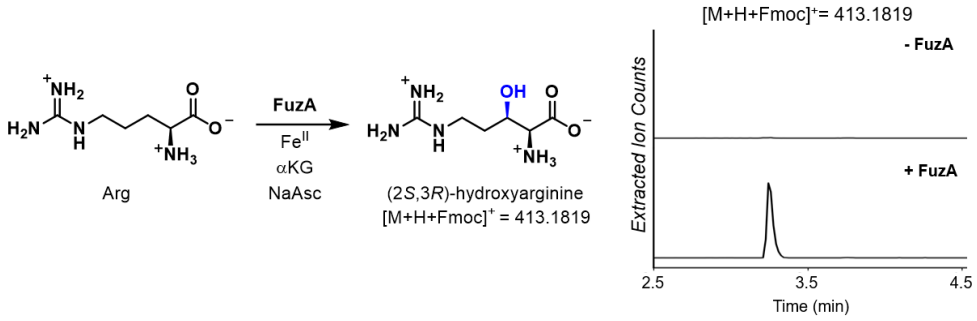
Figure 2.24. A novel BGC containing the requisite genes for hydrazine formation. (A) A conserved biosynthetic gene cluster (BGC) emerged from the tR3D analysis of SSN cluster 15, such as from *Microbispora fusca*, for generating a natural product that could contain a hydrazine moiety. (B) General biosynthetic logic for hydrazine, or N-N, formation [*58,59*], where a hydroxylamine-containing amino acid acts as a nucleophile towards an activated carboxylic acid of another amino acid, and the cupin rearranges the N-O bond to form the N-N bond. The flavin reductase then oxidizes one of the amines to release a hydrazine-containing amino acid.

hydroxylase VioC (35% identical) despite a very similar active site structure [*60*] (**Fig. 2.28**). In addition, this finding supports the hypothesis that this cluster encodes a ncAA-based natural product.

Additional insights were gained from *in vitro* characterizations of FuzC and FuzD. FuzC is a member of the ATP-grasp enzyme family, which typically form amide bonds through phosphorylation of a carboxylic acid followed by nucleophilic attack of an amine. When tested with the twenty standard amino acids and necessary cofactors, we found FuzC formed a dipeptide between arginine and alanine, with MS/MS indicating that arginine is the nitrogen donor to the amide bond **(Fig. 2.29A**). FuzD belongs to the flavin/NAD(P)H dependent family of N-monooxygenases, and we hypothesized that this enzyme would accept lysine and/or ornithine [*61*]. We performed *in vitro* reactions with the necessary cofactors and lysine or ornithine, and by LC/MS-QTOF we detected hydroxylated lysine but no activity towards ornithine, suggesting lysine is the hydroxylamine donor to the hydrazine compound (**Fig. 2.29B**). The remainder of the enzymes in this BGC were hypothesized to act on metabolites produced later in the pathway.

To assess hydrazine formation and eventual product elucidation, we assembled the entire BGC in stages into plasmids for heterologous expression in *E. coli*. When the minimal cassette for hydrazine biosynthesis (FuzD/E/G) was expressed with FuzA and cell supernatant was analyzed by LC/MS-QTOF, we detected the exact mass for hydrazinyl-hydroxyarginine-lysine (**Fig. 2.30A**). Notably, this is the first report of a hydrazine synthetase naturally accepting a ncAA. When FuzA was not present, we detected the expected mass for the hydrazinyl-arginine-lysine dipeptide, demonstrating the hydrazine synthetase FuzG can accept either the ncAA or canonical version of arginine. With purified enzymes, we could reconstitute the hydrazinyl-arginine-lysine dipeptide *in vitro*, indicating these enzymes are sufficient to produce the products we detected *in vivo* (**Fig. 2.30B**).

A



B

Position	^{1}H (ppm)	^{13}C (ppm)	*J* (Hz)
α	3.63	59.22	$J_{\alpha,\beta}$ = 5.1
β	4.1	66.77	$J_{\alpha,\beta}$ = 5.1, $J_{\beta,\gamma}$ = 10.2, $J_{\beta,\gamma}$ = 3.1
γ	1.97, 1.79	31.71	$J_{\beta,\gamma}$ = 10.2, $J_{\beta,\gamma}$ = 3.2, $J_{\gamma,\delta}$ = 6.1, 7.8, 14.1, 15.4
δ	3.36	37.03	$J_{\gamma,\delta}$ = 6.1, 7.8, 14.1, 15.4

Figure 2.25. Discovery of a unique arginine hydroxylase. (A) *In vitro* screening with all 20 amino acids and necessary cofactors revealed FuzA is an arginine hydroxylase. Product was Fmoc-derivatized prior to LC/MS analysis. (B) Summary of all chemical shifts and *J*-values from NMR analysis of the FuzA product suggests *3R* stereochemistry, which is a unique reaction outcome for Fe/αKG arginine hydroxylases.

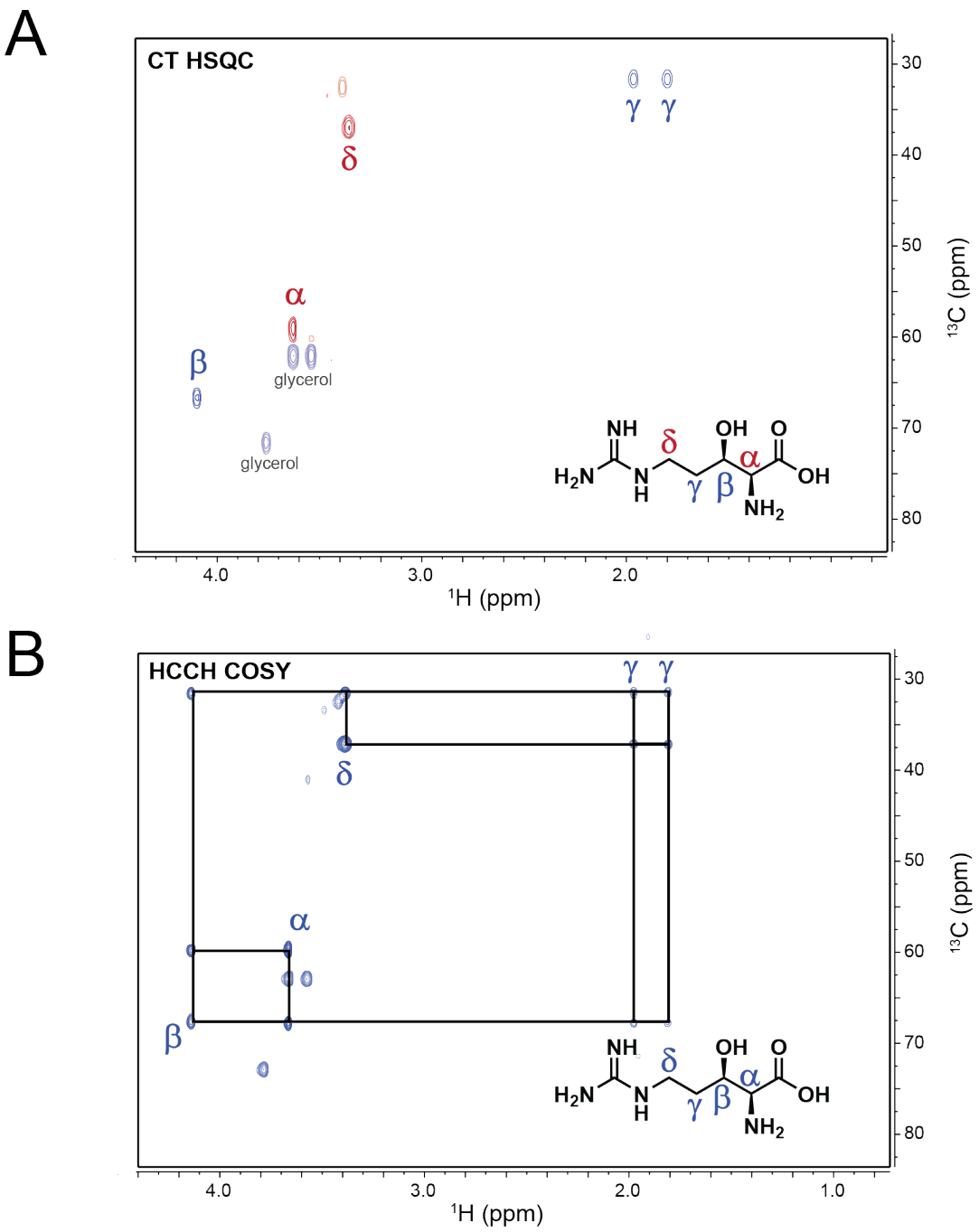
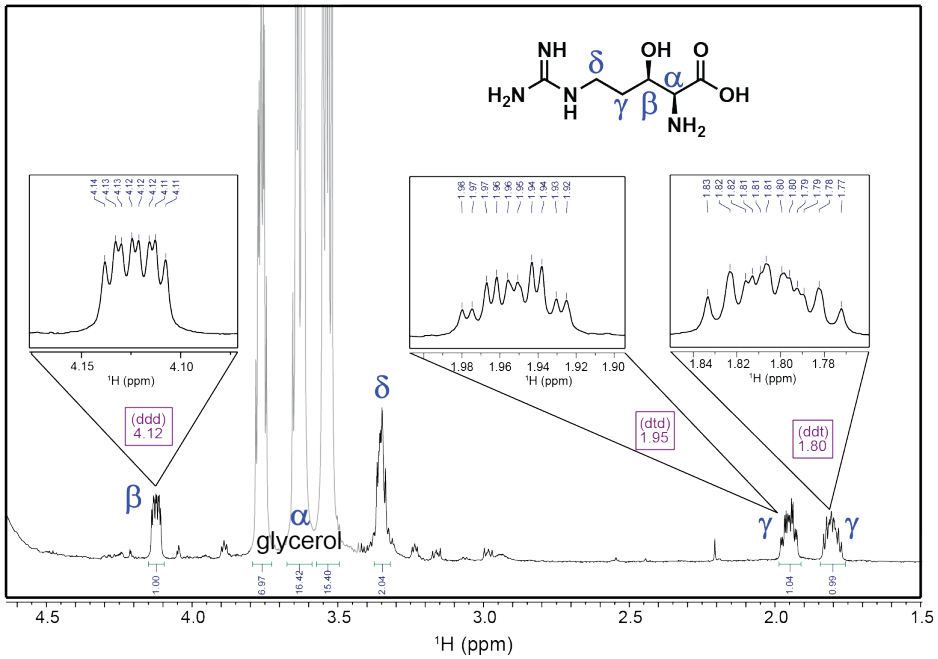
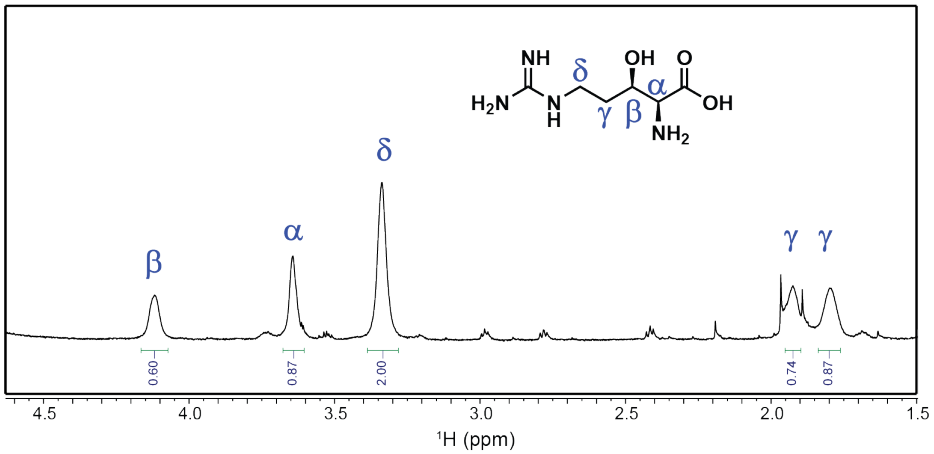


Figure 2.26. 2D NMR analysis of (2*S*,3*R*)-hydroxyarginine. (A) 2D ^{1}H-^{13}C CT-HSQC was used to assign shifts for carbons α-δ. In this experiment, carbons with two neighbors (blue) have the opposite phase of carbons with one or three neighbors (red). (B) 2D ^{1}H-^{13}C HCCH COSY was used to assign the connectivity of carbon/hydrogens α-δ and the site of the hydroxylation as C_{β}, as C_{β}/H_{β} is downfield shifted from C_{α}/H_{α} and shows connectivity with both C_{α}/H_{α} and C_{γ}/H_{γ}s. The HSQC and COSY experiments were recorded at 900 MHz.

A



B



C

Stereoisomer	H_α (ppm)	J (Hz)
FuzA Product	3.63	$J_{\alpha,\beta}$ = 5.1
(2*S*,3*R*)-hydroxyarginine	3.67	$J_{\alpha,\beta}$ = 5.3
(2*S*,3*S*)-hydroxyarginine	3.90	$J_{\alpha,\beta}$ = 2.6

Figure 2.27. 1D NMR analysis of (2*S*,3*R*)-hydroxyarginine. (A) Following IEX purification of the FuzA product, a ^{1}H-NMR spectrum was recorded at 600MHz and multiplets were used to calculate *J*-values. H_α is obscured by residual glycerol. (B) An additional round of purification removed the glycerol and showed the chemical shift for H_α. This experiment was recorded at 700 MHz with solvent suppression. (C) The chemical shift of H_α=3.68 ppm and $J_{\alpha,\beta}$ = 5.1 are more similar to the values for the (2*S*,3*R*) diastereomer of hydroxyarginine, of which there is a synthetic standard [*62*]. This is a novel reaction outcome for a Fe/αKG arginine hydroxylase.

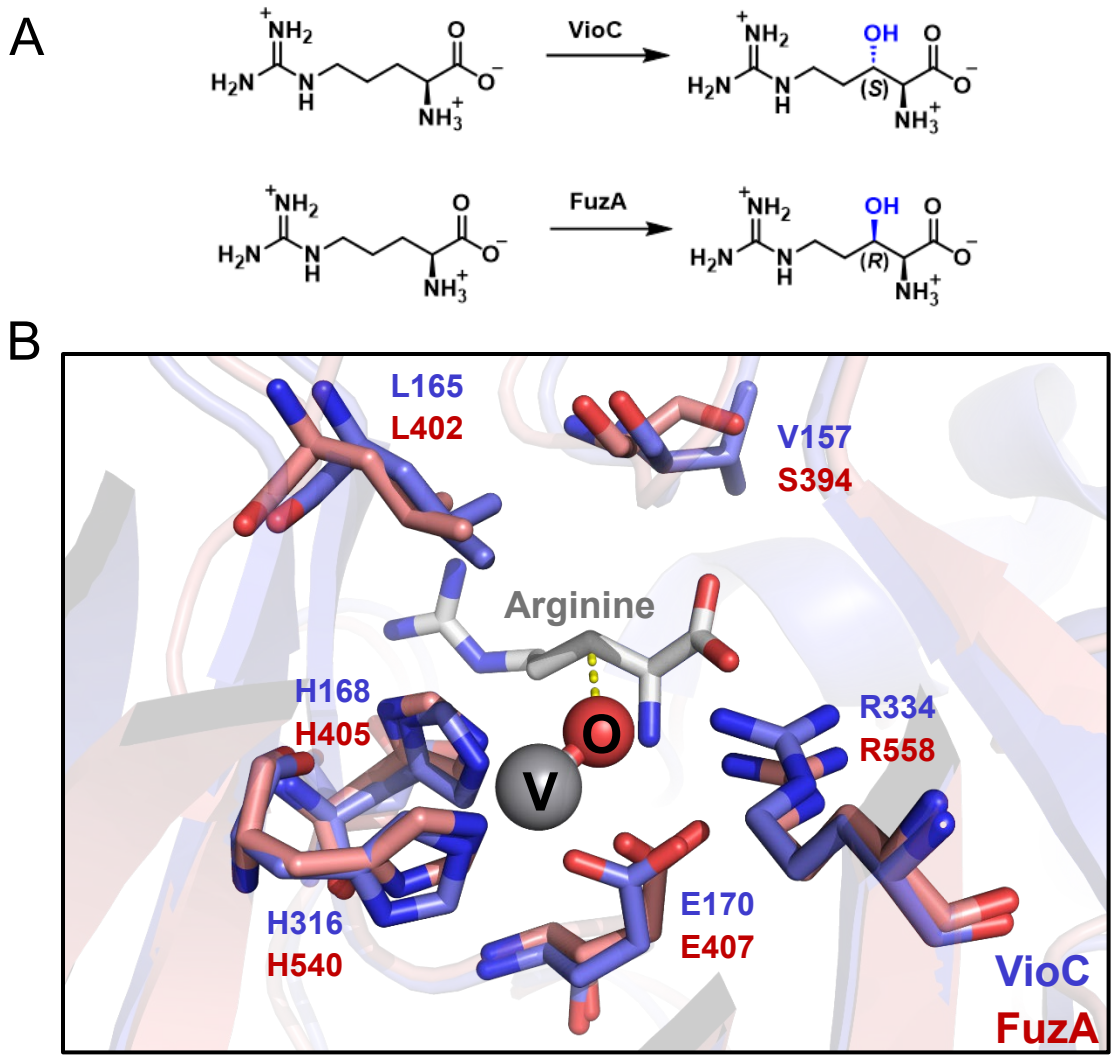


Figure 2.28. Protein structure comparison of FuzA and VioC. (A) VioC catalyzes a hydroxylation to form the alternate diastereomer to FuzA's product [*60,63*]. (B) An Alphafold2-generated structure of FuzA overlaid with a solved crystal structure of VioC shows that while the two enzymes have low sequence similarity (35%), VioC and FuzA have very similar active sites, with the most notable difference being a V to S mutation near the site of hydroxylation. This structure of VioC contains a vanadyl-oxo as a mimic of the iron-oxo reactive species. Despite very similar active site structures, FuzA achieves a different reaction outcome than VioC. Site-directed mutagenesis and would be needed to determine the basis for this alternate selectivity.

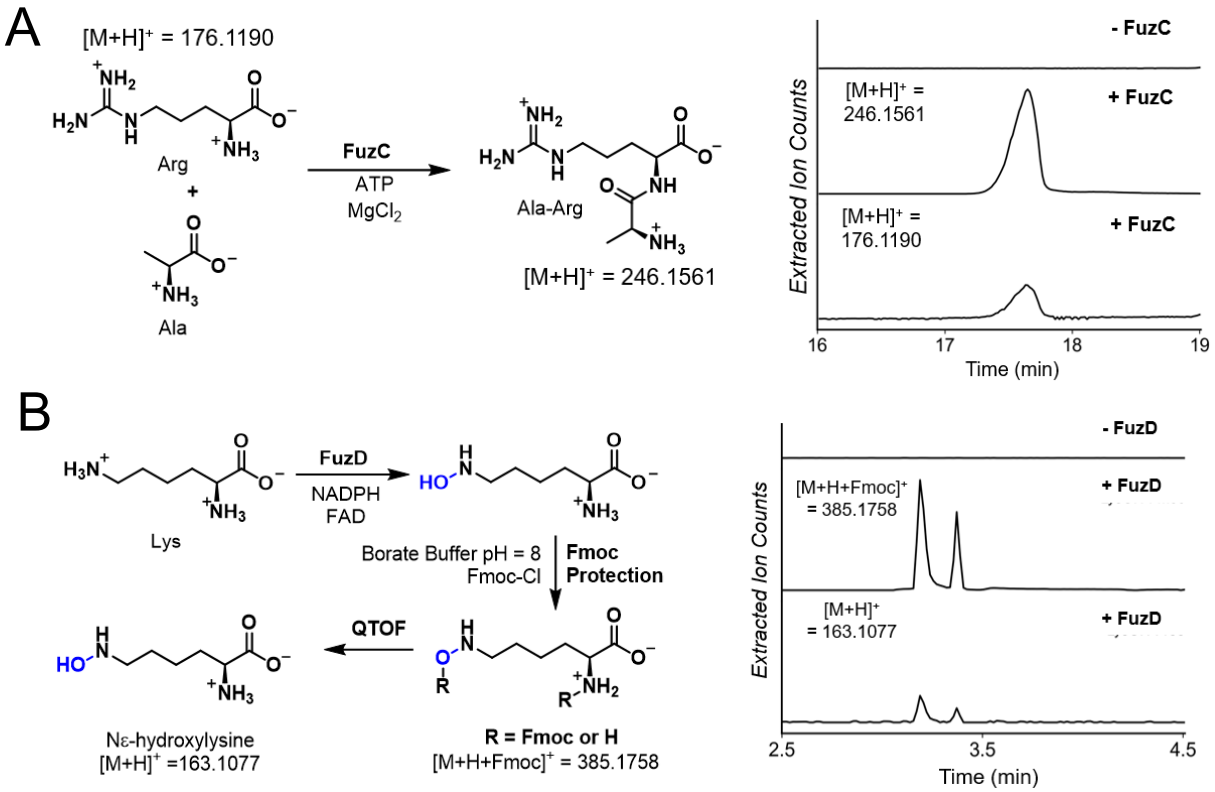


Figure 2.29. *In vitro* characterization of FuzC and FuzD provide insights into the natural product encoded by the Fuz BGC. (A) *In vitro* reaction with the ATP-grasp FuzC showed FuzC is an Arg-Ala ligase. We assigned arginine as the N-donor to the peptide bond as this fragment was detected at the same retention time. (B) FuzD shows monooxygenase activity towards lysine but not ornithine. Fragmentation of the Fmoc group reveals the unprotected mass.

Following hydrazine characterization, we performed comparative metabolomics with the full BGC encoded across three plasmids to elucidate the final product. All of the biosynthetic enzymes were cloned onto pDuet expression vectors (FuzA-G except FuzB), and conditions contained various combination of plasmids to determine which products are unique to different enzymes. The cultures were supplemented with $^{13}C_6$,$^{15}N_4$ arginine and were grown for two days in minimal media. Following harvesting, Fmoc protection, and LC/MS analysis, results were analyzed with MS-Dial to identify unique metabolites and isotopic enrichments.

In the condition with all of the enzymes present, we identified a unique metabolite "A" with $[M+H+Fmoc]^+$ *m/z* = 610.2970 and the molecular formula $C_{15}H_{29}N_7O_5$ (**Fig. 2.31A**). When FuzA was not present and only FuzB-G were expressed, we identified a unique metabolite "B" with $[M+H+Fmoc]^+$ *m/z* = 594.3025 and the formula $C_{15}H_{29}N_7O_4$, which is smaller from metabolite A by a single oxygen atom. In addition, by MS-Dial we detected both the unlabelled and isotopically-labelled masses for these compounds, indicating arginine incorporation (**Fig. 2.31B, Table A2.5**). From these observations, we hypothesized we found both the FuzA-hydroxylated version and unmodified version of the product of this BGC, which we termed fuscazine, and gave descriptors A and B for the hydroxylated and nonhydroxylated versions, respectively.

To identify the core scaffold of the fuscazines, we performed a series of feeding experiments with isotopically-labelled amino acids and analyzed the isotopic enrichments by LC/MS-QTOF of our Fuscazine B-producing strain. We detected arginine incorporation from our original metabolomics

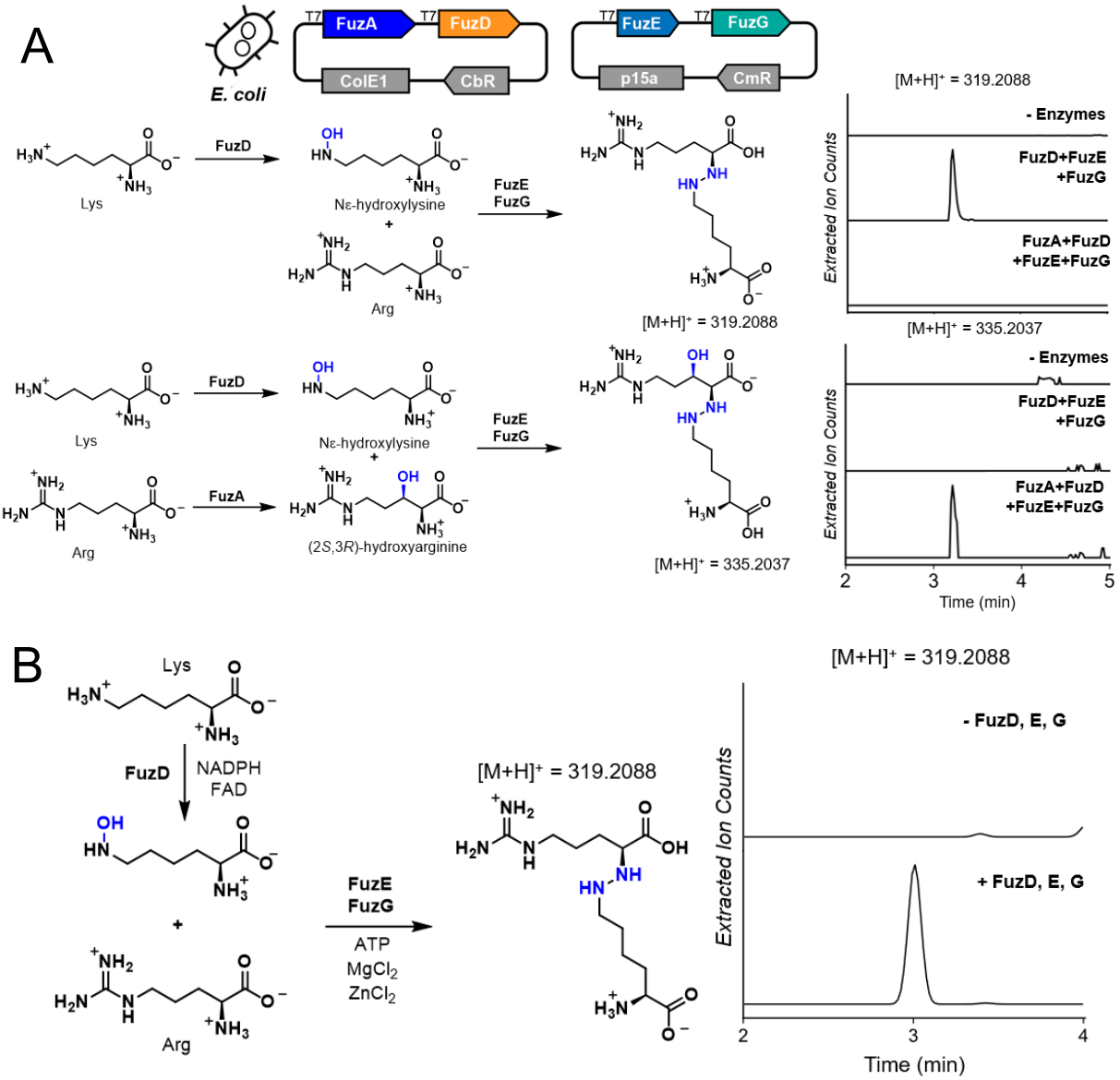
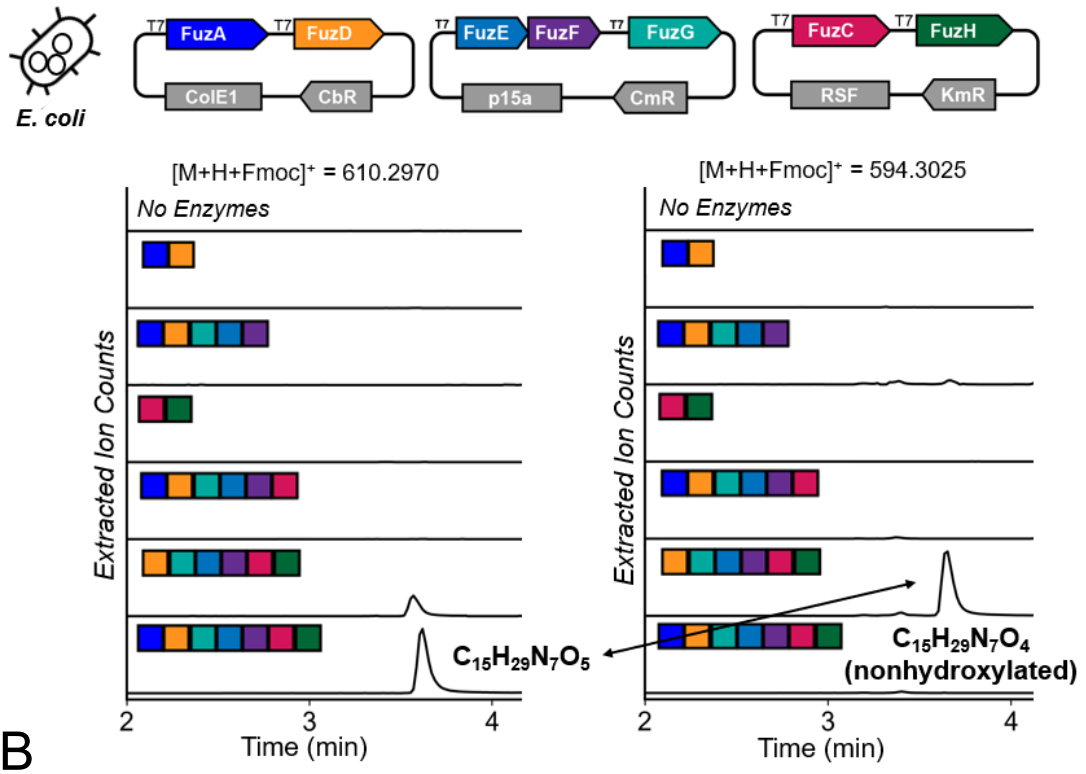


Figure 2.30. *In vivo* detection of hydrazine-containing metabolites from the Fuscazine BGC. (A) The hydrazine-forming machinery of the Fuz BGC was heterologously expressed in *E. coli* and the results show both arginine and (2*S*,3*R*)-arginine are accepted by FuzG to form hydrazinyl-(hydroxy)arginine-lysine dipeptides, depending on if FuzA was present. (B) Reconstitution of the hydrazine-containing dipeptide with purified enzymes from the Fuz BGC indicate FuzD/E/G are sufficient to yield the hydrazine dipeptide detected *in vivo.* Slight differences in retention time are attributed to day-to-day variations in chromatography.

A



B

RT (min)	*m/z*	Type	ID	ANOVA *p*-value	Signal/ Noise	Fold Increase
3.67	604.3123	[M+H+Fmoc]$^+$	Fuscazine B (+$^{13}C_6$,$^{15}N_4$ Arg)	7.41E-02	161493	8099724
3.45	385.1758	[M+H+Fmoc]$^+$	Nε-hydroxy-Lysine	3.94E-06	51405.6	279.58
3.78	411.1913	[M+H+Fmoc]$^+$	Hydrazinyl-Arginine	4.23E-10	42160.5	332.61
4.68	130.0856	[M+H]$^+$	Pipecolic Acid	7.78E-06	35242.5	29.34
3.67	594.3025	[M+H+Fmoc]$^+$	Fuscazine B	1.07E-01	35203.4	875.94
3.6	620.3055	[M+H+Fmoc]$^+$	Fuscazine A (+$^{13}C_6$,$^{15}N_4$ Arg)	1.04E-01	18857.8	303.86
3.61	610.2970	[M+H+Fmoc]$^+$	Fuscazine A	1.62E-01	15354.6	119.19

Figure 2.31. Comparative metabolomics reveals products of the Fuz BGC. (A) The Fuz BGC was cloned onto three plasmids and metabolomics conditions contained various combinations of these plasmids to delineate which products correspond to each enzyme. With all of the enzymes present, we identified a [M+H+Fmoc]$^+$ = 610.2970, and when FuzA was absent, a [M+H+Fmoc]$^+$ = 594.3025 that is smaller by one oxygen. These masses indicate we found both the FuzA-hydroxylated and nonhydroxylated product of this BGC, which we termed Fuscazine (A = hydroxylated, B = nonhydroxylated). (B) Output from MS-Dial of the most abundant metabolites across conditions, revealing arginine incorporation into the fuscazines through the +10 mass increase from isotopic labeling. Components of the fuscazines were in high abundance as well. Fuscazine B is in much higher abundance than Fuscazine A likely due to the abundance of arginine compared to hydroxyarginine in the cell.

experiment, as seen through the 10 Da mass enrichment from $^{13}C_6$, $^{15}N_4$-arginine, indicating arginine is fully incorporated into the fuscazines (**Fig. 2.31B, Fig. 2.32**). We hypothesized alanine was also a component, and when we supplied $^{13}C_3$,$^{15}N_1$-alanine, we detected a mass enrichment of 4 Da, indicating alanine is fully incorporated into the fuscazines (**Fig. 2.32**). In similar logic, we hypothesized lysine would contribute a mass enrichment of 1 Da from the epsilon amine forming the hydrazine bond. When we supplied $^{13}C_6$,$^{15}N_2$-lysine, we detected a 1 Da enrichment supporting hydrazine formation, yet we were intrigued to find additional enrichments of 7 Da and 8 Da (**Fig. 2.32**). This data suggests an entire lysine is incorporated into the fuscazines, with the 1 Da and 7 Da indicating lysine is likely added as two fragments in discrete steps. Altogether, the labelling data shows the fuscazines contain arginine, alanine, and lysine, which account for all of the carbons and nitrogens in the products. The formulas for the fuscazines yield a degree of unsaturation of 4, yet the individual amino acids only contribute 1 degree of unsaturation each. We hypothesized this additional degree is due to a desaturation or cyclization.

We re-examined the possible reactions in this BGC to determine the source of this additional degree of unsaturation. From this, we noted the flavin oxidase produces two products, the hydrazine-containing arginine and an aldehyde-containing derivative of lysine, α-aminoadipic semialdehyde. We hypothesized this lysine derivative can cyclize to generate an imine and a promiscuous *E. coli* imine reductase could form L-pipecolic acid (Pip) *in situ* (**Fig. 2.33**) [*64*]. Production and incorporation of Pip satisfies the +7 enrichment from lysine from the isotopic-labelling study and the 4th degree of unsaturation. Altogether, these results led us to propose the fuscazines are tripeptides with components pipecolic acid, alanine, and 2*S*,3*R*-hydrazinyl-hydroxyarginine (fuscazine A) or 2*S*-hydrazinyl-arginine (fuscazine B) (**Fig. 2.34**).

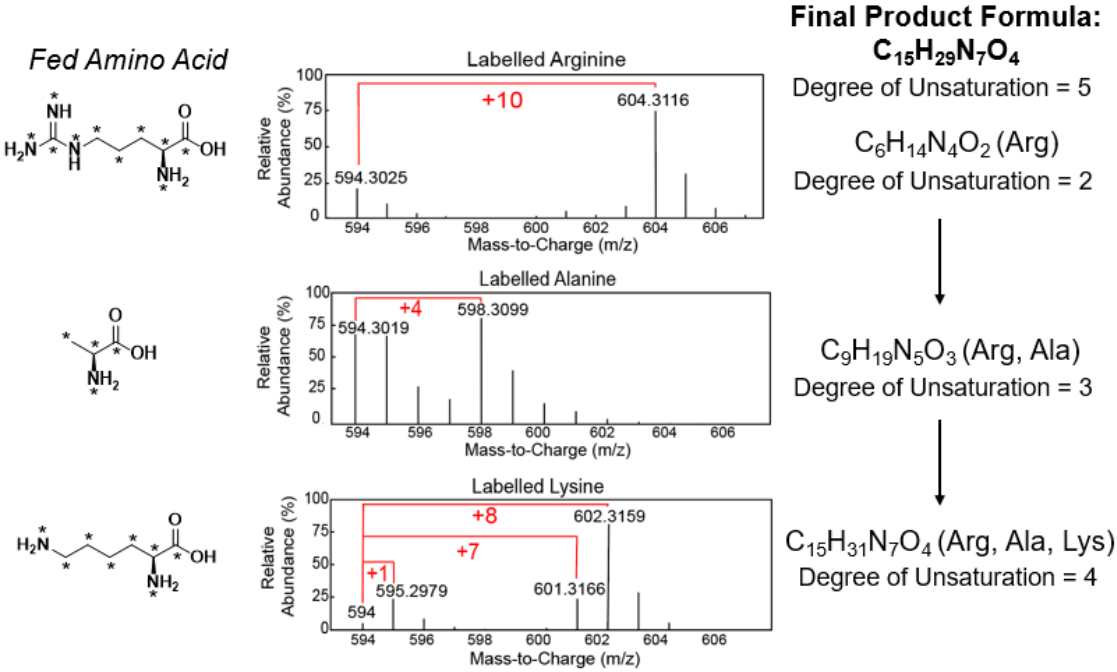


Figure 2.32. Isotopic-labeling study to identify the components of the fuscazines. Supplying a fuscazine A-producing strain with isotopically-labelled amino acids (heavy atoms indicated by asterisk) revealed the components to be arginine, lysine, and alanine. The additional degree of desaturation calculated for the fuscazines compared to these starting materials suggested an additional cyclization or desaturation.

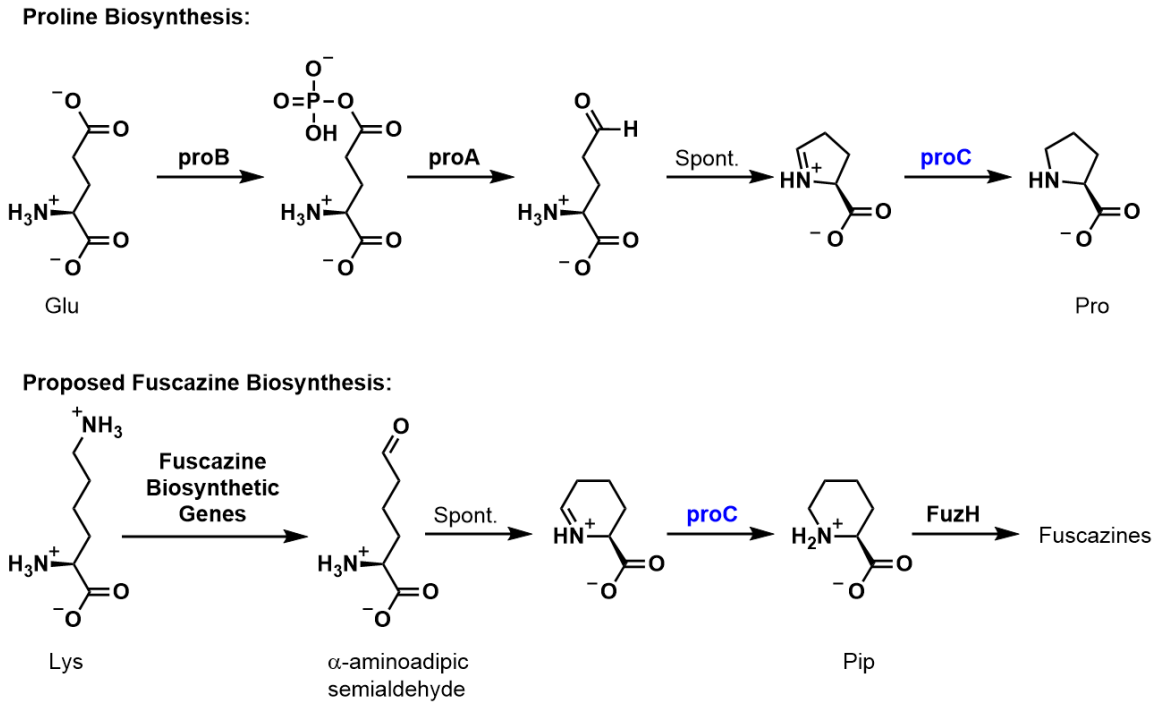


Figure 2.33. Proposed route for pipecolic acid biosynthesis via combination of Fuz cluster and native *E. coli* genes. Pipecolic acid is hypothesized to be generated from lysine by the Fuz BGC through a proline biosynthetic gene proC, which has been shown previously to reduce the cyclized product from α-aminoadipic semialdehyde [*64*].

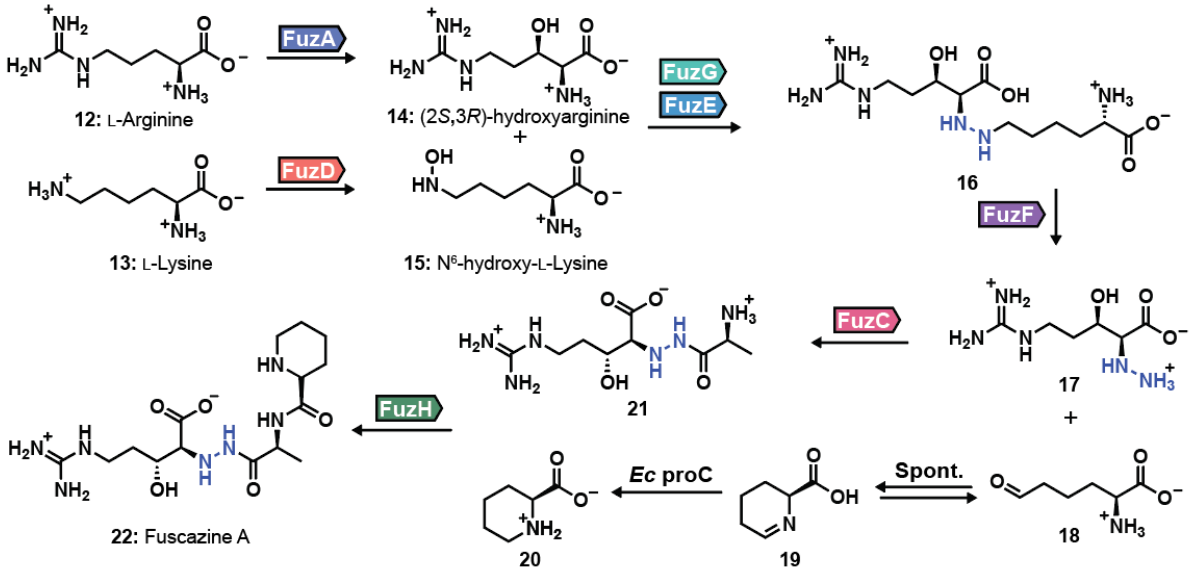


Figure 2.34. Full biosynthetic pathway for Fuscazine A. Beginning from arginine, lysine, and alanine, a combination of hydroxylations, hydrazine formation, and production of pipecolic acid yields the tripeptide fuscazine A. If FuzA is not present, the enzymes are promiscuous enough to form fuscazine B, which contains nonhydroxylated arginine.

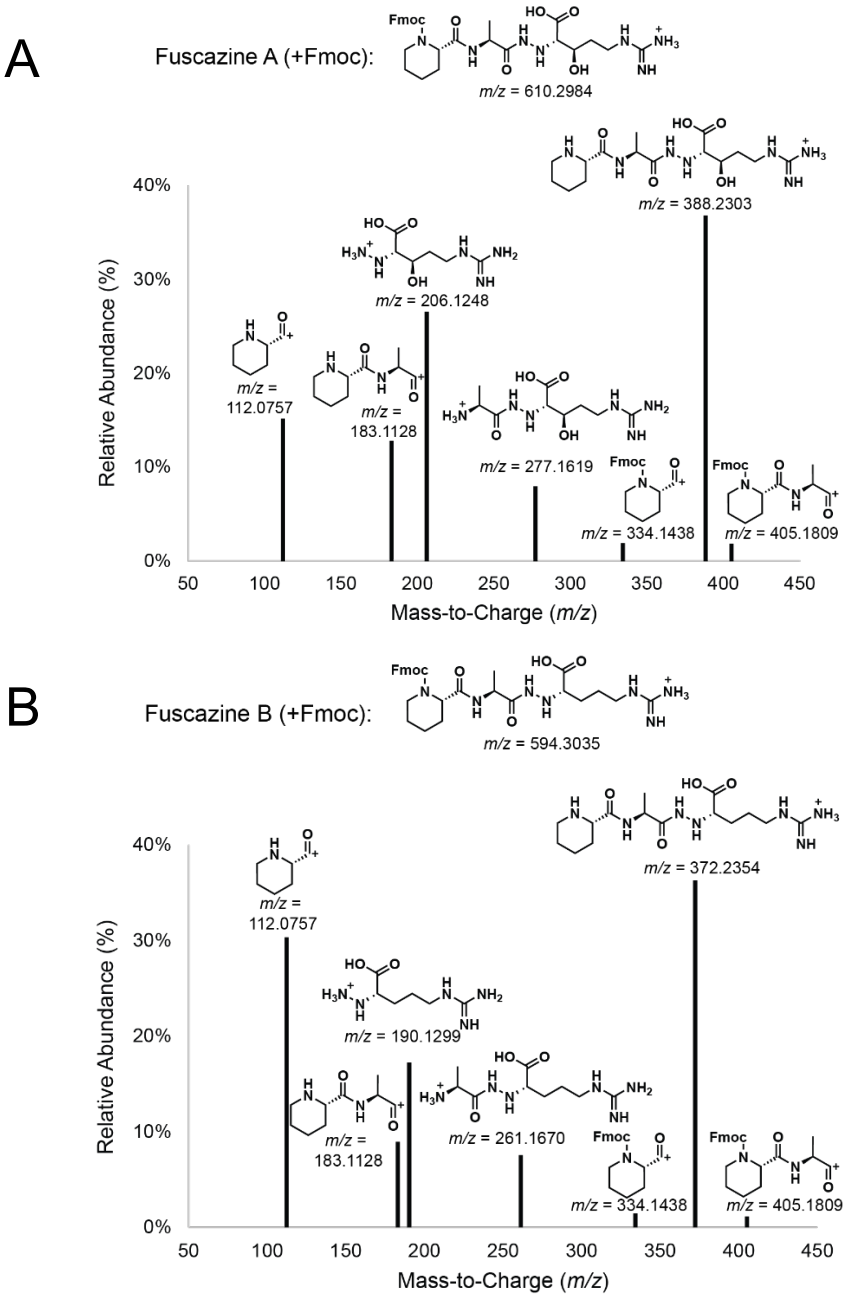


Figure 2.35. Structural analysis of Fuscazines A and B. (A) Fragmentation of Fuscazine A shows connectivity to be pipecolic acid-alanine-hydrazinyl-hydroxyarginine (N to C) and shows hydrazinyl-hydroxyarginine as an independent fragment. Fragments of pipecolic acid-hydroxyarginine were not detected, indicating these two amino acids are not connected. (B) Fragmentation of Fuscazine B shows connectivity as pipecolic acid-alanine-hydrazinyl-arginine (N to C) and shows hydrazinyl-arginine as a fragment.

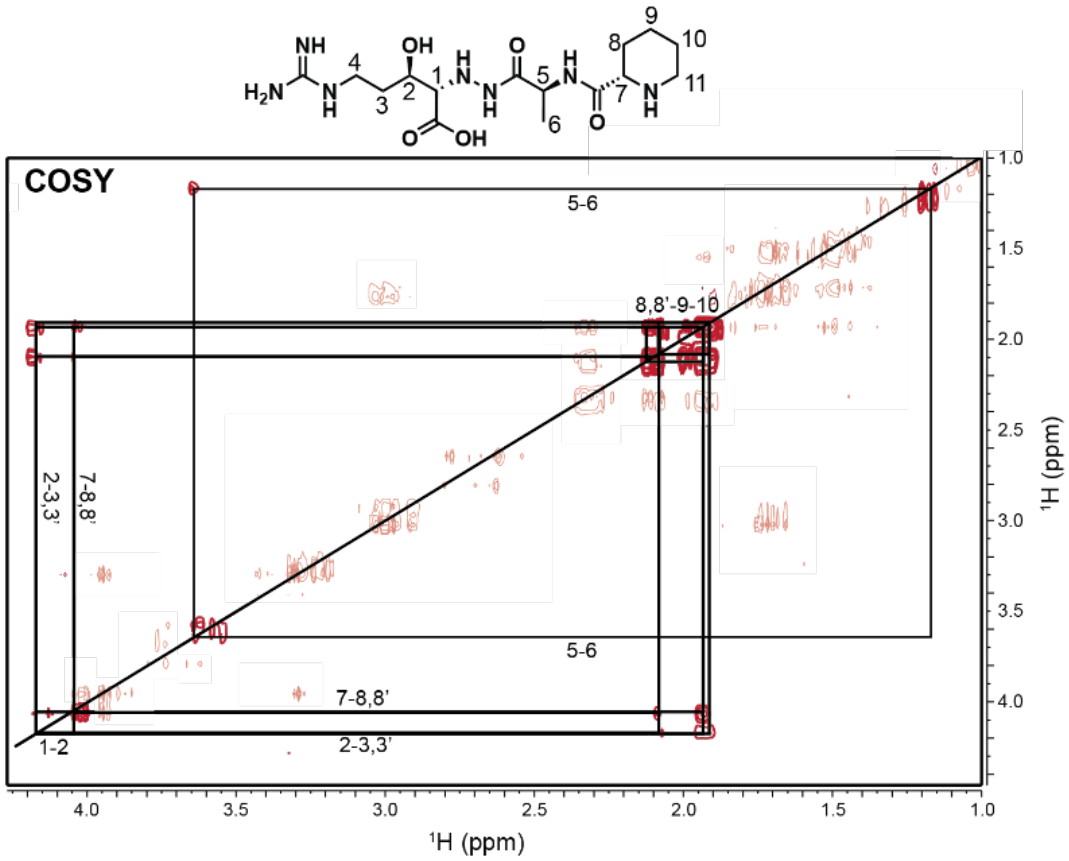


Figure 2.36. NMR analysis of Fuscazine A. 2D COSY experiment shows correlations (highlighted) similar to what we would expect for H_α (1), H_β (2), and H_γs (3) of hydroxyarginine; H_α (5) and H_βs (6) of alanine; and H_α (7), H_βs /H_γs /H_δs (8/9/10) of pipecolic acid, supporting the presence of these three amino acids in Fuscazine A.

To complement the isotopic feeding experiments, we elucidated the connectivity of the fuscazines with LC/MS-MS (**Fig 2.35**). This experiment revealed fragments (N to C) including hydrazinyl-(hydroxy)-arginine, alanine-hydrazinyl-(hydroxy)-arginine, pipecolic acid-alanine, and an N-terminal pipecolic acid. These results led us to propose the connectivity of **22** in **Fig. 2.34**. Notably, no fragments were detected containing hydrazinyl-(hydroxy)-arginine-pipecolic acid, providing evidence that alanine is between these two components in the fuscazines. To supplement this analysis, NMR analysis of a crude purification of Fuscazine A showed shifts for what we would expect of a tripeptide (**Fig. 2.36**).

From our elucidated pathway, several novel enzymatic discoveries were made. The structures of the fuscazines indicate FuzC, an ATP-grasp ligase, can accept a hydrazine nucleophile to form a hydrazide in structure **21**, which is the first known example of this transformation in nature (**Fig. 2.34**). In addition, the final product indicates the remaining ATP-grasp ligase FuzH is a pipecolic acid-peptide ligase (**Fig. 2.34**). Specifically, we propose FuzH accepts the dipeptide alanine-hydrazinyl-arginine as the amine donor to make a peptide bond with pipecolic acid and form the fuscazines. When we performed an *in vitro* reaction with either FuzH or FuzH and FuzC, along with alanine, arginine, and pipecolic acid, we could detect the corresponding mass for the peptide backbone of the fuscazines (**Fig. 2.37A**). This supports our assignment of these enzymes' function. Notably, FuzH can accept both a dipeptide or a single amino acid as the amine donor (**Fig. 2.37A**).

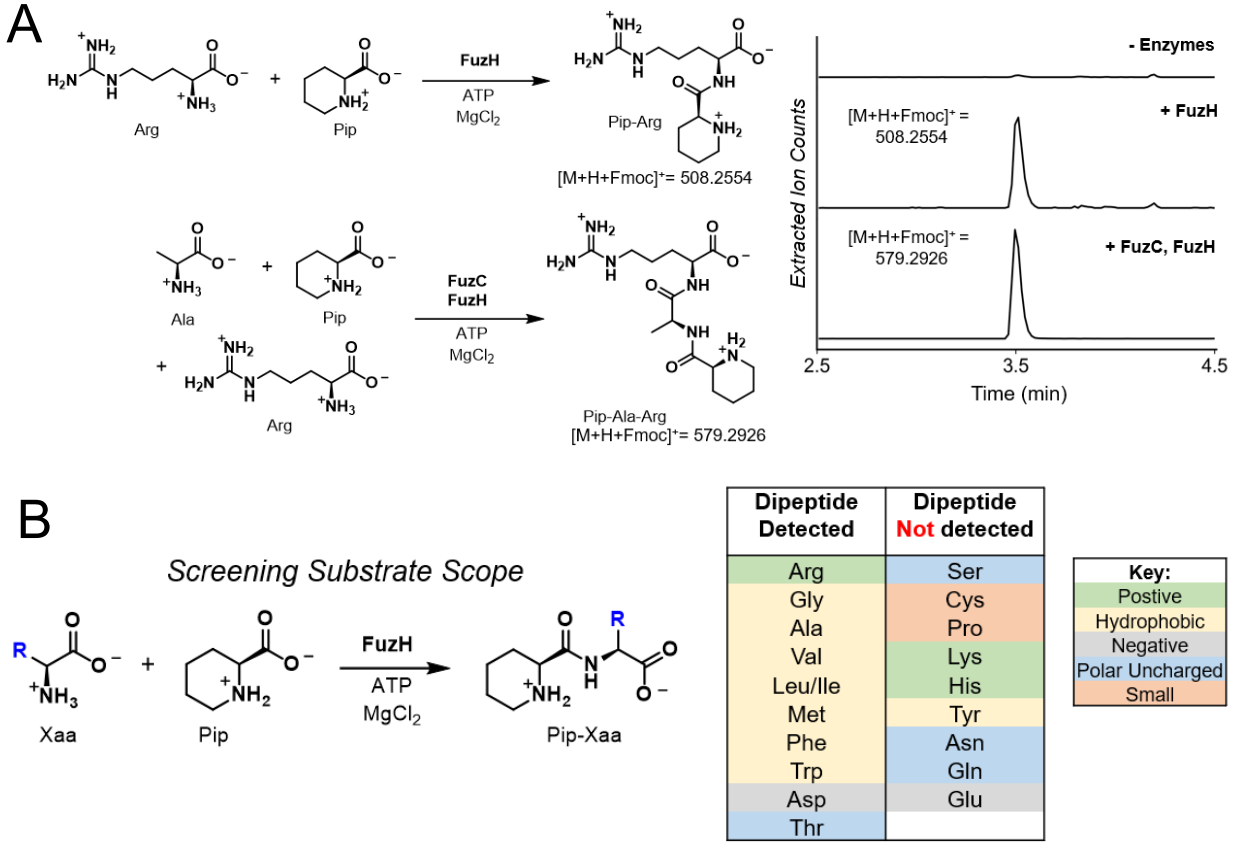


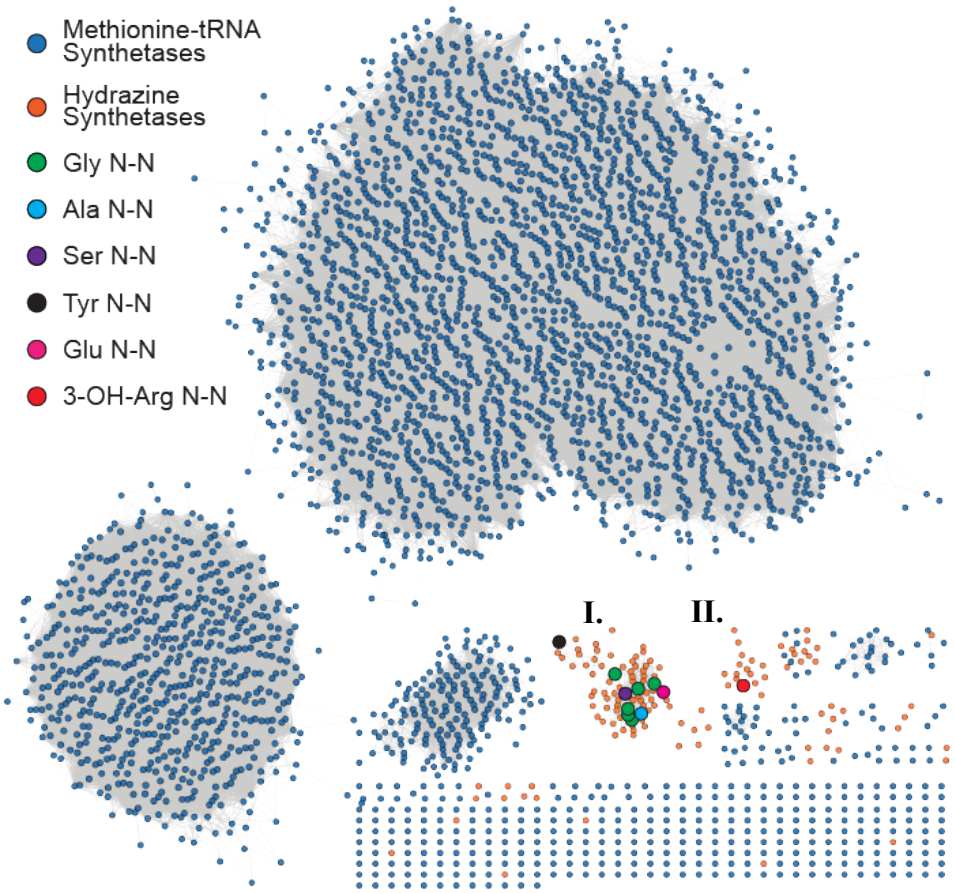
Figure 2.37. *In vitro* reactivities of the ATPG-grasp enzymes from the Fuz cluster. (A) FuzH shows activity as a pipecolic acid (Pip) ligase to Arg, as well as Ala-Arg when FuzC is present. When both FuzC and FuzH are present in one pot, these enzymes can form the peptide backbone of the fuscazines (B) FuzH forms dipeptides between Pip and a range of amino acids when screened with all 20 proteinogenic amino acids. Products were Fmoc-derivatized prior to LC/MS analysis and identified by exact mass.

With the function of FuzH elucidated, we were curious to explore the promiscuity of this enzyme for other amino acids. We performed an *in vitro* substrate screen with all 20 amino acids along with pipecolic acid. We found this enzyme ligated pipecolic acid to a range of amino acids, from glycine to tryptophan, indicating FuzH has a large binding pocket for the amine donor (**Fig. 2.37B**). With the ability of FuzH to ligate amino acids to the N-terminus of peptides or diverse amino acids, we anticipate FuzH is a promising biocatalyst.

Another novel discovery from the Fuz BGC was the elucidation of FuzG, which is the first reported hydrazine synthetase capable of accepting a ncAA. We were curious to explore how this enzyme is distinct from all other hydrazine synthetase discovered to date. For comparison, we visualized the entire MetRS-like family, to which the hydrazine synthetases belong, as a sequence similarity network (SSN) and identified specific clusters in the SSN that contain the hydrazine synthetases (**Fig. 2.38**). Notably, FuzG clusters separately from all other characterized examples, indicating some differences in sequence. To gain an understanding of FuzG's basis for selectivity, we performed a docking experiment with (2*S*,3*R*)-hydroxyarginine (**Fig. 2.39**). From this, we identified Asn312 as the likely residue responsible for β-hydroxy recognition. Finally, we employed the tR3D tool to identify conserved neighboring genes of putative hydrazine synthetases, and this showed a diversity of genomic contexts that could form novel natural products (**Fig. 2.40**).

Figure 2.38. Sequence similarity network of hydrazine synthetases. (A) SSN of PF09334 (Methionine-tRNA Synthetases), annotated with known hydrazine synthetases to date (cluster I) and a new cluster that contains FuzG (cluster II). (B) Quantification of all putative hydrazine synthetases within the MetRS family. (C) A close-up view of cluster I and II. (D) Descriptions of characterized hydrazine synthetases.

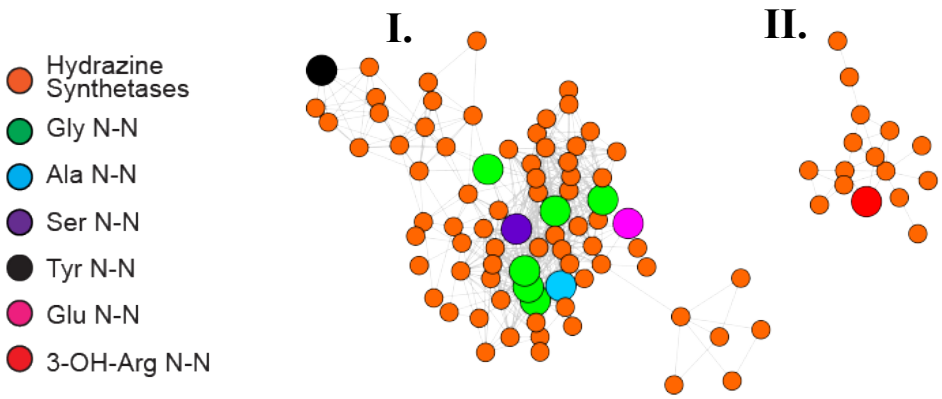
A



B

PF09334 SSN	Nodes (No.)	Represented Sequences (No.
Initial Dataset (accessed Oct. 2022)	4624	84733
Remaining after removing singletons	3731	83840
Found neighboring PF13434 (LysOH) AND PF07883 (Cupin)	135	1226
Known NN cluster I	80	667
Unknown NN cluster II with Fuz Cluster	16	90
Other putative NN-forming clusters	39	469

C



D

Gene ID	Organism	Protein Type	Amino Acid Specificity	Color
Q8KGM6	*Rhizobium loti*	Cupin-MetRS	Glycine	Green
A0A2B9TI29	*Bacillus cereus*	Cupin-MetRS	Glycine	Green
A0A291T5V8	*Streptomyces malaysiensis*	MetRS	Glycine	Green
A0A423LFS6	*Pseudomonas fluorescens*	Cupin-MetRS	Glycine	Green
A0A3S9PCD5	*Streptomyces luteoverticillatus*	MetRS	Glycine	Green
A0A0L0QKX8	*Virgibacillus pantothenticus*	MetRS	Glycine	Green
A0A4P8XS63	*Streptomyces aureofaciens*	Cupin-MetRS	Glycine	Green
A0A126Y2P7	*Streptomyces albidoflavus*	Cupin-MetRS	Alanine	Cyan
A0A552E3D4	*Microcystis aeruginosa*	Cupin-MetRS	Serine	Purple
D5UDN9	*Cellulomonas flavigena*	MetRS	Tyrosine	Black
A0A516ELE7	*Streptomyces candidus*	Cupin-MetRS	Glutamate	Pink
A0A5R8ZBS2	*Microbispora fusca*	MetRS	3*R*-hydroxyarginine	Red

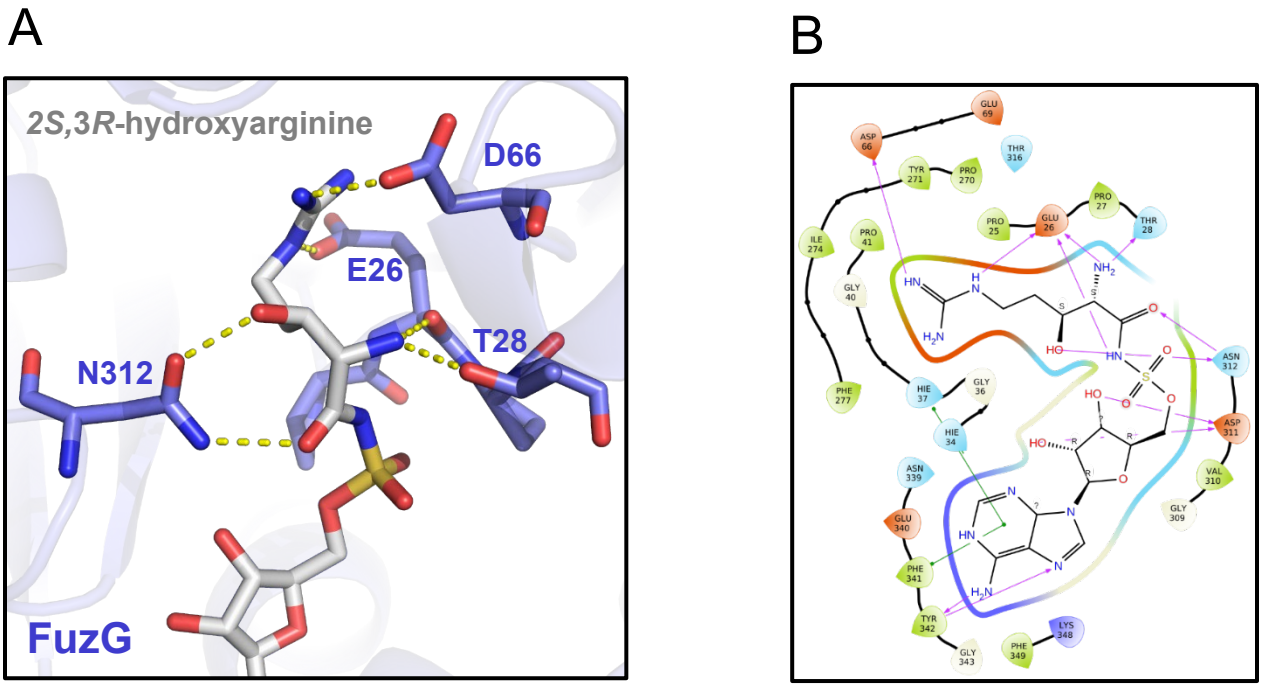
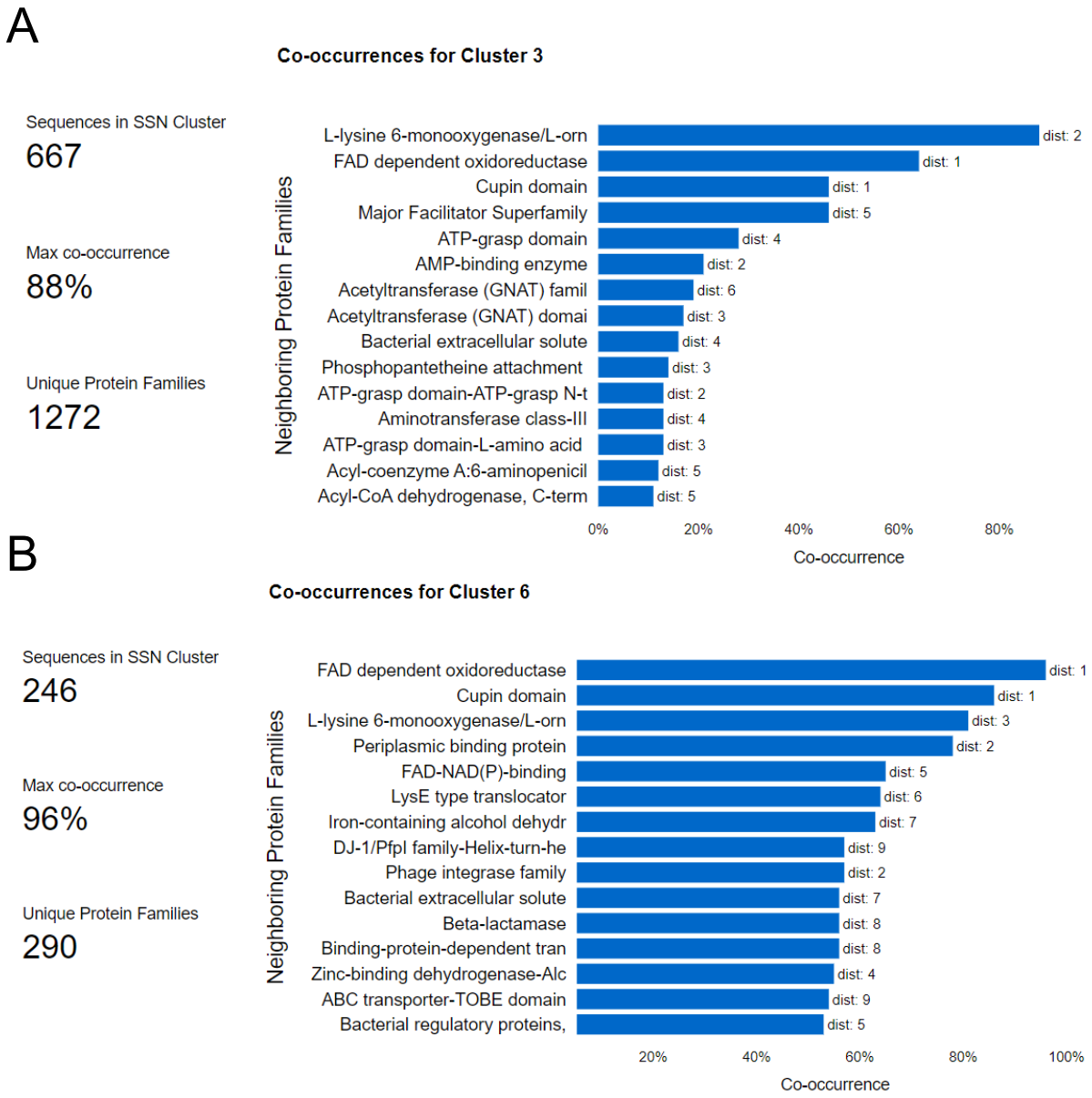


Figure 2.39. Docking analysis of Hydrazine Synthetase FuzG. (A) The active site of hydrazine synthetase FuzG (purple) with docked (2*S*,3*R*)-hydroxyargininyl-adenylate (grey) shows polar interactions with the guanidinium side chain of arginine and interactions with both the alcohol and carbonyl of arginine. (B) 2D interaction shows how Asn312 H-bonds both the α-amine and β-hydroxyl of (2*S*,3*R*)-hydroxyarginine.

Figure 2.40. tR3D Analysis of Hydrazine Synthetases. tR3D visualization of the genomic neighborhoods for the MetRS family to identify hydrazine-forming BGCs. (A) Cluster 3 encompasses all characterized hydrazine synthetases (HydSs) to date, except FuzG (B) Cluster 6 contains 246 uncharacterized HydSs. (C) Cluster 7 contains FuzG as well as 89 other uncharacterized HydSs. (D) Cluster 9 contains shows 67 HydSs with a highly conserved dioxygenase. (E) Cluster 10 shows other unique genomic contexts for HydSs producing unknown natural products.



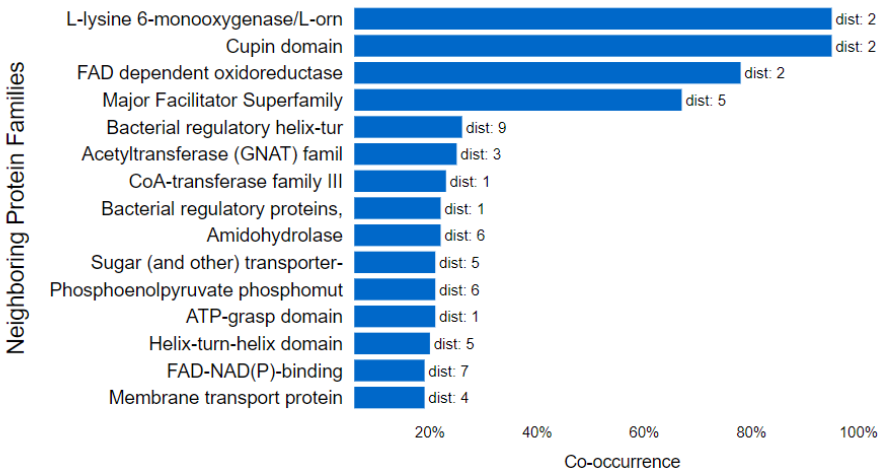
C

Co-occurrences for Cluster 7

Sequences in SSN Cluster
90

Max co-occurrence
95%

Unique Protein Families
348



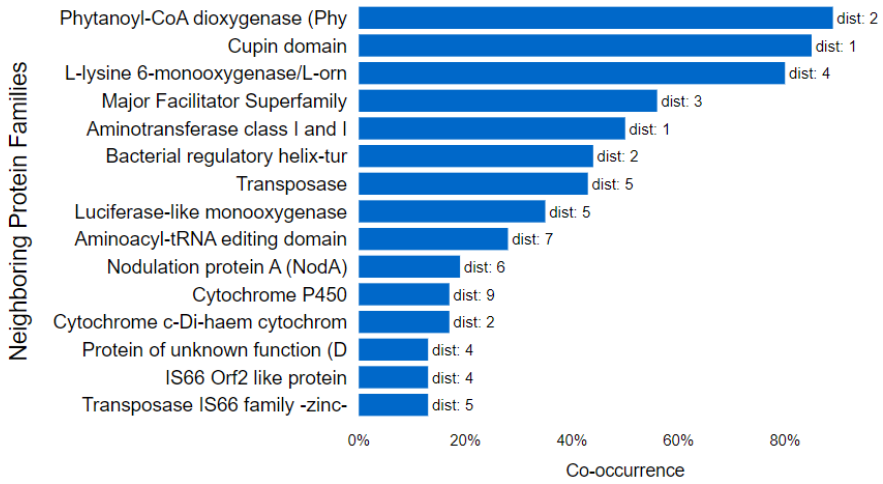
D

Co-occurrences for Cluster 9

Sequences in SSN Cluster
67

Max co-occurrence
89%

Unique Protein Families
173



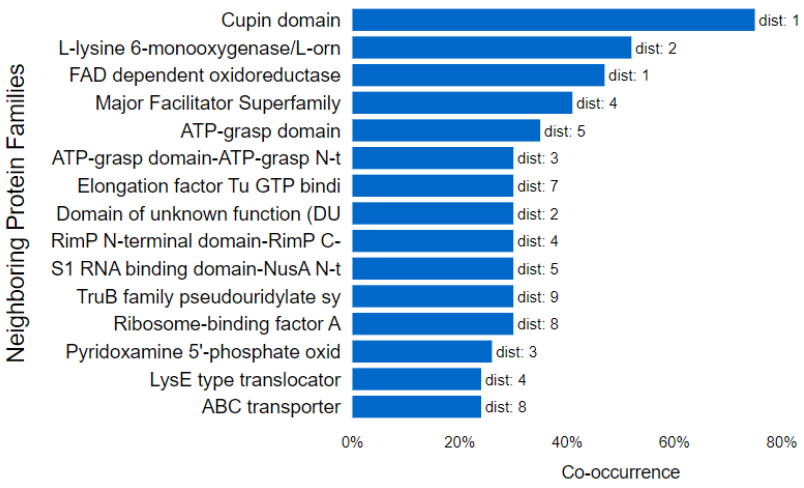
E

Co-occurrences for Cluster 10

Sequences in SSN Cluster
53

Max co-occurrence
75%

Unique Protein Families
194



2.4 Conclusions

With the rapid increase in available genomic information available, genome mining for enzyme and natural product discovery is becoming the standard approach over bioactivity-guided screening. The ability to link BGCs to natural products enables scientists to access the enzymes for applications beyond the natural product itself. With the discovery of tRNA-deacylases as broadly-adopted resistance genes for ncAAs, we demonstrate their fruitfulness as genomic markers for the discovery of new ncAAs and enzymes. Several of the findings, including new reactions and substrates for non-heme hydroxylases and hydrazine synthetases, provides access to previously untouched sequence space with reaction outcomes that would be difficult to predict without the presence of a neighboring tRNA-Deacylase. Altogether, we developed an efficient genome mining approach that led to the discovery of several new ncAAs and a novel hydrazine-containing tripeptide.

Through the combination of tR3D with current genome mining tools, we anticipate the approach taken in this study can be more broadly applied to other enzyme families of interest beyond tRNA-Deacylases. For instance, analyzing the sequence space of hydrazine synthetases via tR3D revealed many untouched sequences and putative BGCs (**Figs. 2.38-2.40**). In addition, we estimate the number of possible hydrazine synthetases in existence is more than 1200, which is nearly double the most recent estimate from a similarity-based genome mining approach [*15*]. With tRNA-deacylases as initial marker genes, we have found other marker genes for nontemplated biosynthesis with potential applications to biocatalysis, including amine-modifying enzymes and ligases. We anticipate ATP-grasp proteins to be especially common in amino-acid based natural products and can likely be engineered given their broad substrate tolerance. Additionally, discovering new enzyme functions from a broader family, in this case the hydrazine synthetases, demonstrates the efficiency with which tR3D can be employed for a desired enzyme family of interest. We anticipate this work will inspire and aid future efforts in enzyme discovery, natural product discovery, and leveraging nature's broad biosynthetic diversity for biocatalysis.

2.5 References

1. Hedges, J. B.; Ryan, K. S. Biosynthetic Pathways to Nonproteinogenic α-Amino Acids. *Chem. Rev.* **2020**, *120* (6), 3161–3209. https://doi.org/10.1021/acs.chemrev.9b00408.
2. Galván, A. E.; Paul, N. P.; Chen, J.; Yoshinaga-Sakurai, K.; Utturkar, S. M.; Rosen, B. P.; Yoshinaga, M. Identification of the Biosynthetic Gene Cluster for the Organoarsenical Antibiotic Arsinothricin. *Microbiol. Spectr.* **2021**, *9* (1), e00502-21. https://doi.org/10.1128/Spectrum.00502-21.
3. Van Cura, D.; Ng, T. L.; Huang, J.; Hager, H.; Hartwig, J. F.; Keasling, J. D.; Balskus, E. P. Discovery of the Azaserine Biosynthetic Pathway Uncovers a Biological Route for α-Diazoester Production. *Angew. Chem. Int. Ed.* **2023**, *62* (28), e202304646. https://doi.org/10.1002/anie.202304646.
4. Kameyama, T.; Takahashi, A.; Matsumoto, H.; Kurasawa, S.; Hamada, M.; Okami, Y.; Ishizuka, M.; Takeuchi, T. Thrazarine, a New Antitumor Antibiotic. I. Taxonomy, Fermentation, Isolation and Biological Properties. *J. Antibiot. (Tokyo)* **1988**, *41* (11), 1561–1567. https://doi.org/10.7164/antibiotics.41.1561.
5. Rao, K. V.; Brooks, S. C.; Kugelman, M.; Romano, A. A. Diazomycins A, B, and C, Three Antitumor Substances. I. Isolation and Characterization. *Antibiot. Annu.* **1959**, *7*, 943–949.
6. Marchand, J. A.; Neugebauer, M. E.; Ing, M. C.; Lin, C.-I.; Pelton, J. G.; Chang, M. C. Y. Discovery of a Pathway for Terminal-Alkyne Amino Acid Biosynthesis. *Nature* **2019**, *567* (7748), 420–424. https://doi.org/10.1038/s41586-019-1020-y.
7. Wegman, M. A.; Janssen, M. H. A.; van Rantwijk, F.; Sheldon, R. A. Towards Biocatalytic Synthesis of β-Lactam Antibiotics. *Adv. Synth. Catal.* **2001**, *343* (6–7), 559–576. https://doi.org/10.1002/1615-4169(200108)343:6/7<559::AID-ADSC559>3.0.CO;2-Z.
8. Kirby, G. W. Biosynthesis of the Morphine Alkaloids. *Science* **1967**, *155* (3759), 170–173. https://doi.org/10.1126/science.155.3759.170.
9. Walsh, C. T.; O'Brien, R. V.; Khosla, C. Nonproteinogenic Amino Acid Building Blocks for Nonribosomal Peptide and Hybrid Polyketide Scaffolds. *Angew. Chem. Int. Ed.* **2013**, *52* (28), 7098–7124. https://doi.org/10.1002/anie.201208344.
10. Navarro-Muñoz, J. C.; Selem-Mojica, N.; Mullowney, M. W.; Kautsar, S. A.; Tryon, J. H.; Parkinson, E. I.; De Los Santos, E. L. C.; Yeong, M.; Cruz-Morales, P.; Abubucker, S.; Roeters, A.; Lokhorst, W.; Fernandez-Guerra, A.; Cappelini, L. T. D.; Goering, A. W.; Thomson, R. J.; Metcalf, W. W.; Kelleher, N. L.; Barona-Gomez, F.; Medema, M. H. A Computational Framework to Explore Large-Scale Biosynthetic Diversity. *Nat. Chem. Biol.* **2020**, *16* (1), 60–68. https://doi.org/10.1038/s41589-019-0400-9.
11. Medema, M. H.; de Rond, T.; Moore, B. S. Mining Genomes to Illuminate the Specialized Chemistry of Life. *Nat. Rev. Genet.* **2021**, *22* (9), 553–571. https://doi.org/10.1038/s41576-021-00363-7.
12. Zaparucha, A.; Berardinis, V. de; Vaxelaire-Vergne, C. Chapter 1 Genome Mining for Enzyme Discovery. **2018**, 1–27. https://doi.org/10.1039/9781788010450-00001.
13. Velásquez, J. E.; van der Donk, W. A. Genome Mining for Ribosomally Synthesized Natural Products. *Curr. Opin. Chem. Biol.* **2011**, *15* (1), 11–21. https://doi.org/10.1016/j.cbpa.2010.10.027.

14. Scherlach, K.; Hertweck, C. Mining and Unearthing Hidden Biosynthetic Potential. *Nat. Commun.* **2021**, *12* (1), 3864. https://doi.org/10.1038/s41467-021-24133-5.
15. Matsuda, K.; Arima, K.; Akiyama, S.; Yamada, Y.; Abe, Y.; Suenaga, H.; Hashimoto, J.; Shinya, K.; Nishiyama, M.; Wakimoto, T. A Natural Dihydropyridazinone Scaffold Generated from a Unique Substrate for a Hydrazine-Forming Enzyme. *J. Am. Chem. Soc.* **2022**, *144* (28), 12954–12960. https://doi.org/10.1021/jacs.2c05269.
16. Mungan, M. D.; Blin, K.; Ziemert, N. ARTS-DB: A Database for Antibiotic Resistant Targets. *Nucleic Acids Res.* **2021**, *50* (D1), D736–D740. https://doi.org/10.1093/nar/gkab940.
17. Kjærbølling, I.; Vesth, T.; Andersen, M. R. Resistance Gene-Directed Genome Mining of 50 Aspergillus Species. *mSystems 4* (4), e00085-19. https://doi.org/10.1128/mSystems.00085-19.
18. Sokabe, M.; Ose, T.; Nakamura, A.; Tokunaga, K.; Nureki, O.; Yao, M.; Tanaka, I. The Structure of Alanyl-tRNA Synthetase with Editing Domain. *Proc. Natl. Acad. Sci.* **2009**, *106* (27), 11028–11033. https://doi.org/10.1073/pnas.0904645106.
19. Kuzmishin Nagy, A. B.; Bakhtina, M.; Musier-Forsyth, K. Chapter Four - Trans-Editing by Aminoacyl-tRNA Synthetase-like Editing Domains. In *The Enzymes*; Ribas de Pouplana, L., Kaguni, L. S., Eds.; Biology of Aminoacyl-tRNA Synthetases; Academic Press, 2020; Vol. 48, pp 69–115. https://doi.org/10.1016/bs.enz.2020.07.002.
20. Perona, J. J.; Gruic-Sovulj, I. Synthetic and Editing Mechanisms of Aminoacyl-tRNA Synthetases. In *Aminoacyl-tRNA Synthetases in Biology and Medicine*; Kim, S., Ed.; Topics in Current Chemistry; Springer Netherlands: Dordrecht, 2014; pp 1–41. https://doi.org/10.1007/128_2013_456.
21. An, S.; Musier-Forsyth, K. Cys-tRNAPro Editing by *Haemophilus Influenzae* YbaK via a Novel Synthetase·YbaK·tRNA Ternary Complex . *J. Biol. Chem.* **2005**, *280* (41), 34465–34472. https://doi.org/10.1074/jbc.M507550200.
22. Sokabe, M.; Okada, A.; Yao, M.; Nakashima, T.; Tanaka, I. Molecular Basis of Alanine Discrimination in Editing Site. *Proc. Natl. Acad. Sci.* **2005**, *102* (33), 11669–11674. https://doi.org/10.1073/pnas.0502119102.
23. Soutourina, O.; Soutourina, J.; Blanquet, S.; Plateau, P. Formation of D-Tyrosyl-tRNATyr Accounts for the Toxicity of D-Tyrosine toward *Escherichia Coli*. *J. Biol. Chem.* **2004**, *279* (41), 42560–42565. https://doi.org/10.1074/jbc.M402931200.
24. Mohler, K.; Ibba, M. Translational Fidelity and Mistranslation in the Cellular Response to Stress. *Nat. Microbiol.* **2017**, *2* (9), 1–9. https://doi.org/10.1038/nmicrobiol.2017.117.
25. Johnson, J. A.; Lu, Y. Y.; Van Deventer, J. A.; Tirrell, D. A. Residue-Specific Incorporation of Noncanonical Amino Acids into Proteins: Recent Developments and Applications. *Curr. Opin. Chem. Biol.* **2010**, *14* (6), 774–780. https://doi.org/10.1016/j.cbpa.2010.09.013.
26. Doublié, S. [29] Preparation of Selenomethionyl Proteins for Phase Determination. In *Methods in Enzymology*; Macromolecular Crystallography Part A; Academic Press, 1997; Vol. 276, pp 523–530. https://doi.org/10.1016/S0076-6879(97)76075-0.
27. Melo Czekster, C.; Robertson, W. E.; Walker, A. S.; Söll, D.; Schepartz, A. In Vivo Biosynthesis of a β-Amino Acid-Containing Protein. *J. Am. Chem. Soc.* **2016**, *138* (16), 5194–5197. https://doi.org/10.1021/jacs.6b01023.
28. Minajigi, A.; Deng, B.; Francklyn, C. S. Fidelity Escape by the Unnatural Amino Acid β-Hydroxynorvaline: An Efficient Substrate for Escherichia Coli Threonyl-tRNA Synthetase

with Toxic Effects on Growth. *Biochemistry* **2011**, *50* (6), 1101–1109. https://doi.org/10.1021/bi101360a.

29. McMurry, J. L.; Chang, M. C. Y. Fluorothreonyl-tRNA Deacylase Prevents Mistranslation in the Organofluorine Producer *Streptomyces Cattleya*. *Proc. Natl. Acad. Sci.* **2017**, *114* (45), 11920–11925. https://doi.org/10.1073/pnas.1711482114.
30. Hauth, F.; Funck, D.; Hartig, J. S. A Standalone Editing Protein Deacylates Mischarged Canavanyl-tRNA^{Arg} to Prevent Canavanine Incorporation into Proteins. *Nucleic Acids Res.* **2023**, *51* (5), 2001–2010. https://doi.org/10.1093/nar/gkac1197.
31. Neugebauer, M. E.; Sumida, K. H.; Pelton, J. G.; McMurry, J. L.; Marchand, J. A.; Chang, M. C. Y. A Family of Radical Halogenases for the Engineering of Amino-Acid-Based Products. *Nat. Chem. Biol.* **2019**, *15* (10), 1009–1016. https://doi.org/10.1038/s41589-019-0355-x.
32. Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R. D.; Bairoch, A. ExPASy: The Proteomics Server for in-Depth Protein Knowledge and Analysis. *Nucleic Acids Res.* **2003**, *31* (13), 3784–3788. https://doi.org/10.1093/nar/gkg563.
33. Chan, K. K. J. Identification and Enzyme Studies of Rare Amino Acid Biosynthesis from *Streptomyces Cattleya*. Thesis, University of St Andrews, 2013. https://research-repository.st-andrews.ac.uk/handle/10023/4478 (accessed 2023-12-13).
34. Zhang, X.; van der Donk, W. A. On the Substrate Specificity of Dehydration by Lacticin 481 Synthetase. *J. Am. Chem. Soc.* **2007**, *129* (8), 2212–2213. https://doi.org/10.1021/ja067672v.
35. Garner, P.; Park, J. M. The Synthesis and Configurational Stability of Differentially Protected .Beta.-Hydroxy-.Alpha.-Amino Aldehydes. *J. Org. Chem.* **1987**, *52* (12), 2361–2364. https://doi.org/10.1021/jo00388a004.
36. Mckillop, A.; Taylor, R. J. K.; Watson, R. J.; Lewis, N. An Improved Procedure for the Preparation of the Garner Aldehyde and Its Use for the Synthesis of N-Protected 1-Halo-2-(R)-Amino-3-Butenes. *Synthesis* **1994**, *1994* (1), 31–33. https://doi.org/10.1055/s-1994-25398.
37. Herold, P. Synthesis of D-Erythro- and D-Threo-Sphingosine Derivatives From L-Serine. *Helv. Chim. Acta* **1988**, *71* (2), 354–362. https://doi.org/10.1002/hlca.19880710208.
38. Triana-Alonso, F. J.; Dabrowski, M.; Wadzack, J.; Nierhaus, K. H. Self-Coded 3′-Extension of Run-off Transcripts Produces Aberrant Products during in Vitro Transcription with T7 RNA Polymerase (∗). *J. Biol. Chem.* **1995**, *270* (11), 6298–6307. https://doi.org/10.1074/jbc.270.11.6298.
39. Goto, Y.; Katoh, T.; Suga, H. Flexizymes for Genetic Code Reprogramming. *Nat. Protoc.* **2011**, *6* (6), 779–790. https://doi.org/10.1038/nprot.2011.331.
40. Gust, B.; Challis, G. L.; Fowler, K.; Kieser, T.; Chater, K. F. PCR-Targeted *Streptomyces* Gene Replacement Identifies a Protein Domain Needed for Biosynthesis of the Sesquiterpene Soil Odor Geosmin. *Proc. Natl. Acad. Sci.* **2003**, *100* (4), 1541–1546. https://doi.org/10.1073/pnas.0337542100.
41. Kong, A. T.; Leprevost, F. V.; Avtonomov, D. M.; Mellacheruvu, D.; Nesvizhskii, A. I. MSFragger: Ultrafast and Comprehensive Peptide Identification in Mass Spectrometry–Based Proteomics. *Nat. Methods* **2017**, *14* (5), 513–520. https://doi.org/10.1038/nmeth.4256.
42. Neidhardt, F. C.; Bloch, P. L.; Smith, D. F. Culture Medium for Enterobacteria. *J. Bacteriol.* **1974**, *119* (3), 736–747.

43. Zallot, R.; Oberg, N.; Gerlt, J. A. The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry* **2019**, *58* (41), 4169–4182. https://doi.org/10.1021/acs.biochem.9b00735.
44. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.
45. Manley, O. M.; Tang, H.; Xue, S.; Guo, Y.; Chang, W.; Makris, T. M. BesC Initiates C–C Cleavage through a Substrate-Triggered and Reactive Diferric-Peroxo Intermediate. *J. Am. Chem. Soc.* **2021**, *143* (50), 21416–21424. https://doi.org/10.1021/jacs.1c11109.
46. McBride, M. J.; Nair, M. A.; Sil, D.; Slater, J. W.; Neugebauer, M. E.; Chang, M. C. Y.; Boal, A. K.; Krebs, C.; Bollinger, J. M. Jr. Substrate-Triggered μ-Peroxodiiron(III) Intermediate in the 4-Chloro-l-Lysine-Fragmenting Heme-Oxygenase-like Diiron Oxidase (HDO) BesC: Substrate Dissociation from, and C4 Targeting by, the Intermediate. *Biochemistry* **2022**, *61* (8), 689–702. https://doi.org/10.1021/acs.biochem.1c00774.
47. Truong, F.; Yoo, T. H.; Lampo, T. J.; Tirrell, D. A. Two-Strain, Cell-Selective Protein Labeling in Mixed Bacterial Cultures. *J. Am. Chem. Soc.* **2012**, *134* (20), 8551–8556. https://doi.org/10.1021/ja3004667.
48. Ignacio, B. J.; Dijkstra, J.; Mora, N.; Slot, E. F. J.; van Weijsten, M. J.; Storkebaum, E.; Vermeulen, M.; Bonger, K. M. THRONCAT: Metabolic Labeling of Newly Synthesized Proteins Using a Bioorthogonal Threonine Analog. *Nat. Commun.* **2023**, *14* (1), 3367. https://doi.org/10.1038/s41467-023-39063-7.
49. Vargas-Rodriguez, O.; Bakhtina, M.; McGowan, D.; Abid, J.; Goto, Y.; Suga, H.; Musier-Forsyth, K. Human Trans-Editing Enzyme Displays tRNA Acceptor-Stem Specificity and Relaxed Amino Acid Selectivity. *J. Biol. Chem.* **2020**, *295* (48), 16180–16190. https://doi.org/10.1074/jbc.RA120.015981.
50. Crothers, D. M.; Seno, T.; Söll, D. G. Is There a Discriminator Site in Transfer RNA? *Proc. Natl. Acad. Sci.* **1972**, *69* (10), 3063–3067. https://doi.org/10.1073/pnas.69.10.3063.
51. Edgar, R. C. MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. *Nucleic Acids Res.* **2004**, *32* (5), 1792–1797. https://doi.org/10.1093/nar/gkh340.
52. Chan, P. P.; Lowe, T. M. GtRNAdb: A Database of Transfer RNA Genes Detected in Genomic Sequence. *Nucleic Acids Res.* **2009**, *37* (suppl_1), D93–D97. https://doi.org/10.1093/nar/gkn787.
53. Bruccoleri, R. E.; Heinrich, G. An Improved Algorithm for Nucleic Acid Secondary Structure Display. *Bioinformatics* **1988**, *4* (1), 167–173. https://doi.org/10.1093/bioinformatics/4.1.167.
54. Vaillancourt, F. H.; Yeh, E.; Vosburg, D. A.; O'Connor, S. E.; Walsh, C. T. Cryptic Chlorination by a Non-Haem Iron Enzyme during Cyclopropyl Amino Acid Biosynthesis. *Nature* **2005**, *436* (7054), 1191–1194. https://doi.org/10.1038/nature03797.

55. Vaillancourt, F. H.; Yin, J.; Walsh, C. T. SyrB2 in Syringomycin E Biosynthesis Is a Nonheme FeII α-Ketoglutarate- and O_2-Dependent Halogenase. *Proc. Natl. Acad. Sci.* **2005**, *102* (29), 10111–10116. https://doi.org/10.1073/pnas.0504412102.

56. Blin, K.; Shaw, S.; Kloosterman, A. M.; Charlop-Powers, Z.; van Wezel, G. P.; Medema, M. H.; Weber, T. antiSMASH 6.0: Improving Cluster Detection and Comparison Capabilities. *Nucleic Acids Res.* **2021**, *49* (W1), W29–W35. https://doi.org/10.1093/nar/gkab335.

57. Wishart, D. S.; Knox, C.; Guo, A. C.; Eisner, R.; Young, N.; Gautam, B.; Hau, D. D.; Psychogios, N.; Dong, E.; Bouatra, S.; Mandal, R.; Sinelnikov, I.; Xia, J.; Jia, L.; Cruz, J. A.; Lim, E.; Sobsey, C. A.; Shrivastava, S.; Huang, P.; Liu, P.; Fang, L.; Peng, J.; Fradette, R.; Cheng, D.; Tzur, D.; Clements, M.; Lewis, A.; De Souza, A.; Zuniga, A.; Dawe, M.; Xiong, Y.; Clive, D.; Greiner, R.; Nazyrova, A.; Shaykhutdinov, R.; Li, L.; Vogel, H. J.; Forsythe, I. HMDB: A Knowledgebase for the Human Metabolome. *Nucleic Acids Res.* **2009**, *37* (Database issue), D603-610. https://doi.org/10.1093/nar/gkn810.

58. Matsuda, K.; Tomita, T.; Shin-ya, K.; Wakimoto, T.; Kuzuyama, T.; Nishiyama, M. Discovery of Unprecedented Hydrazine-Forming Machinery in Bacteria. *J. Am. Chem. Soc.* **2018**, *140* (29), 9083–9086. https://doi.org/10.1021/jacs.8b05354.

59. Zhao, G.; Peng, W.; Song, K.; Shi, J.; Lu, X.; Wang, B.; Du, Y.-L. Molecular Basis of Enzymatic Nitrogen-Nitrogen Formation by a Family of Zinc-Binding Cupin Enzymes. *Nat. Commun.* **2021**, *12* (1), 7205. https://doi.org/10.1038/s41467-021-27523-x.

60. Mitchell, A. J.; Dunham, N. P.; Martinie, R. J.; Bergman, J. A.; Pollock, C. J.; Hu, K.; Allen, B. D.; Chang, W.; Silakov, A.; Bollinger, J. M. Jr.; Krebs, C.; Boal, A. K. Visualizing the Reaction Cycle in an Iron(II)- and 2-(Oxo)-Glutarate-Dependent Hydroxylase. *J. Am. Chem. Soc.* **2017**, *139* (39), 13830–13836. https://doi.org/10.1021/jacs.7b07374.

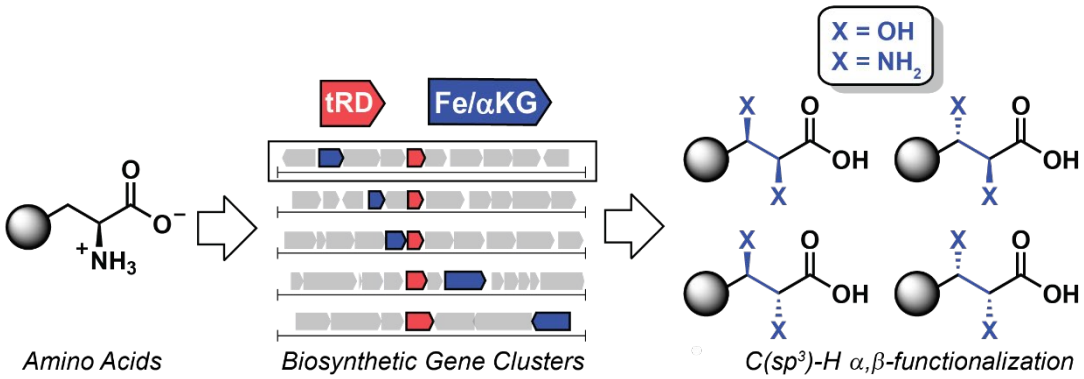
61. Mügge, C.; Heine, T.; Baraibar, A. G.; van Berkel, W. J. H.; Paul, C. E.; Tischler, D. Flavin-Dependent N-Hydroxylating Enzymes: Distribution and Application. *Appl. Microbiol. Biotechnol.* **2020**, *104* (15), 6481–6499. https://doi.org/10.1007/s00253-020-10705-w.

62. Wityak, J.; Gould, S. J.; Hein, S. J.; Keszler, D. A. A 1,3-Dipolar Cycloaddition Route to the 3(R)- and 3(S)-Hydroxy-2(S)-Arginines. *J. Org. Chem.* **1987**, *52* (11), 2179–2183. https://doi.org/10.1021/jo00387a012.

63. Ju, J.; Ozanick, S. G.; Shen, B.; Thomas, M. G. Conversion of (2*S*)-Arginine to (2*S*,3*R*)-Capreomycidine by VioC and VioD from the Viomycin Biosynthetic Pathway of Streptomyces Sp. Strain ATCC11861. *ChemBioChem* **2004**, *5* (9), 1281–1285. https://doi.org/10.1002/cbic.200400136.

64. FUJII, T.; MUKAIHARA, M.; AGEMATU, H.; TSUNEKAWA, H. Biotransformation of L-Lysine to L-Pipecolic Acid Catalyzed by L-Lysine 6-Aminotransferase and Pyrroline-5-Carboxylate Reductase. *Biosci. Biotechnol. Biochem.* **2002**, *66* (3), 622–627. https://doi.org/10.1271/bbb.66.622.

Chapter 3: *Discovery of a Biocatalytic Cascade for Preparation of α,β-Substituted Amino/Hydroxy Acids via C-H Functionalization*



Portions of this work were performed with the following people:

Claire Li contributed towards the elucidation of the Pfrm BGC through protein purification, *in vitro* reactions, and metabolomics.

Helena Roberts-Mataric performed the enzyme docking studies

Madeline Carroll performed *in vitro* screening of the substrate panel

3.1 Introduction

Amino acids are among the most common fragments in natural molecules and are becoming increasingly important building blocks for synthetic molecules [*1–3*]. To meet the high demand for selective amino acid-modifying catalysts, several classes of enzymes including hydrolases, imine reductases, and aminotransferases have become well-established workhorses for important transformations [*4–6*]. However, many of these common biocatalytic transformations rely on pre-installed functional groups, limiting their impact on the synthetic schemes overall [*7*]. C-H functionalization, where functional groups selectively replace C-H bonds, is a promising strategy to increase the efficiency of synthetic routes [*8*]. For amino acids, these transformations are especially challenging due to the indistinguishable C-H bonds along the carbon backbones as well as interference from functional group side chains, requiring protection/deprotection steps [*9*]. Utilizing enzymes capable of C-H functionalization with regio- and stereo-control is a valuable alternative strategy for enabling efficient schemes. Further, the ability to combine several reactions in one pot through biocatalytic cascades could impact process efficiency by eliminating time-consuming purification steps [*10*].

Of the enzyme classes capable of C-H functionalization, the non-heme iron/α-ketoglutarate (Fe/αKG)-dependent dioxygenases have emerged as the predominant choice for transformations of amino acids and similar small molecules [*11*]. Notable examples include the application of a Fe/αKG-dependent hydroxylase in the synthesis of manzacidin C, reducing the total number of synthetic steps from thirteen to five [*12*]. Through discovery and engineering, others utilized the exquisite regio- and stereoselectivity of Fe/αKG-dependent hydroxylases for modifying aromatic acids at different sites in the carbon backbone [*13,14*]. At the process scale, a Fe/αKG-dependent hydroxylase was recently developed for the chemoenzymatic synthesis of belzutifan [*15*]. An engineered pipecolic acid Fe/αKG-dependent hydroxylase replaced a four-step chemical oxidation with one enzymatic step, producing a chiral intermediate in >90% yield at >99% *ee* at the kilogram scale [*16*].

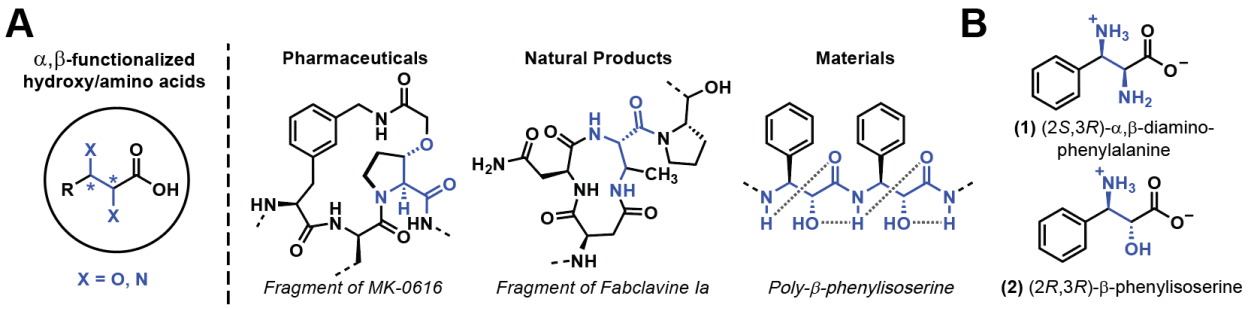


Figure 3.1. Applications of α,β-functionalized amino acids. (A) The multiple stereocenters of α,β-functionalized amino/hydroxy acids provide tunable three-dimensional properties, a scaffold for branching, and unique intramolecular interactions, yet this class of compounds remains relatively underrepresented in pharmaceuticals, natural products, and materials. Demonstrative examples are rare, including the PCSK9 inhibitor peptide macrocycle MK-0616, natural product Fabclavine Ia with a cyclic tripeptide core, and poly-β-phenylisoserine that attains unique, stable helical structures through hydrogen-bonding (small dashes). (B) The free-standing α,β-functionalized amino acids discovered in this study are made through a biosynthetic cascade that includes C-H functionalization by a Fe/αKG-dependent hydroxylase.

With these efforts, biocatalysis is being leveraged to access moieties that are extremely challenging to access synthetically. One interesting class of compounds is the α,β-functionalized acid, which despite being known for its importance in bioactive compounds, has remained relatively under-represented in synthetic pharmaceuticals, likely owing to challenges in synthesizing two vicinal chiral centers [*17*]. The unique structures of α,β-functionalized scaffolds expand the opportunities for molecular design, granting unique 3D properties as well as functional handles for branching or cyclization (**Fig. 3.1A**) [*18*]. Exceptional examples include MK-0616, where a β-hydroxyproline provides three branching points to produce multiple macrocycles in one scaffold [*19*]. Several natural products employ this strategy as well, such as in fabclavine Ia, which contains a unique tricyclic peptide core via an α,β-functionalized acid between peptide and fatty acid branches [*20*]. Whole polymers of α,β-functionalized acids gave been shown to achieve unique and stable structures through the benefits of additional H-bonding contacts [*21*]. Altogether, expanding the pool of accessible α,β-functionalized building blocks could have an impact on molecular design.

Given the challenges in synthesizing these moieties, we envisioned biocatalysis would be a viable route towards generating diverse α,β-functionalized building blocks, with Fe/αKG-dependent C-H functionalization being a route to install stereocenters site-selectively. Herein, we develop a genome mining approach that privileges the discovery of C-H functionalization towards amino acid substrates and leverage this strategy to discover a unique enzymatic pathway for α,β-functionalized acids (**Fig. 3.1B**). Through protein and pathway engineering, we expand the utility of this biocatalytic cascade to access a diverse panel of α,β-functionalized acids. We anticipate the approach developed in this work can enable efficient syntheses of challenging molecules through expanding the stereochemical options for backbone modifications of free-standing acids.

3.2 Materials and Methods

Commercial materials. Luria-Bertani (LB) Broth Miller, LB Agar Miller, Terrific Broth (TB), and glycerol were purchased from EMD Biosciences (Darmstadt, Germany). Carbenicillin (Cb), kanamycin (Km), isopropyl-β-D-thiogalactopyranoside (IPTG), sodium chloride, dithiothreitol (DTT), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), magnesium chloride hexahydrate, ammonium bicarbonate, sodium hydroxide, and dextrose were purchased from Fisher Scientific (Pittsburgh, PA). Phosphoenolpyruvate (PEP), adenosine triphosphate sodium salt (ATP), β-nicotinamide adenine dinucleotide reduced form dipotassium salt (NADH), β-nicotinamide adenine dinucleotide hydrate (NAD), β-nicotinamide adenine dinucleotide phosphate disodium salt (NADP), flavin adenine dinucleotide disodium salt (FAD), lysozyme, magnesium sulfate heptahydrate, calcium chloride, β-mercaptoethanol (βME), sodium phosphate dibasic heptahydrate, sodium phosphate monobasic, dimethylsulfoxide (DMSO), acetonitrile (LC/MS-grade), ammonium formate (LC/MS-grade), ammonium iron (II) sulfate hexahydrate, α-ketoglutaric acid sodium salt, sodium ascorbate, Polymyxin B sulfate salt (PMBS), phenylmethanesulfonyl fluoride (PMSF), all 20 proteinogenic amino acids (L-alanine, L-arginine, L-asparagine, L-aspartic acid, L-cysteine (HCl), L-glutamic acid (potassium salt), L-glutamine, L-glycine, L-histidine, L-isoleucine, L-leucine, L-lysine, L-methionine, L-phenylalanine, L-proline, L-serine, L-serine, L-threonine, L-tryptophan, L-tyrosine, L-valine), $^{13}C_9$,^{15}N-L-α-phenylalanine, L-α-phenylalanine analogs (4-Cl/4-NO_2), hydrocinnamic acid, β-phenyllactic acid, caproic acid, valeric acid, δ-hydroxy-butyric acid, β-alanine, pyruvate kinase/lactate dehydrogenase, anhydrotetracycline hydrochloride, coenzyme A sodium salt hydrate, 24-well plates and Aeraseal covers were purchased from Sigma-Aldrich (St. Louis, MO). PageRuler Plus Prestained Protein Ladder and 4-iodo-L-α-phenylalanine analogs was purchased from ThermoFisherScientific (Waltham, MA). Succinyl-CoA Synthetase was purchased from Megazyme International (Bray, Ireland). Formic acid was purchased from Acros Organics (Morris Plains, NJ). Restriction enzymes, Phusion DNA polymerase, T5 exonuclease, DNase I, and Taq DNA ligase were purchased from New England Biolabs (Ipswich, MA). Deoxynucleotides (dNTPs), were purchased from Invitrogen (Carlsbad, CA). Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA), resuspended at a stock concentration of 100 μM in 10 mM Tris-HCl, pH 8.5, and stored at 4 °C. DNA purification kits and Ni-NTA agarose were purchased from Qiagen (Valencia, CA). SoluLyse reagent was purchased from Genlantis (San Diego, CA). Complete EDTA-free protease inhibitor was purchased from Roche Applied Science (Penzberg, Germany). PD-10 desalting columns were purchased from GE Healthcare Life, (Pittsburg, PA). Amicon Ultra 30,000 MWCO centrifugal concentrators, Benzonase, and Milli-Q Gradient water purification system were purchased from Millipore (Billerica, MA). Acrylamide/bis-acrylamide (30%, 37.5:1), electrophoresis grade sodium dodecyl sulfate (SDS) were purchased from Bio-Rad Laboratories (Hercules, CA). Ring-$^{13}C_6$-L-α-phenylalanine was purchased from Cambridge Isotope Laboratories (Andover, MA). L-α-phenylalanine analogs (4-CN, 4-Br, 4-F, 3-F, thienyl, furyl), β-phenylisoserine analogs [(2*R*,2*R*),(2*R*,3*S*),(2*S*,3*S*)], enantiopure (*S*)-β-phenylalanine and (*R*)-β-phenylalanine were purchased from Chem-Impex (Wood Dale, IL). (2*S*,3*R*)-α,β-diaminophenylalanine was purchased from BLD Pharm (Cincinnati, OH). All chemicals were used as purchased without purification.

Bacterial strains. *Pseudomonas fluorescens* R124 was received as a gift from Prof. Hazel Barton (University of Akron), from which the Pfrm biosynthetic gene cluster (BGC) was cloned. *E. coli*

DH10B-T1^R was used for plasmid construction. *E. coli* BL21 Star (DE3) was used for heterologous protein production and metabolomics experiments.

Construction of plasmids. Gibson ligation [*22*] was used to carry out all plasmid construction using *E. coli* DH10B-T1^R as the cloning host. All PCR amplifications were carried out with Phusion polymerase (New England Biolabs, Ipswich MA) using the oligonucleotides listed in **Table A1.2B**. pET16b-His$_{10}$ was used for protein expression. *pfrm* genes were individually amplified from *P. fluorescens* R124. To insert a gene in pET16b, the plasmid was digested at NdeI and BamHI. For assembly of pET28a.pTet, the BGC was cloned in 3 separate fragments from *P. fluorescens* R124 and the backbone plasmid was digested at NdeI and XhoI. For insertion of PfrmB, primers at the site of mutation (i.e. PfrmB-S73A-Fwd) and at a site in the pET16 backbone (i.e. pET16b-BB-Rev) were used to amplify the entire plasmid into two fragments (including both a Fwd/Rev and a Rev/Fwd pair), which were assembled using Gibson ligation. Following plasmid construction, all cloned inserts were sequenced at Azenta Life Sciences (San Francisco, CA).

Transformation of chemically competent *E. coli*. A 100 µL aliquot of chemically competent *E. coli* is transformed with 1 µL of pure plasmid or full Gibson mixture, 20 µL of KCM (500 mM KCl, 150 mM $CaCl_2$, 250 mM $MgCl_2$), and 160 µL MilliQ H_2O. The cells are incubated on ice for 30 min, followed by heat shock at 42 °C for 90 s. The cells are plated onto agar immediately (Cb) or after 1 h recovery in LB at 37 °C (Km).

Protein expression procedure. Single colonies of *E. coli* BL21 Star (DE3) transformed with plasmid were inoculated into TB (25 mL) containing the appropriate antibiotics and grown overnight at 37°C with shaking at 200 rpm. Overnight cultures were inoculated into 1L of TB with 50 µg/mL antibiotic in a 2.8 L-baffled shake flask to $OD_{600\ nm}$ = 0.05 and grown at 37 °C and 200 rpm until the $OD_{600\ nm}$ reached 0.6 - 0.8. Cultures were cooled on ice for 20 min before inducing with IPTG (0.25 mM) and then cultured for 18 h at 16°C and 200 rpm. Cells were pelleted and stored at -80 °C for storage until purification.

Protein Purification via Ni-NTA Chromatography. Adapted from [*32*]. Frozen cell pellets were thawed and resuspended in 45 ml of lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM imidazole, 20 mM βME and 20% (*v/v*) glycerol, pH 7.5) supplemented with EDTA-free protease inhibitor cocktail (Roche), lysozyme (0.3 mg/mL), and benzonase (5 U/mL). The cell paste was homogenized and then lysed by sonication with a Qsonica Sonicator 3000 (power= 50, 5 *s* on, 25 *s* off, 2min total process time, 0.5-inch tip). The lysate was then centrifuged at 13,000 × *g* for 10 min at 4 °C to separate the soluble and insoluble fractions. The soluble lysate was incubated with Ni-NTA (0.5ml of resin per gram of cell paste) for 45 min at 4 °C, resuspended and loaded onto a column by gravity flow. The column was washed with wash buffer (50 mM HEPES, 300 mM NaCl, 20 mM imidazole, 20 mM βME and 20% (*v/v*) glycerol, pH 7.5) for 15–20 column volumes. The column was then eluted with elution buffer (50 mM HEPES, 300 mM NaCl, 300 mM imidazole, 20 mM βME and 20% (*v/v*) glycerol, pH 7.5). Fractions containing the target protein were pooled according to absorbance at 280 nm and concentrated using an Amicon Ultra spin concentrator (10 or 30-kD MWCO; Millipore). Protein was then exchanged into storage buffer (50 mM HEPES, 100 mM sodium chloride, 10% (*v/v*) glycerol and 1 mM DTT, pH 7.5) using PD-10 desalting columns. Final protein concentrations before storage were estimated using the $\varepsilon_{280\ nm}$ calculated by ExPASy ProtParam [*31*] and measured using a Nanodrop spectrometer. They are as follows: PfrmB: 6.34 mg/mL ($\varepsilon_{280\ nm}$ = 55,460 $M^{-1}cm^{-1}$), PfrmC: 10.62 mg/mL ($\varepsilon_{280\ nm}$ = 31,860 $M^{-1}cm^{-1}$), PfrmD: 16.88 mg/mL ($\varepsilon_{280\ nm}$ = 37,360 $M^{-1}cm^{-1}$), PfrmG: 29.16 mg/mL ($\varepsilon_{280\ nm}$ = 36,440

$M^{-1}cm^{-1}$), and PfrmH: 21.09 mg/mL ($\varepsilon_{280\ nm}$ = 42,400 $M^{-1}cm^{-1}$). All proteins were aliquoted, flash frozen in liquid nitrogen and stored at −80 °C.

Small-scale growth and protein purification of PfrmB mutants. Single colonies of *E. coli* BL21 Star (DE3) transformed with plasmid were inoculated into TB (5 mL) containing the appropriate antibiotics and grown overnight at 37°C with shaking at 200 rpm. Overnight cultures were inoculated into 200mL of TB with 50 µg/mL antibiotic in a 1 L-baffled shake flask to $OD_{600\ nm}$ = 0.05 and grown at 37 °C and 200 rpm until the $OD_{600\ nm}$ reached 0.6 - 0.8. Cultures were cooled on ice for 20 min before inducing with IPTG (0.25 mM) and then cultured for 18 h at 16°C and 200 rpm. Cells were pelleted and stored at -80 °C for storage until purification. Frozen cell pellets were thawed and resuspended in 10 mL of lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM imidazole, 20 mM βME and 20% (*v/v)* glycerol, pH 7.5) supplemented with EDTA-free protease inhibitor cocktail (Roche), lysozyme (1 mg/mL), 1mM phenylmethylsulfonyl fluoride (PMSF), 50 ug/mL Polymyxin B Sulfate (50 µg/mL), $MgCl_2$ (5 mM), $CaCl_2$ (1 mM), and DNase (1 U/mL). The cell paste was homogenized and then lysed by shaking at room temperature (100rpm) for 1 h. The lysate was then centrifuged at 13,000 × *g* for 10 min at 4 °C to separate the soluble and insoluble fractions. The soluble lysate was incubated with Ni-NTA (0.5 mL of resin) for 45min at room temperature, resuspended and loaded onto a column by gravity flow. The column was washed with wash buffer (50 mM HEPES, 300 mM NaCl, 20 mM imidazole, 2 0mM βME and 20% (*v/v)* glycerol, pH 7.5) for 40 column volumes. The column was then eluted with 2.5mL elution buffer (50 mM HEPES, 300 mM NaCl, 300 mM imidazole, 20 mM βME and 20% (*v/v)* glycerol, pH 7.5). The elution was exchanged into storage buffer (50 mM HEPES, 100 mM sodium chloride, 10% (*v/v)* glycerol and 1 mM DTT, pH 7.5) using PD-10 desalting columns. The desalted protein was concentrated using an Amicon Ultra spin concentrator (30-kD MWCO; Millipore). Final protein concentrations before storage were estimated using the $\varepsilon_{280\ nm}$ calculated by ExPASy ProtParam [*23*] and measured by using a Nanodrop spectrometer. They are as follows: PfrmB S73A: 1.06 mg/mL ($\varepsilon_{280\ nm}$ = 55,460 $M^{-1}cm^{-1}$), PfrmB S73A/S74A: 0.97 mg/mL ($\varepsilon_{280\ nm}$ = 55,460 $M^{-1}cm^{-1}$), PfrmB S73A/Y103F: 0.34 mg/mL ($\varepsilon_{280\ nm}$ = 53,970 $M^{-1}cm^{-1}$), PfrmB S73A/Y103A: 0.27 mg/mL ($\varepsilon_{280\ nm}$ = 54,345 $M^{-1}cm^{-1}$), PfrmB S73A/Q226A: 0.25 mg/mL ($\varepsilon_{280\ nm}$ = 55,460 $M^{-1}cm^{-1}$), PfrmB S73A/Q226A: 0.25 mg/mL ($\varepsilon_{280\ nm}$ = 55,460 $M^{-1}cm^{-1}$), PfrmB S73A/Y103F/Q226A: 0.36 mg/mL ($\varepsilon_{280\ nm}$ = 53,970 $M^{-1}cm^{-1}$), PfrmB S73A/Y103A/Q226A: 0.45 mg/mL ($\varepsilon_{280\ nm}$ = 53,970 $M^{-1}cm^{-1}$). All proteins were aliquoted, flash frozen in liquid nitrogen and stored at −80 °C.

General procedure for high-resolution HPLC/MS analysis of polar metabolites with hydrophilic interaction chromatography (HILIC). Adapted from [*32*]. Samples containing polar metabolites were analyzed using an Agilent 1290 UPLC on a SeQuant ZIC-pHILIC (5 µm, 2.1×100 mm; EMD-Millipore) using the following buffers: buffer A (90% water, 10% acetonitrile and 10 mM ammonium formate) and buffer B (90% acetonitrile, 10% water and 10 mM ammonium formate). After 2 min at 100% buffer B, a linear gradient from 0% to 40% buffer A was applied over 15 min followed by a linear gradient from 40% to 60% buffer A over 5 min and a linear gradient from 60% to 70% buffer A over 2 min at a flow rate of 0.2 mL/min. Mass spectra were acquired in positive-ionization or negative-ionization mode using an Agilent 6530 QTOF (Agilent) with the following source and acquisition parameters: Gas temperature = 300 ºC; drying gas = 8 L/min; nebulizer = 35 psig; sheath gas temperature = 275 °C; capillary voltage = 3500 V; fragmentor 175 V; skimmer 65 V; oct 1 RF vpp = 750 V; acquisition rate = 1 spectra/s; acquisition time = 1000 ms/spectrum; min range = 50 *m/z*; max range = 500 *m/z*.

General procedure for high-resolution HPLC/MS analysis of nonpolar metabolites with reverse-phase chromatography. Samples containing nonpolar metabolites were analyzed using an Agilent 1290 UPLC on a Poroshell 120 EC-C18 (5 μm, 2.1 × 30 mm; Agilent) using the following buffers: Buffer A (100% water, 0.1% *v/v* formic acid) and Buffer B (100% acetonitrile (LC/MS grade)). After 1 min at 100% buffer A, a linear gradient from 0% to 100% Buffer B was applied over 5 min was followed by 30 *s* at 100% Buffer B at a flow rate of 0.6 mL/min. Mass spectra were acquired on either positive or negative ionization modes using an Agilent 6530 QTOF (Agilent) with the following source and acquisition parameters: Gas temperature = 300 ºC; drying gas = 8 L/min; nebulizer = 35 psig; sheath gas temperature = 275 °C; capillary voltage = 3500 V; fragmentor 175 V; skimmer 65 V; oct 1 RF vpp = 750 V; acquisition rate = 1 spectra/s; acquisition time = 1000 ms/spectrum; min range = 50 *m/z*; max range = 1700 *m/z*.

General procedure for rapid HPLC/MS analysis of polar metabolites with hydrophilic interaction chromatography (HILIC). Samples containing polar metabolites were analyzed using an Agilent 1290 UPLC and an Ascentis Express HILIC (2.7 μm, 2.1 × 75 mm; Ascentis) using the following buffers: buffer A (90% water, 10% acetonitrile and 10 mM ammonium formate) and buffer B (90% acetonitrile, 10% water and 10 mM ammonium formate). After 1 min at 100% buffer B, a linear gradient from 0% to 75% buffer A over 6 min followed by 1.5 min at 75% A was applied at a flow rate of 0.4 mL/min. Mass spectra were acquired in positive-ionization mode using an Agilent 6530 QTOF (Agilent) with the following source and acquisition parameters: Gas temperature = 300 ºC; drying gas = 8 L/min; nebulizer = 35 psig; sheath gas temperature = 350 °C; capillary voltage = 3500 V; fragmentor 175 V; skimmer 65 V; oct 1 RF vpp = 750 V; acquisition rate = 1 spectra/s; acquisition time = 1000 ms/spectrum; min range = 100 *m/z*; max range = 1700 *m/z*.

General procedure for high-resolution HPLC/MS-MS analysis of polar metabolites with hydrophilic interaction chromatography (HILIC). Samples containing polar metabolites were analyzed using an Agilent 1290 UPLC on a SeQuant ZIC-pHILIC (5μm, 2.1 × 100mm; EMD-Millipore) using the following buffers: buffer A (90% water, 10% acetonitrile and 10mM ammonium formate) and buffer B (90% acetonitrile, 10% water and 10mM ammonium formate). After 2 min at 100% buffer B, a linear gradient from 0% to 40% buffer A over 15min followed by a linear gradient from 40% to 60% buffer A over 5 min followed by a linear gradient from 60% to 70% buffer A over 2 min was applied at a flow rate of 0.2 mL/min. Mass spectra were acquired on either positive ionization modes using an Agilent 6460 QQQ (Agilent) with the following source and acquisition parameters: Gas temperature = 300 ºC; drying gas = 5 L/min; nebulizer = 45 psig; sheath gas temperature = 250 °C; capillary voltage = 3500 V; fragmentor = 135 V; cell accelerator voltage = 7. Ions were monitored by MRM using the following transitions: phenylalanine (*m/z* = 116.1→79.1 @ collision energy = 20 V), β-phenylisoserine (*m/z* = 182.1→91.1 @ collision energy = 10 V). These transitions were determined from a product ion scan.

Sequence similarity network analysis. For generating the sequence similarity network (SSN) of the AlaX family of tRNA-Deacylases, we utilized the EFI-EST tool [*24,25*] for generating a SSN of IPR018163 (tRNA-editing domain) as a Uniref50 dataset containing 128,218 proteins (Sept. 2022) (including fragments). Uniref50 is a representative dataset where each node in the SSN represents proteins that are ≥50% sequence-similar to a representative protein. After generating the initial dataset, we set the sequence alignment score to 43%, which was equivalent to a sequence similarity of 35-45%. A maximum sequence length was specified as 480 amino acids to exclude any Alanine-tRNA Synthetases that have tRNA-editing domains, resulting in 23,260

proteins. This SSN was submitted to the EFI-GNT tool to generate neighborhood data files. The neighborhood distance was set to 10 genes, and the co-occurrence cut-off to 0% to maximize the amount of genomic information available. After completion, we downloaded the Color SSN and visualized this with cytoscape [*26*]. We used the filtering panel to search for TauD-like proteins as a neighboring PFam (PF02668), and this highlighted 36 nodes. We visualized these entries with the EFI-GNT tool and identified 83 TauD-like proteins, 75 of which were being actively maintained by Uniprot [*27*]. These 75 were compared by phylogenetic analysis.

Phylogenetic analysis of TauD-like proteins. In addition to the 75 TauD-like proteins of unknown function identified from the SSN, an additional 25 characterized TauD-like proteins were included for comparison in the phylogenetic analysis. These were identified by searching PF02668 within Uniprot and specifying a reviewed annotation status. We performed a multi-sequence alignment of these 100 sequences with Aliview [*28*], which was the input to generate a phylogenetic tree using the IQTREE Web Server [*29–31*]. This tree was visualized using iTOL [*32*].

***In vitro* screening of enzymatic amino acid modifications.** Enzyme aliquots were thawed on ice and exchanged to an ammonium bicarbonate buffer (50 mM ammonium biocarbonate, 100 mM NaCl, pH 7.5) using BioSpin6® columns (BioRad). Reactions (25 µL) contained all 20 canonical amino acids (L-alanine, L-arginine, L-asparagine, L-aspartate, L-cysteine, L-glutamine, L-glutamate, L-glycine, L-histidine, L-isoleucine, L-leucine, L-lysine, L-methionine, L-phenylalanine, L-proline, L-serine, L-threonine, L-tryptophan, L-tyrosine, L-valine) and 2 nonproteinogenic amino acids (L-ornithine, DL-norleucine) (0.5mM each), and sodium chloride (100 mM) in 50mM ammonium bicarbonate buffer (pH 7.5) along with necessary cofactors. For PfrmB, these cofactors were sodium αKG (5 mM), sodium ascorbate (5 mM) and $(NH_4)_2Fe(SO_4)_2 \cdot 6$ H2O (1 mM). No additional cofactors were added for PfrmC. For PfrmD, PLP (100 µM) and sodium αKG (5 mM) were added. For PfrmG, FAD (200 µM), NAD^+ (500 µM), and $NADP^+$ (500 µM) were added. For PfrmH, ATP (5 mM) and $MgCl_2$ (10mM) were added. Reactions were initiated by addition of enzyme (15 µM final concentration) and allowed to proceed at room temperature before quenching at 2 h and 18 h in 1 volume of methanol with 1% (*v/v*) formic acid, along with control reactions lacking enzyme. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C to pellet precipitated protein, and the supernatant was analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis. Following identification of substrates from the amino acid pool, the reactions were repeated with individual amino acids for confirmation.

M9 minimal media for metabolomics. M9 minimal media made according to the University of Wisconsin-Madison's *E. coli* Genome Project [*33*] was utilized for metabolomics experiments, containing Na_2HPO_4 (47.7 mM), KH_2PO_4 (22.0 mM), NaCl (8.55 mM), NH_4Cl (18.7 mM), glucose (0.4% *w/v*), $MgSO_4$ (2 mM), $CaCl_2$ (100 µM), thiamine (1 mg/mL), and an amino acid mixture diluted from a 5X solution lacking phenylalanine (**Table A2.1**); the recipe is courtesy of the University of Wisconsin-Madison's *E. coli* Genome Project [*33*]. For isotopically-labelled amino acid condition, an equivalent amount of the labelled version ($^{13}C_6$ L-phenylalanine) was supplied exogenously to the complete medium.

Metabolomics for identifying products of Pfrm BGC. Single colonies of *E. coli* BL21 Star (DE3) transformed with pET28a.pTet-PfrmBGC were inoculated in triplicate per condition into LB (5 mL) in a 24-well plate containing kanamycin when appropriate (50 µg/mL) and grown overnight at 37°C with shaking at 200 rpm. Overnight cultures were inoculated into 5 mL of LB

with kanamycin when appropriate (50 μg/mL) in a 24-well plate to $OD_{600\ nm}$ = 0.05 and grown at 37 °C and 200 rpm until the $OD_{600\ nm}$ reached 0.4-0.6. The plate was cooled on ice for 20 min before spinning down at 4,000 × *g* for 10 min at 4 °C. Media was removed, and cell pellets were resuspended in M9 minimal media for metabolomics. Additional components were added, including kanamycin when appropriate (50 μg/mL), anhydrotetracycline (100 ng/mL), and either unlabeled or $^{13}C_6$-L-phenylalanine (50 μM). After the plate was incubated for 72 h at 30°C and 200 rpm, 1 mL aliquots were taken from each condition and centrifuged at 10,000 × *g* for 5 min at room temperature. After decanting the supernatant, pellets were resuspended in 50 μL of methanol with 1% (*v/v*) formic acid, lysed by vortexing for 10 s, and centrifuged at 20,000 × *g* at 4°C for 10 min. The supernatant was analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis in both positive- and negative-ionization modes. Samples were analyzed in triplicate with MS-Dial to identify new metabolites with isotopic labeling in the BGC-expressing strain [*34*].

***In vitro* reconstitution of Pfrm biosynthetic pathway.** Reactions (25 μl) contained amino acids L-phenylalanine and L-valine as the substrates (0.5 mM) and all necessary cofactors for the enzymes in one-pot, which were as follows: sodium αKG (10 mM), sodium ascorbate (5 mM), $(NH_4)_2Fe(SO_4)_2 \cdot 6$ H2O (1 mM), PLP (100 μM), FAD (200 μM), NAD^+ (500 μM), $NADP^+$ (500 μM), ATP (5 mM), $MgCl_2$ (10mM), and NaCl (50mM) in 50mM HEPES buffer (pH 7.5). Reactions were initiated by addition of enzyme (15 μM final concentration PfrmB/D/G/H, 10 μM PfrmC), with each subsequent reaction containing the next additional enzyme in the pathway. The reactions were allowed to proceed at room temperature before quenching at 18 h in 1 volume of acetonitrile. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C to pellet precipitated protein. The supernatant was split 50:50 and derivatized with either Fmoc-Cl or O-Benzylhydroxylamine. For one set of the samples, 20 μL of supernatant was derivatized with 5 μL Fmoc-Cl (10 mM stock solution in DCM) and 6.25 μL Borate buffer (200 mM sodium tetraborate decahydrate, pH 8.2) for 15 min at room temperature to reach a final pH of 8 before quenching in 1 volume of methanol. For the other set of samples, 20 μL of supernatant was derivatized with 20 μL O-Benzylhydroxylamine (100mM stock solution in water) at 42°C for 15 min before quenching in 1 volume of methanol. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C and the supernatant was analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for nonpolar metabolite analysis.

Confirmation of (2*R*,3*R*)-β-phenylisoserine formation by PfrmB. Reactions (15 μL) contained either (*S*)- or (*R*)-β-phenylalanine (0.5 mM), sodium αKG (10 mM), sodium ascorbate (5 mM), $(NH_4)_2Fe(SO_4)_2 \cdot 6$ H2O (1 mM), and NaCl (50mM) in 50mM HEPES buffer (pH 7.5). Reactions were initiated by addition of PfrmB (10 μM final concentration) and allowed to proceed at room temperature before quenching at 18 h in 1 volume of cold methanol with 1% (*v/v*) formic acid. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C to pellet precipitated protein. The supernatant was analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis, along with synthetic standards of (2*R*,3*R*)-β-phenylisoserine, (2*R*,3*S*)-β-phenylisoserine, and (2*S*,3*S*)-β-phenylisoserine to assign stereochemistry.

***In vitro* scale-up of PfrmB Reaction for NMR characterization.** Enzyme aliquots were thawed on ice and exchanged to an ammonium bicarbonate buffer (50mM ammonium bicarbonate, 100mM NaCl, pH 7.5) using BioSpin6® columns (BioRad). The reaction (500 μL) contained $^{13}C_9$,$^{15}N_1$-L-phenylalanine (0.6 mM), sodium αKG (10 mM), sodium ascorbate (5 mM), $(NH_4)_2Fe(SO_4)_2 \cdot 6$ H2O (1 mM) and sodium chloride (100 mM) in 50mM ammonium bicarbonate

buffer (pH 7.5). Reactions were initiated by addition of PfrmB (20 μM final concentration) and PfrmC (25 μM final concentration), and allowed to proceed at room temperature for 18 h before quenching in 2 volumes of methanol with 1% (*v*/*v*) formic acid. Precipitated protein was removed by centrifuging at 20,000 × *g* for 10 min at 4 °C and purified via HPLC.

HPLC purification of (2*R*,3*R*)-β-phenylisoserine. Sample (1.5 mL) was purified with an Agilent 1290 UPLC equipped with a SeQuant ZIC-HILIC column (5 μm, 200 Å, 50 × 21.2 mm; Merck) using the following buffers: buffer A (90% water, 10% acetonitrile and 10 mM ammonium formate) and buffer B (90% acetonitrile, 10% water and 10 mM ammonium formate). After 3 min at 100% buffer B, a linear gradient was applied from 0 to 40% buffer A over 20 min followed by a linear gradient from 40% to 60% buffer A in 7 min, followed by a linear gradient from 60% to 70% buffer A in 3 min at a flowrate of 5 mL/min. Fractions were collected in 2 mL intervals and checked by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the rapid protocol for polar metabolite analysis. Fractions containing the product were collected, pooled, dried down to a solid with N_2, and resuspended in D_2O for NMR analysis.

Kinetic analysis of PfrmB. Steady-state kinetic analysis was performed by monitoring NADH consumption through a coupled assay [*35*]. Reactions (100 μL) contained ATP (2.5 mM), $MgCl_2$ (5 mM), phosphoenolpyruvate (PEP, 1 mM), NADH (0.3 mM), lactate dehydrogenase (LDH, 10 U/ml), pyruvate kinase (PK, 10 U/ml), succinyl-CoA synthetase (SCS, 3.2 U/ml), coenzyme A (1 mM), sodium αKG (1 mM), $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$ (0.2 mM), sodium chloride (10 mM), and sodium ascorbate (2 mM) in 100 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of PfrmB (10.6 μM) in the presence of varying concentrations of (*S*)-β-phenylalanine (0 - 2 mM). Initial rates of NADH consumption were measured by monitoring A_{340} using a SpectraMax M2 Microplate Reader (Molecular Devices) at room temperature. Parameters k_{cat} and K_M were determined using a Hanes-Woolf plot (plotting substrate concentration against substrate concentration/initial rate) and input into the equation below to generate a curve fit for the data.

$$v_0 = \frac{k_{cat}[S]}{K_M + [S]}$$

where v_0 is the initial rate and [*S*] is the substrate concentration.

PfrmG production of D-phenylglycine. Reactions (15 μL) contained (2*R*,3*R*)-β-phenylisoserine (0.5 mM) and NaCl (50 mM) in 50 mM HEPES buffer (pH 7.5). Additional cofactors FAD (200 μM), NAD^+ (500 μM), and $NADP^+$ (500 μM) were either excluded or added to the reaction to determine if they were essential. Reactions were initiated by addition of PfrmG (20 μM final concentration) either with or without the additional cofactors. The reactions were allowed to proceed at room temperature before quenching at 18 h in 1 volume of acetonitrile. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C to pellet protein and 20 μL of supernatant was derivatized with 5 μL Fmoc-Cl (10 mM stock solution in DCM) and 6.25 μL Borate buffer (200 mM sodium tetraborate decahydrate, pH 8.2) for 15 min at room temperature to reach a final pH of 8 before quenching in 1 volume of methanol. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C and the supernatant was analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for nonpolar metabolite analysis.

Determination of PfrmD enantioselectivity. Reactions (15 μL) contained phenylpyruvic acid and L-α-valine (0.5 mM), PLP (100 μM), and NaCl (50 mM) in 50 mM HEPES buffer (pH 7.5).

Reactions were initiated by addition of PfrmD (25 μM final concentration) and were allowed to proceed at room temperature before quenching at 18 h in 1 volume of acetonitrile. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C to pellet precipitated protein. 20 μL of supernatant was derivatized with 5 μL Mosher's acid chloride ((*R*)-(-)-MTPA-Cl) (10mM stock solution in DCM) and and 6.25 μL Borate buffer (200mM in water, pH 8.2) for 15 min at room temperature before quenching in 1 volume of methanol, and this derivatization procedure was repeated for L-α-phenylalanine and D-α-phenylalanine. Borate buffer was prepared from sodium tetraborate decahydrate and titrated with HCl to a pH of 8.2. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C and the supernatant was analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for nonpolar metabolite analysis.

(2*S*,3*R*)-α,β-diaminophenylalanine formation by PfrmD. Reactions (15 μL) contained (2*R*,3*R*)-β-phenylisoserine and L-α-leucine (0.5 mM), FAD (200 μM), NAD^+ (500 μM), $NADP^+$ (500 μM), and NaCl (50 mM) in 100 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of PfrmD and PfrmG (20 μM final concentration). The reactions were allowed to proceed at room temperature before quenching at 18 h in 1 volume of acetonitrile. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C to pellet protein and 20 μL of supernatant was derivatized with 5 μL Fmoc-Cl (10 mM stock solution in DCM) and 6.25 μL Borate buffer (200 mM sodium tetraborate decahydrate, pH 8.2) for 15 min at room temperature to reach a final pH of 8 before quenching in 1 volume of methanol, and this derivatization procedure was repeated for (2*S*,3*R*)-α,β-diaminophenylalanine. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C and the supernatant was analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for nonpolar metabolite analysis.

(2*S*,3*R*)-α,β-diaminophenylalanine-valine formation by PfrmH. Reactions (15 μL) contained synthetic (2*S*,3*R*)-α,β-diaminophenylalanine and L-α-valine (0.5 mM), ATP (5 mM), $MgCl_2$ (15 mM), $NADP^+$ (500 μM), and NaCl (100 mM) in 100 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of PfrmH (25 μM final concentration). The reactions were allowed to proceed at room temperature before quenching at 18 h in 1 volume of cold methanol with 1% (*v/v*) formic acid. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C to pellet precipitated protein. The supernatant was analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis, with the fragmentor voltage adjusted from 175V to 200V to facilitate peptide bond fragmentation.

Biocatalytic cascade reactions with substrate panel of aromatic-substituted phenylalanines. Reactions (15 μl) contained an aromatic L-α-amino acid and L-leucine (0.5 mM) and all necessary cofactors for the enzymes in one-pot, which were as follows: sodium αKG (10 mM), sodium ascorbate (5 mM), $(NH_4)_2Fe(SO_4)_2 \cdot 6$ H2O (1 mM), PLP (100 μM), FAD (200 μM), NAD^+ (500 μM), $NADP^+$ (500 μM), and NaCl (50 mM) in 50 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of enzyme (20 μM final concentration PfrmB/D/G, 5 μM PfrmC), with each subsequent reaction containing the next additional enzyme in the pathway. The reactions were allowed to proceed at room temperature before quenching at 18 h in 1 volume of acetonitrile. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C to pellet protein and 20 μL of supernatant was derivatized with 5 μL Fmoc-Cl (10 mM stock solution in DCM) and 6.25 μL Borate buffer (200 mM sodium tetraborate decahydrate, pH 8.2) for 15 min at room temperature to reach a final pH of 8 before quenching in 1 volume of methanol. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C and the supernatant was analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for nonpolar metabolite analysis. Conversion through the

cascade was calculated by comparing the level of substrate in each reaction condition to a no enzyme control.

Engineering of PfrmB. PfrmB mutants were expressed and purified according to the procedure for small-scale expression and purification via Ni-NTA Chromatography. Reactions (15 µL) contained either *S*-β-phenylalanine or *R*-β-phenylalanine (0.5 mM), sodium αKG (10 mM), sodium ascorbate (5 mM), $(NH_4)_2Fe(SO_4)_2 \cdot 6$ H2O (1 mM), and NaCl (50 mM) in 100 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of enzyme (10 µM final concentration). The reactions were allowed to proceed at room temperature before quenching at 18 h in 1 volume of cold methanol with 1% (*v/v*) formic acid. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C to pellet precipitated protein. The supernatant was analyzed by LC/MS-MS on an Agilent 1290 UPLC-6460 QQQ using the protocol for polar metabolite analysis along with synthetic standards of (2*R*,3*R*)-β-phenylisoserine and (2*R*,3*S*)-β-phenylisoserine to assign stereochemistry. Conversion was calculated by comparing the level of substrate remaining in each reaction condition to a no enzyme control.

C-H hydroxy-functionalization of free acids. Reactions (15 µl) contained the carboxylic acid substrate (1mM) and necessary cofactors which were as follows: sodium αKG (10 mM), sodium ascorbate (5 mM), $(NH_4)_2Fe(SO_4)_2 \cdot 6$ H2O (1 mM), and NaCl (50mM) in 100mM HEPES buffer (pH 7.5) along with catalase (25 U/µL). Reactions were initiated by addition of enzyme (20 µM final concentration PfrmB). The reactions were allowed to proceed at room temperature before quenching at 18 h in 1 volume of cold methanol with 1% (*v/v*) formic acid. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C to pellet precipitated protein and the supernatant was analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis. Conversion was calculated by comparing the level of substrate in each reaction condition to a no enzyme control.

C-H amino-functionalization of free acids. Reactions (15 µl) contained the carboxylic acid substrate (1mM), leucine as the amine donor (20mM), and necessary cofactors for the enzymes in one-pot, which were as follows: sodium αKG (10 mM), sodium ascorbate (5 mM), $(NH_4)_2Fe(SO_4)_2 \cdot 6$ H2O (1 mM), PLP (100 µM), and NaCl (50mM) in 100mM HEPES buffer (pH 7.5) along with catalase (25 U/µL). Reactions were initiated by addition of enzyme (20 µM final concentration PfrmB/D/G), with each subsequent reaction containing the next additional enzyme in the pathway. The reactions were allowed to proceed at room temperature before quenching at 18 h in 1 volume of cold methanol with 1% (*v/v*) formic acid. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C to pellet precipitated protein and the supernatant was analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis. Conversion was calculated by comparing the level of substrate in each reaction condition to a no enzyme control.

NMR spectroscopy. Experiments were recorded on a Bruker Avance II spectrometer operating at 900 MHz and at 298 K The 900 MHz instrument was equipped with a CP TXI cryoprobe. NMR solvent D_2O (99.9%; Sigma-Aldrich) was used without further purification. The instrument was controlled using Topspin (v 3.2) software. Data were processed using MestReNova 14.2.1 using automatic phasing and Bernstein Polynomial Fit baseline correction capabilities. Splitting was determined using the automatic multiplet analysis function with intervention as necessary. Spectral data are reported as follows: chemical shift (multiplicity [singlet (s), broad singlet (bs), doublet (d), triplet (t), quartet (q), pentet (p), multiplet (m), doublet of doublets (dd), doublet of doublet of

doublets (ddd), doublet of triplet of doublets (dtd), doublet of doublet of doublet of doublets (dddd), doublet of triplets (dt), triplet of doublets (td), complex (comp), etc.], coupling constant (Hz), integration). Chemical shifts are reported in ppm (δ), and coupling constants are reported in Hz. Resonances are referenced to solvent residual peaks for $CDCl_3$ (^1H = 7.26 ppm, ^{13}C = 77.2 ppm) or D_2O (^1H = 4.79 ppm).

A constant-time ^1H-^{13}C HSQC experiment (Bruker pulse sequence hsqcctetgpsp) was recorded with carrier frequencies set to 5.48 ppm (^1H) and 40 ppm (^{13}C)) and the spectral widths set to 16 ppm (^1H) and 80 ppm (^{13}C). The carboxyl chemical shift was set to 176 ppm, and the ^1H-^{13}C coupling constant was set to 145 Hz. The recycle delay was set to 1.5 sec. The constant-time evolution period was set to 13.6 ms, to distinguish ^1H-^{13}C pairs with one or three carbon neighbors from those with two carbon neighbors.

Computational modeling and docking. We began by performing a BLAST search of PfrmB in the PDB and identified a similar protein with a solved crystal structure (PDB: 1Y0Z, 28.97% identity). In addition, we attained a model of the enzyme was built using AlphaFold2 [*36*]. These proteins were aligned with the "super" command in PyMOL with a RMSD value of 1.845 Å. Based on the alignment, the binding sites for αKG and the amino acid substrate were identified, and both (*S*)- and (*R*)-β-phenylalanine were rigidly docked into the structure using GLIDE via Maestro 2022-2. Viable ligand poses were selected based on distance of C_α to the iron center, along with proximity of the carboxylic acid to the putative carboxylate-binding arginine 76.

3.3 Results and Discussion

Genome mining for C-H functionalizing enzymes. With an interest in discovering useful enzymes for the installation of functional groups, we began with a bioinformatic effort to identify C-H functionalizing enzymes that act on amino acids as our target substrate. In previous work, we established tRNA-deacylases (tRDs) as genomic markers for biosynthetic gene clusters (BGCs) of amino acid modifications. To leverage this trend for discovering novel enzymatic reactions, we performed a search within the AlaX family of tRDs for neighboring enzymes that perform C-H functionalization. Specifically, we searched for members of the TauD family of Fe/αKG-dependent dioxygenases (PF02668), a subclass of this broader enzyme type that is known to act on amino acids and similar small molecules [*37–39*]. From a sequence similarity network (SSN) of the AlaX family tRDs, 75 TauD-like enzymes were identified within 10 genes upstream or downstream of a tRD (**Fig. 3.2**). While their substrates were likely to be amino acids, enzymes of the TauD family catalyze a broad range of transformations ranging from hydroxylation and epoxidation to decarboxylation and desaturation, making the functions of these 75 enzymes difficult to predict [*40,41*]. To gain insight into their specificities, a phylogenetic tree was constructed with the addition of 25 characterized members of the TauD family (**Fig. 3.3**, **Table A2.3**). From this analysis, we found that about one-half of the unknown TauD-like enzymes clustered with characterized members (Clades A-D). The other half appeared less explored, with clade G only containing one characterized enzyme of fifteen total, a succinate α,β-epoxidase (**Fig. 3.4A**). This finding was noteworthy for being a form of α,β-functionalization in addition to modifying an amino acid-like small molecule.

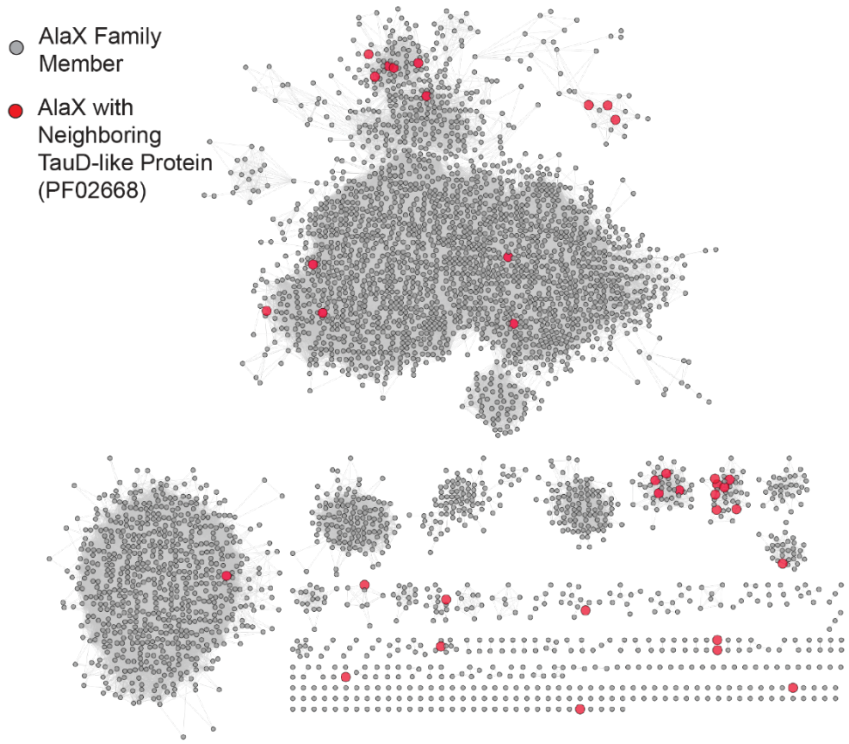


Figure 3.2. Sequence similarity network analysis of AlaX tRD co-occurrence with TauD-like proteins. A sequence similarity network (SSN) of the the AlaX tRNA-Deacylase (tRD) family (IPR018163) (Uniref50 with 43 alignment score) was searched for co-occurrence with TauD-like proteins (PF02668) within 20 neighboring genes upstream and downstream of the AlaX tRD, resulting in 75 total entries (red).

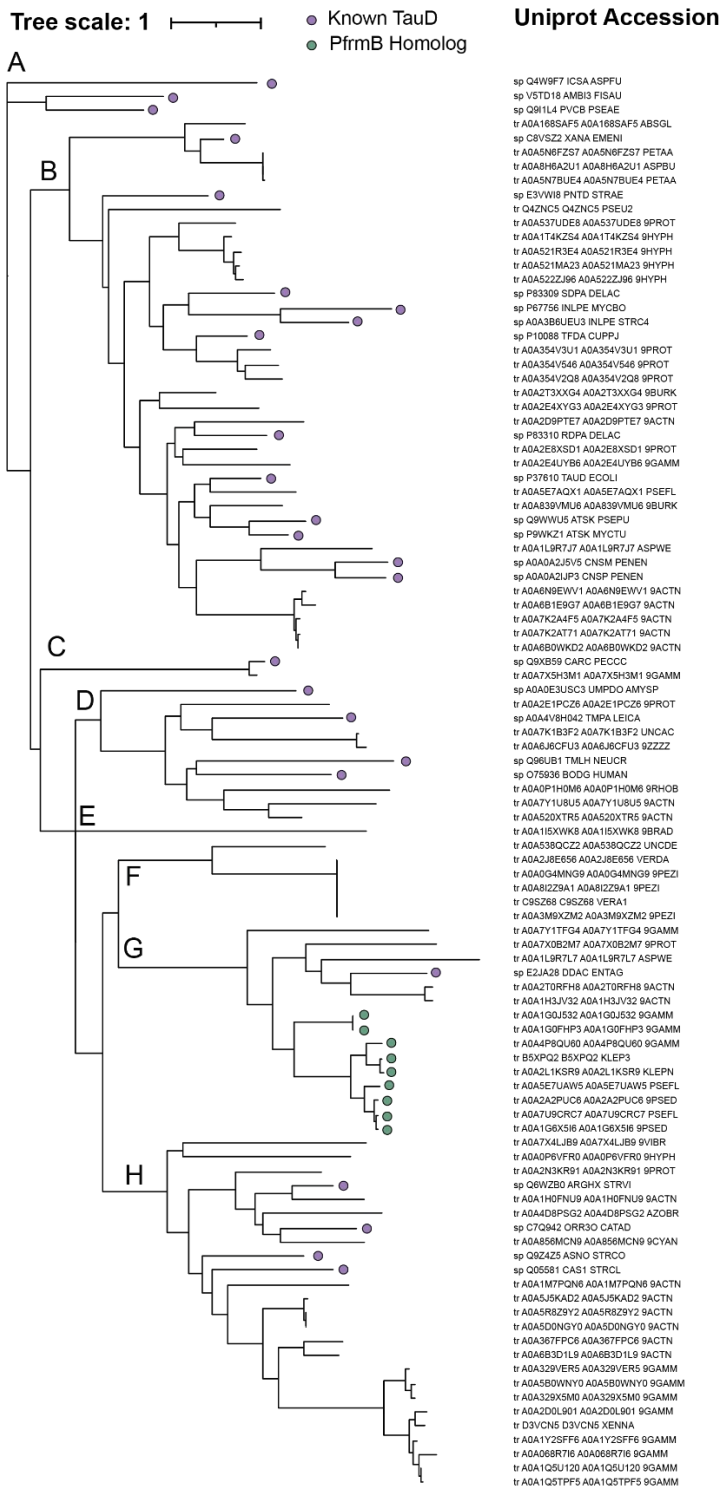


Figure 3.3. Phylogenetic analysis of TauD-like proteins. Phylogenetic relationship between 75 unknown TauD-like proteins that cluster with tRDs compared to 25 known TauD-like proteins (purple dots). PfrmB and its homologs are shown as green dots. Unknown proteins are branches without any associated colored dot.

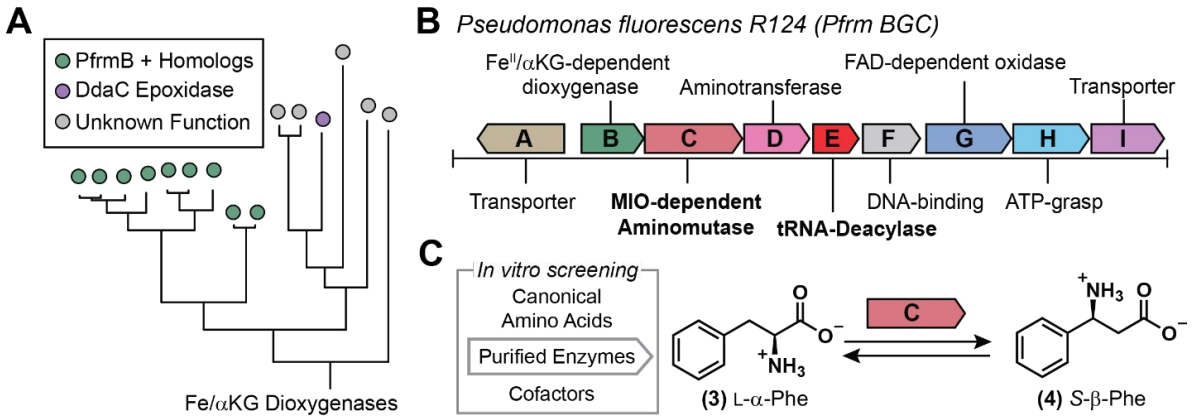


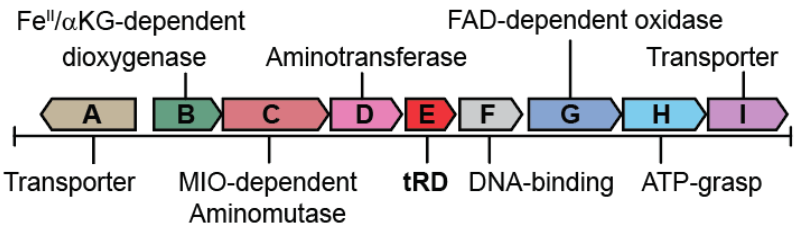
Figure 3.4. Bioinformatic discovery of C-H functionalizing biosynthetic pathway. (A) Phylogenetic analysis of TauD-like Fe/αKG dioxygenases proteins that neighbor tRNA-deacylases revealed a relatively uncharacterized clade grouped with DdaC, which performs α,β-epoxidation of succinate. This suggested PfrmB and its homologs could modify a similar amino acid-like small molecule. (B) The *Pseudomonas fluorescens R124* MIO-containing (Pfrm) BGC contains enzymes that suggested α,β-functionalization of free amino acids. (C) Screening *in vitro* identified promiscuous activities towards canonical amino acids and suggested L-α-phenylalanine was the starting point for the biosynthetic pathway.

Discovery of a promising biosynthetic gene cluster. Within Clade G, we identified a BGC that matched our search criteria (**Fig. 3.4B, Fig. 3.5**). Nine of these fourteen TauD-like proteins shared a conserved group of enzymes, with the most interesting being a 4-methylideneimidazole-5-one (MIO)-dependent aminomutase. Aminomutases are well-established enzymes for biocatalytic applications, capable of transferring the amine group of an aromatic amino acid from C_α to C_β to form a β-amino acid [*42*]. We reasoned this could allow for functionalization at the α-position by other enzymes in the pathway. Additionally, this led us to hypothesize a β-amino acid may be the product of this BGC. Notably, the absence of assembly-line like enzymes, as seen elsewhere in clade G, suggested this pathway was acting on free amino acids (**Fig. A3.4**). We chose the *Pseudomonas fluorescens* R124 MIO-containing (Pfrm) BGC for experimental elucidation.

Despite the natural product of this BGC being unknown, we reasoned this pathway modifies amino acids, enabling us to rapidly assess each enzyme's activity individually to identify the starting point for this BGC (**Fig. 3.4C**). PfrmB, PfrmC, PfrmD, PfrmG, and PfrmH were cloned with N-terminal His_{10} tags, heterologously expressed in *E. coli*, and purified for *in vitro* characterization (**Fig. A3.5**). The MIO-containing enzyme (PfrmC) was previously characterized by others as an (*S*)-selective phenylalanine aminomutase, with equilibrium favoring the (*S*)-β-phenylalanine isomer over L-α-phenylalanine (Phe) or the ammonia-lyase product, *trans*-cinnamic acid [*43*]. When we performed an *in vitro* reaction with purified PfrmC and Phe, we observed the consumption of Phe and the formation of a new peak at the same *m/z* with a different retention time, which we hypothesized to be the formation β-phenylalanine (**Fig. 3.6A**).

Additional insights were gained by examining other enzymes of the Pfrm BGC *in vitro* when screening all 20 proteinogenic amino acids as substrates to purified enzymes to detect modifications. PfrmH, annotated as an ATP-grasp ligase, formed dipeptides between Ala and large aromatic amino acids, including Phe (**Fig. 3.6B**). PfrmD, a PLP-dependent aminotransferase, showed very promiscuous transamination activity towards many amino acids, forming α-keto acids (**Fig 3.6C**). PfrmB and PfrmG showed no activity towards the proteinogenic amino acids. Taken together, these findings suggest Phe is a substrate of the pathway and the product is likely a dipeptide.

A



Protein Name	Gene ID (*PfR124* Strain)	Enzyme Annotation
PfrmA	A0A7U9CV32	Amino Acid Permease
PfrmB	A0A7U9CRC7	Non-heme Iron Dioxygenase
PfrmC	A0A7U9GV55	Aromatic Amino Acid Lyase
PfrmD	A0A7U9CU77	Aminotransferase
PfrmE	A0A7U9CRD7	tRNA-deacylase
PfrmF	A0A7U9CSG0	DNA-binding protein
PfrmG	A0A7U9CRC9	FAD-containing dehydrogenase
PfrmH	A0A7U9GU49	ATP-grasp
PfrmI	A0A7U9CRC4	MFS transporter

B

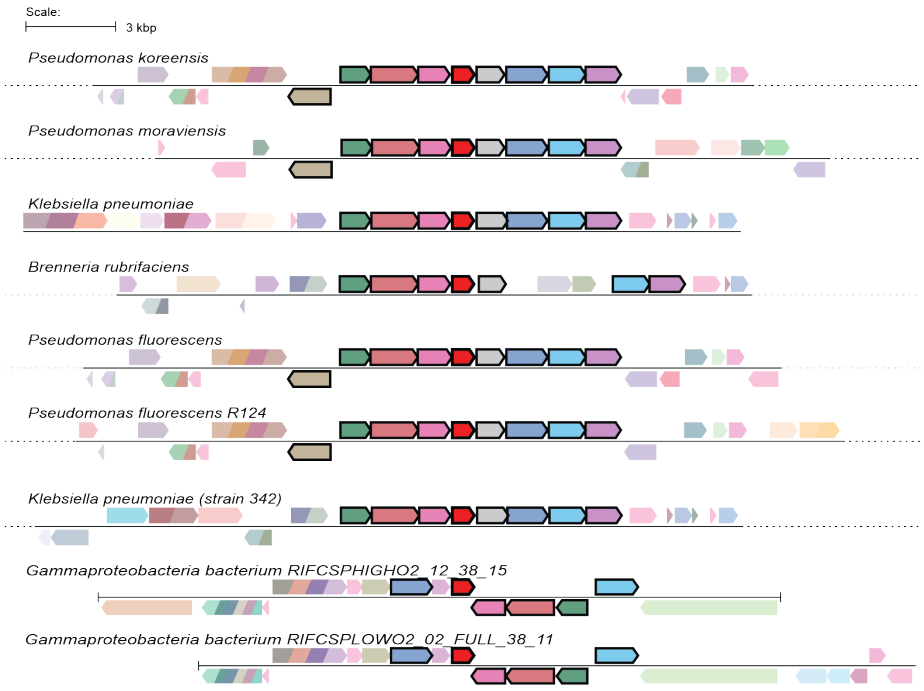


Figure 3.5. Genome mining to identify Pfrm biosynthetic gene cluster in *Pseudomonas fluorescens R124*. Identification of a BGC with uncharacterized enzymes that appear capable of functionalizing free amino acids. (A) Genomic neighborhood of the Pfrm BGC with each Uniprot accession. (B) Nine organisms share a similar biosynthetic gene cluster with a tRD (red).

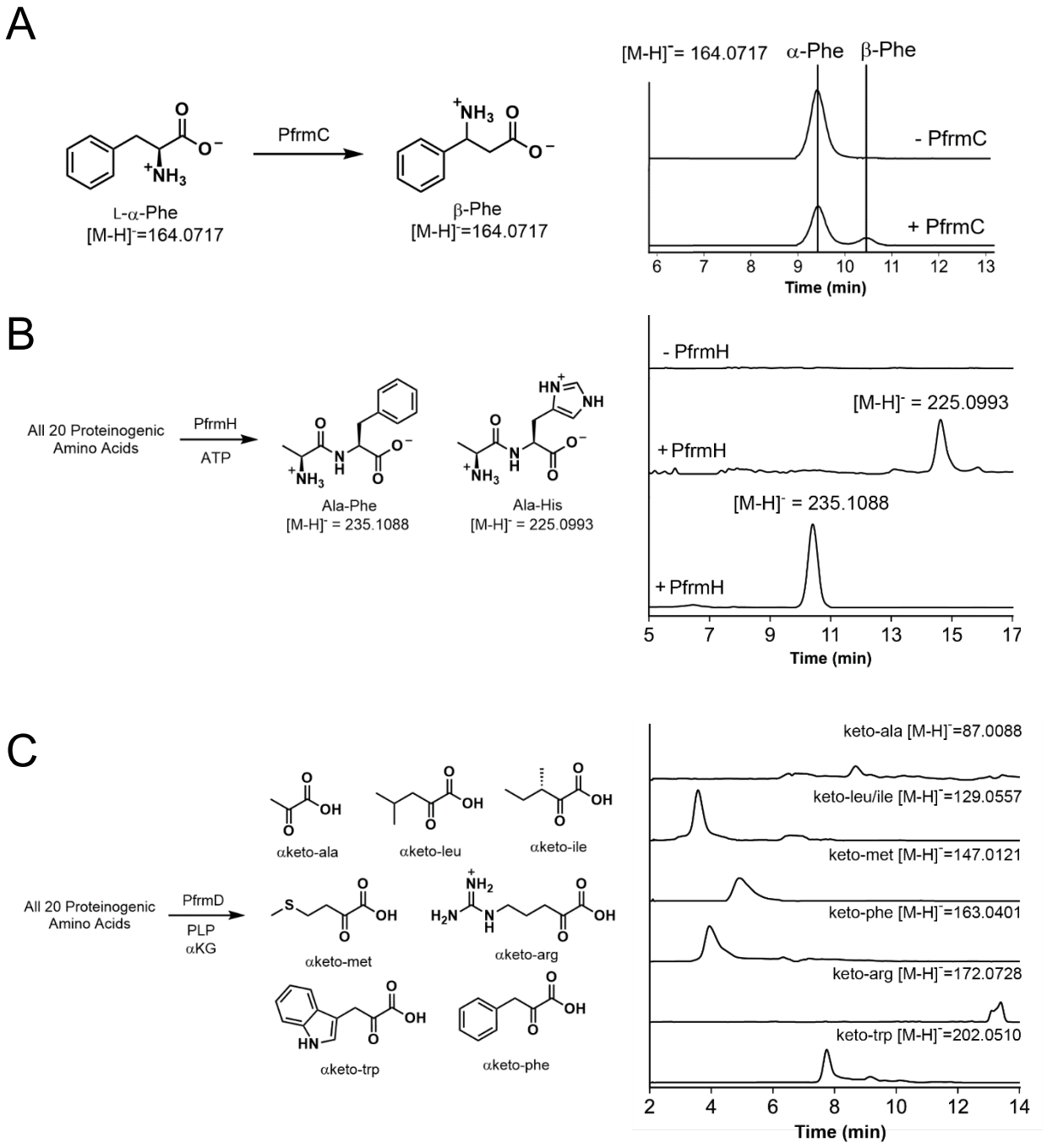
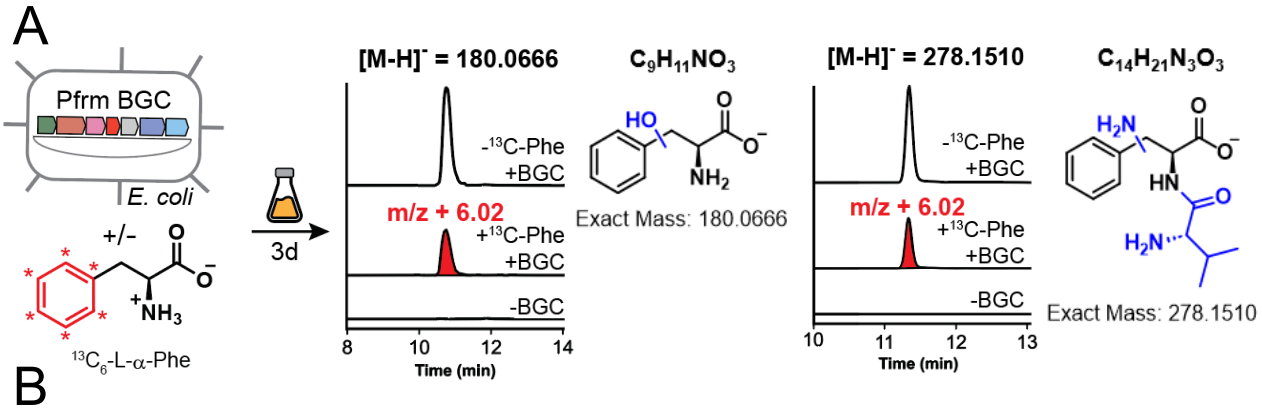


Figure 3.6. *In vitro* screen of Pfrm BGC enzymes with proteinogenic amino acids. (A) Reaction of PfrmC with all 20 amino acids showed aminomutase activity for phenylalanine, where a new peak at the same mass but different retention time indicated a different regioisomer of α-Phe, which we later confirmed to be β-phenylalanine. (B) Reaction of PfrmH with all 20 proteinogenic amino acids revealed dipeptide-forming activity, with a preference of a large hydrophobic amino acid (Phe, His) with a small amino acid (Ala). This finding suggested a dipeptide product from the Pfrm cluster. (C) Reaction of PfrmD with all 20 amino acids revealed a high level of promiscuity for the transaminase reaction. Reactions of PfrmB or PfrmG with all 20 proteinogenic amino acids produced no new products (n.d. = not detectable).

Elucidating the Pfrm biosynthetic pathway. To identify the natural product produced by the Pfrm cluster, we performed a comparative metabolomics experiment. We cloned genes PfrmB-H from *PfR124* into a vector under a single tetracycline promoter and heterologously expressed the entire cluster in *E. coli* BL21 in a minimal media (**Fig. 3.7A**). In addition, we supplied $^{13}C_6$-Phe given our hypothesis that Phe appeared to be a precursor for the Pfrm pathway. After 3 days of growth post-induction, the cell pellets were harvested and the metabolome was analyzed by LC/MS followed by MS-Dial (**Fig. 3.7B, Table A2.6**) [*34*]. From this experiment, two products accumulated with a + 6 mass increase in the $^{13}C_6$-Phe condition, confirming phenylalanine is a component. The most abundant product with *m/z* = 180.066 indicated a hydroxylation of phenylalanine, with a +16 Da increase compared to Phe (**Fig. 3.7A**). The second-most abundant product at *m/z* = 278.1510 corresponds to a molecular formula equivalent to a valine-phenylalanine dipeptide with an NH fragment added, suggesting that transamination of an α-keto acid derived from phenylalanine oxidation may have occurred through the PLP-dependent PfrmD (**Fig. 3.7A**). Interestingly, valine appears to be the substrate for the ATP-grasp enzyme, PfrmH, in *E. coli* rather than alanine, which was preferred *in vitro*.



RT (min)	*m/z*	Type	ID	ANOVA *p*-value	Signal/ Noise	Fold Increase
10.75	180.0666	$[M-H]^-$	β-phenylisoserine	2.92E-05	4771	1246.7
11.33	278.1510	$[M-H]^-$	α,β-diaminophenylalanine-valine	1.88E-05	3144	1412.3
10.75	186.085	$[M-H]^-$	β-phenylisoserine (+$^{13}C_6$-L-phenylalanine)	1.51E-08	1906	1533.97
11.33	234.1599	$[M-H]^-$	α,β-diaminophenylalanine-valine - CO_2	1.42E-05	1539.1	124686
11.32	284.1691	$[M-H]^-$	α,β-diaminophenylalanine-valine (+$^{13}C_6$-L-phenylalanine)	1.04E-05	532.2	492.18
11.33	217.1331	$[M-H]^-$	α,β-diaminophenylalanine-valine - CO_2 - NH_3	9.13E-06	98.9	14.85

Figure 3.7. Metabolomics reveals products of the Pfrm BGC. (A) We identified two compounds from the metabolomics experiment that contained phenylalanine, as indicated by the +6 mass enrichment from $^{13}C_6$-phenylalanine. From the exact mass, we could generate a molecular formula, and by subtracting the atoms of phenylalanine, could hypothesize the modifications. (B) The most abundant compound from metabolomics indicates hydroxylation of phenylalanine, which we later determined to be **(2)** β-phenylisoserine. The second most abundant compound from metabolomics indicates phenylalanine and valine form a dipeptide, with an additional NH fragment, which we later determined to be **(6)** α,β-diaminophenylalanine-valine. Metabolomics was performed in triplicate in M9 minimal media with *E. coli* BL21.

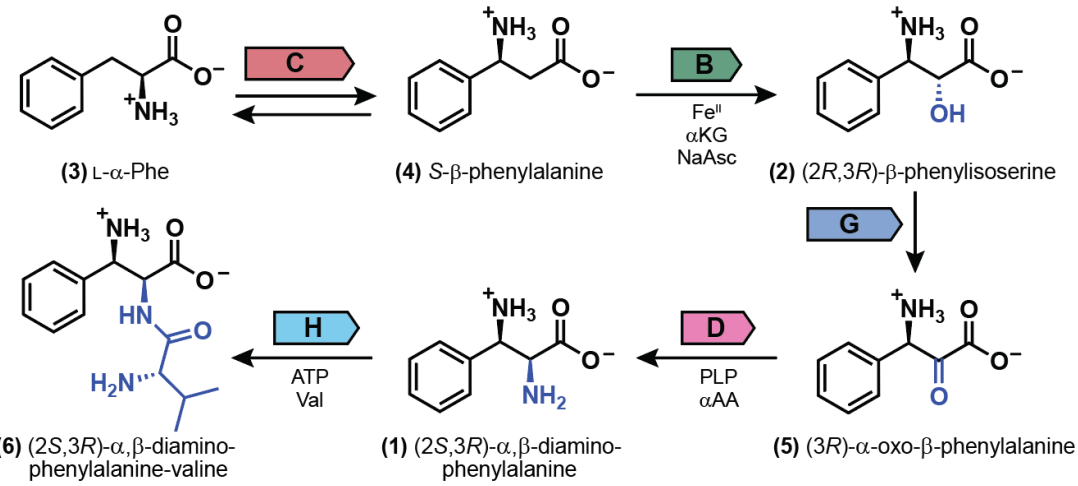


Figure 3.8. Experimental elucidation of the Pfrm biosynthetic pathway. *In vitro* pathway reconstitution revealed a biosynthetic pathway towards α,β-functionalized phenylalanine derivatives.

From *in vitro* and *in vivo* results, a proposal for the biosynthetic pathway can be generated (**Fig. 3.8**). PfrmC catalyzes the first step in the pathway to produce (*S*)-β-phenylalanine (**4**) from Phe. **4** is then hydroxylated by the TauD-like PfrmB. Indeed, *in vitro* reconstitution of PfrmBC with Phe, Fe(II), αKG, and ascorbate produces a compound with an *m/z* that matches the most abundant compound observed by metabolomics (**Fig 3.9A**). Additional assays show that PfrmB demonstrates no observable reaction with Phe but does react directly with β-phenylalanine to produce a hydroxylated product (**Fig. 3.9B**). Based on the likelihood of a transamination step in the pathway, we propose that the hydroxylation takes place at the α-position to produce (2*R*,3*R*)-β-phenylisoserine (**2**). To confirm stereochemistry, we supplied either enantiopure (*S*)- or (*R*)-β-phenylalanine to a reaction with PfrmB and compared the product to synthetic standards. From this experiment, we found that (*S*)-β-phenylalanine is the preferred substrate of PfrmB, with low activity towards (*R*)-β-phenylalanine (**Fig. 3.10**). The predicted stereochemical preference of PfrmC to produce (*S*)-β-phenylalanine results are consistent with other bacterial aminomutases [*42*]. We next isolated **2** by preparative HPLC for characterization by 2D-NMR, which confirmed its assignment as (2*R*,3*R*)-β-phenylisoserine (**Fig. 3.11**).

We propose that **2** is next further oxidized at Cα by FAD-dependent dehydrogenase PfrmG to produce the α-keto acid **5** (**Fig. 3.8**). *In vitro* reconstitution of the PfrmBCG reaction from Phe does indeed produce a product with the expected *m/z* =178.0510 for this oxidation (**Fig. 3.12A**). **2** can be functionalized by *O*-benzylhydroxylamine to produce the corresponding oxime, giving further support to formation of a carbonyl group by PfrmG (**Fig. 3.12B**). Based on the *in vitro* activity of PfrmD with amino acids (**Fig. 3.6C**), the next step is proposed to be a PLP-dependent transamination at Cα to yield α,β-diaminophenylalanine **1** (**Fig 3.13**). Formation of **1** was confirmed by *in vitro* reconstitution of the PfrmBCGD starting with Phe compared to an (2*S*,3*R*)-α,β-diaminophenylalanine authentic standard (**Fig. 3.14**). **1** can be functionalized twice with Fmoc, further confirming the presence of two amine groups. Using phenylpyruvic acid as a model substrate, PfrmD produces Phe, which was confirmed following derivatization to form the Mosher's amide (**Fig. 3.15**). This finding supports the assignment of the stereochemistry of **1** as (2*S*,3*R*), which is equivalent to an L-amino acid and consistent with the activity of the downstream PrfmH on L-amino acids (**Fig. 3.6B**). Finally, PrfmH ligates **1** to an amino acid, which is assigned to be L-valine based on the metabolomics study (**Fig. 3.7A**), producing dipeptide **6** (**Fig. 3.8**, **Fig. 3.16**). The connectivity of **6** was assigned based on the MS/MS fragmentation (**Fig. 3.17**).

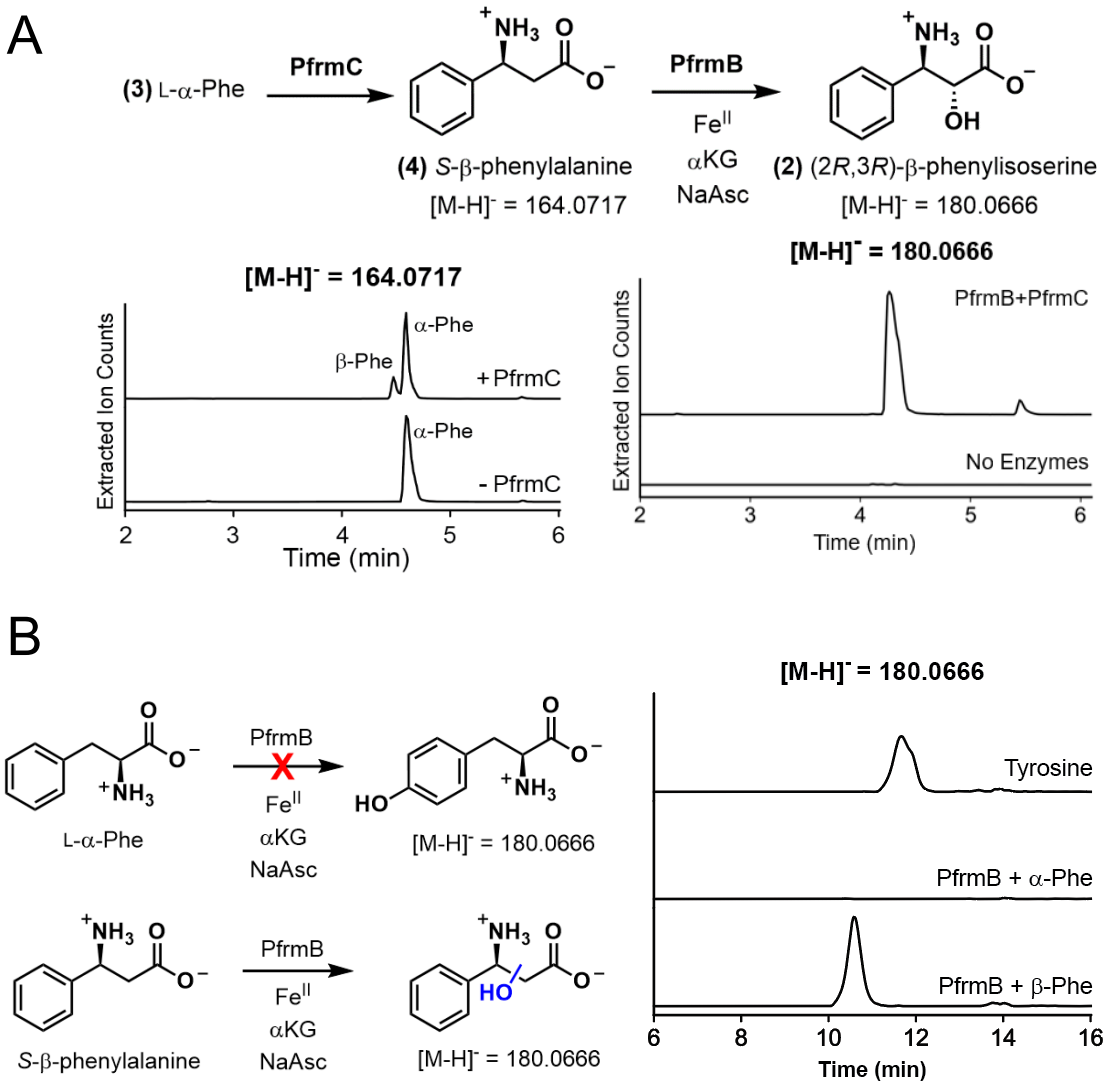


Figure 3.9. *In vitro* reconstitution of β-phenylisoserine biosynthesis. We reconstituted the biosynthetic pathway for the most abundant product detected from metabolomics *in vitro* with purified enzymes beginning from L-α-phenylalanine. (A) PfrmC produces **(4)** *S*-β-Phe (*m/z* = 164.0717) and PfrmB catalyzes C-H oxyfunctionalization of **(4)** *S*-β-phenylalanine to produce **(2)** (2*R*,3*R*)-β-phenylisoserine (*m/z* = 180.0666), which has the same mass as the product detected from metabolomics. These products were protected with Fmoc-Cl and analyzed with high-resolution LC/MS for nonpolar metabolites. The analytes were detected as either the Fmoc-protected form or Fmoc-free form (where the Fmoc fragmented off in the ion source). (B) *In vitro* screen of PfrmB with Phe shows no activity, but PfrmB with *S*-β-phenylalanine produces a compound with the same mass as tyrosine, but a different retention time, indicating a different regioisomer is formed. These products were analyzed with high-resolution LC/MS for polar metabolites.

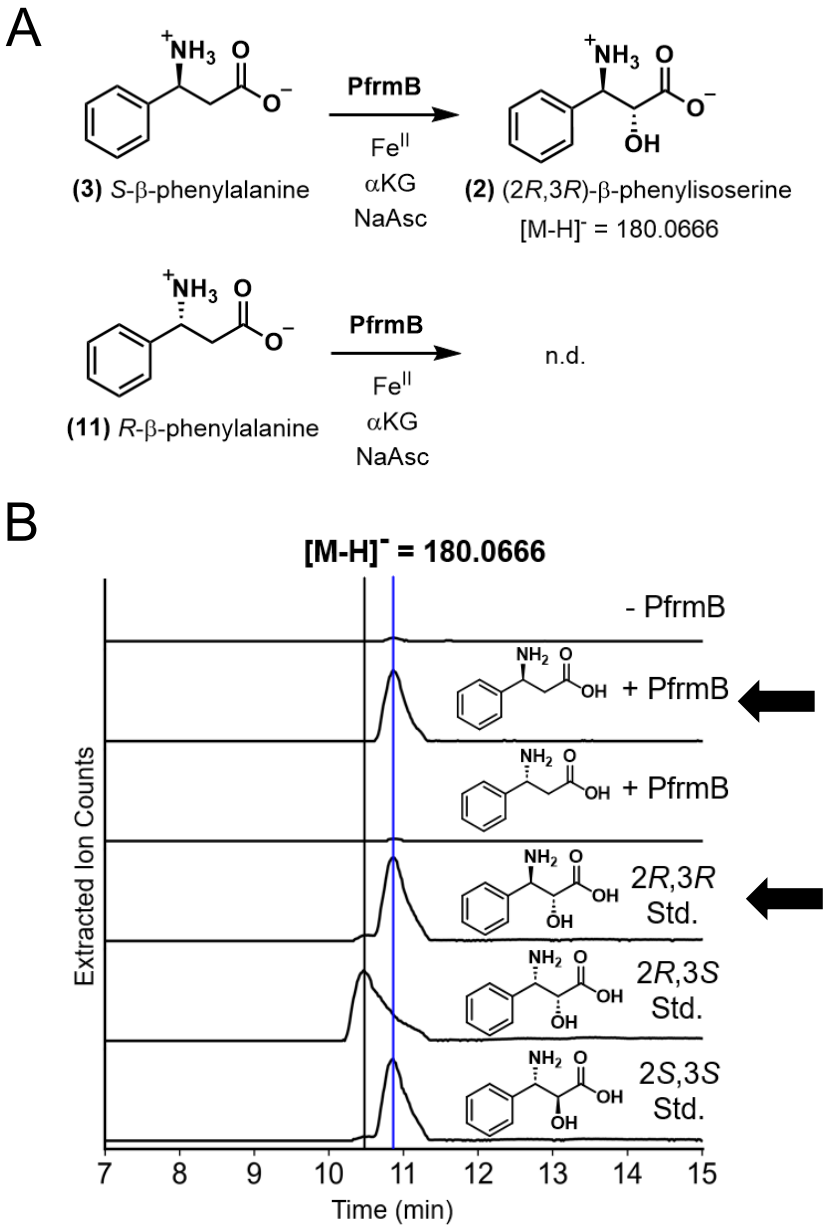
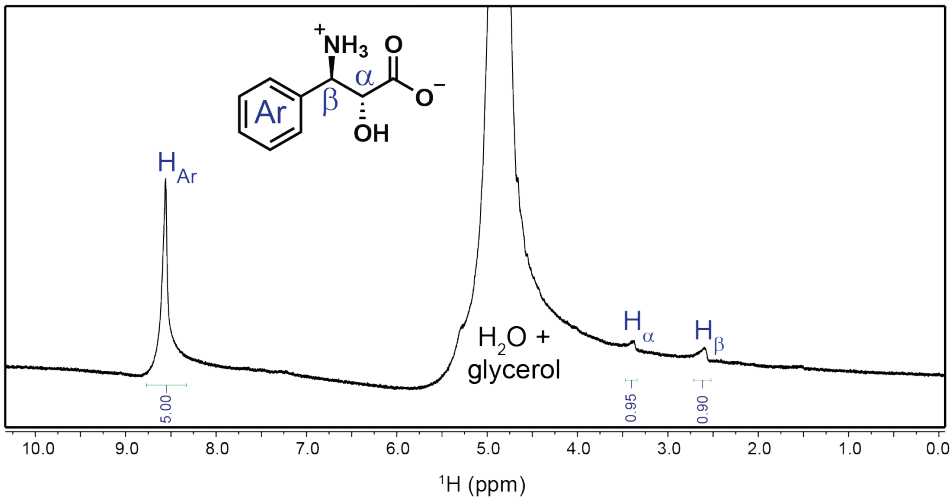
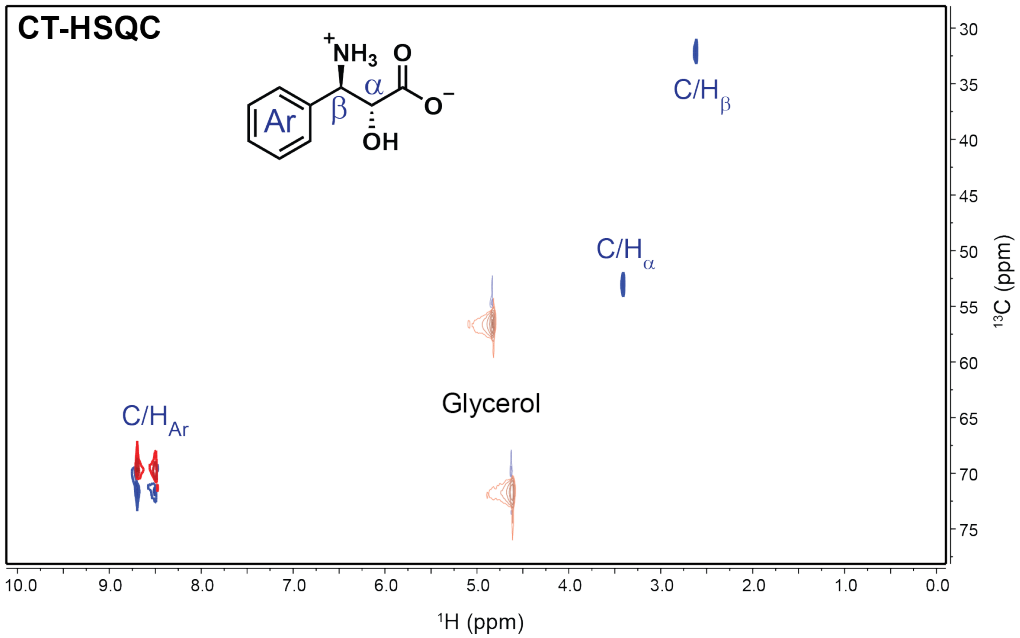


Figure 3.10. PfrmB converts (*S*)-β-phenylalanine to (2*R*,3*R*)-β-phenylisoserine. (A) Reactions of PfrmB with enantiomerically-pure (*S*)- or (*R*)-β-phenylalanine showed (*S*)-β-phenylalanine is the preferred substrate for this enzyme (n.d = not detectable). (B) Comparing the product with chemical standards enabled assignment of the stereochemistry. With the stereochemistry at the 3-position defined, comparing to chemical standards allowed us to assign the stereochemistry at the 2-position as (2*R*). As a note, the designation at the 3 position changes from (*S*) to (*R*) with the addition of the hydroxy moiety.

A



B



C

(2*R*,3*R*)-β-phenylisoserine	^{1}H (ppm)	^{13}C (ppm)
α	3.40	52.9
β	2.63	32.0
Ar	8.5-8.7	70-72

Figure 3.11. NMR characterization of PfrmB product (2*R*,3*R*)-β-phenylisoserine. (A) ^{1}H-NMR spectra of the PfrmB product integrates for the expected 7 hydrogens. (B) 2D ^{1}H-^{13}C CT-HSQC confirms that H_{α} and H_{β} are on separate carbons, indicating both the α- and β-carbons contain functional groups. (C) Assigned chemical shifts from the NMR analysis.

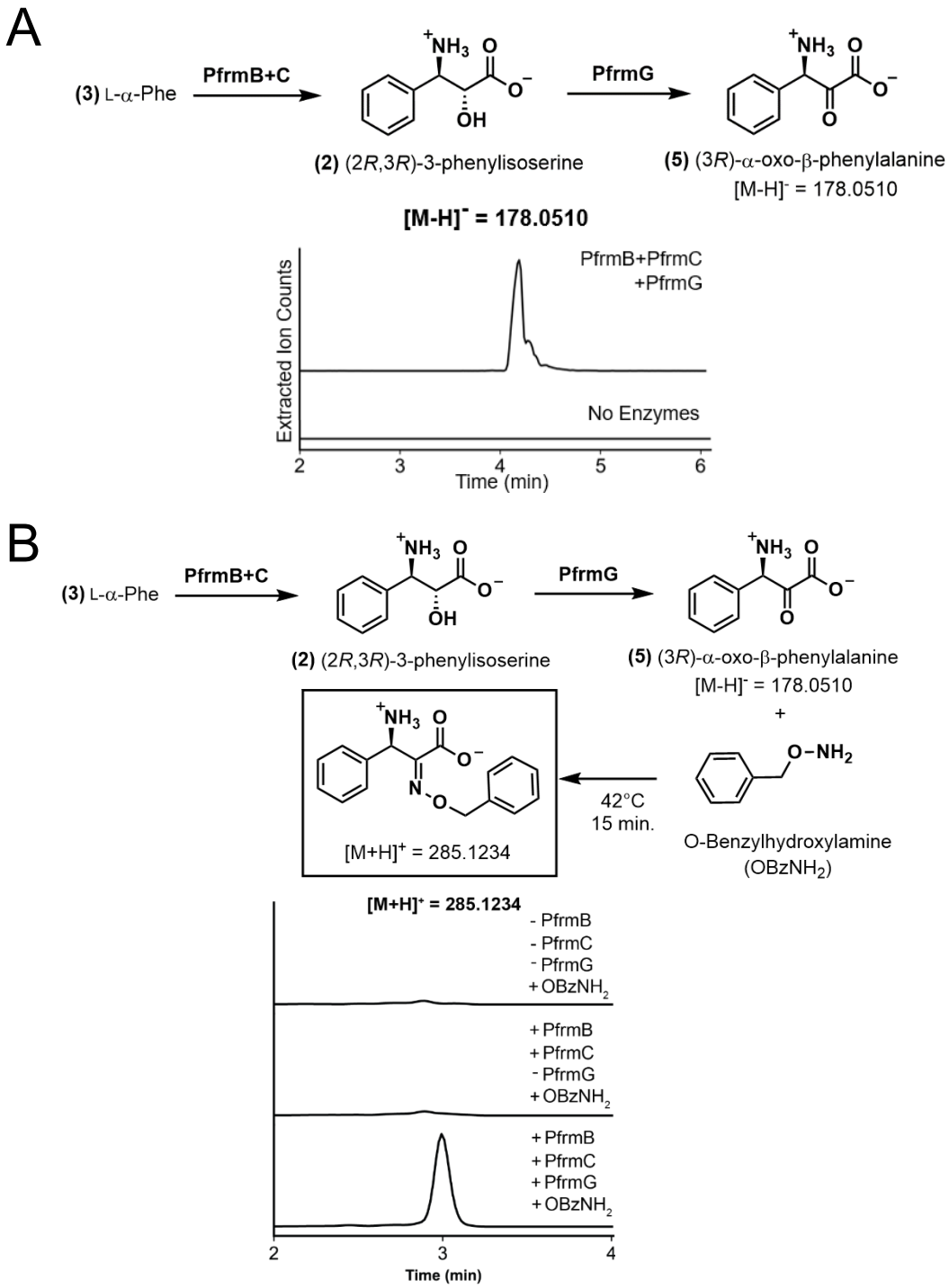


Figure 3.12. *In vitro* reconstitution of PfrmG's activity to produce (3*R*)-α-oxo-β-phenylalanine. (A) In a reaction containing PfrmBCG, the α-keto acid is detected (*m/z* = 178.0510), indicating PrfmG performs an oxidation of **(2)** (2*R*,3*S*)-β-phenylisoserine at the α-alcohol to yield **(5)** (3*R*)-α-oxo-β-phenylalanine. This products was protected with Fmoc-Cl and detected as the Fmoc-free form from ion source fragmentation. (B) To confirm PfrmG produces a ketone, O-benzylhydroxylamine was used to derivatize the product of PfrmG, which is a selective labelling reagent for ketones. We detected this product only when PfrmG was added to a reaction with PfrmB+C.

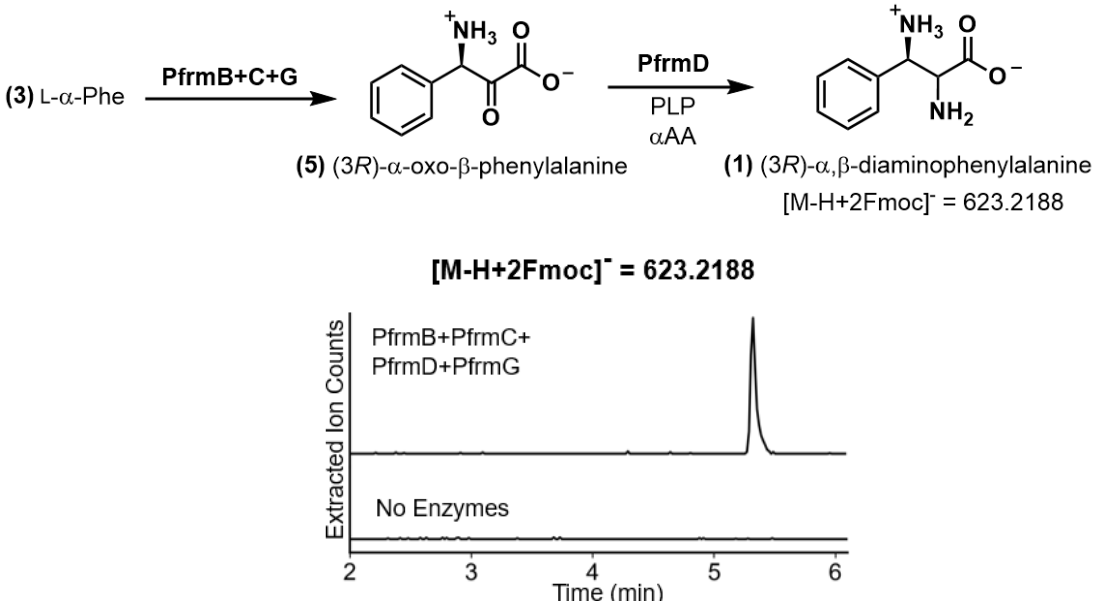


Figure 3.13. ***In vitro*** **reconstitution of α,β-diaminophenylalanine biosynthesis.** In a reaction with PfrmBCDG, a product corresponding to diaminophenylalanine product is detected with 2 Fmoc groups (***m/z*** = 623.2188), indicating PfrmD performs a transamination of **(5)** to yield **(1)** α,β-diaminophenylalanine. From this experiment, the stereochemistry at Cα was unclear.

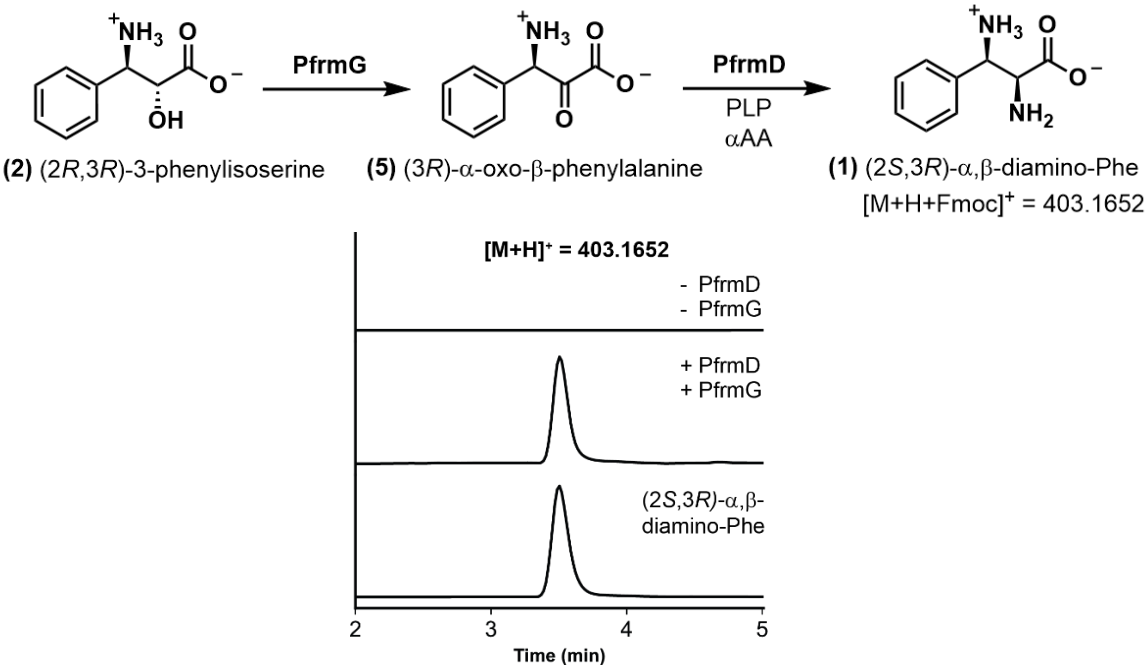


Figure 3.14. PfrmD shows L-amino acid transaminase activity with (3*R*)-α-oxo-β-phenylalanine. A cascade of PfrmG and PfrmD supplied with enantiopure (2*R*,3*R*)-β-phenylisoserine makes (2S,3R)-α,β-diaminophenylalanine, which has an equivalent retention time to the synthetic standard. Since the stereochemistry at C_β is fixed, this data supports the assignment of stereochemistry at C_α as shown.

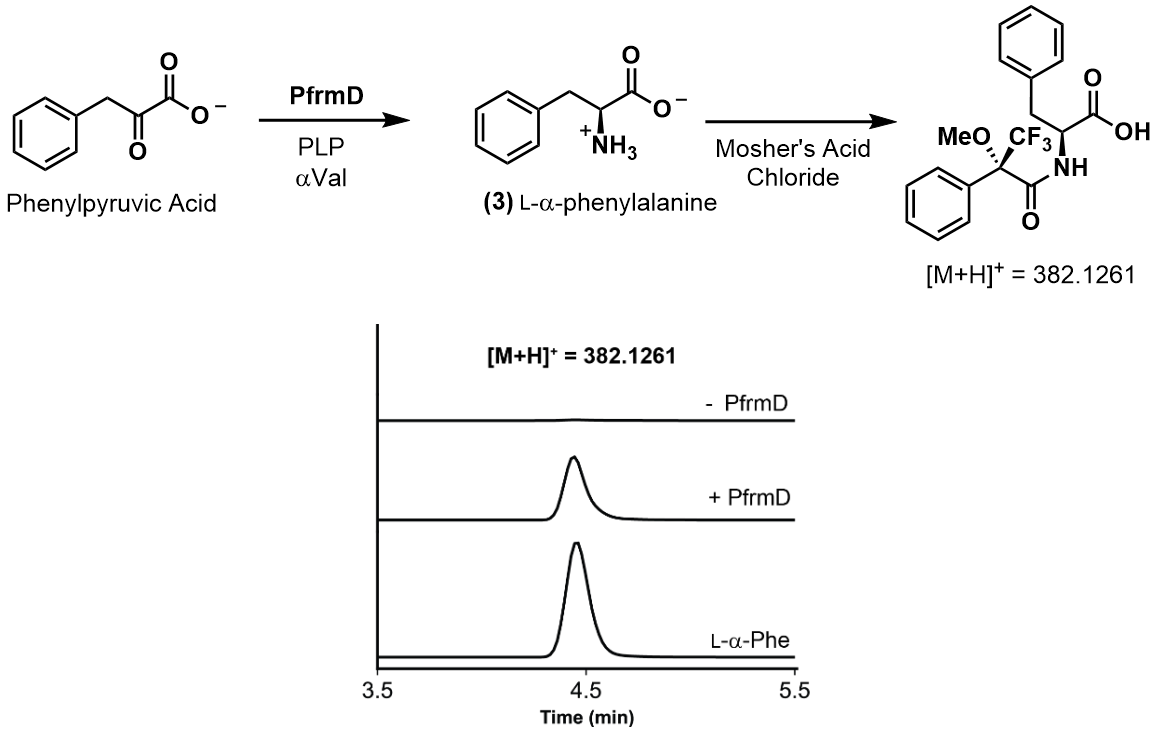


Figure 3.15. PfrmD shows L-amino acid transaminase activity with phenylpyruvic acid. In a reaction with PfrmD and phenylpyruvic acid, following derivatization with a chiral compound (Mosher's Acid Chloride), the product aligns with enantiopure L-α-phenylalanine that has been derivatized according to the same conditions, suggesting these two compounds are identical with identical α-stereochemistry.

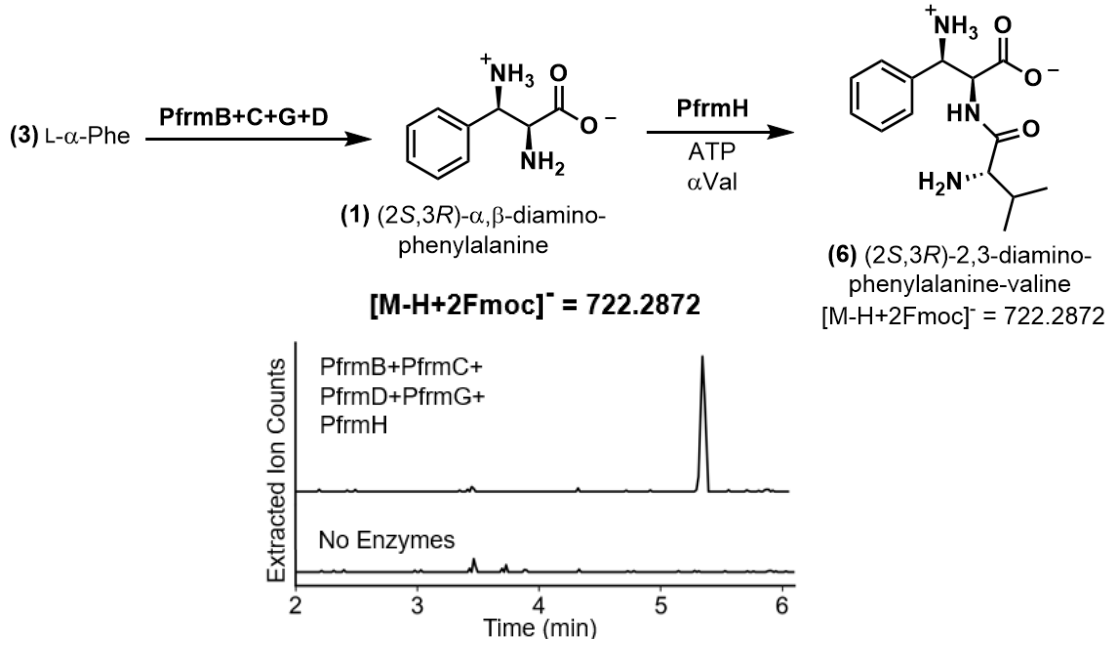


Figure 3.16. PfrmH ligates α,β-diaminophenylalanine to L-α-valine. In an *in vitro* cascade containing PfrmBCGDH, L-α-Phe, and L-α-Val, the mass for diaminophenylalanine-valine is detected with 2Fmoc groups. This molecule is equivalent to what was detected from the metabolomics experiment.

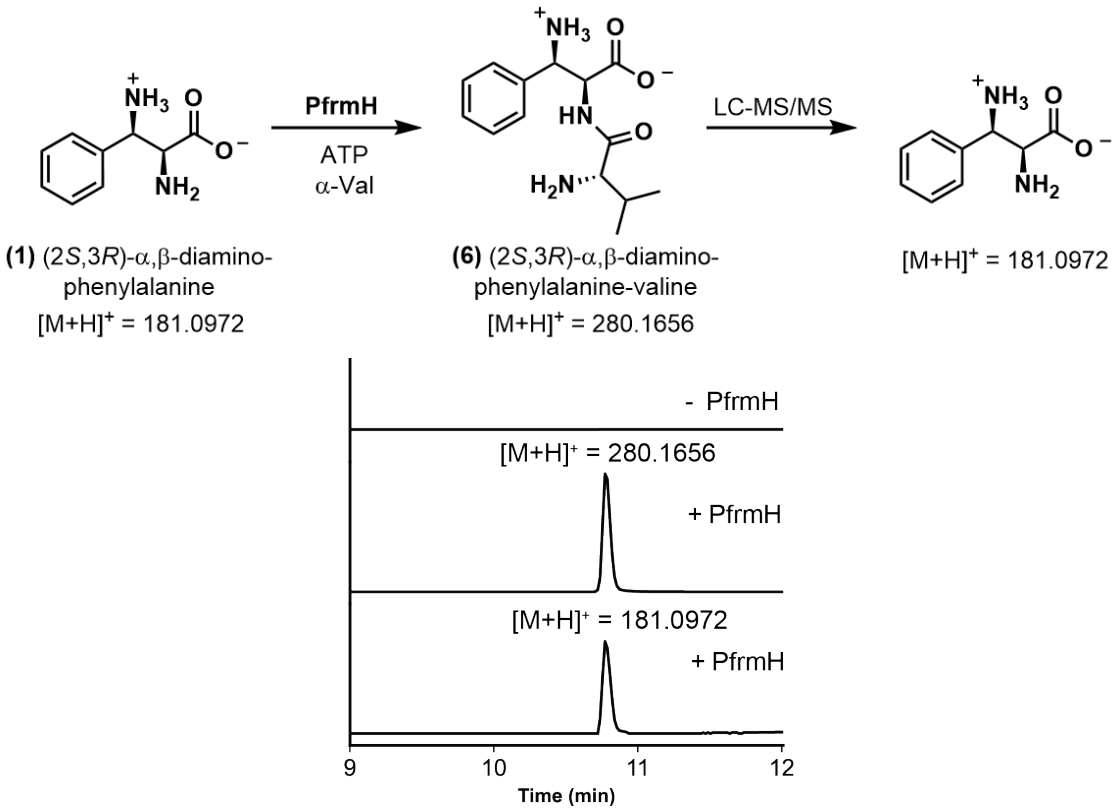
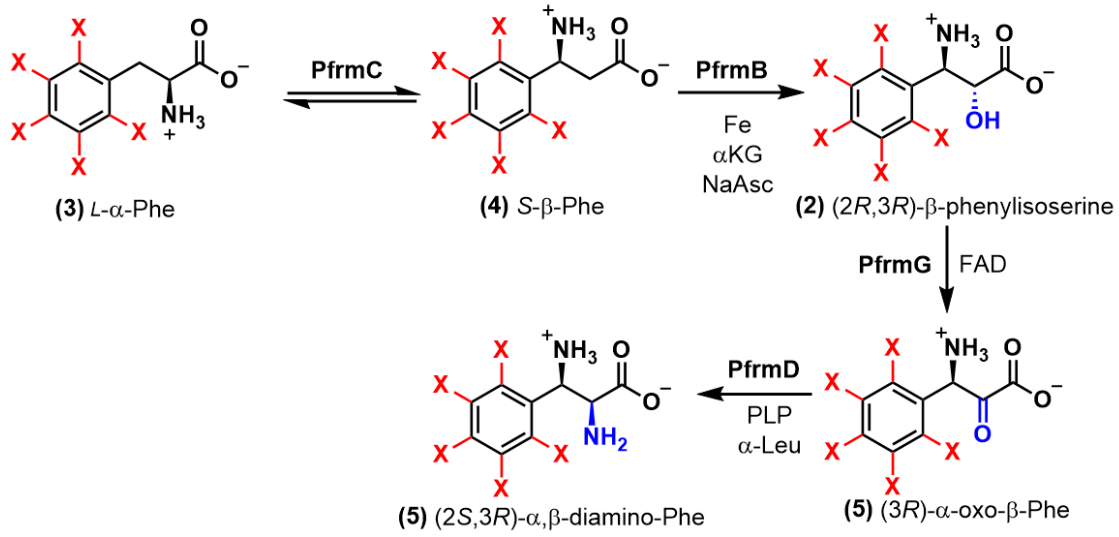


Figure 3.17. Determining the connectivity of the diaminophenylalanine-valine dipeptide. An *in vitro* reaction of PfrmH with synthetic (2*S*,3*R*)-α,β-diaminophenylalanine and L-α-valine produces the final product of the Pfrm pathway ([M+H]$^+$ = 280.1656), which is equivalent to the dipeptide product detected from metabolomics. With LC/MS fragmentation, the diaminophenylalanine component is detected in its entirety as a fragment (at the same retention time as the product), indicating this molecule is the amine donor to the peptide donor, so the connectivity of **(6)** was assigned as shown.

Developing an *in vitro* cascade towards α,β-functionalized acids. The Pfrm pathway provides a new biocatalytic approach for production of β-phenylisoserine **2** as well as discovery of a new amino acid, α,β-diaminophenylalanine **1**. Given the utility of the enzymatic transformations described, we were eager to explore the potential of these enzymes to produce various α,β-functionalized acids in an *in vitro* cascade. Since PfrmC requires an aromatic substrate, we examined a panel of 8 aromatic substituents as well as 2 bioisosteres for the phenyl group, with functional groups that could serve as downstream reaction handles or alter overall molecule properties (**Fig. 3.18**). We found conversion through the cascade to be greater than 20% for 8 of the 10 substrates, including *p*-iodo-phenylalanine, which is the first report of an aminomutase accepting this substrate with broad applications in cross-coupling [*43,44*]. The conversion through the cascade was improved with the addition of PfrmB, likely due to the ability of PfrmB to drive the reversible PfrmC isomerization forward by depletion of β-phenylalanine. As a general trend, overall yields through the four-enzyme cascade decreased as the substituents became larger and more polar. The ability of these substitutions to be tolerated by the cascade likely stems from the occurrence of the transformations along the carbon backbone with the exception of the first isomerization. This could be addressed through engineering efforts to expand the substrate binding pockets of PfrmC and PfrmB, which appear to be the limiting enzymes in the cascade.

A



B

Substrates	Conversion through PfrmC	Conversion through PfrmB	Conversion through PfrmG	Conversion through PfrmD
L-Phe	66.70%	100.00%	92.60%	98.73%
3-F-Phe	39.47%	100.00%	99.92%	98.20%
4-F-Phe	34.84%	85.12%	69.53%	83.13%
4-Cl-Phe	22.67%	64.42%	52.81%	71.82%
Thienyl	23.31%	55.01%	50.85%	64.91%
4-Br-Phe	29.25%	53.34%	33.31%	28.08%
Furyl	33.53%	49.37%	46.59%	42.01%
4-I-Phe	19.74%	31.11%	23.75%	23.62%
L-Tyr	21.81%	24.62%	16.94%	37.70%
4-Cyano-Phe	0.98%	13.99%	5.12%	20.56%
4-Nitro-Phe	6.25%	8.81%	4.05%	9.03%

Figure 3.18. Biocatalytic cascade for aromatic-substituted phenylalanine substrates. (A) Cascade reactions were performed in a series of 5 individual reactions, where each subsequent reaction included the next enzyme in the cascade (including a no enzyme control) with the same concentrations of starting material. (B) Conversion was measured through the entire cascade by consumption of starting material (as compared to the no enzyme control). Conversion increases from column 1 to 2 likely due to PfrmC's reaction being reversible, and PfrmB's reaction being irreversible, driving flux forward. Conversion, in some cases, is slightly higher in columns 4 vs column 3, as PfrmD could perform transamination on the starting amino acid substrate to make an α-keto acid.

Discovering an additional reaction outcome. Beyond the ability of this cascade to form α,β-functionalized acids with a 3-carbon backbone, we found this cascade could produce an alternative product through an oxidative decarboxylation catalyzed by PfrmG to produce D-phenylglycine from β-phenylisoserine (**Fig. 3.19**). PfrmG belongs to a class of oxidases that are flavin-dependent, and as a result, can produce hydrogen peroxide as a side-product [*45*]. Free hydrogen peroxide has been shown previously to catalyze decarboxylation of α-keto acids, allowing PfrmG to produce **7** from **5** [*46*]. D-phenylglycine and its derivatives are an important component of some semisynthetic antibiotics, with synthetic or chemoenzymatic routes that rely on hazardous components like cyanide and strong acid (**Fig. 3.20**). This cascade has the potential to provide an efficient and safe biocatalytic route towards 2-carbon difunctionalized acids like D-phenylglycine.

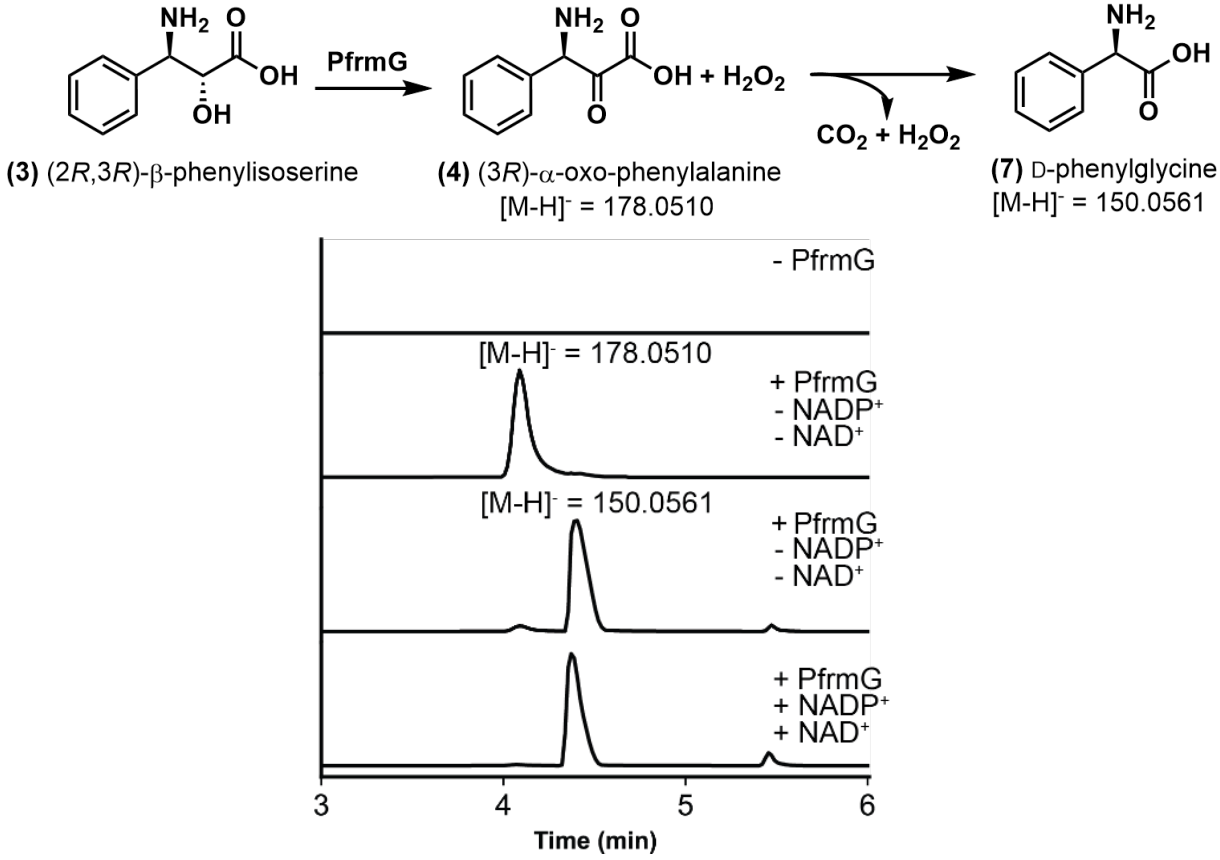


Figure 3.19. PfrmG forms D-phenylglycine from (2*R*,3*R*)-β-phenylisoserine. In an *in vitro* reaction of PfrmG with (2*R*,3*R*)-β-phenylisoserine, LC/MS shows PfrmG catalyzes both oxidation and subsequent decarboxylation to produce a phenylglycine mass. When this reaction was performed with or without nicoteinamides present, the reaction proceeded to phenylglycine, indicating nicotinamides do not act as a hydride acceptor and instead molecular oxygen is reduced by FAD to produce hydrogen peroxide.

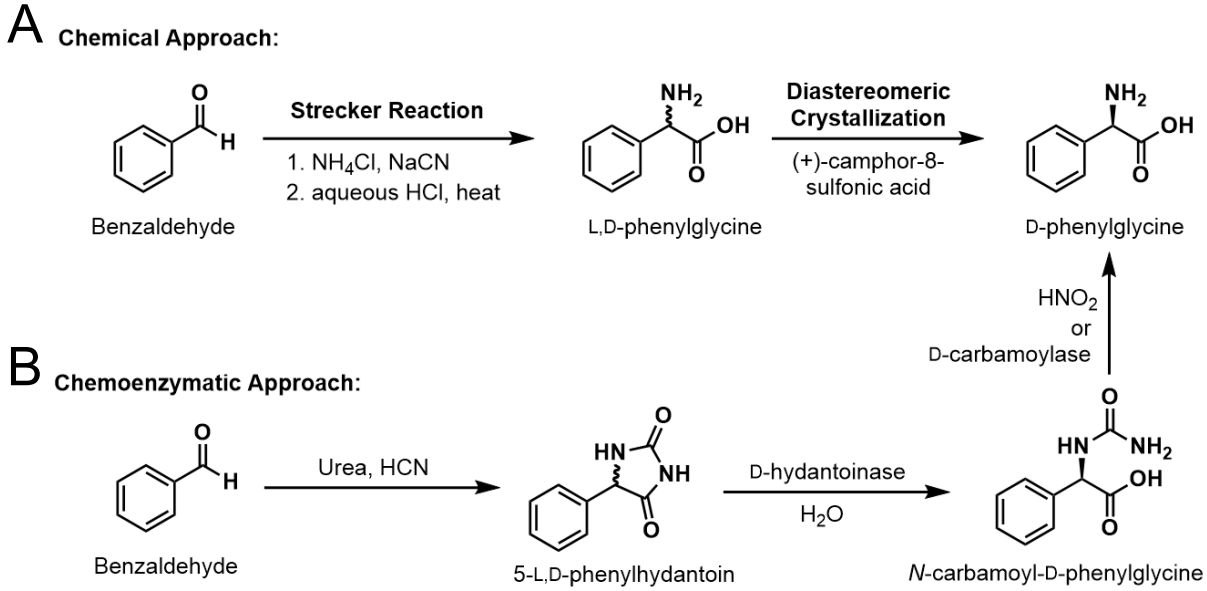


Figure 3.20. Established approaches for the synthesis of D-phenylglycine. For use in β-lactam antibiotic manufacturing, two established methods for D-phenylglycine synthesis include (A) an entirely chemical approach that relies on diastereomeric crystallization of a non-enantioselective Strecker reaction to produce D-phenylglycine or (B) a chemoenzymatic approach that leverages the enantioselectivity of D-hydantoinase to perform a dynamic kinetic resolution by racemization [*47*].

Expanding reaction scope with enzyme engineering. The product of PfrmB, β-phenylisoserine, is an important component of several medicinal natural products (**Fig. 3.21A**) [*48,49*]. However, the only known enzyme capable of forming this moiety requires (*R*)-β-phenylalanine be ligated to baccatin III en route to taxol [*50*]. As the first biocatalyst discovered to synthesize free-standing β-phenylisoserines, we envisioned PfrmB could enable a one-step approach to access alternate diastereomers of β-phenylisoserine (**Fig. 3.21B**). However, WT PfrmB was found to be highly selective towards (*S*)-β-phenylalanine with negligible activity towards (*R*)-β-phenylalanine, indicating protein engineering was necessary (**Fig. 3.10A**). The native reaction catalyzed by PfrmB is relatively fast (k_{cat} = 570 s^{-1}) compared to similar enzymes, suggesting this enzyme might retain catalytic activity after engineering **(Fig. 3.22)** [*14*].

As mentioned previously, PfrmB shows very little activity towards (*R*)-β-phenylalanine, so we performed a docking study for both the native substrate (*S*)-β-phenylalanine and desired substrate (*R*)-β-phenylalanine with a PfrmB AlphaFold2 model to predict the basis for substrate selectivity (**Fig. 3.23**) [*36*]. This study revealed a likely substrate binding pocket, with several polar residues that we postulate could interact with the β-amine, including S73, S74, Y103, and Q226. We hypothesized by removing these interactions in changing these residues to alanine, the alternate amine stereochemistry at the β-position would be better accommodated.

We inserted alanines at these sites and identified variants with improved activity. When we assayed the activity of PfrmB S73A towards (*R*)-β-phenylalanine, activity increased by about an order of magnitude compared to WT PfrmB, producing (2*S*,3*S*)-β-phenylisoserine (**Fig. 3.24A,B**). Combining S73A with Q226A enabled the production of the alternate diastereomer, (2R,3*S*)-β-phenylisoserine, although at the expense of catalytic activity (**Fig. 3.24C,D**). Additional alterations of polar residues in the active site led to either a further decrease in activity or undetectable activity (**Fig. 3.25**). From this effort, we were pleased to see PfrmB's activity increased for an unnatural substrate from near zero to enable stereochemical control over PfrmB's reaction scope, and we anticipate that directed evolution could provide additional benefits by increasing catalytic activity.

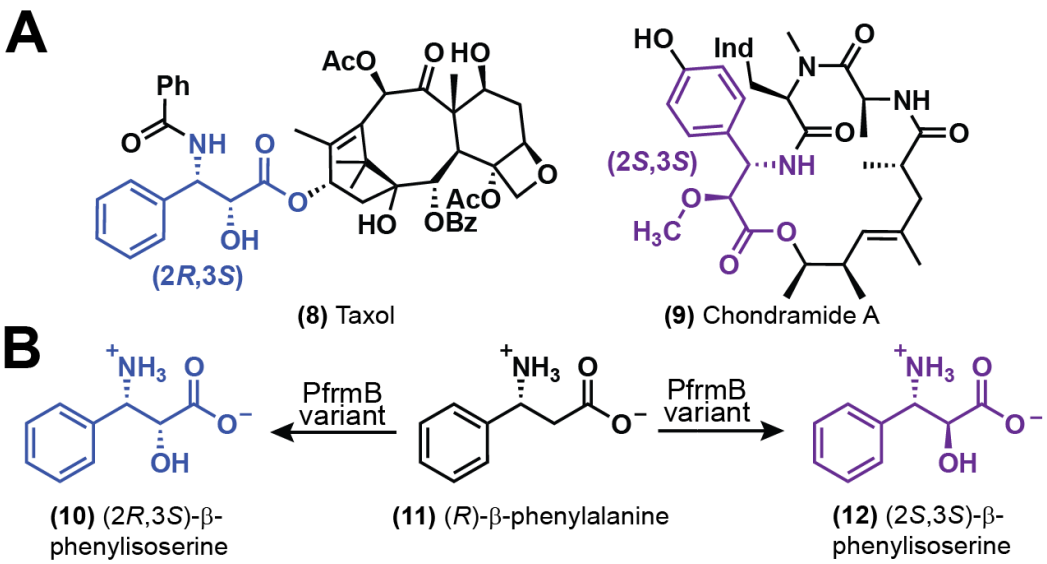
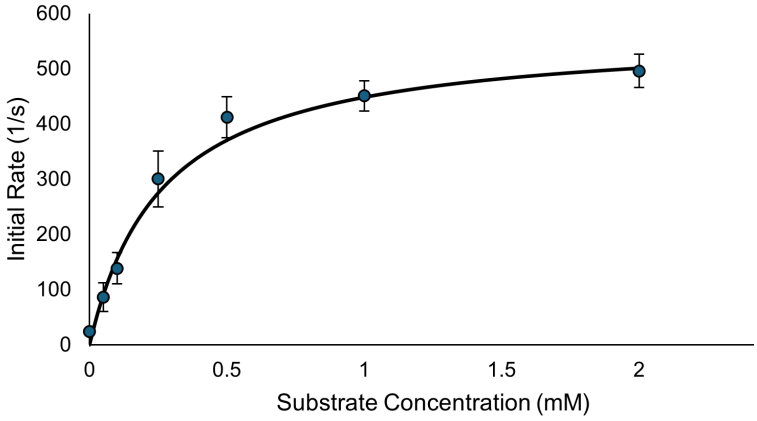


Figure 3.21. Expanding PfrmB's reaction scope to access useful alternate diastereomers of β-phenylisoserine. (A) Medicinal natural products with alternative diastereomers of β-phenylisoserine. (B) The desired fragments for PfrmB to produce. WT PfrmB has marginal activity towards *(R)*-β-phenylalanine, so protein engineering was begun to achieve these transformations.



	k_{cat} (s^{-1})	K_M (mM)	k_{cat} / K_M (mM^{-1} s^{-1})
PfrmB	570 ± 49	0.270 ± .072	2200 ± 360

Figure 3.22. Kinetics of WT PfrmB for (*S*)-β-phenylalanine. The rate of succinate formation by WT PfrmB was monitored by change in A_{340} using an NADH-coupled assay with succinyl-CoA synthetase, pyruvate kinase, and lactate dehydrogenase [*35*]. Reactions were initiated by the addition of PfrmB (10.6 mM) in the presence of varying concentrations of *S*-β-phenylalanine substrate (0 – 2mM). Initial rates of NADH consumption were measured by monitoring A_{340} using a SpectraMax M2 Microplate reader (Molecular Devices) at room temperature.

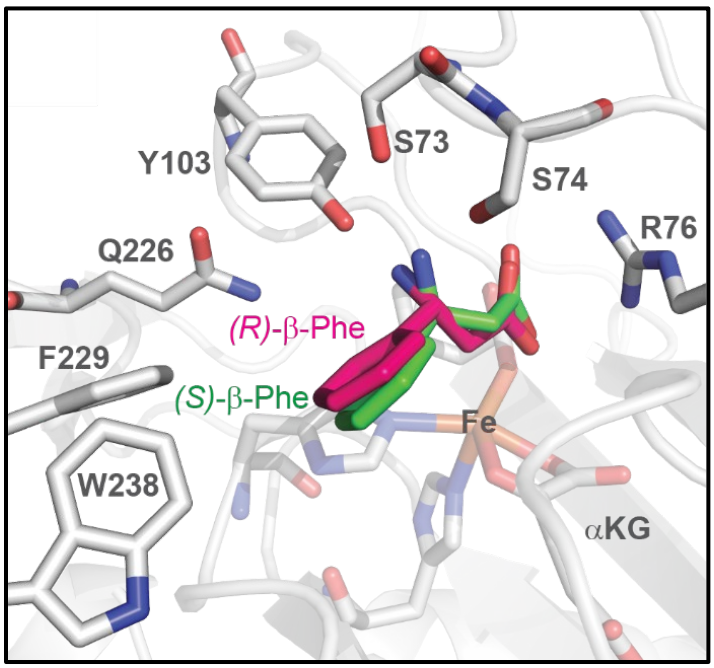


Figure 3.23. Docking study of PfrmB with (*R*)- and (*S*)-β-phenylalanine. PfrmB's active site docked with the native substrate (*S*)-β-phenylalanine (green) and desired substrate (*R*)-β-phenylalanine (pink) shows several polar side chains near the amine, such as S73, S74, Y103, and Q226, which were targeted for engineering. R76 likely binds the substrate's carboxylate and could play a role in constraining C-H functionalization to the α-position.

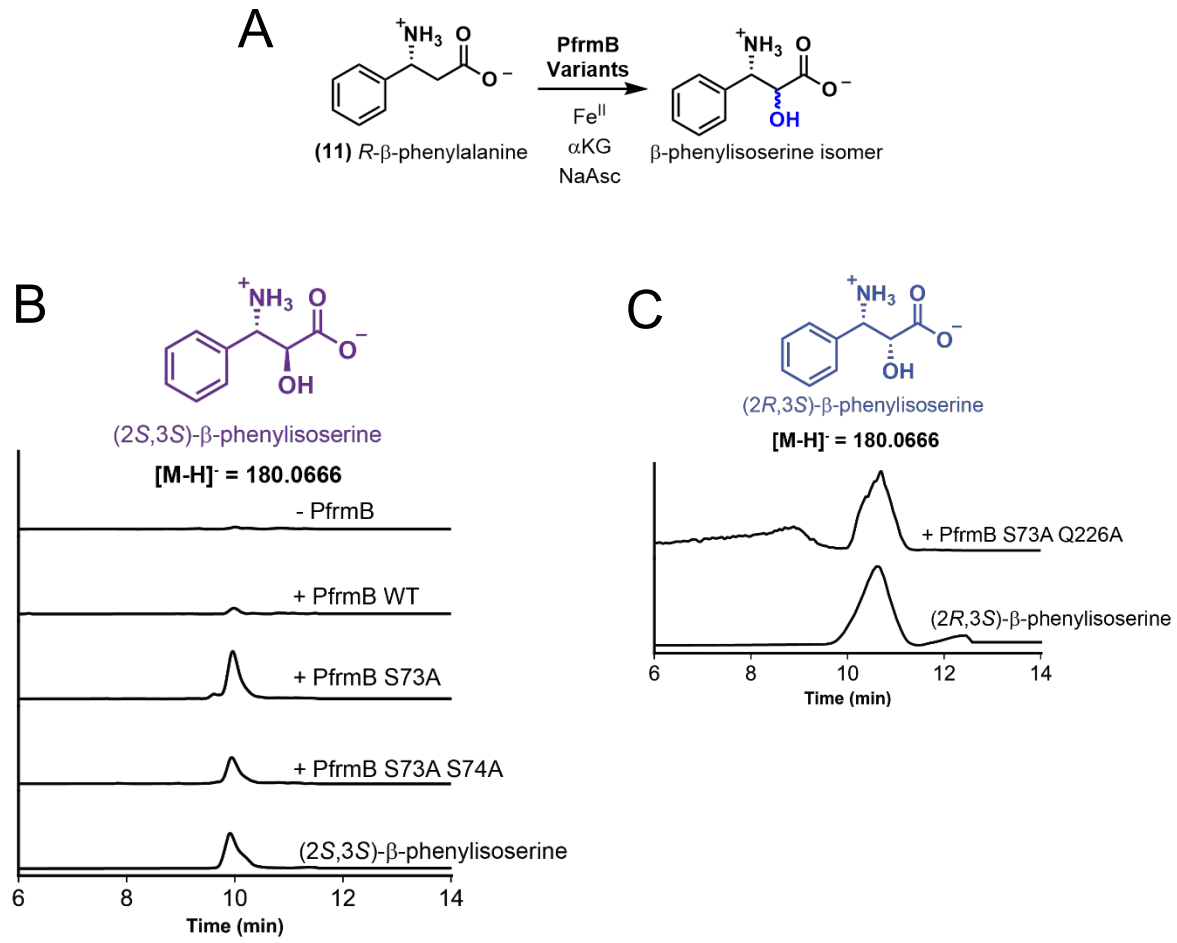


Figure 3.24. Rational engineering of PfrmB to accept (*R*)-β-phenylalanine. (A) Engineering of PfrmB was necessary to expand substrate scope towards (*R*)-β-phenylalanine. All variants were tested under identical conditions for their ability to accept (*R*)-β-phenylalanine as a substrate (B) Wildtype PfrmB shows very little activity towards (*R*)-β-phenylalanine and changing serine 73 to an alanine noticably improved this activity. By retention time, we assigned the product as (2*S*,3*S*)-β-phenylisoserine. (C) From the S73A mutant, an additional mutation of Q226A altered the reaction outcome, where the product peak aligned with (2*R*,3*S*)-β-phenylisoserine. (D) Summary of results for most productive mutants, which increased activity from near 0 and enabled access to alternate diastereomers of β-phenylisoserine.

A

Desired Substrate: 3*R*-β-phenylalanine

Sequence	Conversion	Fold-Change from WT	Diastereochemical Outcome
WT	1.28%	1.00	**(2*S*,3*S*)**
S73A	9.93%	7.78	**(2*S*,3*S*)**
S73A Q226A	3.97%	3.11	**(2*R*,3*S*)**

Native Substrate: 3*S*-β-phenylalanine

Sequence	Conversion	Fold-Change from WT	Diastereochemical Outcome
WT	99.8%	1.00	(2*R*,3*R*)
S73A	99.5%	0.997	(2*R*,3*R*)
S73A Q226A	6.87%	0.0689	(2*R*,3*R*)

B

Desired Substrate: 3*R*-β-phenylalanine

Variant	Conversion	Diastereochemical Outcome
S73A Y103F	n.d.	N/A
S73A Y103A	n.d.	N/A
S73A Y103A Q226A	n.d.	N/A
S73A Y103F Q226A	n.d.	N/A
S73A S74A	5.15%	(2*S*,3*S*)

Native Substrate: 3*S*-β-phenylalanine

Variant	Conversion	Diastereochemical Outcome
S73A Y103F	9.95%	(2*R*,3*R*)
S73A Y103A	n.d.	N/A
S73A Y103A Q226A	n.d.	N/A
S73A Y103F Q226A	n.d.	N/A
S73A S74A	99.8%	(2*R*,3*R*)

Figure 3.25. Compiled results from PfrmB rational engineering. (A) The S73A mutation maintained native activity towards *S*-β-phenylalanine and expanded activity towards *R*-β-phenylalanine, to produce (2*S*,3*S*)-β-phenylisoserine. The additional glutamine to alanine mutation severely reduced native activity but altered the reaction outcome to (2*S*,3*R*)-β-phenylisoserine. (B) Additional mutations either provided no additional benefits or eliminated activity altogether.

Exploiting enzyme promiscuity for C-H functionalization of free acids. Given the ability of the PfrmB hydroxylase to accept various substrates via PfrmC, we thought that it might be possible to directly produce amino acids from simpler starting materials by C-H functionalization. Towards this goal, we began by testing whether PfrmB can perform hydroxylation of hydrocinnamic acid and found that this substrate was quantitatively converted to the corresponding α-hydroxy acid (**Fig. 3.26, 3.27B**). Furthermore, PfrmGD were very active towards β-phenyllactic acid to yield the L-α-amino acid (**3**). These findings indicated the β-amino group is not absolutely required for these enzymes and that this cascade could be utilized for synthesizing amino acids directly from carboxylic acids. Additionally, it may be possible to perform two cycles of this cascade PfrmC to synthesize the α,β-disubstituted amino acids from aromatic acids instead of the amino acid (**Fig. 3.27A**). To further explore the substrate promiscuity of PfrmB, we examined if non-aromatic substrates could also be accepted (**Table 3.1**). We found that cyclohexanepropionic acid is accepted by PfrmB at unexpectedly high conversion (55.9%) with high conversion to the α-amino acid by PfrmGD (88.6%) (**Fig. 3.27C**). In addition, we were surprised to find PfrmB is promiscuous towards a variety of smaller carboxylic acids with either nonpolar or polar side chains. We found all of them were accepted to various degrees to form the associated α-hydroxy acid (**Fig. 3.27D**). Altogether, the enzymes in this cascade are promising biocatalysts for performing some challenging reactions very efficiently where the use of alternative transaminases could afford amino acids of either stereochemistry.

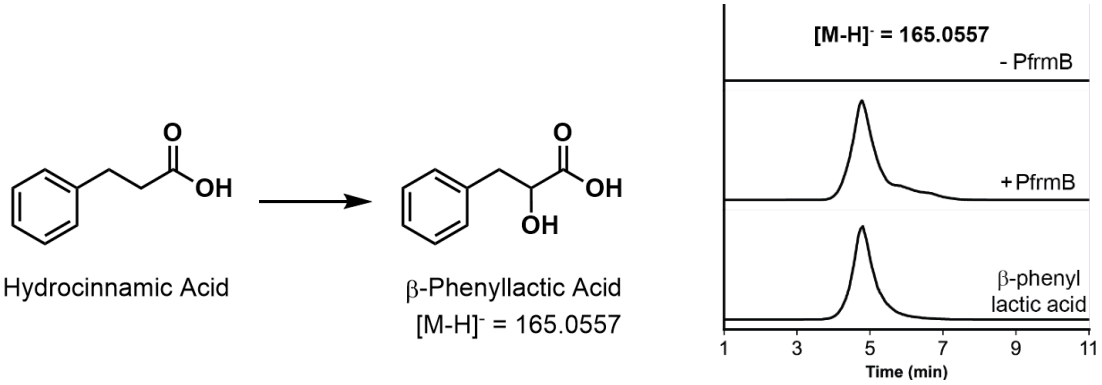


Figure 3.26. C-H functionalization of hydrocinnamic acid with PfrmB. Wild-type PfrmB can hydroxylate at the α-position of hydrocinnamic acid to yield β-phenyllactic acid, which was confirmed with a synthetic standard. This demonstrates the β-amino group is not a required binding determinant for PfrmB turnover.

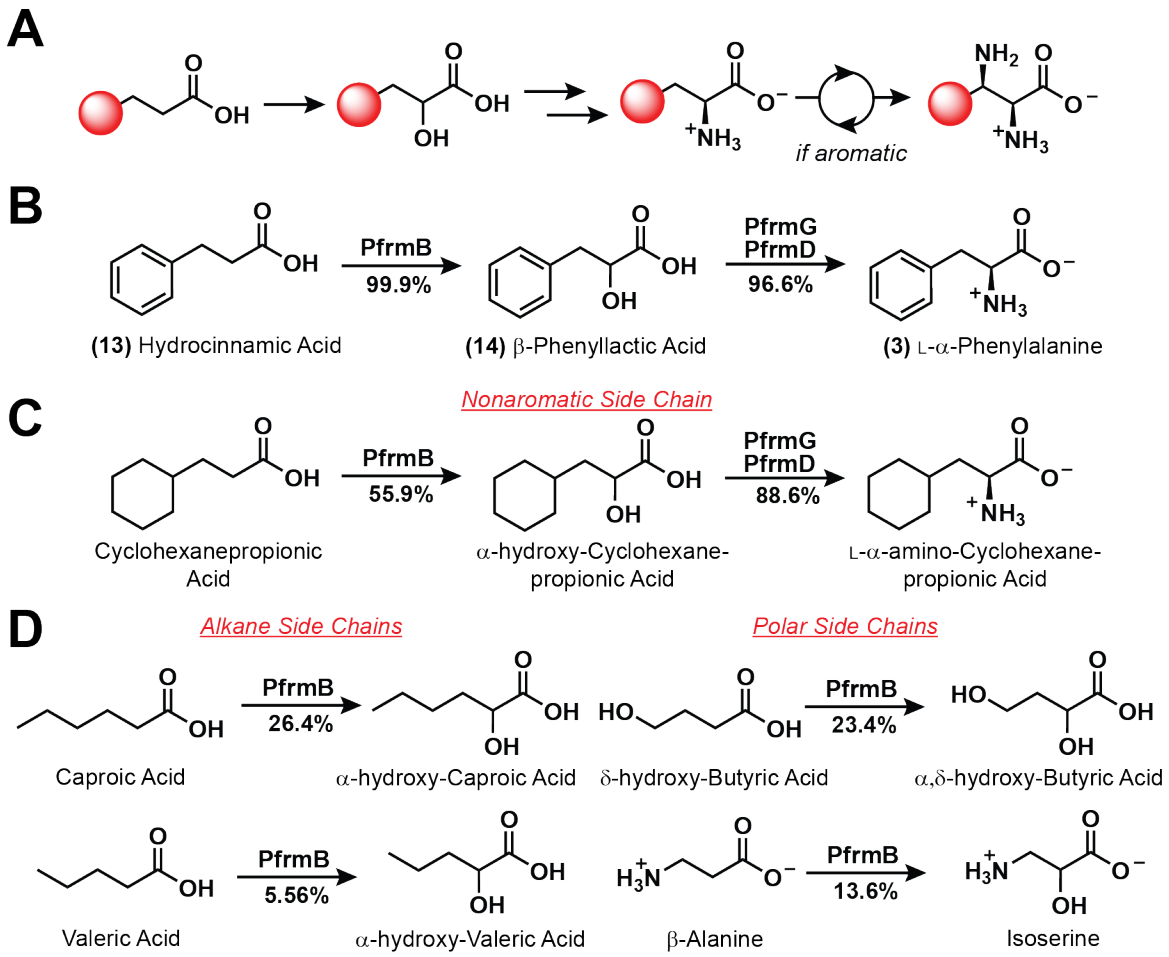


Figure 3.27. Biocatalytic cascade to α-hydroxy and α,β-amino acids from diverse carboxylic acids. (A) Through C-H activation, carboxylic acids can be functionalized at the α and β positions through the biocatalytic cascade discovered in this work. (B) The Pfrm cascade has very high activity towards hydrocinnamic acid to synthesize L-α-phenylalanine. (C) The Pfrm cascade can functionalize at the α-position of a nonaromatic carboxylic acid, with conversion through PfmB being the limiting component. (D) PfrmB shows activity towards alkane carboxylates and polar carboxylates, where conversion was calculated by decrease in starting material. Regiochemistry is assigned at the α-position through a combination of synthetic standards, acceptance by downstream enzymes, and enzyme active site structure.

Substrate	Conversion through PfrmB (%)	Conversion through PfrmGD (%)
Hydrocinnamic Acid	99.9	96.6
Cyclohexanepropionic acid	55.9	88.6
Caproic Acid	26.4	-
δ-hydroxy-butyric acid	23.4	-
β-alanine	13.6	-
Valeric Acid	5.56	-
3,3-dimethyl-butyric acid	0.413	-

Table 3.1. Summary of PfrmB promiscuity towards carboxylic acid substrates. Conversion from *in vitro* reactions of these substrates with PfrmB was measured by decrease in substrate peak as detected by LC/MS. Only hydrocinnamic acid and cyclohexanepropionic acid were tested with downstream enzymes PfrmGD.

3.4 Conclusions

As amino acids continue to be important scaffolds for building molecule complexity, access to a toolbox of active and selective catalysts is becoming increasingly important. By discovering biocatalysts in the context of biosynthetic gene clusters, chemists can leverage whole cascades for synthesizing complex molecules in one-pot. Here, we utilized a novel genome mining approach to discover a biocatalytic cascade for producing α,β-difunctionalized amino acids through C-H functionalization. We leveraged this to efficiently synthesize valuable molecules entirely from enzymes, further expanding the current repertoire of biocatalysts for C-H functionalization of amino acids and accessing fragments that have not previously been accessible from biocatalysis alone.

Several opportunities present themselves for expanding upon this discovery. By replacing enzymes in this cascade, alternative reaction outcomes can increase the stereochemical options. For instance, replacing PfrmD with a promiscuous D-amino acid transaminase could double the number of possible products from this cascade [*53*]. In addition, the α-keto acid could be a starting point for branching at the amine with stereochemical control using imine reductases to generate more complex structures [*54*]. Beyond this cascade, many other Fe/αKG-dioxygenase-containing BGCs remain unexplored. In addition, the discovery approach developed in this work, where a class of enzymes was searched within the dataset of tRDs to focus the discovery towards amino acid modifications, could be easily replicated with other enzyme families of interest, such as N-monooxygenases, ATP-grasp ligases, or radical SAMs. We anticipate this approach can be readily applied to expand the utility of biocatalysis for challenging transformations.

3.5 References

1. Lehninger, A. L.; Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry*; Macmillan, 2005.
2. Blaskovich, M. A. T. Unusual Amino Acids in Medicinal Chemistry. *J. Med. Chem.* **2016**, *59* (24), 10807–10836. https://doi.org/10.1021/acs.jmedchem.6b00319.
3. Henninot, A.; Collins, J. C.; Nuss, J. M. The Current State of Peptide Drug Discovery: Back to the Future? *J. Med. Chem.* **2018**, *61* (4), 1382–1414. https://doi.org/10.1021/acs.jmedchem.7b00318.
4. Wandrey, C.; Liese, A.; Kihumbu, D. Industrial Biocatalysis: Past, Present, and Future. *Org. Process Res. Dev.* **2000**, *4* (4), 286–290. https://doi.org/10.1021/op9901011.
5. Mangas-Sanchez, J.; France, S. P.; Montgomery, S. L.; Aleku, G. A.; Man, H.; Sharma, M.; Ramsden, J. I.; Grogan, G.; Turner, N. J. Imine Reductases (IREDs). *Current Opinion in Chemical Biology* **2017**, *37*, 19–25. https://doi.org/10.1016/j.cbpa.2016.11.022.
6. Guo, F.; Berglund, P. Transaminase Biocatalysis: Optimization and Application. *Green Chemistry* **2017**, *19* (2), 333–360. https://doi.org/10.1039/C6GC02328B.
7. Zetzsche, L. E.; Chakrabarty, S.; Narayan, A. R. H. The Transformative Power of Biocatalysis in Convergent Synthesis. *J. Am. Chem. Soc.* **2022**, *144* (12), 5214–5225. https://doi.org/10.1021/jacs.2c00224.
8. Ren, X.; Fasan, R. Engineered and Artificial Metalloenzymes for Selective C–H Functionalization. *Current Opinion in Green and Sustainable Chemistry* **2021**, *31*, 100494. https://doi.org/10.1016/j.cogsc.2021.100494.
9. Aguilar Troyano, F. J.; Merkens, K.; Anwar, K.; Gómez-Suárez, A. Radical-Based Synthesis and Modification of Amino Acids. *Angewandte Chemie International Edition* **2021**, *60* (3), 1098–1115. https://doi.org/10.1002/anie.202010157.
10. Bruggink, A.; Schoevaart, R.; Kieboom, T. Concepts of Nature in Organic Synthesis: Cascade Catalysis and Multistep Conversions in Concert. *Org. Process Res. Dev.* **2003**, *7* (5), 622–640. https://doi.org/10.1021/op0340311.
11. Zwick, C. R. I.; Renata, H. Overview of Amino Acid Modifications by Iron- and α-Ketoglutarate-Dependent Enzymes. *ACS Catal.* **2023**, *13* (7), 4853–4865. https://doi.org/10.1021/acscatal.3c00424.
12. Zwick, C. R. I.; Renata, H. Remote C–H Hydroxylation by an α-Ketoglutarate-Dependent Dioxygenase Enables Efficient Chemoenzymatic Synthesis of Manzacidin C and Proline Analogs. *J. Am. Chem. Soc.* **2018**, *140* (3), 1165–1169. https://doi.org/10.1021/jacs.7b12918.
13. Mao, Y.; Zhang, W.; Fu, Z.; Liu, Y.; Chen, L.; Lian, X.; Zhuo, D.; Wu, J.; Zheng, M.; Liao, C. Versatile Biocatalytic C(sp^3)−H Oxyfunctionalization for the Site- Selective and Stereodivergent Synthesis of α- and β-Hydroxy Acids. *Angewandte Chemie International Edition* **2023**, *62* (33), e202305250. https://doi.org/10.1002/anie.202305250.
14. Lian, X.; Mao, Y.; Fu, Z.; Zhang, W.; Chen, J.; Zhuo, D.; Zheng, M.; Wu, J.; Liao, C. Siteselective and Enantiocomplementary C(sp^3)–H Oxyfunctionalization for Synthesis of α-Hydroxy Acids. *ACS Catal.* **2024**, 4463–4470. https://doi.org/10.1021/acscatal.4c00398.
15. DiRocco, D. A.; Zhong, Y.-L.; Le, D. N.; McCann, S. D.; Hethcox, J. C.; Kim, J.; Kolev, J. N.; Kosjek, B.; Dalby, S. M.; McMullen, J. P.; Gangam, R.; Morris, W. J. Evolution of a Green and Sustainable Manufacturing Process for Belzutifan: Part 1─Process History and

Development Strategy. *Org. Process Res. Dev.* **2024**, *28* (2), 404–412. https://doi.org/10.1021/acs.oprd.3c00408.

16. Cheung-Lee, W. L.; Kolev, J. N.; McIntosh, J. A.; Gil, A. A.; Pan, W.; Xiao, L.; Velásquez, J. E.; Gangam, R.; Winston, M. S.; Li, S.; Abe, K.; Alwedi, E.; Dance, Z. E. X.; Fan, H.; Hiraga, K.; Kim, J.; Kosjek, B.; Le, D. N.; Marzijarani, N. S.; Mattern, K.; McMullen, J. P.; Narsimhan, K.; Vikram, A.; Wang, W.; Yan, J.-X.; Yang, R.-S.; Zhang, V.; Zhong, W.; DiRocco, D. A.; Morris, W. J.; Murphy, G. S.; Maloney, K. M. Engineering Hydroxylase Activity, Selectivity, and Stability for a Scalable Concise Synthesis of a Key Intermediate to Belzutifan. *Angewandte Chemie International Edition* **2024**, *63* (13), e202316133. https://doi.org/10.1002/anie.202316133.
17. Viso, A.; Fernández de la Pradilla, R.; García, A.; Flores, A. α,β-Diamino Acids: Biological Significance and Synthetic Approaches. *Chem. Rev.* **2005**, *105* (8), 3167–3196. https://doi.org/10.1021/cr0406561.
18. Ziora, Z.; Skwarczynski, M.; Kiso, Y. Medicinal Chemistry of α-Hydroxy-β-Amino Acids. In *Amino Acids, Peptides and Proteins in Organic Chemistry*; John Wiley & Sons, Ltd, 2011; pp 189–245. https://doi.org/10.1002/9783527631827.ch6.
19. Kingwell, K. Macrocycle Drugs Serve up New Opportunities. *Nature Reviews Drug Discovery* **2023**, *22* (10), 771–773. https://doi.org/10.1038/d41573-023-00152-3.
20. Fuchs, S. W.; Grundmann, F.; Kurz, M.; Kaiser, M.; Bode, H. B. Fabclavines: Bioactive Peptide–Polyketide-Polyamino Hybrids from Xenorhabdus. *ChemBioChem* **2014**, *15* (4), 512–516. https://doi.org/10.1002/cbic.201300802.
21. Motorina, I. A.; Huel, C.; Quiniou, E.; Mispelter, J.; Adjadj, E.; Grierson, D. S. Phenylisoserine: A Versatile Amino Acid for the Construction of Novel β-Peptide Structures. *J. Am. Chem. Soc.* **2001**, *123* (1), 8–17. https://doi.org/10.1021/ja002700+.
22. Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat Methods* **2009**, *6* (5), 343–345. https://doi.org/10.1038/nmeth.1318.
23. Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R. D.; Bairoch, A. ExPASy: The Proteomics Server for in-Depth Protein Knowledge and Analysis. *Nucleic Acids Res* **2003**, *31* (13), 3784–3788. https://doi.org/10.1093/nar/gkg563.
24. Oberg, N.; Zallot, R.; Gerlt, J. A. EFI-EST, EFI-GNT, and EFI-CGFP: Enzyme Function Initiative (EFI) Web Resource for Genomic Enzymology Tools. *Journal of Molecular Biology* **2023**, *435* (14), 168018. https://doi.org/10.1016/j.jmb.2023.168018.
25. Zallot, R.; Oberg, N.; Gerlt, J. A. The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry* **2019**, *58* (41), 4169–4182. https://doi.org/10.1021/acs.biochem.9b00735.
26. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* **2003**, *13* (11), 2498–2504. https://doi.org/10.1101/gr.1239303.
27. The UniProt Consortium. UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Research* **2023**, *51* (D1), D523–D531. https://doi.org/10.1093/nar/gkac1052.

28. Larsson, A. AliView: A Fast and Lightweight Alignment Viewer and Editor for Large Datasets. *Bioinformatics* **2014**, *30* (22), 3276–3278. https://doi.org/10.1093/bioinformatics/btu531.
29. Kalyaanamoorthy, S.; Minh, B. Q.; Wong, T. K. F.; von Haeseler, A.; Jermiin, L. S. ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates. *Nat Methods* **2017**, *14* (6), 587–589. https://doi.org/10.1038/nmeth.4285.
30. Nguyen, L.-T.; Schmidt, H. A.; von Haeseler, A.; Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* **2015**, *32* (1), 268–274. https://doi.org/10.1093/molbev/msu300.
31. Hoang, D. T.; Chernomor, O.; von Haeseler, A.; Minh, B. Q.; Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* **2018**, *35* (2), 518–522. https://doi.org/10.1093/molbev/msx281.
32. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v5: An Online Tool for Phylogenetic Tree Display and Annotation. *Nucleic Acids Research* **2021**, *49* (W1), W293–W296. https://doi.org/10.1093/nar/gkab301.
33. Neidhardt, F. C.; Bloch, P. L.; Smith, D. F. Culture Medium for Enterobacteria. *J Bacteriol* **1974**, *119* (3), 736–747.
34. Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M. MS-DIAL: Data-Independent MS/MS Deconvolution for Comprehensive Metabolome Analysis. *Nat Methods* **2015**, *12* (6), 523–526. https://doi.org/10.1038/nmeth.3393.
35. Luo, L.; Pappalardi, M. B.; Tummino, P. J.; Copeland, R. A.; Fraser, M. E.; Grzyska, P. K.; Hausinger, R. P. An Assay for Fe(II)/2-Oxoglutarate-Dependent Dioxygenases by Enzyme-Coupled Detection of Succinate Formation. *Analytical Biochemistry* **2006**, *353* (1), 69–74. https://doi.org/10.1016/j.ab.2006.03.033.
36. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.
37. Westendorf, A.; Benndorf, D.; Müller, R. H.; Babel, W. The Two Enantiospecific Dichlorprop/α-Ketoglutarate-Dioxygenases from *Delftia Acidovorans* MC1 – Protein and Sequence Data of RdpA and SdpA. *Microbiological Research* **2002**, *157* (4), 317–322. https://doi.org/10.1078/0944-5013-00164.
38. Drake, E. J.; Gulick, A. M. Three-Dimensional Structures of *Pseudomonas Aeruginosa* PvcA and PvcB, Two Proteins Involved in the Synthesis of 2-Isocyano-6,7-Dihydroxycoumarin. *Journal of Molecular Biology* **2008**, *384* (1), 193–205. https://doi.org/10.1016/j.jmb.2008.09.027.
39. Strieker, M.; Kopp, F.; Mahlert, C.; Essen, L.-O.; Marahiel, M. A. Mechanistic and Structural Basis of Stereospecific Cβ-Hydroxylation in Calcium-Dependent Antibiotic, a Daptomycin-Type Lipopeptide. *ACS Chem. Biol.* **2007**, *2* (3), 187–196. https://doi.org/10.1021/cb700012y.

40. Bollinger Jr., J. M.; Chang, W.; Matthews, M. L.; Martinie, R. J.; Boal, A. K.; Krebs, C. Mechanisms of 2-Oxoglutarate-Dependent Oxygenases: The Hydroxylation Paradigm and Beyond. **2015**. https://doi.org/10.1039/9781782621959-00095.
41. Hausinger, R. P. Fe(II)/α-Ketoglutarate-Dependent Hydroxylases and Related Enzymes. *Critical Reviews in Biochemistry and Molecular Biology* **2004**, *39* (1), 21–68. https://doi.org/10.1080/10409230490440541.
42. Turner, N. J. Ammonia Lyases and Aminomutases as Biocatalysts for the Synthesis of α-Amino and β-Amino Acids. *Current Opinion in Chemical Biology* **2011**, *15* (2), 234–240. https://doi.org/10.1016/j.cbpa.2010.11.009.
43. Csuka, P.; Juhász, V.; Kohári, S.; Filip, A.; Varga, A.; Sátorhelyi, P.; Bencze, L. C.; Barton, H.; Paizs, C.; Poppe, L. Pseudomonas Fluorescens Strain R124 Encodes Three Different MIO Enzymes. *ChemBioChem* **2018**, *19* (4), 411–418. https://doi.org/10.1002/cbic.201700530.
44. Ratnayake, N. D.; Wanninayake, U.; Geiger, J. H.; Walker, K. D. Stereochemistry and Mechanism of a Microbial Phenylalanine Aminomutase. *J. Am. Chem. Soc.* **2011**, *133* (22), 8531–8533. https://doi.org/10.1021/ja2030728.
45. Ratnayake, N. D.; Liu, N.; Kuhn, L. A.; Walker, K. D. Ring-Substituted α-Arylalanines for Probing Substituent Effects on the Isomerization Reaction Catalyzed by an Aminomutase. *ACS Catal.* **2014**, *4* (9), 3077–3090. https://doi.org/10.1021/cs500474s.
46. Ho, C. C.; Olding, A.; Smith, J. A.; Bissember, A. C. Nuances in Fundamental Suzuki–Miyaura Cross-Couplings Employing [Pd(PPh3)4]: Poor Reactivity of Aryl Iodides at Lower Temperatures. *Organometallics* **2018**, *37* (11), 1745–1750. https://doi.org/10.1021/acs.organomet.8b00189.
47. Massey, V. Activation of Molecular Oxygen by Flavins and Flavoproteins. *Journal of Biological Chemistry* **1994**, *269* (36), 22459–22462. https://doi.org/10.1016/S0021-9258(17)31664-2.
48. Lopalco, A.; Dalwadi, G.; Niu, S.; Schowen, R. L.; Douglas, J.; Stella, V. J. Mechanism of Decarboxylation of Pyruvic Acid in the Presence of Hydrogen Peroxide. *Journal of Pharmaceutical Sciences n/a* (n/a). https://doi.org/10.1002/jps.24653.
49. Wegman, M. A.; Janssen, M. H. A.; van Rantwijk, F.; Sheldon, R. A. Towards Biocatalytic Synthesis of β-Lactam Antibiotics. *Advanced Synthesis & Catalysis* **2001**, *343* (6–7), 559–576. https://doi.org/10.1002/1615-4169(200108)343:6/7<559::AID-ADSC559>3.0.CO;2-Z.
50. Sharma, S.; Lagisetti, C.; Poliks, B.; Coates, R. M.; Kingston, D. G. I.; Bane, S. Dissecting the Paclitaxel-Microtubule Association: Quantitative Assessment of the 2′-OH Group. *Biochemistry* **2013**, *52* (13), 2328–2336. https://doi.org/10.1021/bi400014t.
51. Sasse, F.; Kunze, B.; Gronewold, T. M. A.; Reichenbach, H. The Chondramides: Cytostatic Agents From Myxobacteria Acting on the Actin Cytoskeleton. *JNCI: Journal of the National Cancer Institute* **1998**, *90* (20), 1559–1563. https://doi.org/10.1093/jnci/90.20.1559.
52. Sanchez-Muñoz, R.; Perez-Mata, E.; Almagro, L.; Cusido, R. M.; Bonfill, M.; Palazon, J.; Moyano, E. A Novel Hydroxylation Step in the Taxane Biosynthetic Pathway: A New Approach to Paclitaxel Production by Synthetic Biology. *Front Bioeng Biotechnol* **2020**, *8*, 410. https://doi.org/10.3389/fbioe.2020.00410.
53. Galkin, A.; Kulakova, L.; Yamamoto, H.; Tanizawa, K.; Tanaka, H.; Esaki, N.; Soda, K. Conversion of α-Keto Acids to d-Amino Acids by Coupling of Four Enzyme Reactions.

Journal of Fermentation and Bioengineering **1997**, *83* (3), 299–300. https://doi.org/10.1016/S0922-338X(97)80997-X.

54. Yao, P.; Marshall, J. R.; Xu, Z.; Lim, J.; Charnock, S. J.; Zhu, D.; Turner, N. J. Asymmetric Synthesis of N-Substituted α-Amino Esters from α-Ketoesters via Imine Reductase-Catalyzed Reductive Amination. *Angewandte Chemie International Edition* **2021**, *60* (16), 8717–8721. https://doi.org/10.1002/anie.202016589.

Chapter 4: *Efforts towards Elucidating Unknown Functions of Biosynthetic Genes using tRDs and Engineering Site-Specific Incorporation*

Portions of this work were performed with the following people:

Claire Li contributed towards the elucidation of the *Phal* BGC through protein purification, *in vitro* reactions, and metabolomics.

Tavis Chen contributed towards the elucidation of the *Pant* BGC through protein purification, *in vitro* reactions, and metabolomics.

Isabel Piper contributed towards the engineering of site-specific incorporation of propargylglycine through plasmid assembly and FACS-based screening.

4.1 Introduction

Inherent to many enzyme discovery approaches is a reliance on homology to known enzymes [*1*]. This is often a successful strategy because while metabolism is incredibly diverse in its reaction scope, it is generally understood that new enzyme function evolves from some known activity [*2–4*]. However, this approach makes genome mining more challenging for poorly annotated or entirely unannotated genes within biosynthetic gene clusters (BGCs); these enzymes, often assigned domains of unknown function (DUF), are challenging to study in isolation but could perform novel reactions [*5,6*]. To help elucidate the roles of these enzymes, determining the activities of neighboring genes can shed light [*7*]. For instance, the products of neighboring enzymes could be substrates for the DUF or vice versa. Altogether, gene neighborhood information can aid the discovery of entirely new enzyme types.

With the establishing of tRNA-deacylases (tRDs) as a class-defining marker of amino acid modifications, we were interested in deploying this genome mining approach for the discovery and elucidation of new enzyme function. A particular advantage of utilizing tRNA-deacylases as genome mining beacons is they not only constrain the substrate scope to amino acids but also are agnostic to the types of transformations that are possible. In this chapter, I detail our efforts to elucidate the functions of poorly annotated or entirely unannotated enzymes by using tRDs as genomic markers. Specifically, we identified two biosynthetic pathways that both appear to be performing modifications to amino acids but through very different strategies. In addition to peptide-forming ATP-grasps and C-H activating Fe/αKG dioxygenases, these BGCs contain unique enzymes that are poorly annotated or completely unannotated. Through a combination of *in vitro* screening and *in vivo* metabolomics, we have begun efforts to elucidate their function. In this work, we also look beyond the AlaX family of tRDs to the YbaK/INS family. Using our established analysis pipeline, we identified a large amount of biosynthetic diversity in this family and selected several pathways for experimental characterization.

With the discovery of several biosynthetic pathways for noncanonical amino acids (ncAAs), we also explored the opportunity to synthesize and incorporate a ncAA *in tandem* in a host organism. Specifically, we developed a platform for engineering site-specific incorporation of the terminal alkyne-containing amino acid propargylglycine using fluorescence-activated cell-sorting (FACS). This approach, if fully developed, would enable a clickable probe to be installed into proteins without needing to supply the ncAA exogenously, with applications in chemical biology and catalysis. Towards this goal, we have begun engineering the biosynthetic pathway for Pra biosynthesis to improve enzyme solubility. Taken together, these efforts could enable a new amino acid to be added to the genetic code.

4.2 Materials and Methods

Commercial materials. Luria-Bertani (LB) Broth Miller, LB Agar Miller, Terrific Broth (TB), 2xYT Broth, and glycerol were purchased from EMD Biosciences (Darmstadt, Germany). Carbenicillin (Cb), kanamycin (Km), chloramphenicol (Cm), isopropyl-β-D-thiogalactopyranoside (IPTG), sodium chloride, dithiothreitol (DTT), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), magnesium chloride hexahydrate, sodium hydroxide, and dextrose were purchased from Fisher Scientific (Pittsburgh, PA). Phosphoenolpyruvate (PEP), adenosine triphosphate sodium salt (ATP), β-nicotinamide adenine dinucleotide reduced form dipotassium salt (NADH), β-nicotinamide adenine dinucleotide phosphate reduced form disodium salt (NADPH), flavin adenine dinucleotide disodium salt (FAD), lysozyme, magnesium sulfate heptahydrate, calcium chloride, β-mercaptoethanol (βME), sodium phosphate dibasic heptahydrate, sodium phosphate monobasic, dimethylsulfoxide (DMSO), acetonitrile (LC/MS-grade), ammonium formate (LC/MS-grade), ammonium iron (II) sulfate hexahydrate, α-ketoglutaric acid sodium salt, sodium ascorbate, copper sulfate, Tris(3-hydroxypropyltriazolylmethyl)amine (THPTA), L-propargylglycine (Pra), all 20 proteinogenic amino acids (L-alanine, L-arginine, L-asparagine, L-aspartic acid, L-cysteine (HCl), L-glutamic acid (potassium salt), L-glutamine, L-glycine, L-histidine, L-isoleucine, L-leucine, L-lysine, L-methionine, L-phenylalanine, L-proline, L-serine, L-serine, L-threonine, L-tryptophan, L-tyrosine, L-valine), $^{13}C_6$-L-leucine, $^{13}C_6$,^{15}N-L-isoleucine, pyruvate kinase/lactate dehydrogenase, anhydrotetracycline hydrochloride, 24-well plates and Aeraseal covers were purchased from Sigma-Aldrich (St. Louis, MO). PageRuler Plus Prestained Protein Ladder was purchased from ThermoFisherScientific (Waltham, MA). Formic acid was purchased from Acros Organics (Morris Plains, NJ). Restriction enzymes, Phusion DNA polymerase, T5 exonuclease, and Taq DNA ligase were purchased from New England Biolabs (Ipswich, MA). Deoxynucleotides (dNTPs), were purchased from Invitrogen (Carlsbad, CA). Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA), resuspended at a stock concentration of 100 μM in 10 mM Tris-HCl, pH 8.5, and stored at 4 °C. DNA purification kits and Ni-NTA agarose were purchased from Qiagen (Valencia, CA). Complete EDTA-free protease inhibitor was purchased from Roche Applied Science (Penzberg, Germany). PD-10 desalting columns were purchased from GE Healthcare Life, (Pittsburg, PA). Amicon Ultra 10,000 MWCO and 30,000 MWCO centrifugal concentrators, Benzonase, and Milli-Q Gradient water purification system were purchased from Millipore (Billerica, MA). Acrylamide/bis-acrylamide (30%, 37.5:1) and electrophoresis grade sodium dodecyl sulfate (SDS) were purchased from Bio-Rad Laboratories (Hercules, CA). $^{13}C_5$,^{15}N-L-valine was purchased from Cambridge Isotope Laboratories (Andover, MA). CalFluor-488-Azide was purchased from Vector Laboratories (Newark, CA). GeneMorph II Error-prone polymerase kit was purchased from Agilent (Santa Clara, CA). 4-Cl-allylglycine was purchased from Chemspace (Monmouth Junction). All chemicals were used as purchased without purification.

Bacterial strains. *E. coli* DH10B-T1^R was used for plasmid construction. *E. coli* BL21 Star (DE3) was used for heterologous protein production and metabolomics experiments. *Pantoea sp. GM01* was received as a gift from Dr. Dale Pelletier (Oak Ridge National Laboratory), from which the *Pant* biosynthetic gene cluster (BGC) was cloned.

Construction of plasmids. Gibson ligation [*8*] was used to carry out all plasmid construction using *E. coli* DH10B-T1^R as the cloning host. All PCR amplifications were carried out with Phusion polymerase (New England Biolabs, Ipswich MA) using the oligonucleotides listed in

Table A1.3B. pET16b-His_{10} was used for protein expression. *Pant* genes were individually amplified from *Pantoea sp. GM01*. To insert a gene in pET16b, the plasmid was digested at NdeI and BamHI. For assembly of pET28a.pTet-PantBGC, the BGC was cloned in 2 separate fragments from *Pantoea sp. GM01* and the backbone plasmid was digested at NdeI and XhoI. pATLIC was digested at XmaI for mCherry insertion. Following plasmid construction, all cloned inserts were sequenced at Azenta Life Sciences (San Francisco, CA).

Transformation of chemically competent *E. coli*. A 100 µL aliquot of chemically competent *E. coli* is transformed with 1 µL of pure plasmid or full Gibson mixture, 20 µL of KCM (500 mM KCl, 150 mM $CaCl_2$, 250 mM $MgCl_2$), and 160 µL MilliQ H_2O. The cells are incubated on ice for 30 min, followed by heat shock at 42 °C for 90 s. The cells are plated onto agar immediately (Cb) or after 1 h recovery in LB at 37 °C (Km).

Protein expression procedure. Single colonies of *E. coli* BL21 Star (DE3) transformed with plasmid were inoculated into TB (25 mL) containing the appropriate antibiotics and grown overnight at 37°C with shaking at 200 rpm. Overnight cultures were inoculated into 1 L of TB with 50 µg/mL antibiotic in a 2.8 L-baffled shake flask to OD_{600} = 0.05 and grown at 37 °C and 200 rpm until the OD_{600} reached 0.6 - 0.8. Cultures were cooled on ice for 20 min before inducing with IPTG (0.25 mM) and then cultured for 18 h at 16°C and 200 rpm. Cells were pelleted and stored at -80 °C for storage until purification.

Protein Purification via Ni-NTA Chromatography. Adapted from [*32*]. Frozen cell pellets were thawed and resuspended in 45 ml of lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM imidazole, 20 mM βME and 20% (*v/v*) glycerol, pH 7.5) supplemented with EDTA-free protease inhibitor cocktail (Roche), lysozyme (0.3 mg/mL), and benzonase (5 U/mL). The cell paste was homogenized and then lysed by sonication with a Qsonica Sonicator 3000 (power= 50, 5 s on, 25 s off, 2min total process time, 0.5-inch tip). The lysate was then centrifuged at 13,000 × *g* for 10 min at 4 °C to separate the soluble and insoluble fractions. The soluble lysate was incubated with Ni-NTA (0.5ml of resin per gram of cell paste) for 45 min at 4 °C, resuspended and loaded onto a column by gravity flow. The column was washed with wash buffer (50 mM HEPES, 300 mM NaCl, 20 mM imidazole, 20 mM βME and 20% (*v/v*) glycerol, pH 7.5) for 15–20 column volumes. The column was then eluted with elution buffer (50 mM HEPES, 300 mM NaCl, 300 mM imidazole, 20 mM βME and 20% (*v/v*) glycerol, pH 7.5). Fractions containing the target protein were pooled according to absorbance at 280 nm and concentrated using an Amicon Ultra spin concentrator (10 or 30-kD MWCO; Millipore). Protein was then exchanged into storage buffer (50 mM HEPES, 100 mM sodium chloride, 10% (*v/v*) glycerol and 1 mM DTT, pH 7.5) using PD-10 desalting columns. Final protein concentrations before storage were estimated using the $\varepsilon_{280\ nm}$ calculated by ExPASy ProtParam [*31*] and measured using a Nanodrop spectrometer. They are as follows: PhalA: 1.84 mg/mL ($\varepsilon_{280\ nm}$ = 55,920 $M^{-1}cm^{-1}$), PhalD: 1.76 mg/mL ($\varepsilon_{280\ nm}$ = 44,810 $M^{-1}cm^{-1}$), PhalF: 13.9 mg/mL ($\varepsilon_{280\ nm}$ = 33,350 $M^{-1}cm^{-1}$), PantB: 3.21 mg/mL ($\varepsilon_{280\ nm}$ = 35,410 $M^{-1}cm^{-1}$), PantC: 3.43 mg/mL ($\varepsilon_{280\ nm}$ = 50,880 $M^{-1}cm^{-1}$), PantD: 1.45 mg/mL ($\varepsilon_{280\ nm}$ = 54,430 $M^{-1}cm^{-1}$), PantE: 11.2 mg/mL ($\varepsilon_{280\ nm}$ = 35,410 $M^{-1}cm^{-1}$). All proteins were aliquoted, flash frozen in liquid nitrogen and stored at −80 °C.

General procedure for high-resolution HPLC/MS analysis of polar metabolites with hydrophilic interaction chromatography (HILIC). Adapted from [*32*]. Samples containing polar metabolites were analyzed using an Agilent 1290 UPLC on a SeQuant ZIC-pHILIC (5 µm, 2.1×100 mm; EMD-Millipore) using the following buffers: buffer A (90% water, 10% acetonitrile and 10 mM ammonium formate) and buffer B (90% acetonitrile, 10% water and 10 mM

ammonium formate). After 2 min at 100% buffer B, a linear gradient from 0% to 40% buffer A was applied over 15 min followed by a linear gradient from 40% to 60% buffer A over 5 min and a linear gradient from 60% to 70% buffer A over 2 min at a flow rate of 0.2 mL/min. Mass spectra were acquired in positive-ionization or negative-ionization mode using an Agilent 6530 QTOF (Agilent) with the following source and acquisition parameters: Gas temperature = 300 ºC; drying gas = 8 L/min; nebulizer = 35 psig; sheath gas temperature = 275 °C; capillary voltage = 3500 V; fragmentor 175 V; skimmer 65 V; oct 1 RF vpp = 750 V; acquisition rate = 1 spectra/s; acquisition time = 1000 ms/spectrum; min range = 50 *m/z*; max range = 500 *m/z*.

General procedure for high-resolution HPLC/MS analysis of nonpolar metabolites with reverse-phase chromatography. Samples containing nonpolar metabolites were analyzed using an Agilent 1290 UPLC on a Poroshell 120 EC-C18 (5 μm, 2.1 × 30 mm; Agilent) using the following buffers: Buffer A (100% water, 0.1% *v/v* formic acid) and Buffer B (100% acetonitrile (LC/MS grade)). After 1 min at 100% buffer A, a linear gradient from 0% to 100% Buffer B was applied over 5 min was followed by 30 s at 100% Buffer B at a flow rate of 0.6 mL/min. Mass spectra were acquired on either positive or negative ionization modes using an Agilent 6530 QTOF (Agilent) with the following source and acquisition parameters: Gas temperature = 300 ºC; drying gas = 8 L/min; nebulizer = 35 psig; sheath gas temperature = 275 °C; capillary voltage = 3500 V; fragmentor 175 V; skimmer 65 V; oct 1 RF vpp = 750 V; acquisition rate = 1 spectra/s; acquisition time = 1000 ms/spectrum; min range = 50 *m/z*; max range = 1700 *m/z*.

General procedure for rapid HPLC/MS analysis of polar metabolites with hydrophilic interaction chromatography with ZIC-HILIC. Samples containing polar metabolites were analyzed using an Agilent 1290 UPLC on a SeQuant ZIC-HILIC (5 μm, 2.1 × 50 mm; Sigma-Aldrich) using the following buffers: buffer A (90% water, 10% acetonitrile and 10 mM ammonium formate) and buffer B (90% acetonitrile, 10% water and 10 mM ammonium formate). After 0.2 min at 100% buffer B, a linear gradient from 0% to 33% buffer A was applied over 1.80 min followed by a linear gradient from 33% to 60% buffer A over 1 min followed by a 1 min at 60% A at a flow rate of 0.6 mL/min. Mass spectra were acquired in positive-ionization mode using an Agilent 6530 QTOF (Agilent) with the following source and acquisition parameters: Gas temperature = 300 ºC; drying gas = 8 L/min; nebulizer = 35 psig; sheath gas temperature = 275 °C; capillary voltage = 3500 V; fragmentor 175 V; skimmer 65 V; oct 1 RF vpp = 750 V; acquisition rate = 1 spectra/s; acquisition time = 1000 ms/spectrum; min range = 100 *m/z*; max range = 1700 *m/z*.

General procedure for rapid HPLC/MS analysis of polar metabolites with hydrophilic interaction chromatography with Express HILIC. Samples containing polar metabolites were analyzed using an Agilent 1290 UPLC and an Ascentis Express HILIC (2.7 μm, 2.1 × 75 mm; Ascentis) using the following buffers: buffer A (90% water, 10% acetonitrile and 10 mM ammonium formate) and buffer B (90% acetonitrile, 10% water and 10 mM ammonium formate). After 1 min at 100% buffer B, a linear gradient from 0% to 75% buffer A was applied over 6 min followed by 1.5 min at 75% A at a flow rate of 0.4 mL/min. Mass spectra were acquired in positive-ionization mode using an Agilent 6530 QTOF (Agilent) with the following source and acquisition parameters: Gas temperature = 300 ºC; drying gas = 8 L/min; nebulizer = 35 psig; sheath gas temperature = 350 °C; capillary voltage = 3500 V; fragmentor 175 V; skimmer 65 V; oct 1 RF vpp = 750 V; acquisition rate = 1 spectra/s; acquisition time = 1000 ms/spectrum; min range = 100 *m/z*; max range = 1700 *m/z*.

Bioinformatic method for identifying Fe/αKG dioxygenases within the AlaX Family. For generating the sequence similarity network (SSN) of the AlaX family of tRNA-Deacylases, we utilized the EFI-EST tool for generating a SSN of IPR018163 (tRNA-editing domain) as a Uniref50 dataset containing 128,218 proteins (Sept. 2022) (including fragments) [*10*]. Uniref50 is a representative dataset where each node in the SSN represents proteins that are ≥50% sequence-similar to a representative protein. After generating the initial dataset, we set the sequence alignment score to an arbitrary value of 35%, which was equivalent to a sequence similarity of 30-40%. A maximum sequence length was specified as 480 amino acids to exclude any Alanine-tRNA Synthetases that have tRNA-editing domains, resulting in 23,260 proteins. The SSN was visualized with Cytoscape, and the characterized AlaXs were mapped onto the SSN [*11*]. We found that the majority of these proteins were in the same cluster, so the sequence alignment score was increased until it reached 43%, which was the highest value before most characterized AlaXs split from each other. This SSN was submitted to the EFI-GNT tool to generate neighborhood data files. The neighborhood distance was set to 10 genes, and the co-occurrence cut-off to 0% to maximize the amount of genomic information available. After completion, we downloaded the Color SSN, SSN Cluster GNN Hub-Node file, and the Genomic Neighborhood Diagrams (GNDs) file.

We used the Color SSN for the next round of analysis. This SSN contained 23,260 proteins represented as 3,676 nodes. To clean the data for analysis, we removed entries with poor neighborhood information resulting from short reads by only keeping nodes with >6 neighbors. Additionally, we removed nodes where every represented sequence is a fragment. Empirically, these sequences often had reads starting in the middle of the editing domain AlaX, and likely belong to a full AlaRSs. This resulted in 1,879 nodes (51% of original) and 20,123 AlaX proteins (87% of original).

For the next stage of the analysis, we uploaded the SSN Cluster GNN Hub-Node file to Cytoscape and downloaded the table as a CSV. This is the input file for the tR3D data visualization tool. Using tR3D, we identified the most common protein families to co-occur with the deacylases for each cluster of the SSN. We also used the EFI-GND tool for viewing neighborhood data directly. From this analysis, we generated a list of protein families that participate in primary metabolism, filtered for all AlaX proteins that co-occur with these genes, and removed them from further analysis. This yielded a much smaller SSN with 844 nodes (23% of original) and 2,675 genomes (11.5% of original) that likely contain mostly secondary metabolism-related BGCs. This SSN was resubmitted to the EFI-GNT tool using the same specifications as above to generate neighborhood data files. The neighborhood distance was set to 10 genes, and the co-occurrence cut-off to 0% to maximize the amount of genomic information available. After completion, we downloaded the Color SSN and visualized this with Cytoscape. We used the filtering panel to search for all Fe/αKG protein families as neighboring PFams (PF02668, PF13640, PF13532, PF05721, or PF10014), and this highlighted 50 nodes. We visualized these entries with the EFI-GNT tool and identified 16 highly-conserved BGCs.

Bioinformatic method for identifying secondary metabolism within the YbaK/INS Family. For generating the sequence similarity network (SSN) of the YbaK/INS family of tRNA-Deacylases, we utilized the EFI-EST tool for generating a SSN of PF04073 (tRNA-editing domain) as a Uniref50 dataset containing 88,098 proteins (Dec. 2022) (including fragments) [*10*]. Uniref50 is a representative dataset where each node in the SSN represents proteins that are ≥50% sequence-similar to a representative protein. After generating the initial dataset, we set the sequence alignment score to a value of 39%, which was equivalent to a sequence similarity of 30-40%. A maximum sequence length was specified as 300 amino acids to exclude any Prolyl-tRNA

Synthetases that have tRNA-editing domains, resulting in 53,310 proteins. The SSN was visualized with Cytoscape, and the characterized YbaK/INS were mapped onto the SSN [*11*]. This SSN was submitted to the EFI-GNT tool to generate neighborhood data files. The neighborhood distance was set to 10 genes, and the co-occurrence cut-off to 0% to maximize the amount of genomic information available. After completion, we downloaded the Color SSN, SSN Cluster GNN Hub-Node file, and the Genomic Neighborhood Diagrams (GNDs) file.

We used the Color SSN for the next round of analysis. This SSN contained 53,310 proteins represented as 3,184 nodes. To clean the data for analysis, we removed entries with poor neighborhood information resulting from short reads by only keeping nodes with >6 neighbors. Additionally, we removed nodes where every represented sequence is a fragment. Empirically, these sequences often had reads starting in the middle of the editing domain, and likely belong to a full ProRSs. This resulted in 2,527 nodes (79.4% of original) and 52,516 AlaX proteins (98.5% of original).

For the next stage of the analysis, we uploaded the SSN Cluster GNN Hub-Node file to Cytoscape and downloaded the table as a CSV. This is the input file for the tR3D data visualization tool. Using tR3D, we identified the most common protein families to co-occur with the deacylases for each cluster of the SSN. We also used the EFI-GND tool for viewing neighborhood data directly. From this analysis, we generated a list of protein families that participate in primary metabolism, filtered for all YbaK/INS proteins that co-occur with these genes, and removed them from further analysis. This yielded a smaller SSN with 1,422 nodes (44.7% of original) and 10,395 genomes (19.5% of original) that could contain predominantly secondary metabolism-related BGCs. This SSN was resubmitted to the EFI-GNT tool using the same specifications as above to generate neighborhood data files. The neighborhood distance was set to 10 genes, and the co-occurrence cut-off to 0% to maximize the amount of genomic information available. After completion, we downloaded the Color SSN and visualized this with Cytoscape. We used the filtering panel to search for class-defined secondary metabolism as well as other conserved BGCs with the tR3D tool.

Bioinformatic analysis of the flavin *N*-monooxygenase family IPR038732. The SSN of the IPR038732 family of *N*-monooxygenases containing PhalD was generated with the EFI-EST tool [*10*] using a sequence-similarity cutoff of 75%. This was visualized using Cytoscape, and known members were highlighted (CreE: A0A0J2KL70, HpyO: Q8PDQ6, PhalD: A0A0F5VE59).

***In vitro* screening of enzymatic amino acid modifications.** Reactions (25 µl) contained one or a subset of the 20 canonical amino acids (L-alanine, L-arginine, L-asparagine, L-aspartate, L-cysteine, L-glutamine, L-glutamate, L-glycine, L-histidine, L-isoleucine, L-leucine, L-lysine, L-methionine, L-phenylalanine, L-proline, L-serine, L-threonine, L-tryptophan, L-tyrosine, L-valine) (0.5 mM each), sodium chloride (100 mM) in 50 mM HEPES buffer (pH 7.5) along with necessary cofactors. For PhalA and PantE, these cofactors were sodium αKG (5 mM), sodium ascorbate (5 mM) and $(NH_4)_2Fe(SO_4)_2 \cdot 6$ H2O (1 mM). For PhalD, FAD (200 µM) and NADPH (500 µM) were added. For PhalF and PantC, ATP (5 mM) and $MgCl_2$ (15 mM) were added. Reactions were initiated by addition of enzyme(s) (25 µM final concentration) and allowed to proceed at room temperature before quenching at 18 h in 1 volume of acetonitrile. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C to pellet protein. 20 µL of supernatant was derivatized with 5 µL Fmoc-Cl (10 mM stock solution in DCM) and 6.25 µL Borate buffer (200 mM sodium tetraborate decahydrate, pH 8.2) for 15 min at room temperature to reach a final pH of 8 before quenching in 1 volume of methanol. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C and the

supernatant was analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for nonpolar metabolite analysis.

For reactions with only PhalD, resolution of peaks was improved by quenching samples in methanol with 1% (*v/v*) formic acid, along with control reactions lacking enzymes. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C to pellet protein, and the supernatant was analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for rapid polar metabolite analysis with ZIC-HILIC.

For reactions with PhalD and PhalF, resolution of peaks was somewhat improved by quenching samples in methanol with 1% (*v/v*) formic acid, along with control reactions lacking enzymes. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C to pellet protein, and the supernatant was analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for rapid polar metabolite analysis with Express HILIC.

M9 minimal media for metabolomics and BesB engineering. M9 minimal media prepared according to the University of Wisconsin-Madison's *E. coli* Genome Project [*12*] was utilized for metabolomics experiments, containing Na_2HPO_4 (47.7 mM), KH_2PO_4 (22.0 mM), NaCl (8.55 mM), NH_4Cl (18.7 mM), glucose (0.4% *w/v*), $MgSO_4$ (2 mM), $CaCl_2$ (100 µM), thiamine (1 mg/mL), and an amino acid mixture diluted from a 5X solution (**Table A2.1**). For isotopically-labelled amino acid conditions, an equivalent amount of the labelled version ($^{13}C_6$-L-leucine, $^{13}C_6$,^{15}N-L-isoleucine, $^{13}C_5$,^{15}N-L-valine) was supplied exogenously to the complete medium.

Metabolomics for identifying products of the Phal BGC. Single colonies of *E. coli* BL21 Star (DE3) transformed with pET16b-PhalD and/or pACYC-Duet-PhalF-only and/or pACYC-Duet-PhalF-PhalA were inoculated in triplicate per condition into LB (5 mL) in a 24-well plate containing carbenicillin and/or chloramphenicol when appropriate (50 µg/mL) and grown overnight at 37°C with shaking at 200 rpm. Overnight cultures were inoculated into 5 mL of LB with carbenicillin and/or chloramphenicol when appropriate (50 µg/mL) in a 24-well plate to $OD_{600\ nm}$ = 0.05 and grown at 37 °C and 200 rpm until the $OD_{600\ nm}$ reached 0.4-0.6. The plate was cooled on ice for 20 min before spinning down at 4,000 × *g* for 10 min at 4 °C. Media was removed, and cell pellets were resuspended in M9 minimal media for metabolomics. Additional components were added, including carbenicillin and/or chloramphenicol when appropriate (50 µg/mL), IPTG (0.25 mM), and isotopically-labelled amino acid as appropriate ($^{13}C_6$-L-leucine/$^{13}C_6$,^{15}N-L-isoleucine/$^{13}C_5$,^{15}N-L-valine) (equivalent to concentration in the amino acid mixture). After the plate was incubated for 48 h at 30°C and 200 rpm, 1 mL aliquots were taken from each condition and centrifuged at 10,000 × *g* for 5 min at RT. After decanting the supernatant, pellets were resuspended in 50 µL of methanol with 1% (*v/v*) formic acid, lysed by vortexing for 10 s, and centrifuged at 20,000 × *g* at 4°C for 10 min. Supernatant was analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis in both positive- and negative-ionization modes. Samples were analyzed in triplicate with MS-Dial to identify new metabolites with isotopic labeling in the BGC-expressing strain [*13*].

Metabolomics for identifying products of the Pant BGC. Single colonies of *E. coli* BL21 Star (DE3) transformed with pET28a.pTet-PantBGC were inoculated in triplicate per condition into LB (5 mL) in a 24-well plate containing kanamycin when appropriate (50 µg/mL) and grown overnight at 37°C with shaking at 200 rpm. Overnight cultures were inoculated into 5 mL of LB with kanamycin when appropriate (50 µg/mL) in a 24-well plate to $OD_{600\ nm}$ = 0.05 and grown at 37 °C and 200 rpm until the $OD_{600\ nm}$ reached 0.4-0.6. The plate was cooled on ice for 20 min before spinning down at 4,000 × *g* for 10 min at 4 °C. Media was removed, and cell pellets were

resuspended in M9 minimal media for metabolomics. Additional components were added, including kanamycin when appropriate (50 μg/mL) and anhydrotetracycline (100 ng/mL). After the plate was incubated for 48 h at 30°C and 200 rpm, 1 mL aliquots were taken from each condition and centrifuged at 10,000 × *g* for 5 min at RT. After decanting the supernatant, pellets were resuspended in 50 μL of methanol with 1% (*v/v*) formic acid, lysed by vortexing for 10 s, and centrifuged at 20,000 × *g* at 4°C for 10 min. Supernatant was analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis in both positive- and negative-ionization modes. Samples were analyzed in triplicate with MS-Dial to identify new metabolites with isotopic labeling in the BGC-expressing strain [*13*].

N-monooxygenation of L-valinol. Reactions (25 μL) contained L-valinol (5 mM), FAD (200 μM), NADPH (10 mM), sodium chloride (100 mM) in 50 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of PhalD (15μM final concentration) and allowed to proceed at room temperature before quenching at 18 h in 1 volume of acetonitrile. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C to pellet protein. 20 μL of supernatant was derivatized with 5 μL Fmoc-Cl (10 mM stock solution in DCM) and 6.25 μL Borate buffer (200 mM sodium tetraborate decahydrate, pH 8.2) for 15 min at room temperature to reach a final pH of 8 before quenching in 1 volume of methanol. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C and the supernatant was analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for nonpolar metabolite analysis.

Chemoenzymatic synthesis of diazeniumdiolate-containing valine. Reactions (25 μL) were run in duplicate and contained L-valine (2.5 mM), FAD (200 μM), NADPH (5 mM), sodium chloride (100 mM) in 50mM HEPES buffer (pH 7.5). A 100 mM stock of diethylamine NONOate was prepared in 10 mM NaOH and stored at -80°C until use. Reactions were initiated by addition of PhalD (15 μM final concentration) and allowed to proceed at room temperature for 2 h before diethylamine NONOate was added at a concentration of 5 mM to one set of the duplicate reactions. The reactions were allowed to proceed at room temperature for another 16 h before quenching samples in methanol with 1% (*v/v*) formic acid, along with control reactions lacking enzymes. Samples were centrifuged at 20,000 × *g* for 10 min at 4 °C to pellet protein, and the supernatant was analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for rapid polar metabolite analysis with ZIC-HILIC.

Kinetic analysis of PhalF. Steady-state kinetic analysis was performed by monitoring NADH consumption through a coupled assay [*14*]. Reactions (100 μL) contained ATP (2.5 mM), $MgCl_2$ (5 mM), phosphoenolpyruvate (PEP, 1 mM), NADH (0.3 mM), lactate dehydrogenase (LDH, 10 U/mL), pyruvate kinase (PK, 10 U/mL), and sodium chloride (10 mM) in 100 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of PhalF (10 μM) in the presence of combinations of amino acids including valine, leucine, and isoleucine (10 mM) Initial rates of NADH consumption were measured by monitoring A_{340} using a SpectraMax M2 Microplate Reader (Molecular Devices) at room temperature. Steady-state kinetic analysis was performed by monitoring NADH consumption through a coupled assay [*14*].

Computational modeling of proteins. Protein models were acquired from the Uniprot [*15*] database using the accessions for each protein, which are listed in **Table A2.4**. The protein models were generated with AlphaFold2. The PDB files were downloaded directly from Uniprot and were visualized with PyMol.

Engineering of chAlaRS. Site-saturation libraries were assembled via Gibson [*8*] from PCR reactions using partially overlapping primers that replicated the entire plasmid as a linear strand. The plasmid for chAlaRS engineering was pCDF-chAlaRS-proKtRNA^{Ala}. The library was mixed with pATLIC-UAG-mCherry-containing electrocompetent *E. coli* DH10B cells by vortexing, and after cooling on ice for 10 min, was electroporated and immediately transferred to 1 mL of 2X YT media. After 1 h of shaking at 37°C at 200 rpm, the cells were transferred into 5 mL LB media supplemented with carbenicillin (Cb) and streptomycin (Sp) (50 μg/mL). The subsequent day, the library was diluted 1:50 into 5 mL fresh LB media. When the $OD_{600\,nm}$ reached 0.5, the cells were induced with IPTG (0.25 mM), arabinose (0.2% *v/v*), and Pra were supplied exogenously (5 mM). Following an overnight growth at 37°C, samples (1 mL) were collected, and centrifuged at 10,000 × *g* at room temperature for 2 min. The supernatant was decanted and the cell pellets washed once with PBS before resuspending in 1mL 1X PBS. The Cu-AAC reaction was performed with 90 μL of cells and 10 μL of master mix containing $CuSO_4$ (0.5 mM), THPTA (1.5 mM), Calfluor 488 Azide (10 μM), and sodium ascorbate (5 mM). The reaction was allowed to proceed at 4°C in the dark for 1.5 h at 4°C. The reaction (100 μL) was then centrifuged at 10,000 × *g* at 4 ºC for 2 min, decanted, and the pellet washed once with PBS (100 μL). Following resuspension in PBS (100 μL), cell suspension (10 μL) was diluted with PBS (1 mL) for FACS-based analysis and sorting, which was performed at the Flow Cytometry Core Facility (UC Berkeley, Weill Hall).

Examining chAlaRS promiscuity towards small ncAAs. dsDNA oligos for *in vitro* tRNA transcription containing a T7 promoter and corresponding tRNA were synthesized by IDT. PCR using Phusion polymerase with forward and reverse primers (primer name reads acceptor stem 5′ to 3′) was used to amplify DNA from the synthetic oligos to obtain high-concentration (150-200 ng/μL) DNA used for *in vitro* transcription. The primers for tRNA^{Ala} contained a methylguanosine as the penultimate base on the 3′ end to minimize RNA polymerase run-off. The extended primer length ($\geq$ 60 nts) eliminated the need for a dsDNA template [*16*]. The HiScribe T7 High Yield RNA Synthesis Kit (NEB) was used for *in vitro* transcription, with the recommended conditions for short RNAs. Transcription reactions contained 1 μg of DNA template in 20 μL total reaction volume and were allowed to proceed overnight at 37 °C. tRNA was purified from these reactions directly using miRNeasy purification kit (QIAGEN) following manufacturer protocols. For aminoacylation reactions with L-alanine (Ala), L-propargylglycine (Pra), L-allylglycine (Alg), or L-4-Cl-allylglycine (Cl-Alg), the reaction contained chAlaRS (15 μM), ATP (5 mM), $MgCl_2$ (10 mM), amino acid (4 mM), HEPES (40 mM, pH 7.5) and tRNA (4 μL) with nuclease-free water added to a final volume of 15 μL. Prior to beginning the reaction, tRNA was first refolded according to a literature protocol [*17*]. Briefly, HEPES buffer and water were added to the transcription reaction, the sample was heated to 98°C on a thermocycler for 2 min, and slowly cooled to room temperature over 5 min. $MgCl_2$ was added, and the sample was incubated at room temperature for 5 min, followed by 3 min on ice. The rest of the components were added and acylation was initiated by addition of enzyme. The reaction was incubated for 4 h at 37 °C. In an approach adapted from [*29*], samples (10 μL) were taken at 4 h intervals and added to 20 μL of RNaseA (1 U/μL) in 200 mM ammonium acetate. Protein was removed by precipitation with 5% TCA (*v/v*) and centrifugation at 14,000 × *g* for 10 min. 5 μL of each sample was analyzed by LC/MS-MS on an Agilent 1290 UPLC-6530 QQQ using the protocol for aminoacyl-tRNA analysis with hydrophobic chromatography.

General procedure for high-resolution HPLC/MS-MS analysis of aminoacyl-tRNA with reverse-phase chromatography. In an approach adapted from [*29*], adenylate samples were

analyzed using an Agilent 1290 UPLC on a Poroshell 120 EC-C18 (5 μm, 2.1 × 30 mm; Agilent) using the following buffers: Buffer A (100% water, 10 mM ammonium formate) and Buffer B (100% acetonitrile (LC/MS grade)). after 50 s at 95% buffer A, a linear gradient from 5% to 30% Buffer B was applied over 40 s followed by a linear gradient to 100% Buffer B over 90 s at a flow rate of 0.7 mL/min. Mass spectra were acquired on either positive ionization modes using an Agilent 6460 QQQ (Agilent) with the following source and acquisition parameters: Gas temperature = 250 ºC; drying gas = 5 L/min; nebulizer = 45 psig; sheath gas temperature = 250 °C; capillary voltage = 3500 V; fragmentor = 60 V; collision energy = 0 V; cell accelerator voltage = 7. Ions were monitored by MRM using the following transitions: Ala-adenylate (*m/z* = 339.1→136.1), Pra-adenylate (*m/z* = 363.4→136.1), Alg-adenylate (*m/z* = 365.4→136.1), 4-Cl-Alg-adenylate (*m/z* = 399.8→136.1)

Engineering of BesB. Error-prone libraries were assembled via Gibson [*8*] from PCR reactions of the entire BesB gene replicated with an error-prone polymerase optimized to install 0-4.5 mutations/kb using the GeneMorph II kit (Agilent), transformed into chemically-competent BL21, plated onto LB agar with carbenicillin (Cb) (50 μg/mL), and incubated overnight at 37°C. Individual colonies were inoculated into individual wells of a 96-well plate containing 200 μL of LB with the appropriate antibiotic and grown overnight. Dense cultures were diluted 20-fold into 96-well plates (2-mL, v-bottom) containing 1 mL of LB, covered with Aeraseal film (Sigma) and grown to mid-log phase ($OD_{600\ nm}$ = 0.6–0.8) before chilling the plates on ice for 30 min. Cells were pelleted by centrifugation at 4,000 × *g* for 20 min and then resuspended in 1 mL of M9 medium supplemented with 2 mM Cl-allylglycine, 0.25 mM IPTG and the appropriate antibiotics. Cells were grown for 2 d at 16 °C with shaking at 225 rpm. Note that at the end of the 2 d, some of the cells had settled in the bottom of the plate despite shaking during growth. The remaining cells were then pelleted by centrifugation. The supernatant (50 μL) was transferred to a fresh 96-well black plate with a clear bottom (Corning). Fifty microliters of freshly prepared click solution (CalFluor 488 azide (1 μM), $CuSO_4$ (0.5 mM), THPTA (0.1 mM), sodium ascorbate (5 mM)) was then added to the cell supernatant and incubated for 15 min in the dark. The fluorescence of the samples was measured using a SynergyMx microplate reader (BioTek) at room temperature (λ_{ex} = 485 nm, λ_{em} = 528 nm).

4.3 Results and Discussion

Genome mining for C-H functionalizing enzymes. In an effort to discover more natural products and useful enzymatic reactions, we continued the strategy began in Chapter 3 of searching for tRDs neighboring oxidative enzymes, this time including more families of the non-heme iron/α-ketoglutarate-dependent (Fe/αKG) dioxygenases beyond the TauD members from Chapter 3. Beginning with the AlaX SSN that was cleaned of primary metabolism in Chapter 2, we searched for genomic reads that contain at least one of five well-studied families of Fe/αKG-dependent dioxygenases within 10 genes upstream or downstream of an AlaX, which revealed a total of 50 entries at the Uniref50 level (**Fig. 4.1**). Uniref50 sequences are representative of a group of similar proteins with sequence identity $\geq$ 50%, where these 50 entries represent 331 total sequences. Searching for features of what we would expect for secondary metabolism, such as neighboring transporters, regulators, or genes like what were tabulated in Chapter 2, we identified 16 unique types of biosynthetic gene clusters (BGCs) (**Fig. 4.2, Table 4.1**). A majority of these BGCs showed the Fe/αKG enzyme in close proximity to the tRD, so we postulated there could be some relationship in their function, such as was seen in the methionine hydroxylase co-clustered with the selective deacylase from Chapter 2. Several of these BGCs have been characterized in previous chapters, but of the ones that were unknown, a conserved gene cassette piqued our interest for containing a Fe/αKG dioxygenase belonging to the 2OG family (PF13640) adjacent to an AlaX in a variety of genomic contexts (**Bold**, **Table 4.1**). We hypothesized exploring one of these BGCs could provide insights into the other BGCs that share this cassette.

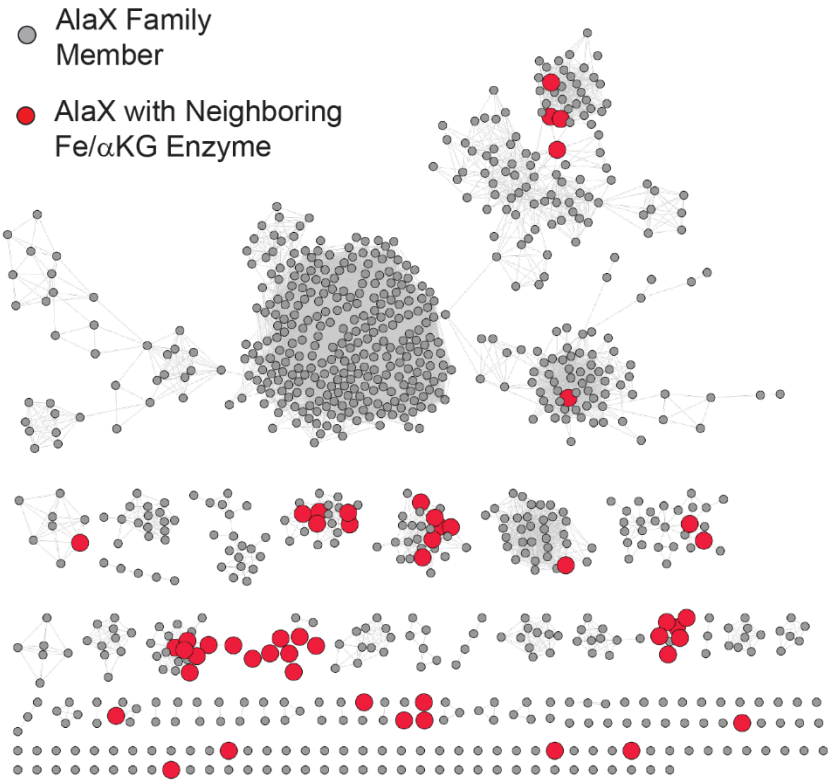


Figure 4.1. Sequence Similarity Network (SSN) of AlaX tRDs in secondary metabolism with Fe/αKG oxygenases. SSN of the AlaX tRNA deacylases cleaned of primary metabolism and searched for neighboring Fe/αKG enzymes, resulting in 50 highlighted nodes representing 331 sequences (red).



Figure 4.2. Biosynthetic gene clusters (BGCs) containing both Fe/αKG-dependent oxygenases and AlaX tRNA-Deacylases. 16 representative biosynthetic gene clusters were identified to contain both an AlaX and a Fe/αKG-dependent dioxygenase with some similarities and differences in their genomic organization. BGCs from SSN clusters 24, 32, and 132 appeared similar in their gene ordering, with the 2OG (green) hypothesized to perform a similar function across the different genomic contexts.

Protein Family	Representative tRD (Gene ID)	Function
TauD (PF02668)	A0A1H3JVZ4	Unknown
	A0A4D8PSJ1	Unknown
	A0A5R8Z9P6	Arg C_3 hydroxylase
	A0A856MBM9	Unknown
	A0A7U9CRD7	β-Phe C_2 hydroxylase
	A0A0P6VYU1	Unknown
	A0A1I5XWL6	Unknown
2OG (PF13640)	A0A1D8ITP8	Met C_3 hydroxylase
	A0A2T1FSD1	Unknown
	D0I6N6	Unknown
	A0A1Z2SF43	Unknown
	A0A7H0HZI1	Unknown
2OG (PF13532)	A0A0F7HGQ0	Unknown
	A0A241W1K3	Unknown
2OG (PF10014)	A0A0P7JQA8	Unknown
PhyH (PF05721)	A0A0F9YCM3	Thr halogenase

Table 4.1. Representative Accessions for AlaX tRNA-Deacylases with neighboring Fe/αKG-dependent oxygenases. From the SSN of the AlaX tRNA Deacylases, 16 unique types of BGCs were identified that contained both an AlaX and a Fe/αKG dioxygenase within 10 genes upstream/downstream, along with function if known. **Bold** indicates a group of similar BGCs that were selected for experimental exploration in this work.

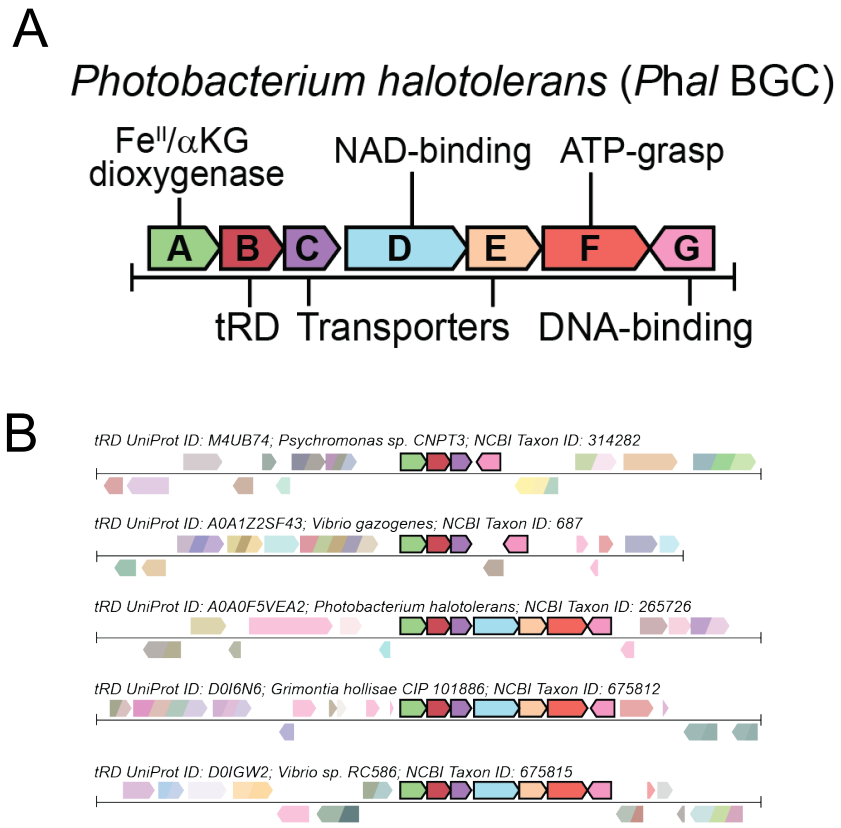


Figure 4.3. Identification of an oxygenase-containing biosynthetic gene cluster in *Photobacterium halotolerans*. (A) A BGC was identified that contains three putative biosynthetic enzymes, including a Fe/αKG-dependent dioxygenase PhalA, a NAD-binding protein PhalD, and an ATP-grasp protein PhalF, in addition to a deacylase, transporters, and a DNA-binding protein. (B) Three organisms contain the full biosynthetic gene cluser while two others contain a fragment with only the oxygenase and deacylase within this clade (Cluster 32 of the AlaX SSN).

Identification of the *Phal* BGC containing a poorly annotated enzyme. With the identification of this conserved gene cassette, we selected a BGC that contained this cassette along with additional interesting enzymes for experimental characterization. The *Phal* BGC is shared across 5 organisms with one of two levels of gene conservation (**Fig. 4.3**). Two of the five organisms contain only the gene cassette, including a Fe/αKG-dependent dioxygenase, an AlaX tRNA-deacylase, a transporter, and a DNA-binding domain, leading us to hypothesize the dioxygenase forms the ncAA that gives rise to the presence of the AlaX editing protein within the cluster. In the other three organisms that share this core cassette, an additional 3 genes are present, including a NAD-binding protein, an ATP-grasp protein, and another transporter. While the function of the NAD-binding protein was unclear from annotation alone, the ATP-grasp protein suggested a dipeptide natural product, as seen in Chapters 2 and 3. We chose the complete BGC from *Photobacterium halotolerans* (*Phal* BGC) for experimental elucidation.

We began our efforts towards biosynthetic pathway discovery with *in vitro* reactions to determine individual enzyme activity. Following heterologous expression and purification from *E. coli* with N-terminal His_{10} tags, we performed reactions with individual proteinogenic amino acids and necessary cofactors (**Fig. A3.6**). With purified PhalA, we detected hydroxylation activity towards leucine and isoleucine with a *m/z* of $[M+H+Fmoc]^+ = 370.1649$ (**Fig. 4.4**). Of the two substrates, we found leucine was the preferred substrate (>99% conversion). To determine regio- and stereoselectivity, the product would need to be isolated for structural characterization. Preliminarily, we hypothesize the site of hydroxylation to be the β-position based on the lower reactivity of PhalA towards isoleucine (<1% conversion) where there is more steric bulk at the β-position. In addition, when isoleucine is supplied as a substrate, we detected the mass corresponding to aziridination or desaturation (**Fig 4.5**). This is a previously described reaction outcome of Fe/αKG-dioxygenases with this particular substrate, which highlights how reaction outcomes can be altered by substrate properties [*18*]. Owing to the much higher conversion of Leu over Ile, we hypothesized that Leu is incorporated into this pathway's natural product.

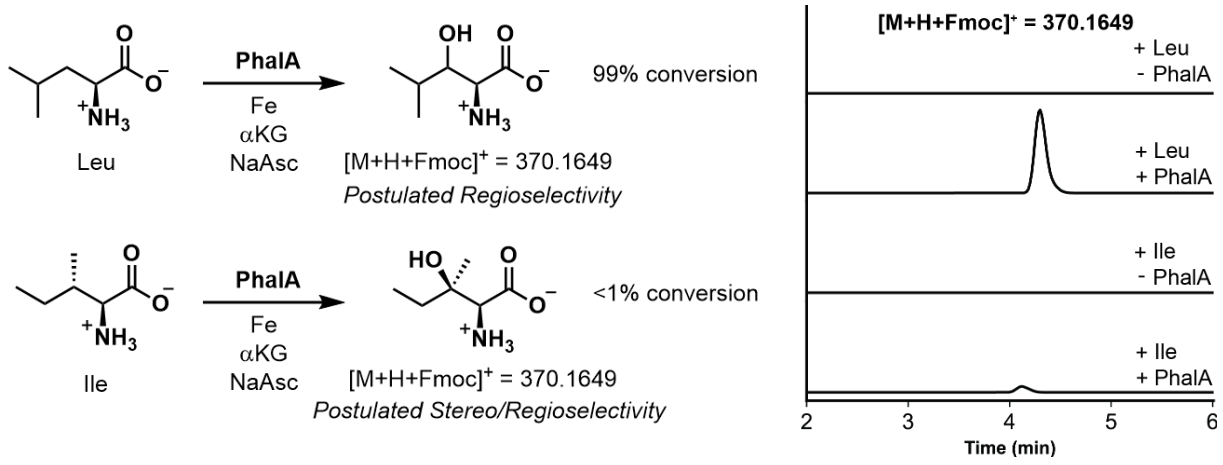


Figure 4.4. Identification of PhalA hydroxylation of leucine (Leu) and isoleucine (Ile). PhalA shows hydroxylation activity towards leucine (Leu) and isoleucine (Ile) in endpoint *in vitro* reactions with individual amino acids, with much higher activity towards leucine over isoleucine. Regioselectivity is tentatively assigned at the β-position.

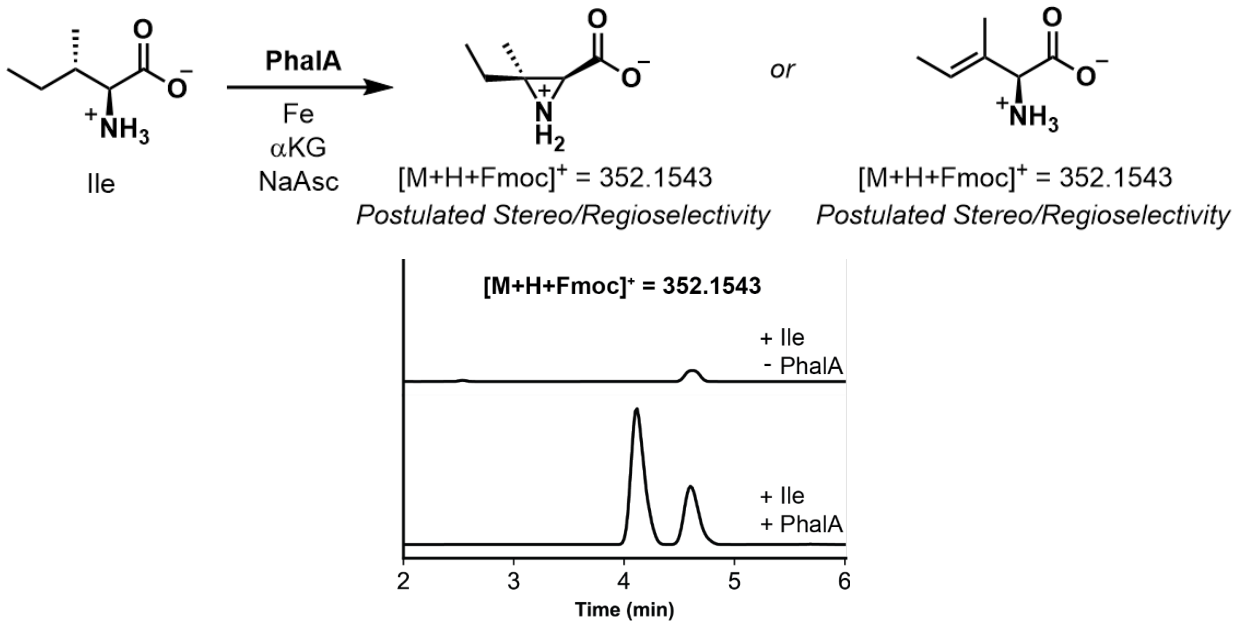


Figure 4.5. Possible aziridination of Ile by PhalA. In addition to hydroxylation activity, aziridination or desaturation activity is detected by PhalA towards Ile in this endpoint *in vitro* reaction. This activity is quite low (<1% conversion) and further structural characterization would be necessary to definitively determine the product outcome of this reaction.

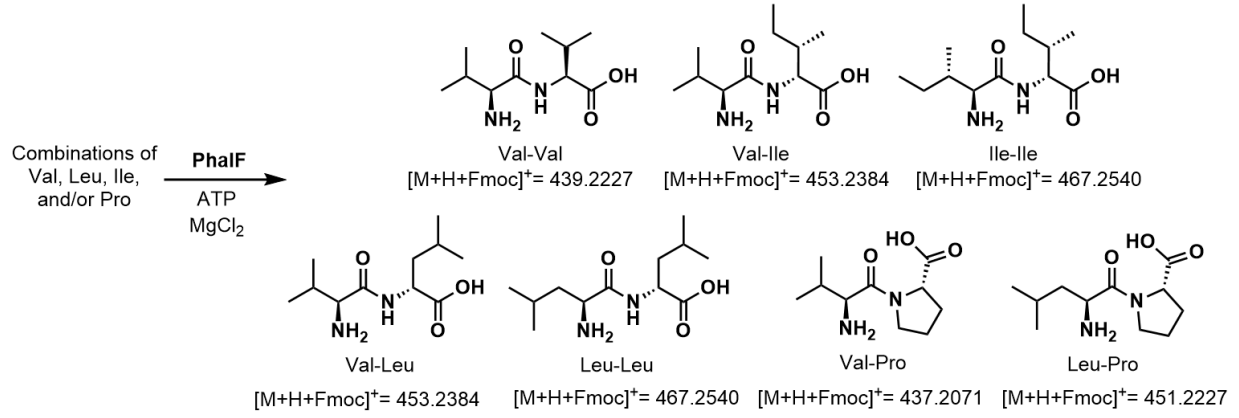


Figure 4.6. Small dipeptides produced by PhalF. Products detected by LC/MS of PhalF with various combinations of four proteinogenic amino acids. Connectivity shown is proposed, requiring confirmation with MS/MS fragmentation.

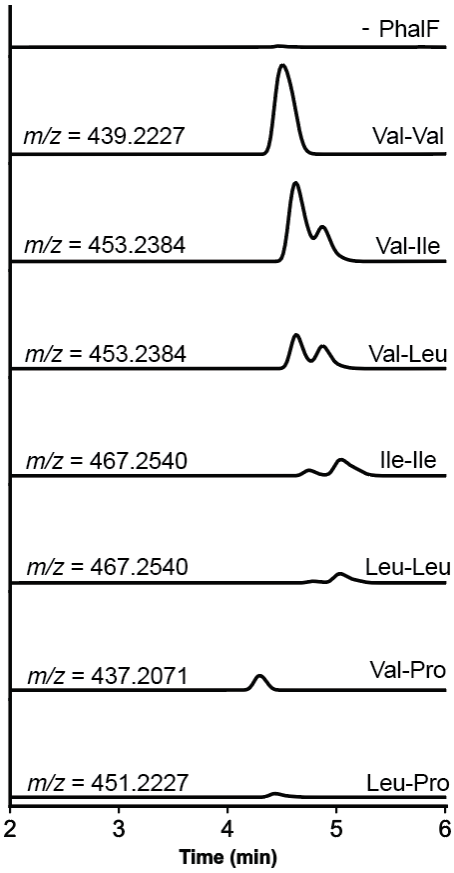


Figure 4.7. Relative abundance of PhalF dipeptide products. From an endpoint assay, relative intensities of the dipeptide products from reactions with PhalF as detected by LC/MS give an approximate indication of substrate preference by PhalF. Scale is relative to the most abundant product shown (Val-Val). Peak doubling is likely due to either the amine or amide being protected by Fmoc. For this figure, *m/z* is equivalent to $[M+H]^+$.

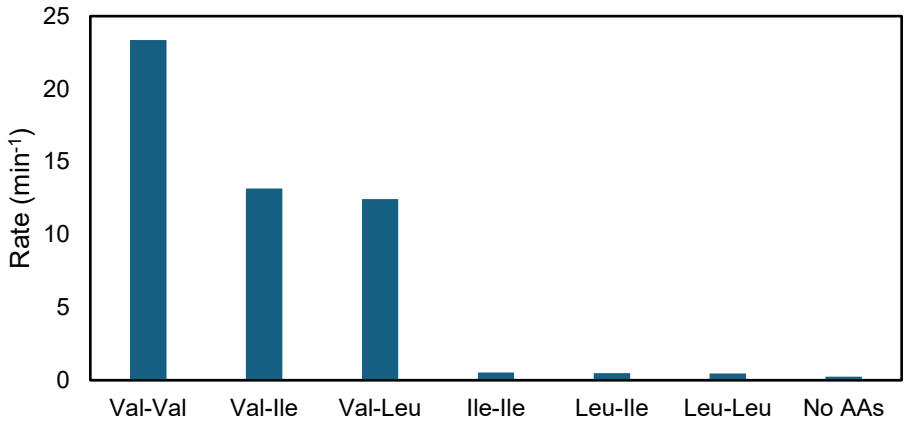


Figure 4.8. Continuous assay of PhalF shows preference for valine. The rate of dipeptide formation was measured for various combinations of two small aliphatic amino acids, with Val-Val-forming activity being the highest. With the rate going down by half when the amount of valine was reduced by half, activity towards Leu and Ile is likely lower than background Val-Val dipeptide formation. Rate of reaction was measured by change in A_{340} using an NADH-coupled assay with pyruvate kinase and lactate dehydrogenase [*19*].

To gain more insight into the natural product produced by the *Phal* BGC, we performed a similar screen with various combinations of the proteinogenic amino acids, purified PhalF, and necessary cofactors. We detected several new masses, indicating promiscuity (**Fig. 4.6, 4.7**). The most abundant product was a Val-Val dipeptide, followed by a Val-Ile/Leu, Ile/Leu-Ile/Leu, and Val/Leu-Pro. From this, we concluded PhalF can accept small nonpolar amino acids and taken together with PhalA's specificity, we expected the *Phal* BGC product to include both leucine and valine. We examined PhalF's substrate preference using a steady state kinetics assay and found valine to be the preferred substrate to such a degree that testing combinations with other amino acids was obscured by likely background formation of Val-Val dipeptides (**Fig. 4.8**). The steady state findings are in accordance with the *in vitro* endpoint studies, where Val-Val dipeptide formation is the predominant product.

Metabolomics of the *Phal* BGC. From our *in vitro* findings, we reasoned this cluster likely produces a hydroxylated dipeptide of aliphatic amino acids. To determine the natural product of this BGC, we performed a comparative metabolomics experiment through heterologous expression of the *Phal* BGC in *E. coli* (**Fig. 4.9**). We examined various combinations of the biosynthetic enzymes encoded across several unique plasmids to determine what products required all three enzymes. Guided by our *in vitro* results, we supplied one of three isotopically-labelled aliphatic amino acids to determine if they are incorporated: $^{13}C_5$,^{15}N valine; $^{13}C_6$ leucine; or $^{13}C_6$,^{15}N isoleucine. From this experiment, several products were detected by MS-Dial (**Table 4.2, Table A2.7**) and we could reason through these findings to determine the activity of PhalD, the poorly-annotated enzyme.

The most abundant products from the metabolomics experiment were dipeptides containing one of the three labelled amino acids, one unlabeled amino acid, and an additional oxygen (**Fig. 4.10**). We reasoned the unlabeled amino acid in these dipeptides is proline based on the additional degree of unsaturation needed to satisfy the predicted chemical formulas of these compounds. Notably, these products were detectable when only PhalD and PhalF were present, suggesting PhalD is responsible for the additional oxygen in these products, where PhalD has some monooxygenase activity. In addition to these products, we detected dipeptides Val-Val and Val-Ile with hydroxylations in conditions that contained PhalD and PhalF (**Fig. 4.11**). While these products were less abundant than the other detected products, they were in better agreement with our *in vitro* findings. To our surprise, no unique products were found when all three Phal enzymes were present (condition #1), despite PhalA activity being detectable *in vivo* (**Fig. 4.12**).

Elucidation of the function of PhalD. Beginning from a poor annotation as simply NAD-binding, we could determine the function of PhalD from this metabolomics experiment. Given hydroxylated products were detected when only PhalD and PhalF were present, we reasoned PhalD is responsible for the hydroxylation and assigned it as a FAD/NAD(P)H monooxygenase. This class of enzymes oxygenate at heteroatoms, suggesting hydroxylation at the α-amine. We attempted reconstituting some of the products detected *in vivo* by supplying the necessary cofactors for flavin-dependent monooxygenation (NADPH and free flavin) along with ATP for PhalF. We could detect hydroxylated Val-Val and Val-Leu/Ile dipeptides in endpoint reactions (**Fig. 4.13**).

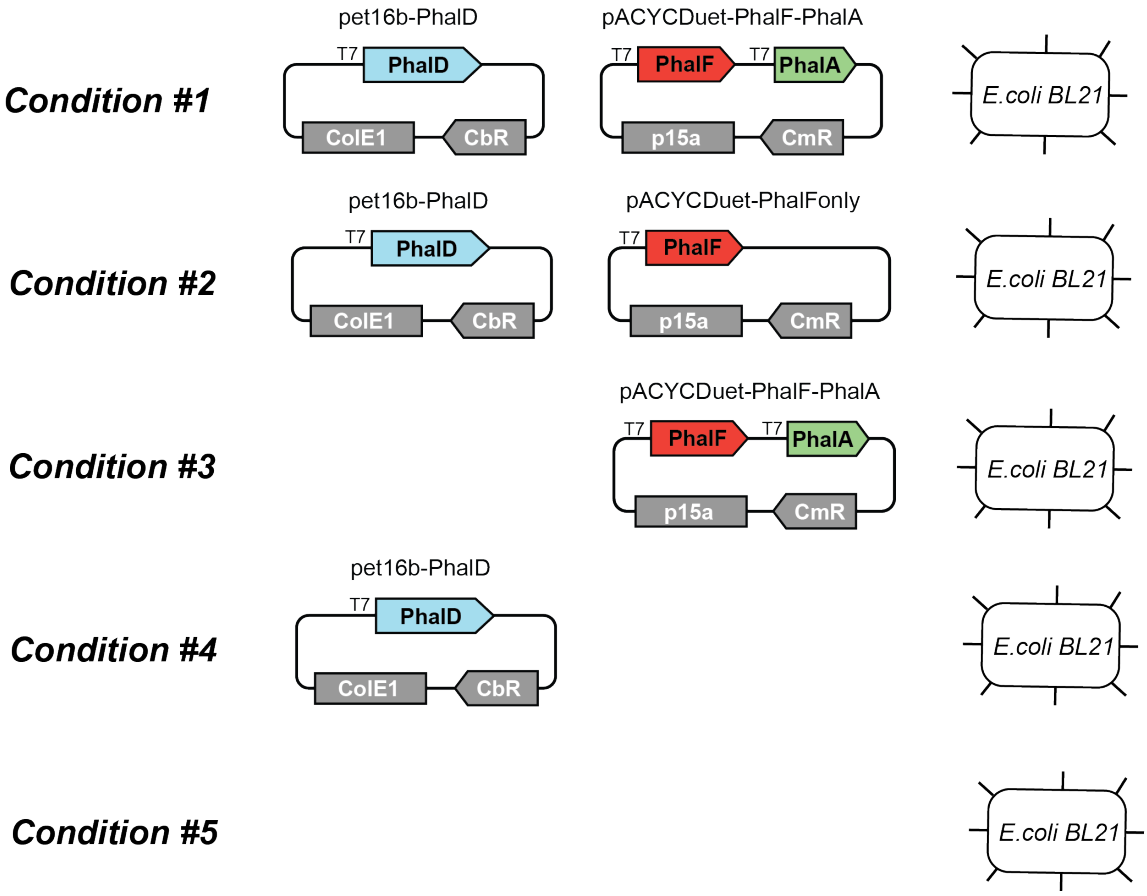
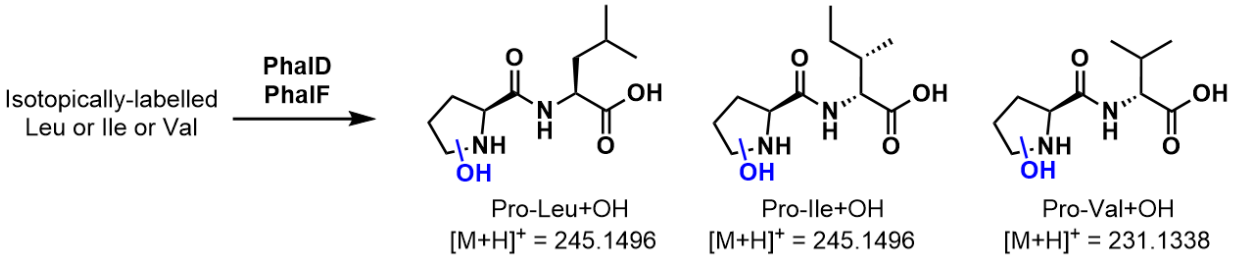


Figure 4.9. Experimental design for comparative metabolomics of the Phal BGC. Through various combinations across five conditions, the roles of the *Phal* BGC enzymes were assessed *in vivo*. Each of the conditions included one of 3 isotopically-labelled amino acids: $^{13}C_5$,^{15}N valine or $^{13}C_6$ leucine or $^{13}C_6$,^{15}N isoleucine, resulting in 15 conditions overall.

RT (min)	*m/z*	Type	ID	ANOVA *p*-value	Signal/ Noise	Fold Increase
6.37	252.1670	$[M+H]^+$	Pro-Ile+OH ($_{+13}C_6$,^{15}N-Ile)	5.61E-01	4488.8	329.21
7.28	231.1338	$[M+H]^+$	Pro-Val+OH	1.43E-01	2826.2	402.96
7.29	237.1480	$[M+H]^+$	Pro-Val+OH (+$^{13}C_5$,^{15}N-Val)	5.66E-01	1735.4	119
6.87	251.169	$[M+H]^+$	Pro-Leu+OH (+$^{13}C_6$-Leu)	2.09E-01	675.9	50.56
6.38	245.1496	$[M+H]^+$	Pro-Ile+OH	1.11E-01	456.9	72.11
6.81	245.1496	$[M+H]^+$	Pro-Leu+OH	1.23E-01	244.5	46.23
7.48	245.1765	$[M+H]^+$	Val-Val+OH (+2x $^{13}C_5$,^{15}N-Val)	9.34E-02	230.4	13.75
6.70	247.1652	$[M+H]^+$	Val-Ile+OH	4.01E-02	46.3	6.36

Table 4.2. Summary of results from metabolomics experiment. From data analysis with MS-Dial, unique metabolites were identified, along with isotopic enrichments. These metabolites were detected in conditions with or without PhaIA and only required PhaIF and PhaID to be present.



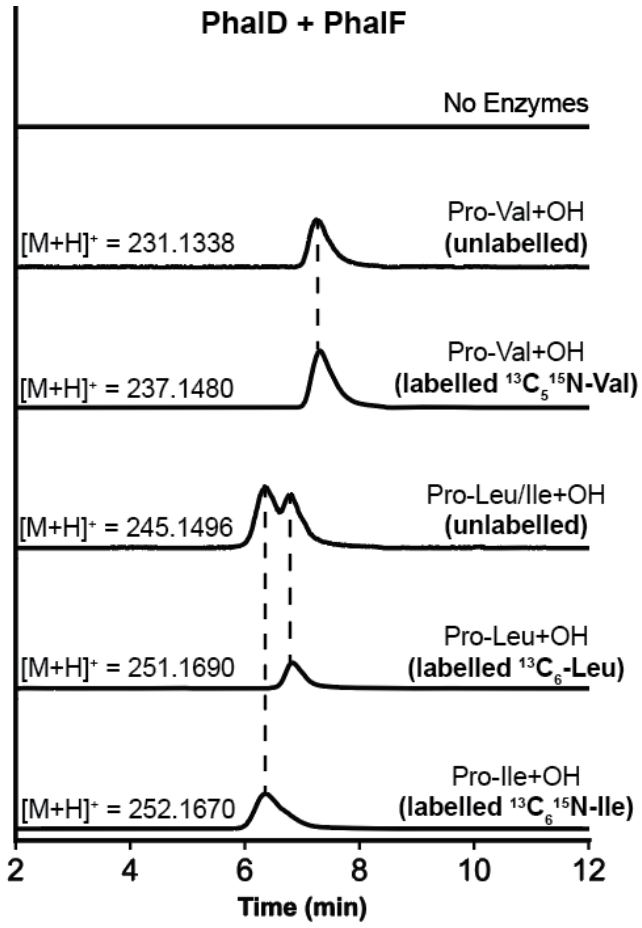


Figure 4.10. PhaIF and PhaID form hydroxylated dipeptides with proline *in vivo*. From metabolomics conditions where PhaIF and PhaID are expressed, we detect masses for various dipeptides with an additional oxygen. Valine incorporation was not detected in the dipeptides containing leucine or isoleucine, and from the additional degree of unsaturation, this experiment indicated proline was likely the ligated amino acid. Detection of hydroxylation indicated PhaID is likely a monooxygenase.

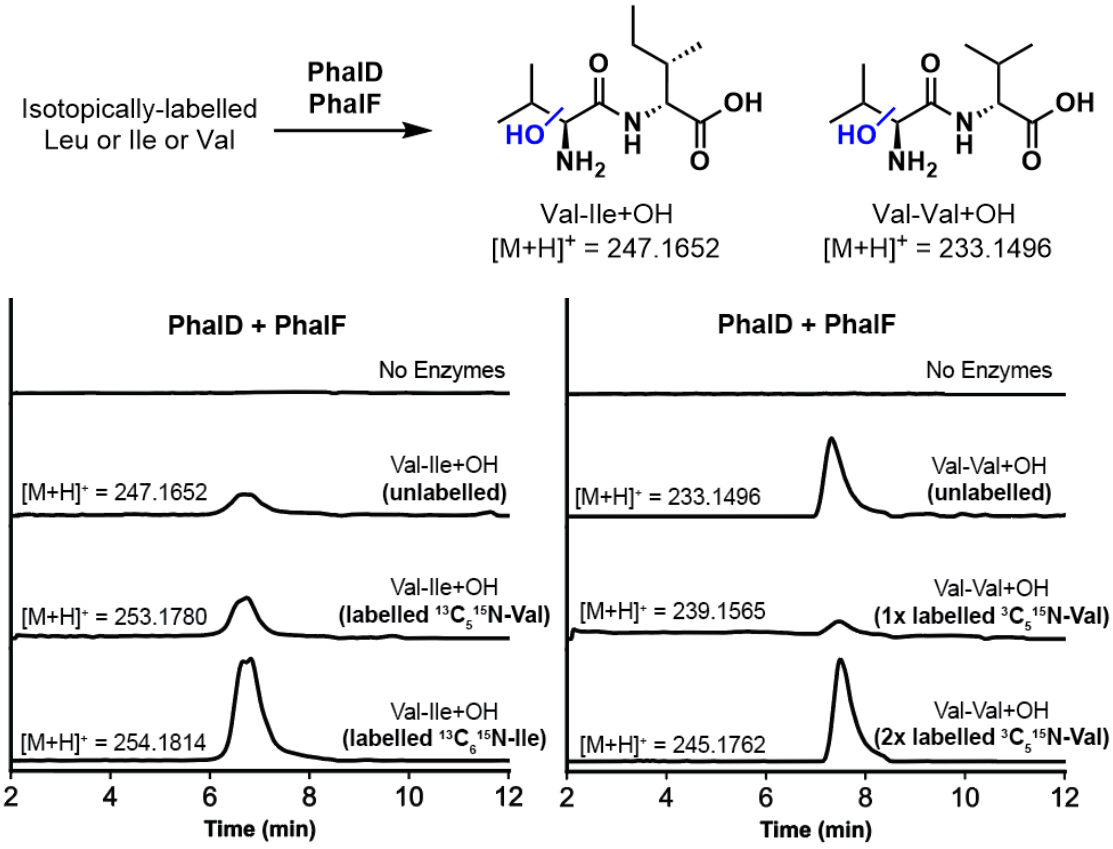


Figure 4.11. PhaIF and PhaID form hydroxylated dipeptides with valine and valine/isoleucine *in vivo*. From metabolomics conditions where PhaIF and PhaID are expressed, we detected masses for dipeptides with valine and isoleucine with an additional hydroxylation *in vivo*. The regiochemistry and connectivity are not clear from this experiment.

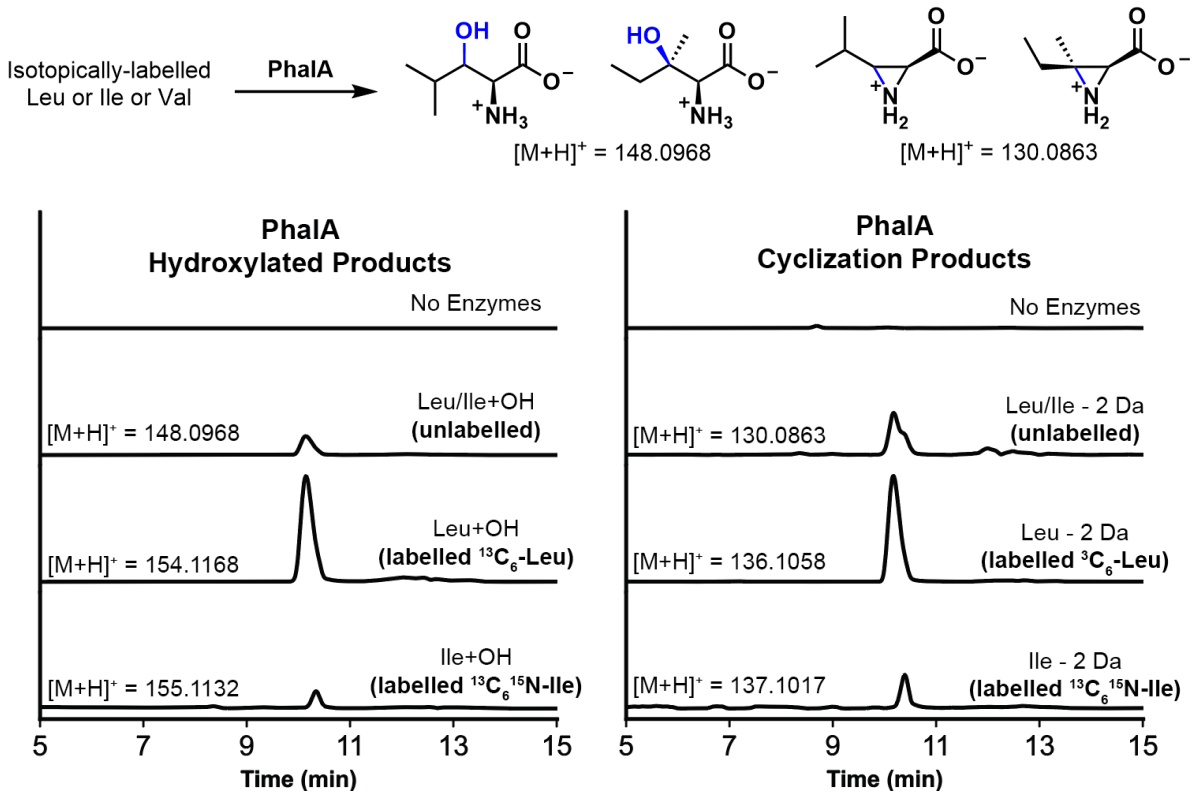


Figure 4.12. PhaIA hydroxylates leucine and isoleucine *in vivo*. In accordance with *in vitro* findings, PhaIA showed hydroxylation and cyclization activity towards leucine and isoleucine in metabolomics conditions containing PhaIA. Despite the enzyme clearly being active, no products were detected that required all 3 enzymes PhaIA, PhaID, and PhaIF. The regioselectivity is unclear from this experiment.

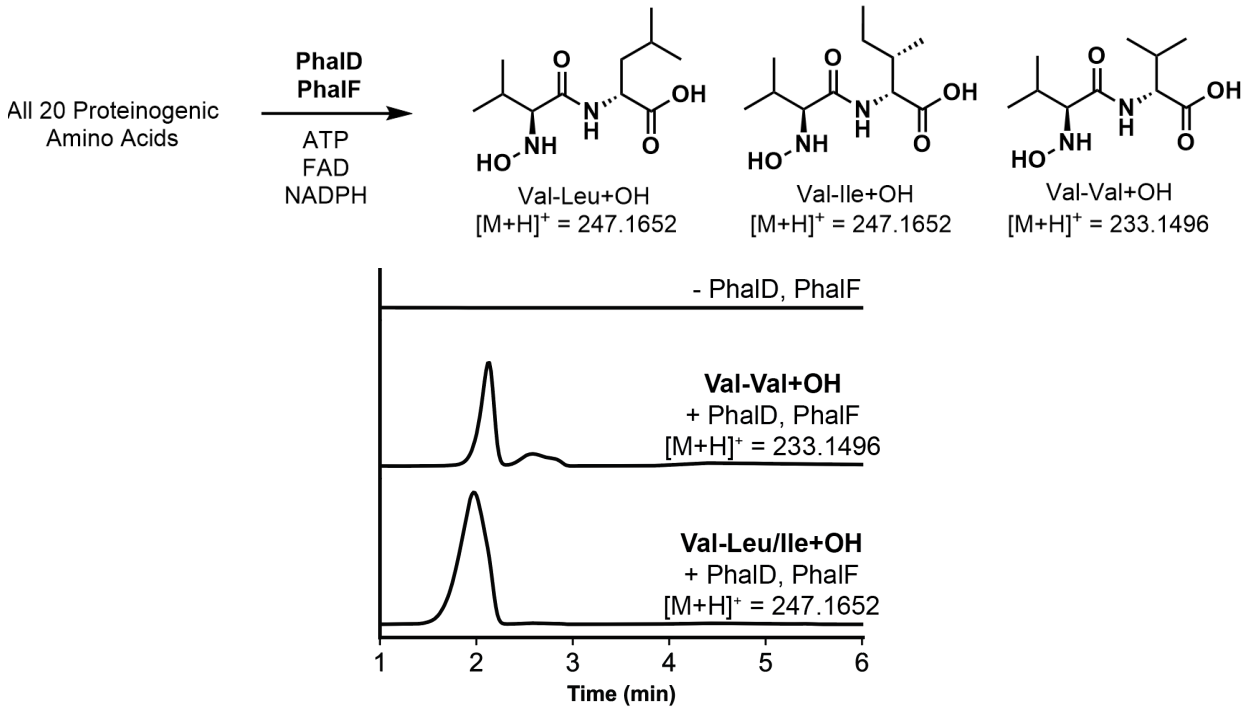


Figure 4.13. *In vitro* reconstitution of Val-Val and Val-Ile hydroxylated peptides. With all 20 proteinogenic amino acids as substrates, along with cofactors for *N*-monooxygenation (FAD, NADPH), we could detect the Val-Val and Val-Ile dipeptides that were seen in the metabolomics experiment. As both isoleucine and leucine were present in the reaction, the mass we detected at *m/z* = 247.1652 could refer to both or either Val-Ile/Leu.

Despite proline-containing dipeptides being the most abundant from the metabolomics experiment, proline incorporation into dipeptides was not detected. However, since we did detect hydroxylated dipeptides of Val-Val and Val-Leu/Ile, we reasoned we were detecting monooxygenase activity by PhalD. When we examined the activity of PhalD individually, we found it could accept single amino acids as substrates, with the highest detectable activity towards valine (**Fig 4.14**). Interestingly, we could not detect activity towards proline, but this could be due to the analytical method requiring amine protection and this being challenging for the less-nucleophilic hydroxylamine.

Since first identifying the *Phal* BGC, the annotation of PhalD was updated to belong to the CreE-like family of NADP-dependent monooxygenases. CreE performs 3 successive oxidations of the α-amine of aspartate to form a nitro group. This comparison supported our assignment of PhalD as a *N*-monooxygenase. We therefore tested whether PhalD can carry out multiple oxidations on valine, the preferred substrate from our *in vitro* studies. When we performed a reaction PhalD with flavin and NADPH under stoichiometric levels of NADPH, we found we could detect all intermediates of the reaction (**Fig. 4.15**). Given the detection of only one oxidation *in vivo*, we previously only searched for one oxidation in our *in vitro* studies even with excess NADPH. This new finding of PhalD catalyzing 3 successive oxidations *in vitro* could be the preferred reaction outcome or due to nonenzymatic oxidation following formation of the hydroxylamine [*20*]. Notably, CreE is one of only two other characterized enzymes in this protein family (15,016 total proteins), indicating this enzyme family is drastically underexplored and the discovery of PhalD's unique substrate and reaction scope is an important contribution (**Fig. 4.16**).

Biocatalytic approaches to prepare diazeniumdiolates using PhalD. With the discovery of PhalD as a new enzyme to form hydroxylamines, we examined the literature for how this moiety may be important in natural products. We were intrigued to find that *N*-hydroxylated valine and similar amino acids are privileged scaffolds for some important bioactive compounds. In the biosynthesis of valdiazen and fragin for example, valine is tethered to a carrier protein for *N*-hydroxylation and eventual formation of a diazeniumdiolate moiety (**Fig. 4.17A**) [*21*]. Using a very similar biosynthetic logic, L-alanosine is biosynthesized from diaminopropionic acid [*22,23*] (**Fig. 4.17B**). The diazeiumdiolate moiety can affect cell metabolism, with L-alanosine even being explored for clinical applications [*24*]. Reconstituting the full biosynthesis of these molecules has been unsuccessful, likely due to the challenges of working with these large modular proteins [*25*]. With the discovery of PhalD, we were curious if *N*-oxidation of free-standing valine with a single enzymatic step could offer an alternative route towards synthesizing these valuable molecules. With this in view, we developed schemes to combine PhalD with other chemical and enzymatic steps to produce these molecules in biocatalytic cascades (**Fig. 4.18A-C**).

The outstanding challenge in constructing these compounds is the lack of consensus on how the diazeniumdiolate moiety is formed. Generally, it is agreed that nitrite is reduced to nitric oxide, which spontaneously reacts with the hydroxylamine to form the diazeniumdiolate through a nitrosonium ion (**Fig 4.18D**) [*23*]. However, the means by which nitrite is reduced *in vivo* is an open question. Some BGCs for these molecules contain nitrate and nitrite reductases, but a majority of these BGCs do not [*26*]. This has led several groups to propose co-clustered cupin

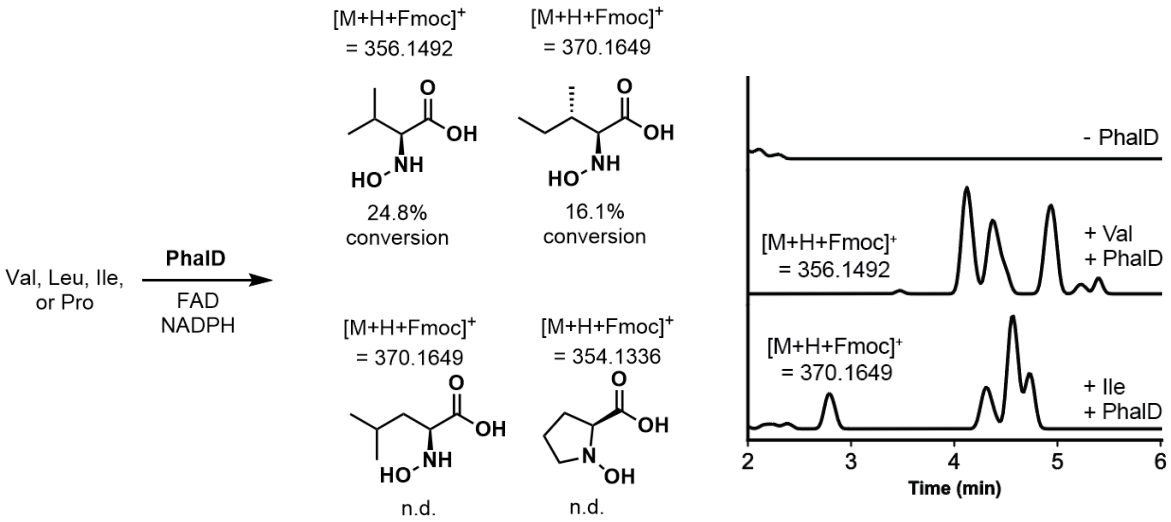


Figure 4.14. PhaID shows preference for valine *in vitro*. The *N*-monooxygenase activity of PhaID was evaluated *in vitro* with individual amino acids. PhaID showed activity towards valine and isoleucine, with poorly resolved peaks likely due to poor protection of the hydroxylamines by Fmoc-Cl.

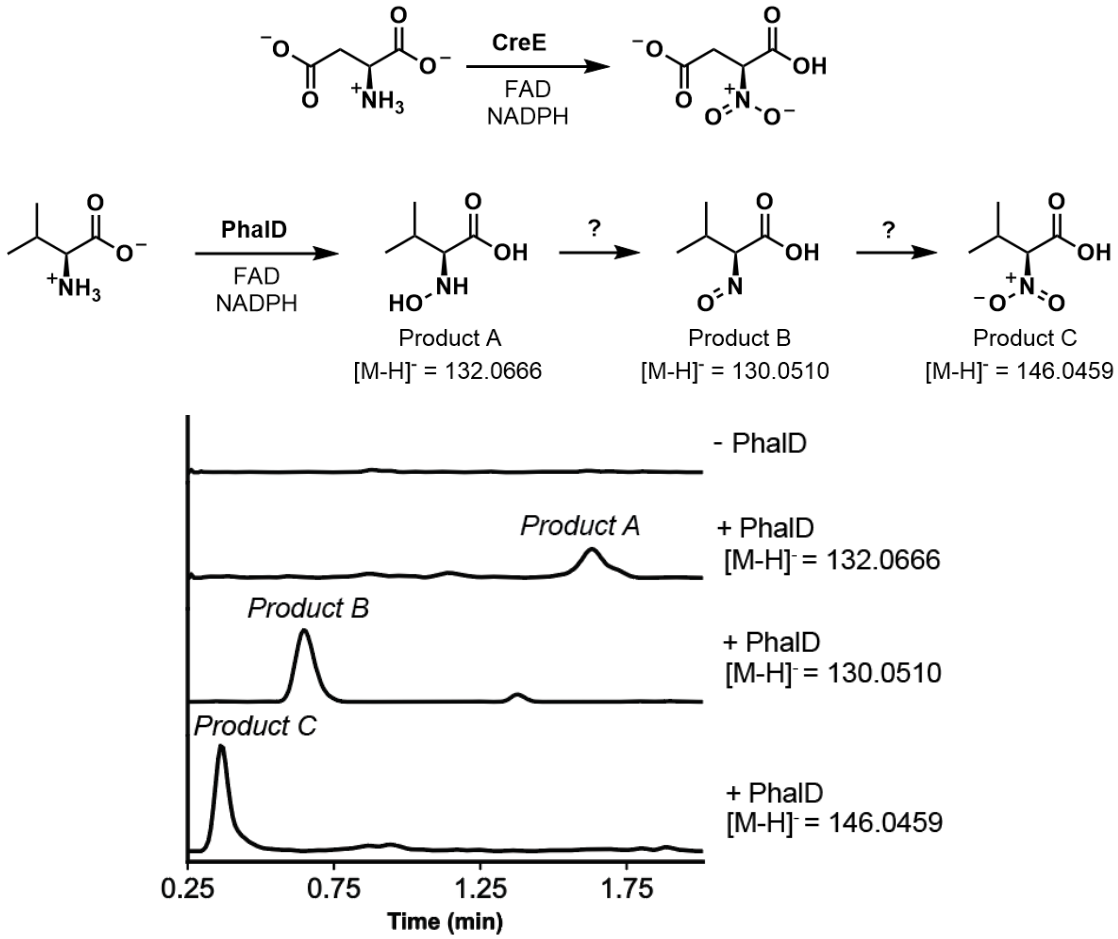


Figure 4.15. PhalD can produce both hydroxylated and nitro valine. In an *in vitro* reaction with stoichiometric levels of NADPH and valine, all oxidation products of valine can be detected from PhalD. With the single oxidation detected *in vitro*, this finding suggests that while PhalD may be capable of 3 successive oxidations, it is unclear if this is enzymatic or nonenzymatic. PhalD may have a different reaction outcome preference than CreE, one of the only other characterized members of this protein family.

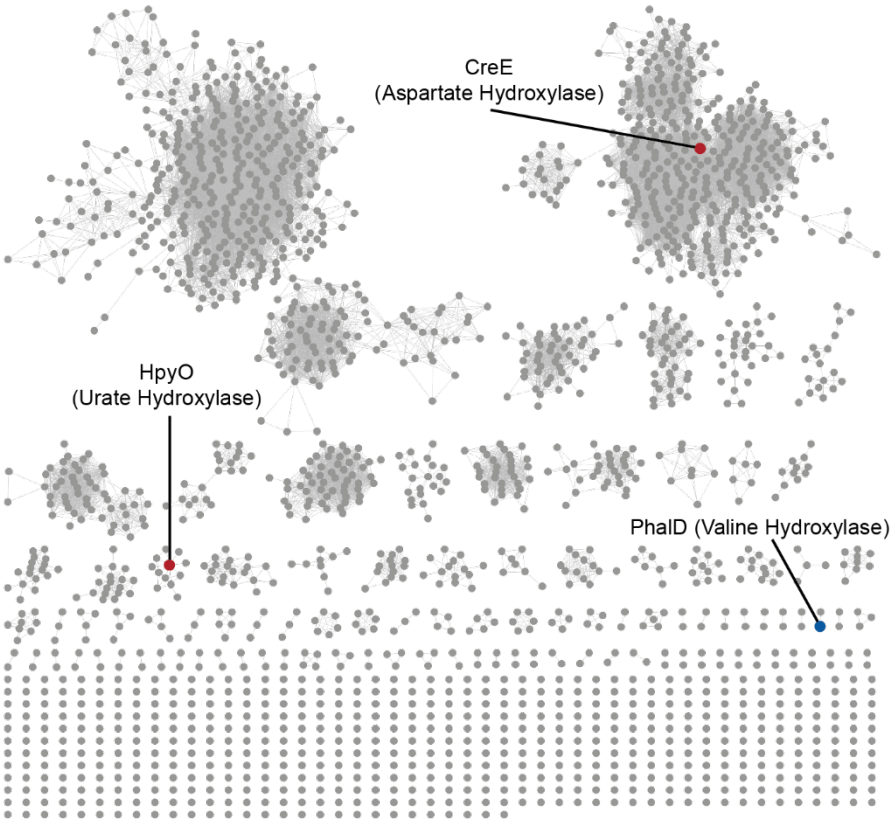


Figure 4.16. Sequence-Similarity Network of Flavin-Dependent *N*-Monooxygenases. Sequence similarity network of a new class of flavin-dependent monooxygenases (IPR038732), with only two previously characterized members. This is a recently discovered protein family with 15,000 members.

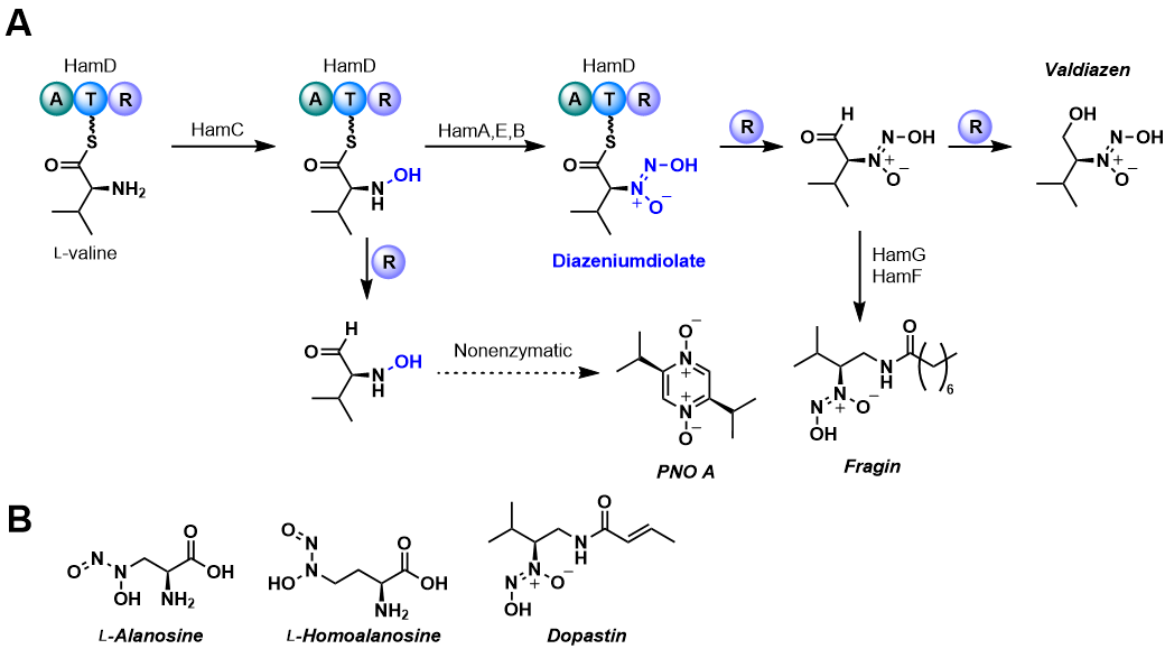


Figure 4.17. Valine is a building block for diazeniumdiolate-containing natural products. (A) Proposed biosynthetic pathways found in the literature for several hydroxylamine- or diazeniumdiolate-containing natural products that begin with valine. (B) Additional diazeniumdiolate-containing molecules that follow a similar biosynthetic logic as valdiazen and fragin.

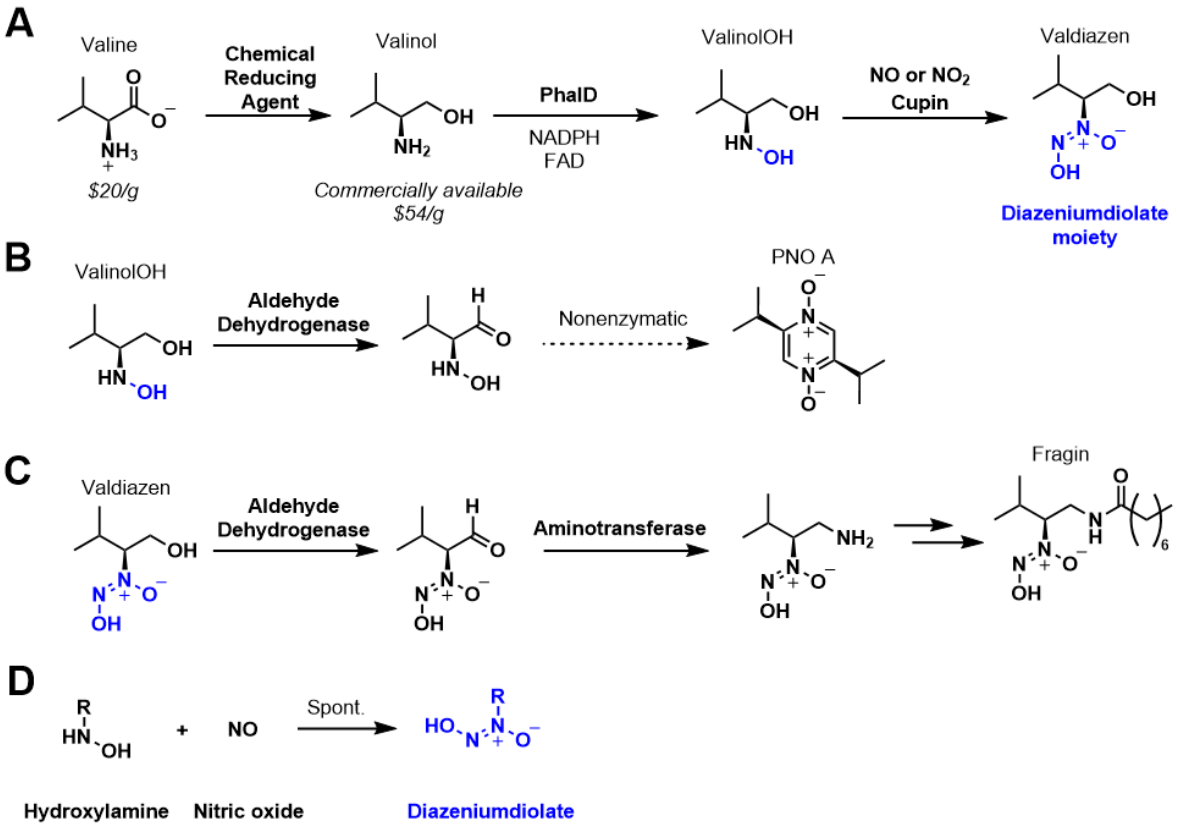


Figure 4.18. Biocatalytic approaches toward diazeniumdiolate-containing compounds. (A) A chemoenzymatic approach to synthesizing diazeniumdiolate-containing valdiezen using PhaID following reduction of valine. (B) An enzymatic approach to synthesize PNO A, which can spontaneously assemble following aldehyde formation. (C) An approach to synthesize Fragin from Valdiezen by intercepting the aldehyde with an aminotransferase and additional enzymes from the BGC for Fragin. (D) Synthetic approach to synthesize diazeniumdiolates from nitric oxide and hydroxylamine.

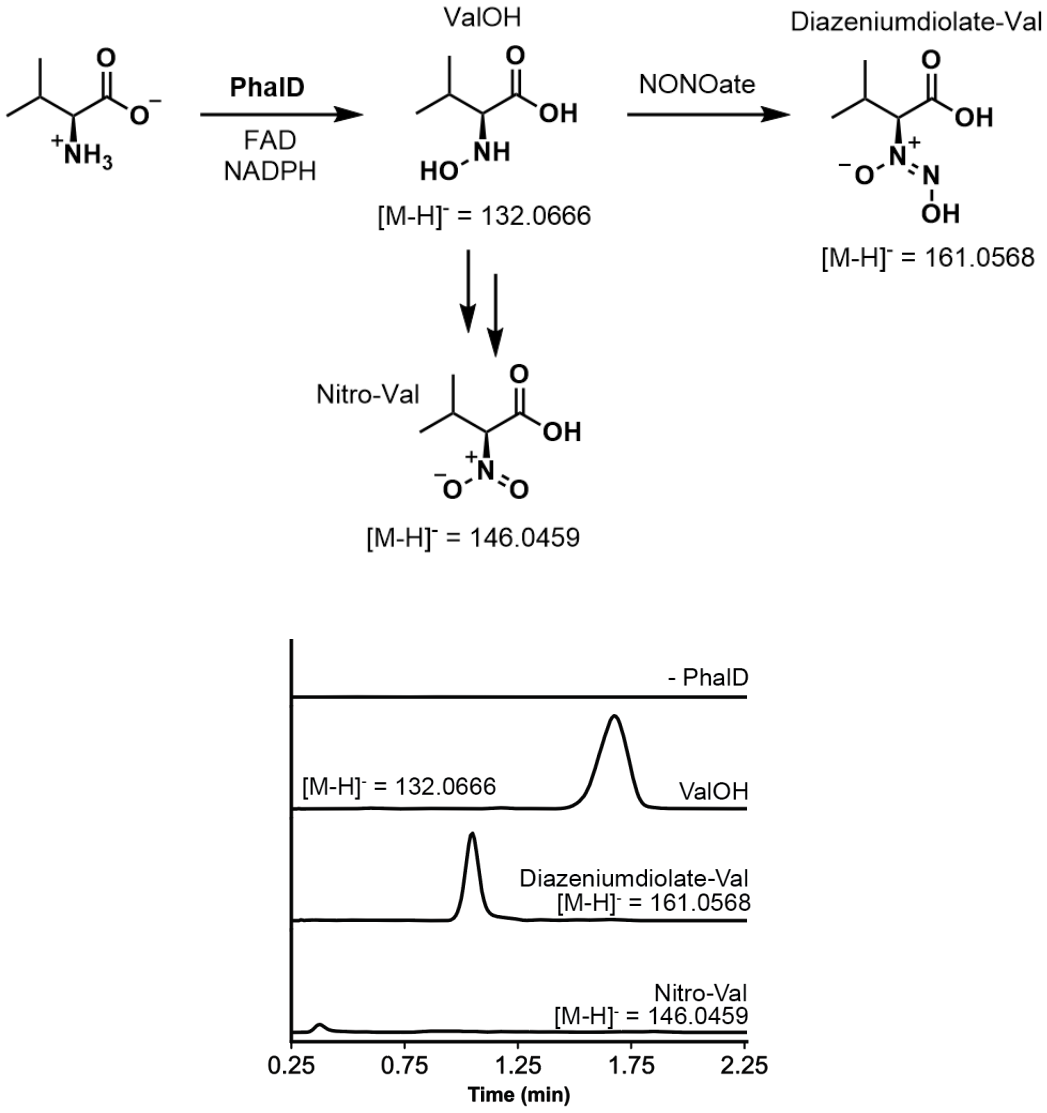


Figure 4.19. Diazeniumdiolate formation with PhalD and diethylamine NONOate. Following hydroxylamine formation with PhalD, diethylamine NONOate was added to produce nitric oxide *in situ*, which formed the expected mass for diazeniumdiolate-Val. By comparison, the amount of Nitro-Val was much lower than ValOH as seen by LC/MS. This reaction was done at 5x the level of NADPH and valine as previous *in vitro* experiments with PhalD and valine.

proteins are responsible for reducing nitrite but no experimental evidence has been shown to date. Nitric oxide can alternatively be made chemically, where ascorbic acid can reduce nitrite at low pH, but this rate goes to 0 at physiological pH [*27*]. As an alternative strategy, diethylamine NONOate spontaneously releases NO at physiological pH and has been shown to produce the diazeniumdiolate moiety previously [*23*]. However, this reagent is unstable when exposed to air at room temeperature, making it less amenable to a biocatalytic process. With these different routes in view, we began exploring both the chemical and biological approaches to synthesizing the diazeniumdiolate moiety.

In order to assess the potential for utilized PhalD for preparation of the diazeniumdiolate of valine, we utilized a chemical approach that would also assist to establish an analytical detection of the diazeniumdiolate moiety. We scaled up a reaction with PhalD and valine, and then supplied dethylamine NONOate to the reaction after 2 hours. From this experiment, we could detect the diazeniumdiolate-containing product by mass (**Fig 4.19**). We also found that scaling the reaction five-fold improved the relative ratio between the hydroxylamine product and the nitro product.

With the ability to form the diazeniumdiolate, we examined if PhalD is promiscuous towards L-valinol to provide the precursor for Valdiezen (**Fig. 4.18A**). To examine if PhalD can accept this substrate, we performed an *in vitro* reaction with L-valinol and PhalD and gathered timepoints at 2 and 18 h. From this, we detected a time-dependent increase in the abundance of the hydroxylamine product, showing that our scheme has some potential (**Fig. 4.20**). As a note, the analytical methodology here depends on the positive ion of the amine for detection, as there is not a carboxylic acid present to provide a negative charge. If the diazeniumdiolate version of valinol is synthesized, it may be challenging to detect this product with the analytical procedures we have established thus far. The logical next step would be to attempt diazeniumdiolate formation with diethylamine NONOate following L-valinol *N*-oxidation by PhalD.

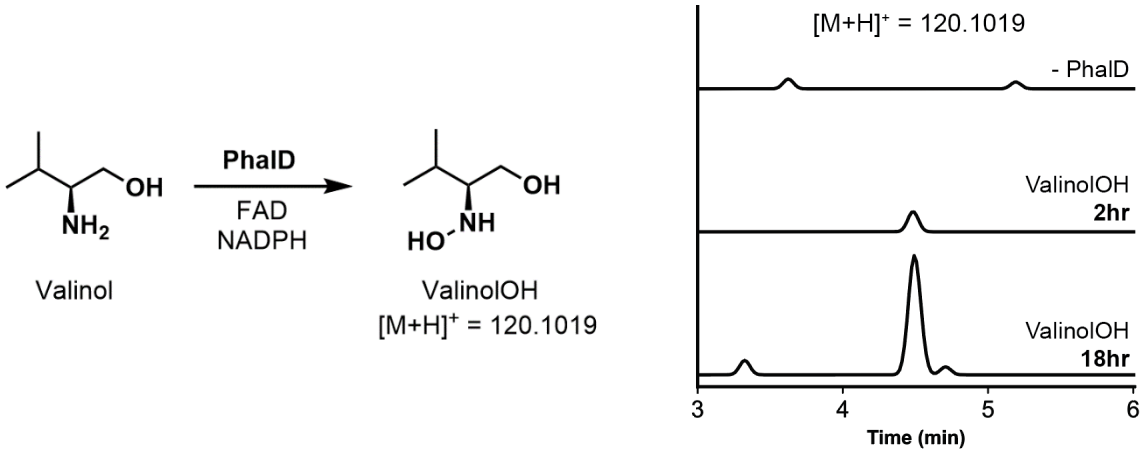


Figure 4.20. *N*-monooxygenation of valinol by PhalD. PhalD appears to accept to valinol as a substrate, with the hydroxylamine product detected as the Fmoc-deriviatized mass. This could be leveraged in our designed pathway for valdiezen synthesis.

Efforts to develop biocatalytic methods for preparation of a diazeniumdiolate. With a chemoenzymatic route in hand for production of the diazeniumdiolate moiety using PhalD with NONOate, we next approached the possibility of providing the NO enzymatically. We examined Alphafold2 structures of the cupin family proteins from both the L-alanosine and fragin biosynthetic gene clusters to see if there are conserved sequence motifs as these enzymes are

predicted to utilize amino acid-derived substrates (**Fig. 4.21**) [*28*]. Characterized nitrite reductases rely on copper or iron in the active sites for catalysis [*29,30*]. Enzymes that are selective for performing one reduction to nitric oxide have been found to utilize Cu, whereas enzymes that carry out the full reduction to ammonia have been found to utilize heme cofactors. We identified residues that could potentially bind metals within the characteristic β-barrel fold in both AlnA and HamB (His, Glu, Tyr) (**Fig. 4.21**) These cupin proteins could be a new form of nitrite reductases, though it is uncertain how many and which metals they bind, as well as their preferred electron donor. As a note, a similar class of cupin proteins was recently shown to form the nitrogen-nitrogen bonds in hydrazine-containing amino acids with a catalytic Zn [*31*].

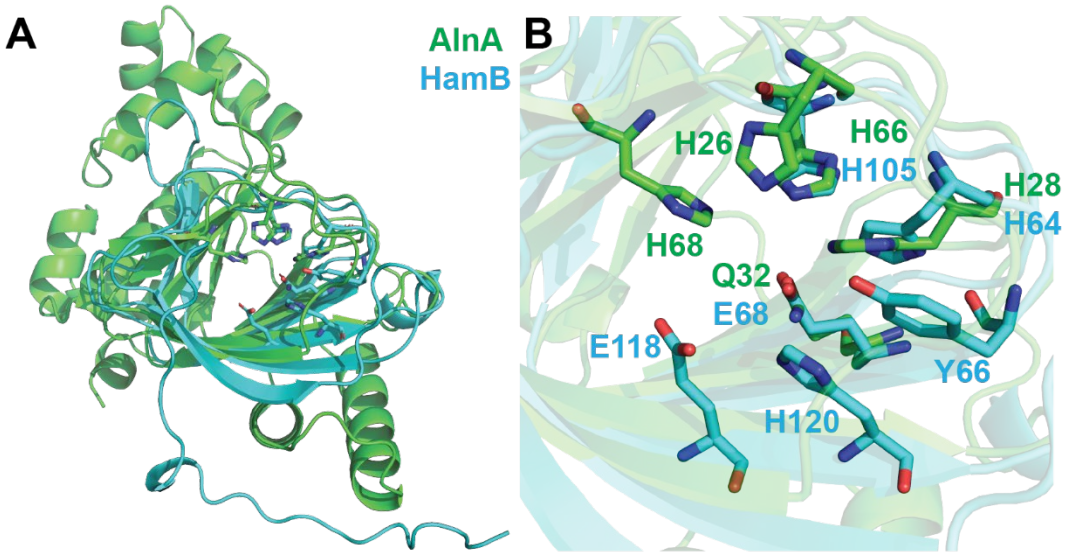


Figure 4.21. Structural models of putative nitrite reductases. AlphaFold2 models of cupin proteins from diazeniumdiolate-producing BGCs contain the characteristic β-barrels of the cupin family with putative metal-binding residues for catalyzing nitrite reduction.

Following this *in silico* modeling, we began to explore the activity of these proteins experimentally. We began with heterologous expression and purification of N-terminal His_{10}-tagged Hamb and C-terminal His_6-tagged AlnA. From this, we attained a high yield of HamB but considerably less for AlnA. As a note, AlnA is a didomain protein with a DNA-binding domain that is C-terminal to the Cupin domain. For this experiment, we only expressed the N-terminal Cupin portion of the protein, which could have an impact on protein stability and explain the lower yield. In experiments with PhalD, valine, sodium nitrite, and either AlnA or HamB, no diazeniumdiolate product was detected. Because of this, we hypothesize adding a metal and a reducing agent such as sodium ascorbate may be necessary for the reaction to proceed. In addition, if product were to form, the product may degrade over time as the diazeniumdiolate releases NO, so measuring production via LC/MS needs to be done soon after the reaction completes [*32*].

Developing routes to additional diazeniumdiolate-containing natural products. With developments in synthesizing diazeniumdiolate moieties from valine, we explored if this strategy could be readily applied towards other natural products. Namely, L-alanosine and L-homoalanosine are derived from the diamino acids diaminopropionic acid (DAP) and diaminobutyric acid

(DABA), respectively. We envisioned by discovering *N*-hydroxylases for these molecules, we could bypass the need for the carrier-protein approach found in known BGCs (**Fig. 4.22A**). A N_{γ}-monooxygenase is known for DABA but none have been discovered for DAP [*33*]. Fortunately, DAP is biosynthesized according to a conserved logic, which we could use as marker genes for genome mining [*22*]. We examined the family of lysine and ornithine *N*-monooxygenases (PF13434) and discovered a conserved BGC containing the DAP biosynthetic genes proximal to both a putative N_{β}-monooxygenase as well as a putative N_{α}-monooxygenase (representative Uniprot ID: A0A0N0FVR4) (**Fig. 4.22B**). We anticipate the approach developed to access valdiezen and fragin could easily be transferred to the biocatalytic synthesis of L-alanosine and its analogs. This scheme could be further scaled through the use of a NADPH recycling system for a biocatalytic synthesis of these molecules.

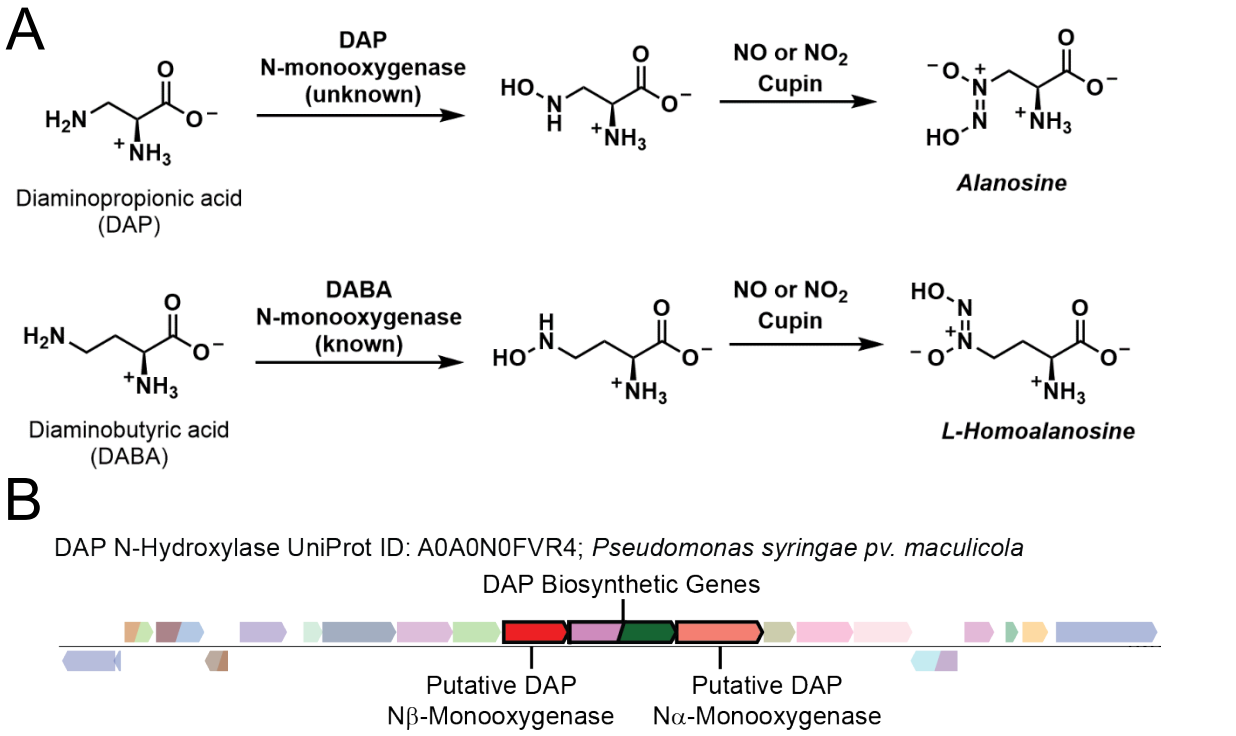


Figure 4.22. Discovery of *N*-monooxygenases for L-alanosine synthesis. (A) Schemes for the biocatalytic synthesis of diazeniumdiolate-containing diamino acids through enzyme discovery. (B) Discovery of a putative DAP Nβ-monooxygenase (PF13434) and Nα-monooxygenase (PF13454) through co-occurrence of a DAP biosynthetic gene cassette (PF00291 + PF02423).

Expanding tR3D to the YbaK/INS family of tRDs. To this point, tR3D had only been applied to the AlaX family of tRNA deacylases. This was largely a product of chance, as BesG belonged to the AlaX family and thus was the family chosen first for analysis. There are several other families of tRNA deacylases, with one of the largest being the YbaK/INS family (PF04073) [*34*]. This family includes editing proteins for the proline-tRNA synthetase as well as the editing protein for 4-fluorothreonine (F-Thr), indicating this family encompasses both primary and secondary metabolism. We were interested in applying the same analytical methodology as we did for the AlaX family, with the goal of quantifying how much of this family is potentially involved in *s*ary metabolism. To do this, we followed the same general procedure for SSN cleaning and analysis, resulting in the identification of several interesting biosynthetic gene clusters.

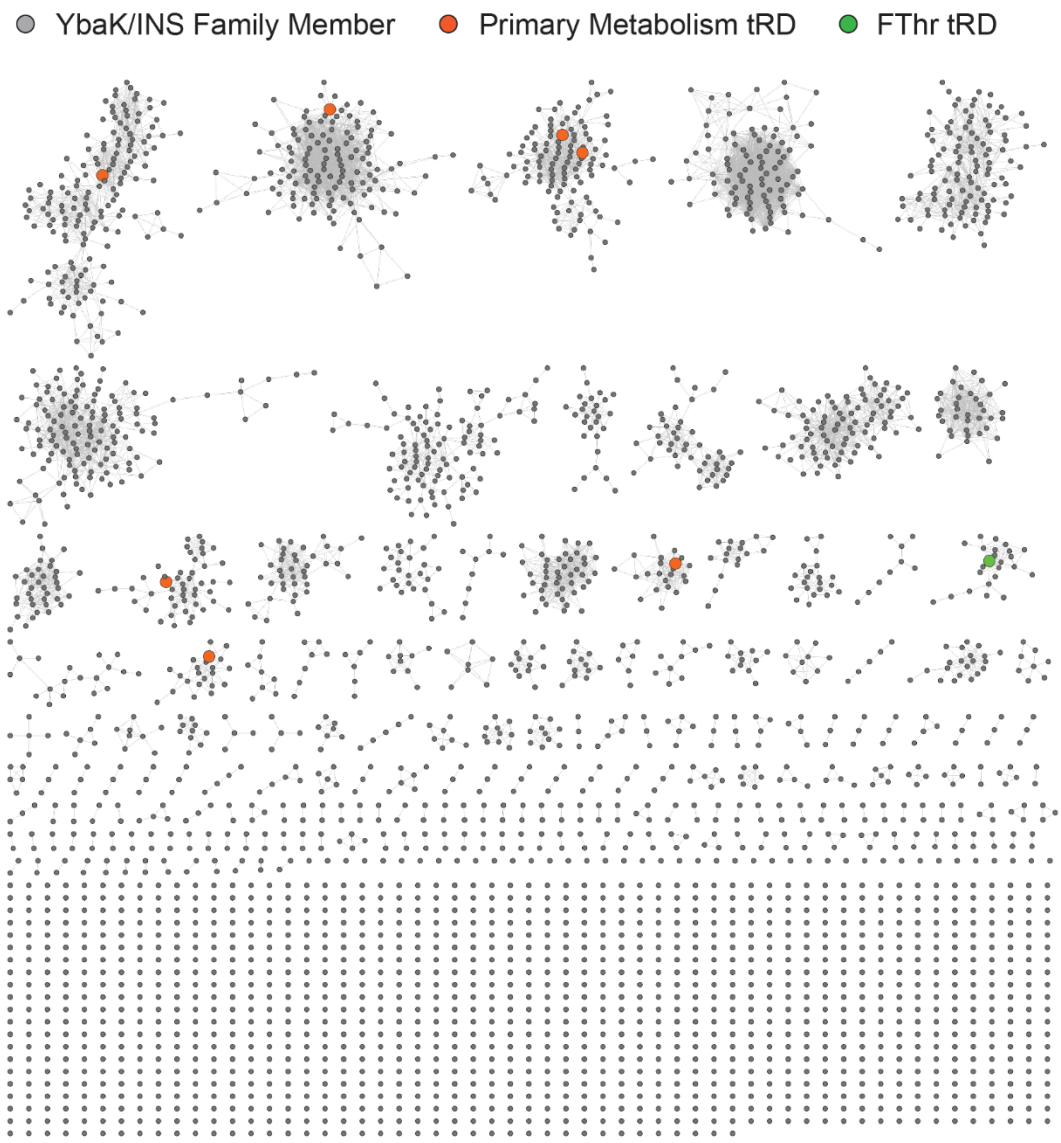


Figure 4.23. Sequence similarity network (SSN) of YbaK/INS family. SSN of the YbaK/INS family (PF04073) of tRNA-deacylases (tRDs) with characterized tRDs mapped to their corresponding cluster.

Gene ID	SSN Cluster #	Protein (PDB if available)	Primary or Sary
P0AAR3	1	*E. coli* Cys YbaK (pdb: 2dxa)	Primary
P45202	1	H*aemophilus influenzae* YbaK (pdb:1DBU)	Primary
Q5SHN1	4	*Thermus thermophilus* ProX (pdb:2z0k)	Primary
A0A0H3C2S2	5	*Caulobacter vibrioides* ProX (pdb: 5vxb)	Primary
P64483	8	*E. coli* Ser/Thr ProXp-y	Primary
P64484	8	*E. coli* ProX	Primary
Q9L4Q7	9	*Acetoanaerobium sticklandii* ProX	Primary
A6NEY8	17	H*omo sapiens* ProX	Primary
F8JMF2	22	*Streptomyces cattleya* FThr editing protein	Sary

Table 4.3. Characterized members of the YbaK/INS family. Characterized tRDs from the YbaK/INS family along with their role in either primary or secondary metabolism.

To begin the analysis of the Ybak/INS family, we produced a SSN using the EFI tool and mapped characterized tRDs to this network (**Fig. 4.23, Table 4.3**) [*10*]. This network contained more than twice the number of proteins as the AlaX family (53,310 YbaK/INS members vs. 23,260 AlaX members) (**Table 4.4**). Visually, the SSN appeared more fragmented than the AlaX family even though the sequence similarity score cut off was lower (39 vs. 43), so a rough assignment of primary metabolism was less clear. The primary metabolism functions include YbaK proteins that edit Cys-tRNA^{Pro} and ProX proteins that edit Ala-tRNA^{Pro}, where these two types separate at this alignment score. Many of the SSN clusters contain no characterized member, so we examined how genomic neighborhood data can be utilized to group these clusters by function.

Description	No.
Total Nodes	3184
Total YbaK/INS Sequences Represented	53310
Minimum Length Filter (AAs)	0
Maximum Length Filter (Aas)	300
Sequence Similarity Score Cutoff (%)	39
Neighborhood Size (upstream and downstream)	10
Co-occurrence Lower Limit (%)	0

Table 4.4. Statistics for YbaK/INS family SSN generation. Summary statistics from the YbaK/INS family SSN generation (PF04073) (accessed November 2022), which encompasses more than 50,000 proteins, which is twice as large as the AlaX family.

For the neighborhood-based analysis, we first needed to clean the dataset before proceeding to analyze this network with the tR3D tool. We found that relatively few sequences were fragments or resulted from poor genomic reads, as only 800 sequences were removed after two cleaning steps (**Tables 4.5-4.6**). Trimming the dataset of primary metabolism-associated tRDs was very impactful, resulting in 19.5% of the dataset remaining for further analysis (**Tables 4.7-4.8**). Some of the most common protein families associated with primary metabolism included RNA synthetic genes, DNA polymerase, and aminoacyl-tRNA synthetases (aaRSs). With this dataset, we could now search for *s*ary metabolism, which had been enriched in this dataset.

Description	Total Nodes (No.)	Represented Sequences (No.)
≤ 6 Neighboring Genes	580	682
>6 Neighboring Genes	2604	52628

Table 4.5. Filtering SSN of short/poor genomic reads. Resulting sequences from cleaning the YbaK/INS family of short genomic reads with poor neighborhood information, of which there were very few.

Description	Total Nodes (No.)	Represented Sequences (No.)
Every sequence is a fragment	77	112
At least one sequence is complete	2527	52516

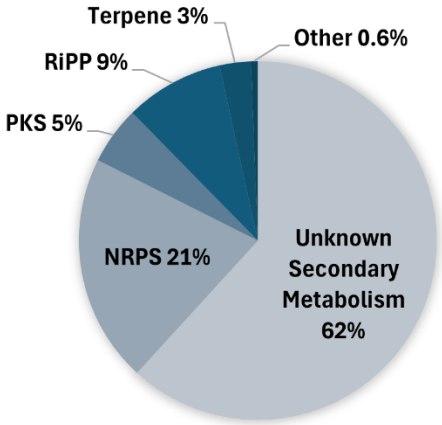
Table 4.6. Filtering SSN of fragments. Resulting sequences from cleaning the YbaK/INS family of fragments, of which there were very few. The resulting SSN was examined for primary metabolism genes.

Pfam	Description	Nodes Removed (No.)	Sequences Removed (No.)
PF00745	Glutamyl-tRNA^{Glu} reductase	32	6614
PF07733	Bacterial DNA polymerase III alpha NTPase domain	153	14394
PF00849	RNA pseudouridylate synthase	126	9279
PF01612	3'-5' exonuclease	38	2721
PF00749	tRNA-synt_1c (E,Q)	54	2999
PF00133	tRNA-synt_1 (I,L,M,V)	81	1233
PF06144	DNA polymerase III, delta subunit	20	248
PF13393	Histidyl-tRNA synthetase	34	1144
PF01926	50S ribosome-binding GTPase	125	3210
PF03548	Outer membrane lipoprotein carrier protein LolA	32	351
PF01084	Ribosomal protein S18	13	1082
PF00453	Ribosomal_L20	14	801
PF00448	Flagellar Biosynthesis protein	16	1341
A6NEY8	Neighbors of Human ProX	12	601
	Remaining Singletons	389	389
	Total	**1139**	**46407**

Table 4.7. Common primary metabolism genes that co-occur with YbaK/INS tRDs. Primary metabolism genes neighboring YbaK/INS tRDs were removed to enrich the dataset for secondary metabolism.

Description	Total Count (No.)
Remaining Nodes	1422
Remaining Genomes (% of original)	10,395 (19.5%)

Table 4.8. Remaining YbaK/INS entries after filtering steps. The number of remaining nodes and sequences following cleaning the YbaK/INS tRDs, which gives a rough approximation of how many tRDs could be involved in secondary metabolism.



Pfam ID	Description	Selected Nodes (No.)	Represented Sequences (No.)
NRPS			
PF00668	Condensation Domain	22	562
PF00501	AMP-binding enzyme	153	2888
PKS			
PF00109	Beta-ketoacyl synthase N-terminal	33	570
RiPP			
PF05147	Lanthionine synthetase C-like protein	5	68
PF00733	Lasso Peptide Cyclase/Asparagine Synthase	19	426
PF13471	Lasso Peptide Protease/Transglutaminase	1	4
PF05402	RiPP Recognition Element/PqqD	8	86
PF02624	YeaO Cyclodehydratase	8	754
Terpene			
PF01397	Squalene/phytoene synthase	14	396
Other			
PF00195	Chalcone and stilbene synthases, N-terminal domain	3	163
PF02797	Chalcone and stilbene synthases, C-terminal domain	4	167
PF00432	Prenyltransferase [Indole Alkaloids]	1	12
	Total (Unique):	**214**	**3973**

Figure 4.24. Overall analysis of secondary metabolism in filtered YbaK/INS SSN. All forms of class-defined secondary metabolism neighboring members of the YbaK/INS tRDs, with the largest contributors being NRPSs and RiPPs. All sequences that do not contain class-defined secondary metabolism genes were assigned “unknown secondary metabolism” and make up 62% of the entire dataset including 6,422 sequences.

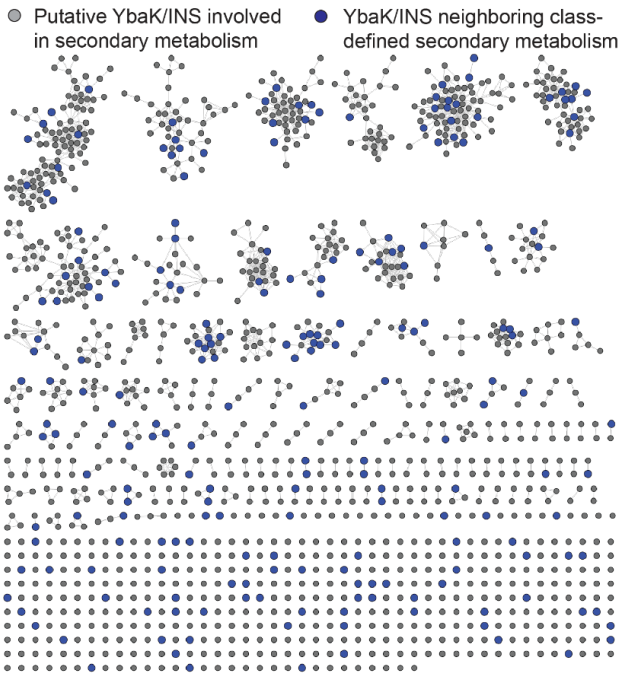


Figure 4.25. Filtered SSN enriched for secondary metabolism. Visualization of the YbaK/INS tRD SSN cleaned of primary metabolism and hypothesized to contain secondary metabolism, with nodes corresponding to tRDs neighboring class-defined secondary metabolism genes.

Searching the YbaK/INS family of tRDs for novel BGCs. Similar to the approach taken for the AlaX family, we searched for class-defined secondary metabolism in our trimmed dataset. We could identify many forms of secondary metabolic genes, with the most abundant being NRPSs, followed by RiPPs, followed by PKSs (**Fig. 4.24, Fig. 4.25**). The YbaK/INS family of tRDs appeared to include much more class-defined secondary metabolism than the AlaX family with a sum of 38%; the AlaX family only contained 16.5% templated metabolism. Overall, from this analysis we speculate the YbaK/INS family contains a large amount of secondary metabolism.

Intrigued by the abundance of class-defined secondary metabolism, we examined several of the SSN clusters in more detail and one in the RiPP family was particularly interesting. RiPPs, or ribosomally translated and post-translationally modified peptides, have a typical biosynthetic logic relying on five main components [*35*]. Namely, a peptide substrate, a protease, a transporter, any number of tailoring enzymes, and a RiPP recognition element (RRE) (**Fig. 4.26**). The RiPP recognition element is a peptide binding component of RiPP-producing enzymes that facilitates an enzyme/peptide complex [*36*]. Owing to its ubiquity, this RiPP recognition element has been used for genome mining for discovering new RiPPs [*37,38*]. For what caught our attention, we identified a BGC containing a YbaK/INS tRD neighboring a protein with RRE homology but lacking any other neighboring genes characteristic of typical RiPP biosynthesis (**Fig. 4.27**). Instead, there were 3 biosynthetic genes more typical of the BGCs explored in this work, including a Fe/αKG dioxygenase, an ATP-grasp ligase, and an NTN hydrolase (**Table A2.4**). The lack of the other typical features of RiPPs suggested a new type of biosynthetic logic. We hypothesized this biosynthetic pathway does not require a template-driven biosynthetic logic, but rather performs chemistry on a free-standing peptide. This BGC also appeared highly conserved among 41 organisms, leading us to suspect it could be an important bioactive natural product that requires all of the enzymes identified (**Fig. 4.28**). We chose the BGC from *Pantoea sp. GM01* (*Pant* BGC) for experimental characterization.

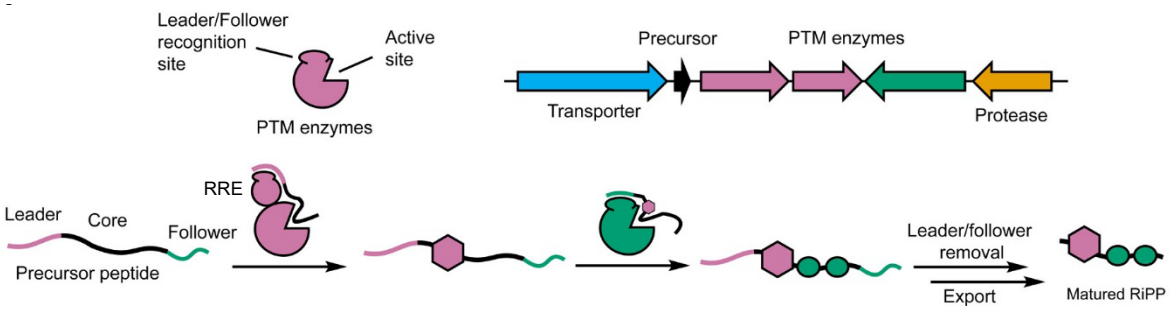


Figure 4.26. Overview of RiPP biosynthesis. General biosynthetic logic of RiPPs contain 5 key elements: a precursor peptide, a protease, a transporter, tailoring enzymes, and a RiPP-recognition element (RRE) that can facilitie peptide/enzyme interactions. This figure was taken from [*35*].

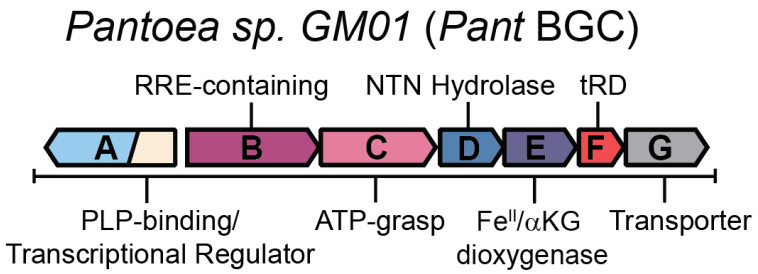


Figure 4.27. Identification of a RiPP-like cluster in *Pantoea sp. GM01*. Identification of a unique BGC containing an RRE-encoding gene but lacking other typical elements of RiPP BGCs. Instead, enzymes characteristic of free amino acid modifications were identified, suggesting this cluster synthesizes a natural product not templated by a precursor peptide.

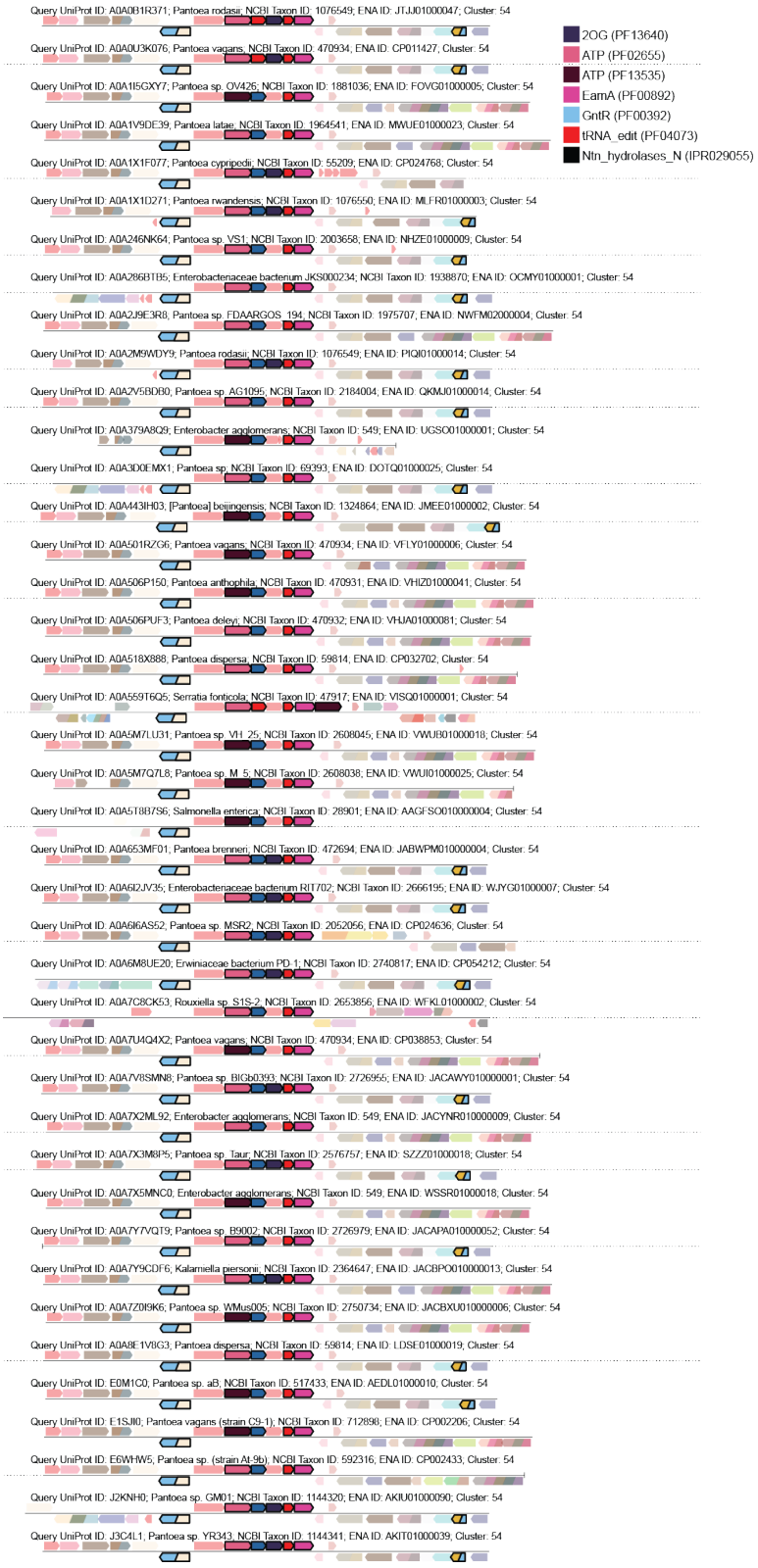


Figure 4.28. Highly conserved RiPP-like BGC. Identification of a unique BGC containing an RRE-encoding gene highly conserved across 41 organisms. Its high level of conservation is suggestive all genes identified are important for the related natural product.

Examining a poorly annotated protein PantB. Examining the RRE-containing protein (PantB) more closely, we were intrigued to find the RRE as a component of a very large protein with no annotated function. To evaluate whether PantB might be performing a novel function, we examined the structure generate with AlphaFold2 and identified several key features [*28*] (**Fig. 4.29**). The RRE forms the N-terminus, with three α-helices and one β-sheet that we would hypothesize to be the peptide-binding site. Adjacent to this is a region with several internal cysteines, as well as other metal binding residues. This is suggestive of an iron-sulfur cluster and some form of redox activity. Interestingly, there are only 3 cysteines, suggesting that if a 4Fe-4S cluster is present, the fourth iron could hypothetically coordinate the substrate or a different metal-binding residue. From the model, it appears the C-terminal arginine comes into the putative active site and is well-positioned to coordinate, forming a site-differentiated cluster. While the function of this protein is not clear from modeling alone, we expected the functions of neighboring enzymes could provide insights.

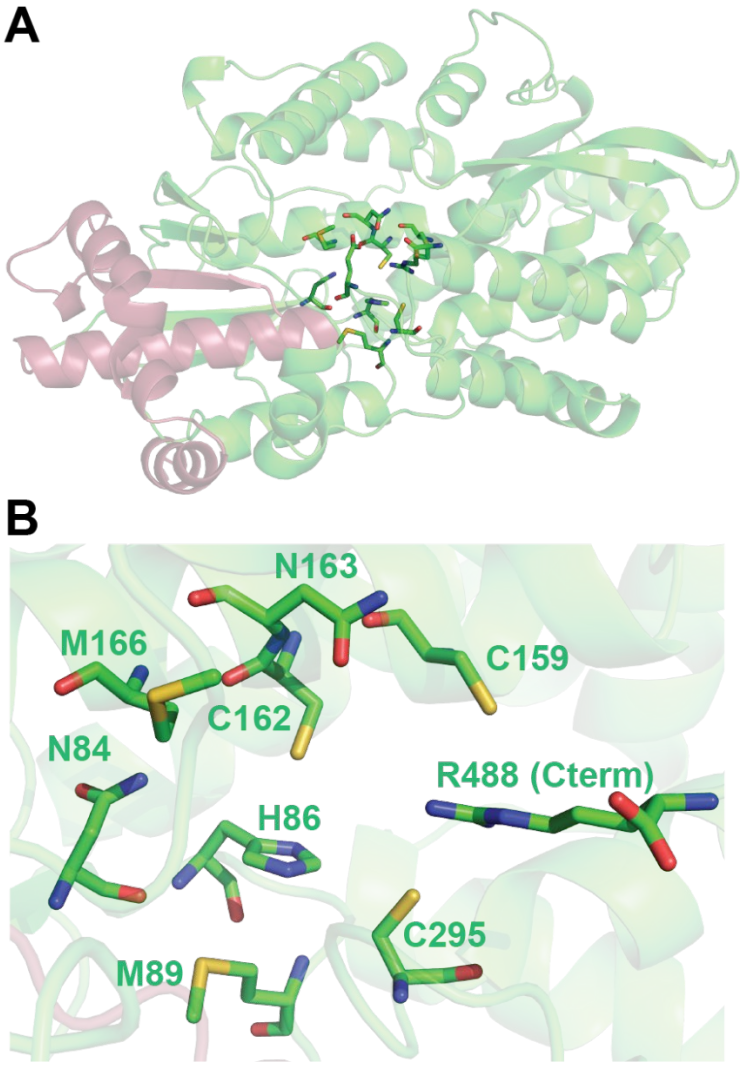


Figure 4.29. Structural model of PantB. (A) PantB Alphafold2 model with the RRE-encoding domain highlighted in maroon. This region appears to be adjacent to the protein active site. (B) Zoomed-in image of the putative active site of PantB, with 3 cysteines and a C-terminal arginine suggesting an iron-sulfur cluster-containing active site. Additional residues capable of metal-binding are shown as well, indicating some form of redox activity.

Elucidating the function of enzymes in the *Pant* BGC. The additional enzymes in the *Pant* BGC were examined *in vitro* to provide insights into the resulting natural product after purifying from heterologous expression with N-terminal His_{10} tags (**Fig. A3.6**). When we performed a reaction with all 20 proteinogenic amino acids, purified ATP-grasp PantC, and necessary cofactors, we detected a glutamate-valine dipeptide as well as a much more abundant peptide corresponding to either glutamate-asparagine or glutamine-aspartate, which have identical masses (**Fig. 4.30**). When we performed a reaction with all 20 proteinogenic amino acids, purified Fe/αKG dioxygenase PantE, and necessary cofactors, we did not detect any modifications of single amino acids. However, when PantE and PantC were present in the same reaction, we did detect a hydroxylation of these dipeptides (**Fig. 4.31**). From these experiments, we could confirm that these enzymes were indeed active towards amino acids and the corresponding natural product could contain one of these dipeptides.

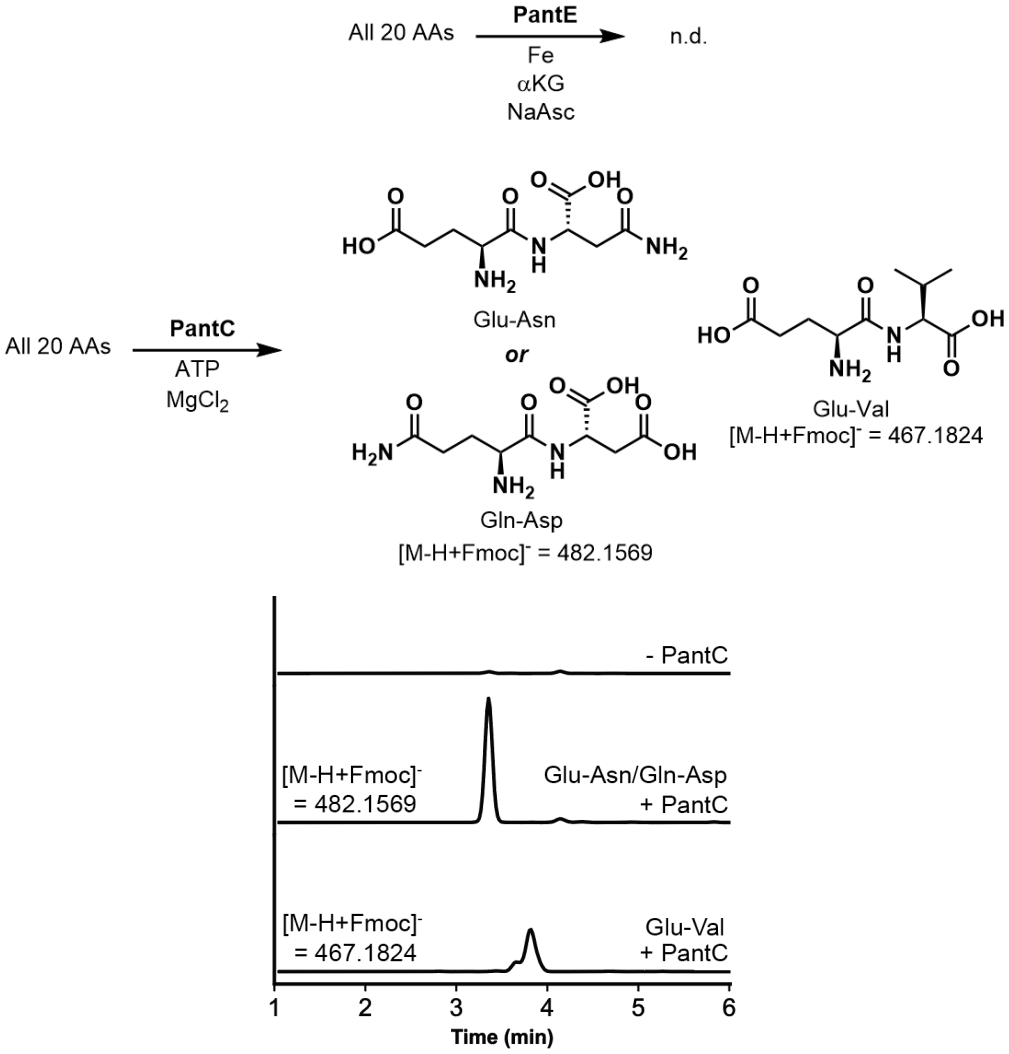


Figure 4.30. *In vitro* screening of PantC and PantE with proteinogenic amino acids. While PantE showed no reaction towards individual amino acids, PantC showed dipeptide-forming activity towards several polar amino acids and valine. Due to the mass being identical for Glu-Asn and Gln-Asp dipeptides, at this point it was unclear which amino acids were components of these products. The connectivity of these dipeptides is proposed but not confirmed.

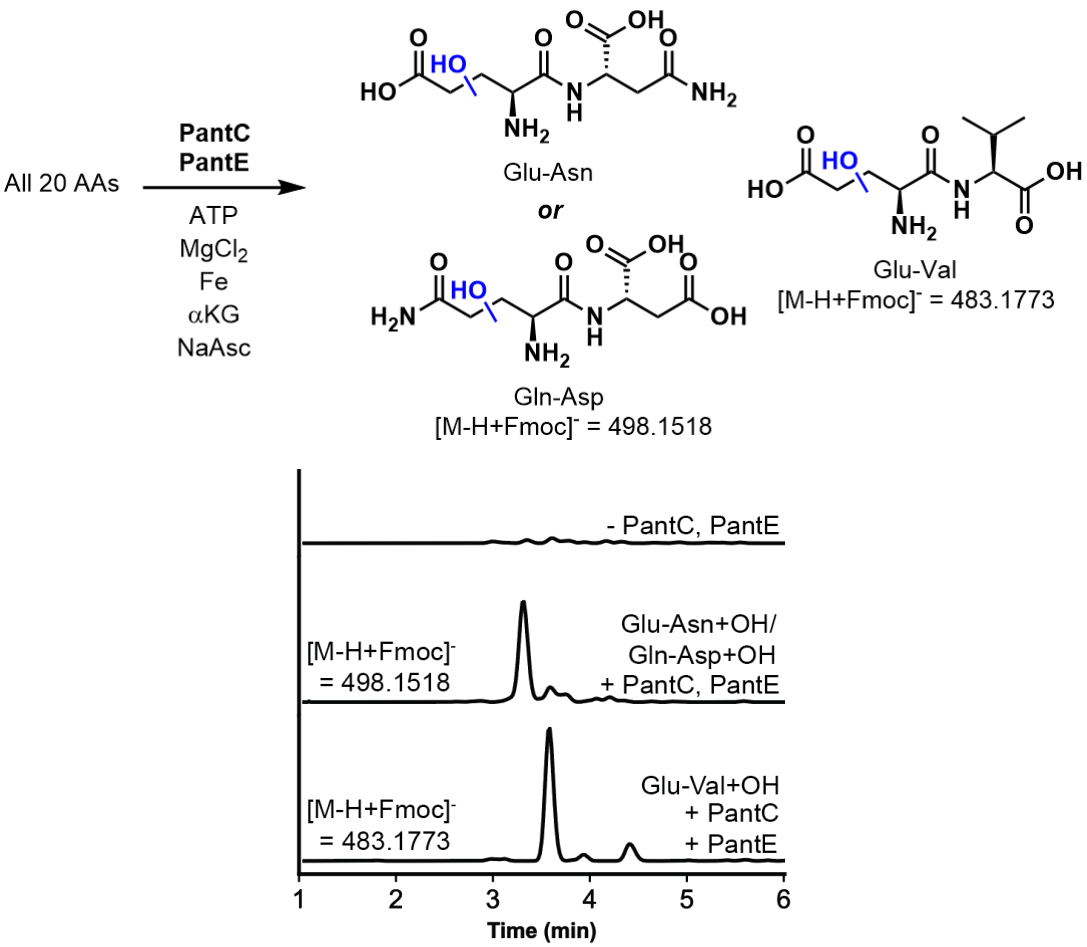


Figure 4.31. Combining PantC and PantE yields unique products. When PantC and PantE are combined in a one-pot reaction, we detected hydroxylated dipeptides, indicating both PantC and PantE are active *in vitro*. The connectivity of the peptides and the regioselectivity of oxidation have not been determined.

To parse the substrate selectivity of PantE, separate reactions were performed with Glu and Asn or Gln and Asp (**Fig. 4.32**). We only detected the dipeptide product when Glu and Asn were present, indicating this is the preferred product of the reaction. Despite these amino acid substrates being similar, this demonstrates some selectivity of PantE for glutamate. Interestingly, the NTN hydrolase encoded in this cluster, PantD, belongs to the same enzyme family as EgtC, which is a cysteine-glutamate peptidase. Altogether, these findings support glutamate as an important substrate in the synthesis of this natural product.

Performing metabolomics on the *Pant* BGC. To understand how all of these enzymes function in tandem to produce a final natural product, we performed a comparative metabolomics experiment. We expressed the enzymes PantB-F heterologously in *E. coli* in minimal media and searched for new accumulated metabolites in the BGC-expressing strain. We identified two enriched products, one being very abundant and producing several interesting fragments (**Fig. 4.33, Table 4.9, Table A2.8**). While the structures are proposed, isotope labelling experiments could provide insights into the molecular makeup of this compound. We detected another product that is more desaturated and smaller by one oxygen, albeit in far lower abundance (**Fig. 4.34**). Altogether, we identified two possible hits from the metabolomics experiment that could enable future *in vitro* reconstitution and enzyme discovery.

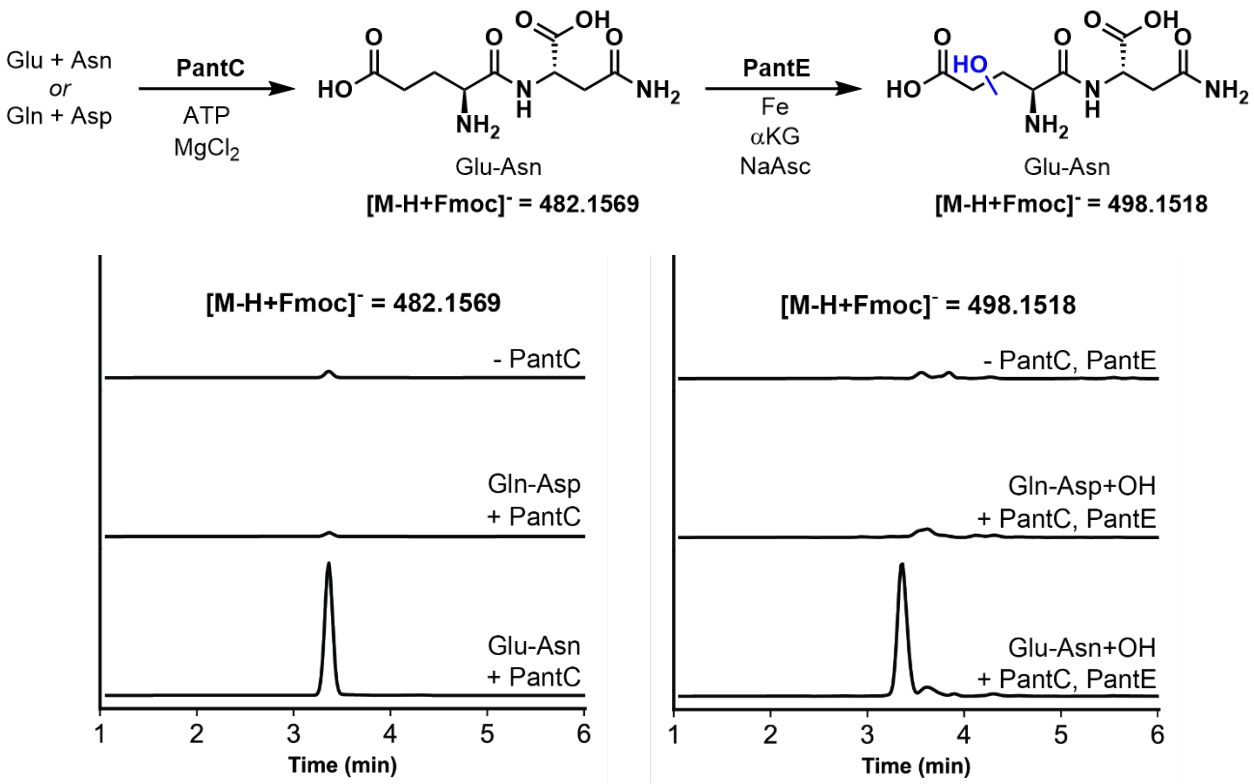


Figure 4.32. PantE hydroxylates a Glu-Asn dipeptide. Reactions performed with either Glu/Asn or Gln/Asp demonstrated that Glu and Asn are substrates for PantC. When PantE along with necessary cofactors is supplied to this reaction, hydroxylation is detected. As the hydroxylated individual amino acids were not detected, this suggests that PantE requires a peptide substrate.

RT (min)	*m/z*	Type	Chemical Formula	ANOVA *p*-value	Signal/Noise	Fold Increase
12.57	191.1021	$[M+H]^+$	$C_7H_{14}N_2O_4$	6.79E-04	6572.5	106.65
11.83	173.0921	$[M+H]^+$	$C_7H_{12}N_2O_3$	1.14E-03	648.4	68.85
12.57	62.0598	$[M+H]^+$	C_2H_7NO	1.01E-03	341.2	48.04
12.57	128.0695	$[M+H]^+$	$C_6H_9NO_2$	5.64E-04	82.9	19.99
12.57	174.0761	$[M+H]^+$	$C_7H_{11}NO_4$	1.21E-03	67	6.51

Table 4.9. Summary of results from metabolomics experiment of Pant BGC. From data analysis with MS-Dial, unique metabolites were identified, with many fragments detected at RT = 12.57 min. Formulas were generated using Agilent's Qualitative Analysis Formula Calculator and were confirmed to have an integer value for degree of unsaturation.

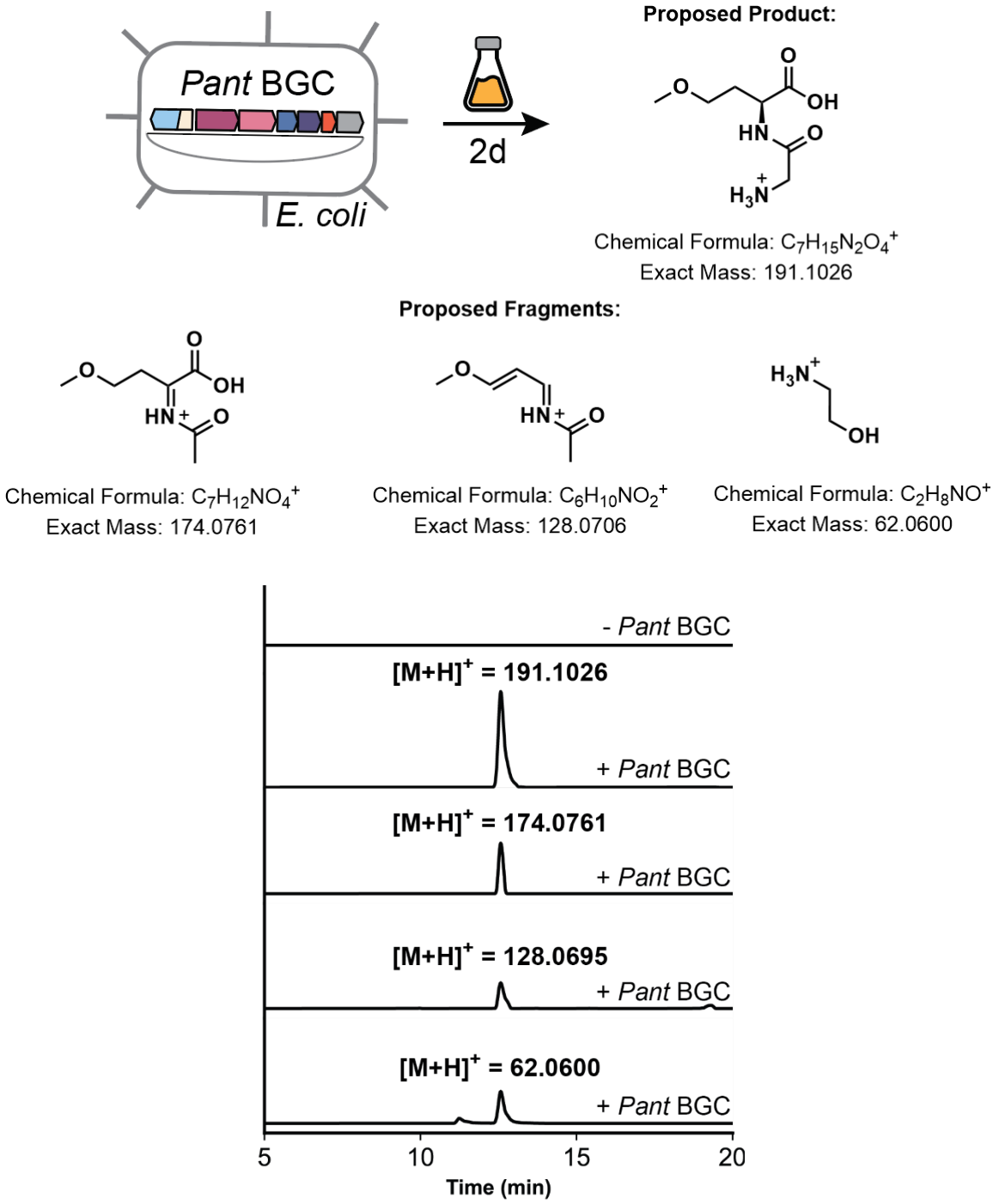


Figure 4.33. Most abundant new metabolite from Pant BGC metabolomics. Most abundant mass from the metabolomics experiment and its fragments. All structures are proposed, where isotopic feeding experiments would be helpful for determining the components of these products.

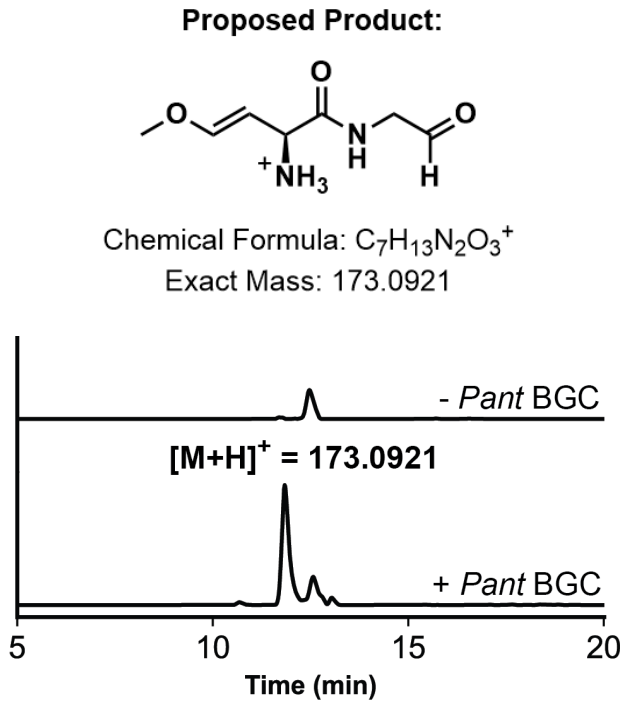


Figure 4.34. Additional hit from metabolomics of Pant BGC. An additional product identified from the comparative metabolomics experiment performed with the *Pant* BGC, with an additional desaturation and one oxygen smaller than the most abundant product detected.

Biosynthetic hypothesis for the *Pant* pathway. With putative structures drawn from the chemical formulas pulled from the metabolomics, we attempted to draw out a biosynthetic pathway (**Fig. 4.35**). Given PantC's preference for glutamate, we hypothesize this is the starting point for the biosynthetic gene cluster. Following ligation to glycine, **1** is produced. PantE performs a hydroxylation on the carbon backbone of glutamate to form **2**. PantD then performs a decarboxylation following a nucleophilic attack at the carbon hydroxylated by PantE to form **3**. PantD is a NTN-hydrolase, where a member of this family catalyzes a similar reaction [*39*]. PantB, the RRE-containing protein, then performs a methyltransferase reaction to form **4**. This pathway was inspired by a pathway with some similarities, including a Fe/αKG-dependent hydroxylation of glutamate [*40*]. As there is relatively little data on the components of this product and its connectivity, this is a very preliminary product structure and biosynthetic route. PantB seems to be capable of more than a simple methyltransferase reaction, for instance. Fortunately, isotopically-labelled amino acids are readily available for labelling studies.

From the analysis in this work, several other putative biosynthetic pathways have been identified, and could be fruitful if explored experimentally. These are shown in (**Fig. 4.36**). One example of an interesting BGC contains a Fe/αKG-dependent hydroxylase adjacent to genes for diaminopropionic acid biosynthesis (**Fig. 4.37**). As there is only one C(sp^3) carbon to hydroxylate, we hypothesize this to be a route towards a threonine analog, where the methyl group is replaced with an amine. While this does need to be explored experimentally to be sure, there are many BGCs like this one that could be explored to yield new molecules and enzymes.

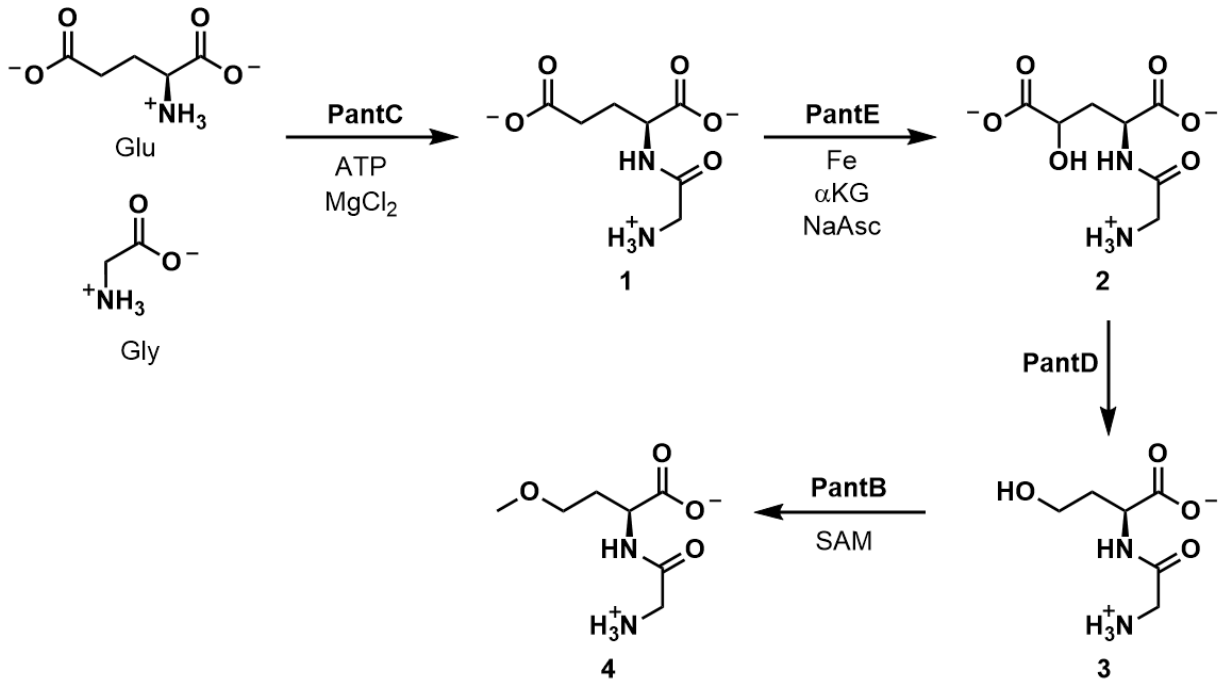


Figure 4.35. Proposed biosynthetic pathway of the Pant BGC. Proposed biosynthetic pathway towards the most abundant metabolite detected from the metabolomics experiment of the PantBGC. The structures of these compounds are preliminary.



Figure 4.36. BGCs with YbaK/INS tRDs. Additional YbaK/INS tRDs in putative secondary metabolism BGCs. From the bioinformatic analysis, there could be up to 10,000 BGCs that are unexplored.

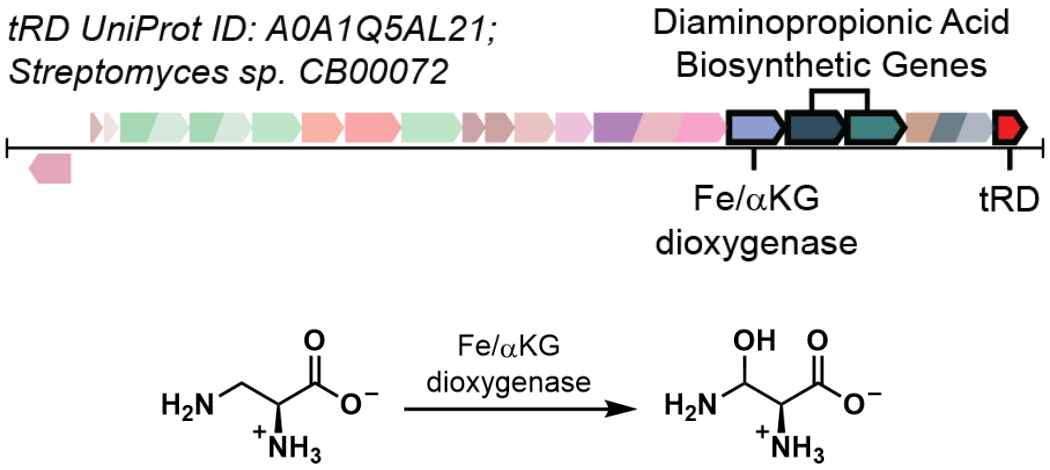
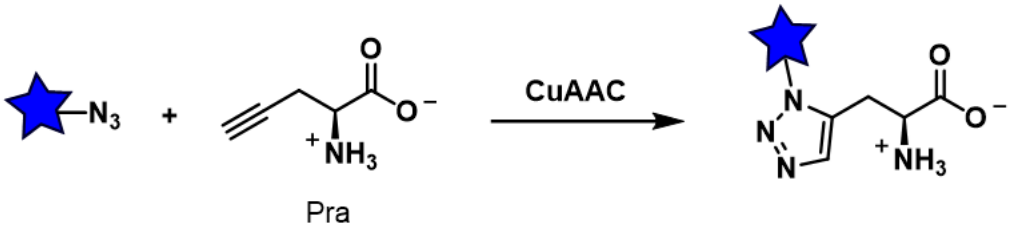


Figure 4.37. Genome mining to identify a DAP C_β hydroxylase. A putative assignment of function to a tRD-containing BGC. The presence of a Fe/αKG-dioxygenase proximate to diaminopropionic acid biosynthetic genes suggests a hydroxylation to produce an amine-containing version of threonine.

Engineering site-selective incorporation of propargylglycine. As a component to this exploration into ncAA biosynthesis, we were interested in developing an approach to biosynthesize and incorporate a ncAA site-specifically within the same host organism. This was especially exciting following the discovery of the biosynthetic pathway for propargylglycine (Pra) and β-ethynylserine (βes), which contain terminal alkynes. In contrast to βes, Pra does not incorporate into proteins on its own, leading others to engineering aaRSs to incorporate Pra in place of other amino acid residues [*41*]. This has proved useful for chemical biological applications through copper-catalyzed azide-alkyne "click" reaction (CuAAC) (**Fig. 4.38**). We aimed to advance beyond residue-specific incorporation by engineering site-specific incorporation of Pra.



Pra

Figure 4.38. Overview of CuAAC reaction with terminal alkyne of Pra. Propargylglycine (Pra) is a useful chemical biology tool by utilizing the alkyne for "click" chemistry. We explored the ability to both biosynthesize and incorporate this amino acid in a host organism.

To enable site-specific incorporation, we needed to identify an orthogonal aaRS and engineer selectivity for Pra. While orthogonal aaRSs have been established for large hydrophobic ncAAs, such as the *M. jannashchii* TyrRS or *M. mazei* PylRS, Pra would require an aaRS with a smaller active site [*42,43*]. Fortunately, others developed a chimeric system that combines the orthogonality of the PylRS system with the substrate preferences of other aaRSs (**Fig. 4.39**) [*44*]. One aaRS we were particularly interested in exploring was the chimeric alanine-tRNA synthetase (chAlaRS), as the alanine-tRNA synthetase has known promiscuity for serine, which is a similar size to Pra (**Fig. 4.40**) [*45*]. We reasoned this synthetase could have some base promiscuity towards Pra and could then be engineered to be selective towards it.

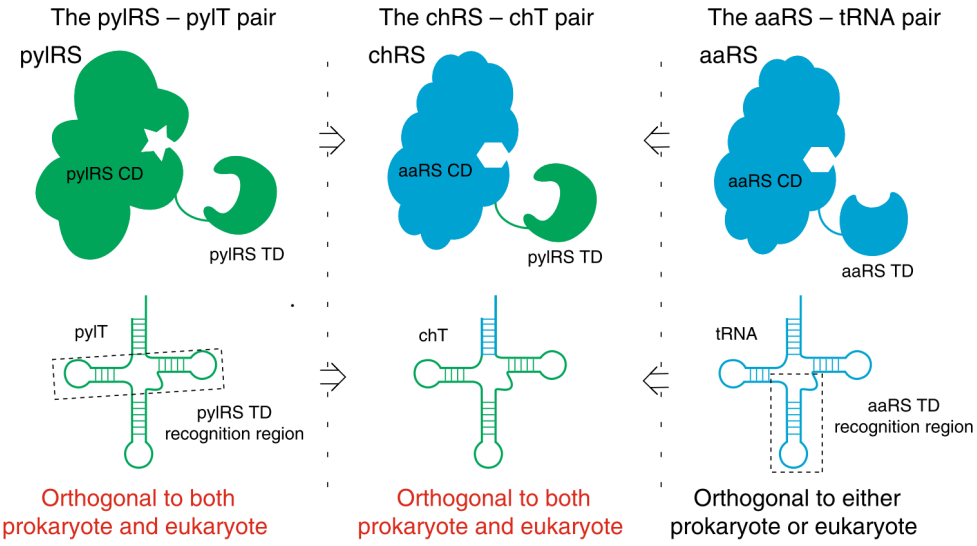


Figure 4.39. Chimeric design of orthogonal aaRSs. A chimeric aaRS containing various active sites of *E. coli* aaRS with the orthogonal tRNA binding domain of the PylRS. We hypothesized this system would be a good starting point for engineering site-selective incorporation. This figure was taken from [*44*].

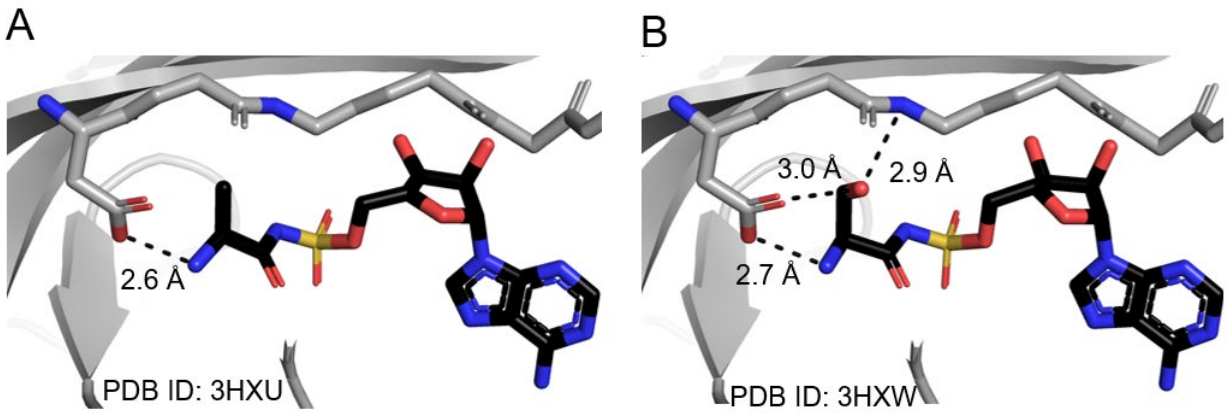


Figure 4.40. Binding pocket for Ala-adenylate in *Ec*AlaRS. Crystal structures of the *E. coli* alanine-tRNA synthetase shows a binding pocket large enough to bind not only (A) alanine but (B) serine as well [*45*]. We hypothesized the enzyme may have some promiscuity towards Pra, which is a similar size to serine.

Figure 3.8. Experimental elucidation of the Pfrm biosynthetic pathway. *In vitro* pathway reconstitution revealed a biosynthetic pathway towards α,β-functionalized phenylalanine derivatives.

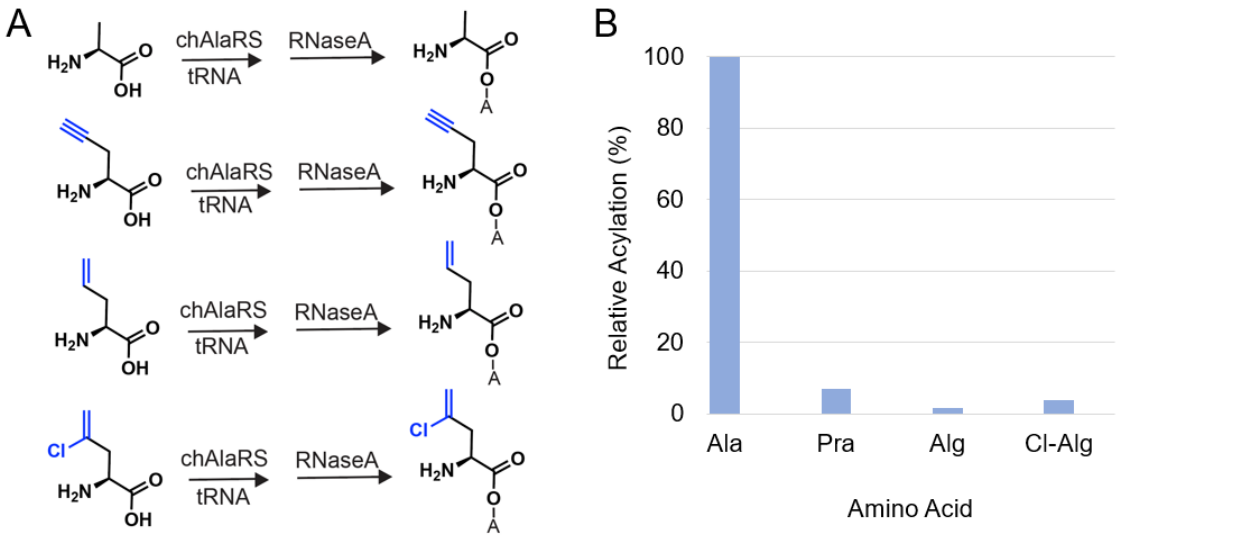


Figure 4.41. *In vitro* promiscuity of chAlaRS for "clickable" ncAAs. Purified chAlaRS was screened *in vitro* with various "clickable" amino acids with the RNase assay to assess WT enzyme promiscuity [*46*]. Of the substrates tested, Pra appeared to be the best substrate with 7% activity compared to Ala.

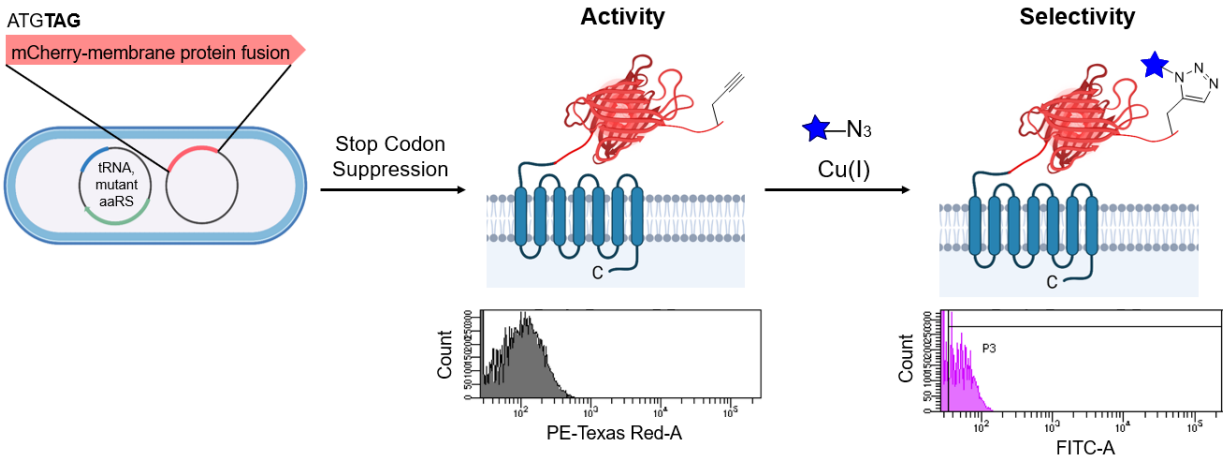


Figure 4.42. Overview of protein display method for FACS-based engineering. Protein display strategy for chAlaRS engineering combines a red fluorescent protein (RFP) mCherry for readout of enzyme variant activity with a N-terminal stop codon and copper-click to measure Pra incorporation. After expressing mCherry and the chAlaRS library in DH10B cells in LB overnight at 37°C, cells are washed and clicked with an azide-containing fluorophore. The two fluorescent signals of mCherry and the fluorophore can be read in separate channels using FACS.

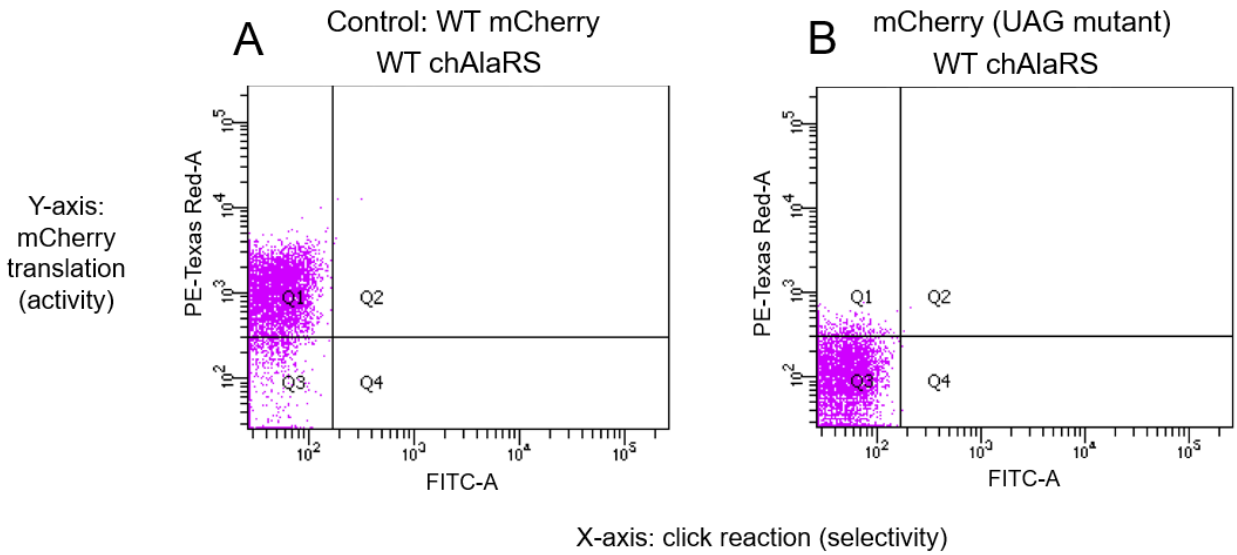


Figure 4.43. Results of FACS with WT chAlaRS. (A) Testing the protein display method with WT chAlaRS and WT mCherry shows the DH10B cells are producing detectable levels of fluorescent protein. (B) Inserting the amber codon in mCherry decreases this signal about an order of magnitude. No alkyne incorporation (FITC signal) is detected in either case. Axes are fluorescence-intensity and dots are cells. This experiment defined the quartiles for engineering.

Designing an engineering approach for chAlaRS. To examine the ability of the chAlaRS to accept Pra as a substrate, we cloned, expressed, and purified chAlaRS with a N-terminal His_{10} tag and analyzed its ability to acylate tRNA *in vitro*. We utilized the RNase assay developed in this laboratory for its high sensitivity and found chAlaRS showed 7% conversion to Pra-tRNA compared to alanine, which is the native substrate (**Fig. 4.41**). This baseline promiscuity supported that this enzyme was a good starting point for engineering. With this insight, we began developing a high-throughput method for engineering.

Whereas typical aaRS engineering leverages antibiotic resistance for selection, we hypothesized we could take advantage of the terminal alkyne to use click chemistry for engineering. Towards this goal, we built a protein display approach that includes a red-fluorescent protein (RFP) with an amber codon near the N-terminus for Pra incorporation (**Fig. 4.42**). Specifically, the RFP mCherry is fused to a N-terminal signaling peptide for export and a C-terminal translocator transmembrane protein that anchors mCherry to the surface. The amber codon variant contains the amber codon (TAG) at the second residue of mCherry, which is an unstructured region we reasoned would accommodate alternate amino acids. An chAlaRS variant's ability to read through the amber codon correlates with the amount of RFP produced, enabling us to quantify enzyme activity. With incorporation of Pra at this amber codon and subjecting the cells to a click reaction, we could quantify the selectivity of the enzyme variants towards Pra incorporation over canonical amino acids. With this system, we could efficiently screen bacteria containing variants of chAlaRS using Fluorescence-Activate Cell Sorting (FACS).

Before introducing mutations into chAlaRS, we examined the promiscuity and activity of the WT enzyme with the FACS-based assay (**Fig. 4.43**). The two fluorescent channels we monitored were PE-Texas Red-A for RFP and FITC-A for CalFluor attached by CuAAC. With the WT sequence of mCherry, we found a majority of cells were in the first quartile, indicating RFP expression and no Pra incorporation, as we would expect. With the amber codon installed, we found the amount of RFP decreased by about an order of magnitude with no change in FITC intensity. The drop in RFP produced is likely a result of competition with release factors, which could be addressed with a RF1 knockout strain or improving enzyme activity with engineering.

Mutagenesis and screening of chAlaRS variants. With the system appearing to work with WT chAlaRS, we could measure both enzyme variant activity and selectivity simultaneously at high-throughput using FACS-based sorting. Four residues around the active site, M92, W170, V214, and G237, were selected for site-saturation mutagenesis (**Fig. 4.44**). The individually most impactful site to alter appeared to be W170, which increased the FITC signal about 3-fold, indicating Pra may have been incorporated (**Fig. 4.45**). Site-saturation at both V214 and M92 increased fluoresence as well. Cells in Quadrant 2 from these experiments were sorted by FACS. After sequencing these cells, we mapped the mutations onto the active site of chAlaRS to rationalize their effect on amino acid selectivity.

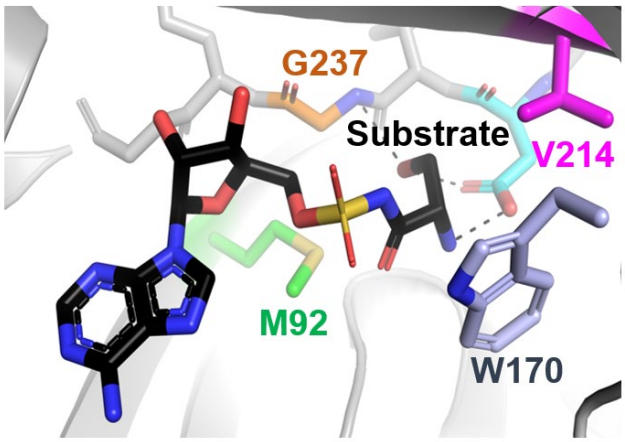


Figure 4.44. Engineering approach for chAlaRS. chAlaRS active site with residues labelled that were subject to site-saturation mutagenesis.

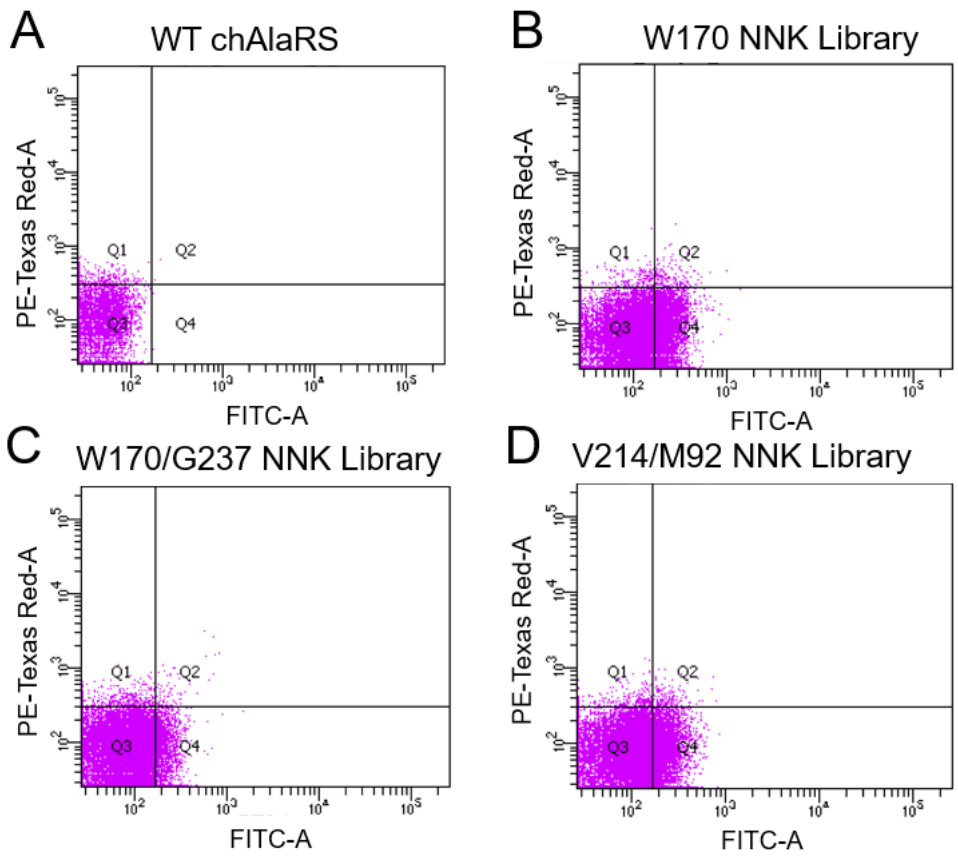


Figure 4.45. Results from engineering of chAlaRS. FACS readouts from site-saturation from several positions within chAlaRS appeared to increase activity towards Pra, as monitored by an increase in FITC signal compared to WT. W70 appears to be the most impactful single mutation, and other combinations also improved Pra incorporation.

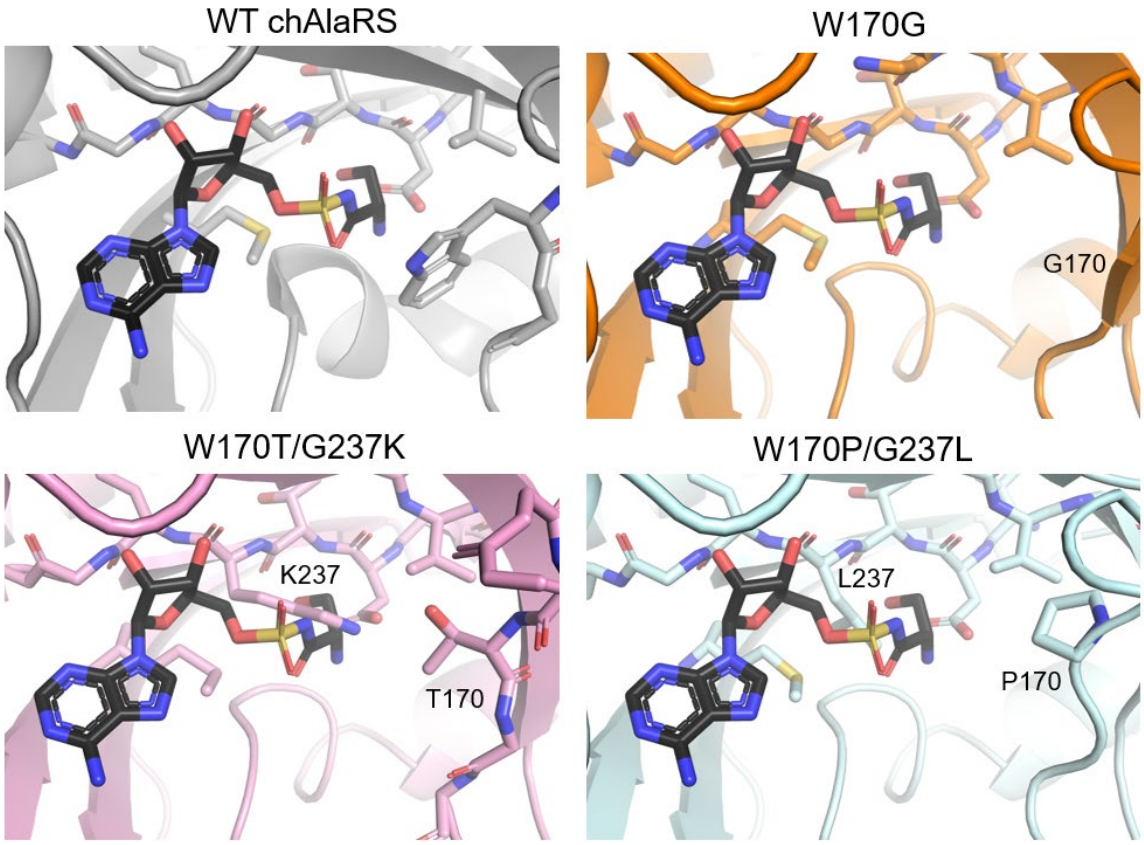


Figure 4.46. chAlaRS variants sorted with FACS. Sequences of the top-performing variants from FACS experiments show much smaller residues in place of W170 and larger residues in place of G237.

Examining the different variants revealed drastic changes in the active site of chAlaRS (**Fig. 4.46**). Trp170 was converted to a much smaller amino acid, such as glycine or proline. Glycine was converted to a much larger amino acid, such as lysine or leucine. These would have been difficult to predict by rational engineering, showing the utility of this FACS-based approach. To confirm that these mutations are beneficial, these mutants would need to be tested *in vitro* with the RNase assay we developed as well as with a cytosolic reporter protein that could be purified and its exact mass be measured. As only one round of engineering was performed, additional engineering is likely needed to improve the activity and selectivity of chAlaRS for Pra.

4.4 Conclusions

By employing tRNA-deacylases as a marker gene of amino acid modifications, we could gain insights into the functions of poorly annotated and completely unannotated enzymes. This is achievable through elucidating the function of neighboring genes and with the fundamental insight that amino acids are the target substrate for these transformations. Following the discovery of PhalD, we anticipate this enzyme can be leveraged for the biocatalytic synthesis of diazeniumdiolate-containing natural products. Further experiments are needed to explore the role of co-clustered cupin proteins potentially capable of reducing nitrite. With the discovery of a new protein PantB, we potentially identified a new form of biosynthetic logic, where RiPP-like enzymes are used on a nontemplated peptide substrate. Isotope-feeding experiments are essential for gaining insights into the components of this new natural product.

In a separate effort, we began engineering an aminoacyl-tRNA synthetase for the selective incorporation of Pra, a terminal alkyne amino acid. By leveraging a fluorescent protein reporter along with the clickable terminal alkyne, we utilized FACS to identify variants with improved activity towards Pra. In combination with the improved biosynthetic pathway for Pra, we are en route to developing a system that enables incorporation and biosynthesis in the same organism. Further rounds are needed to improve the catalytic efficiency of the synthetase, as the *in vivo* concentration from the biosynthetic pathway will be much lower than what can be supplied exogenously.

4.5 References

1. Blin, K.; Shaw, S.; Kloosterman, A. M.; Charlop-Powers, Z.; van Wezel, G. P.; Medema, M. H.; Weber, T. antiSMASH 6.0: Improving Cluster Detection and Comparison Capabilities. *Nucleic Acids Res.* **2021**, *49* (W1), W29–W35. https://doi.org/10.1093/nar/gkab335.
2. Soskine, M.; Tawfik, D. S. Mutational Effects and the Evolution of New Protein Functions. *Nat. Rev. Genet.* **2010**, *11* (8), 572–582. https://doi.org/10.1038/nrg2808.
3. Andersson, D. I.; Jerlström-Hultqvist, J.; Näsvall, J. Evolution of New Functions De Novo and from Preexisting Genes. *Cold Spring Harb. Perspect. Biol.* **2015**, *7* (6), a017996. https://doi.org/10.1101/cshperspect.a017996.
4. Jensen, R. A. Enzyme Recruitment in Evolution of New Function. *Annu. Rev. Microbiol.* **1976**, *30* (1), 409–425. https://doi.org/10.1146/annurev.mi.30.100176.002205.
5. Goodacre, N. F.; Gerloff, D. L.; Uetz, P. Protein Domains of Unknown Function Are Essential in Bacteria. *mBio* **2013**, *5* (1), e00744-13. https://doi.org/10.1128/mBio.00744-13.
6. Bateman, A.; Coggill, P.; Finn, R. D. DUFs: Families in Search of Function. *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.* **2010**, *66* (Pt 10), 1148–1152. https://doi.org/10.1107/S1744309110001685.
7. Yee, D. A.; Niwa, K.; Perlatti, B.; Chen, M.; Li, Y.; Tang, Y. Genome Mining for Unknown–Unknown Natural Products. *Nat. Chem. Biol.* **2023**, *19* (5), 633–640. https://doi.org/10.1038/s41589-022-01246-6.
8. Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat. Methods* **2009**, *6* (5), 343–345. https://doi.org/10.1038/nmeth.1318.
9. Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R. D.; Bairoch, A. ExPASy: The Proteomics Server for in-Depth Protein Knowledge and Analysis. *Nucleic Acids Res.* **2003**, *31* (13), 3784–3788. https://doi.org/10.1093/nar/gkg563.
10. Zallot, R.; Oberg, N.; Gerlt, J. A. The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry* **2019**, *58* (41), 4169–4182. https://doi.org/10.1021/acs.biochem.9b00735.
11. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13* (11), 2498–2504. https://doi.org/10.1101/gr.1239303.
12. Neidhardt, F. C.; Bloch, P. L.; Smith, D. F. Culture Medium for Enterobacteria. *J. Bacteriol.* **1974**, *119* (3), 736–747.
13. Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M. MS-DIAL: Data-Independent MS/MS Deconvolution for Comprehensive Metabolome Analysis. *Nat. Methods* **2015**, *12* (6), 523–526. https://doi.org/10.1038/nmeth.3393.
14. Luo, L.; Pappalardi, M. B.; Tummino, P. J.; Copeland, R. A.; Fraser, M. E.; Grzyska, P. K.; Hausinger, R. P. An Assay for Fe(II)/2-Oxoglutarate-Dependent Dioxygenases by Enzyme-

Coupled Detection of Succinate Formation. *Anal. Biochem.* **2006**, *353* (1), 69–74. https://doi.org/10.1016/j.ab.2006.03.033.

15. The UniProt Consortium. UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **2023**, *51* (D1), D523–D531. https://doi.org/10.1093/nar/gkac1052.
16. Triana-Alonso, F. J.; Dabrowski, M.; Wadzack, J.; Nierhaus, K. H. Self-Coded 3′-Extension of Run-off Transcripts Produces Aberrant Products during in Vitro Transcription with T7 RNA Polymerase (∗). *J. Biol. Chem.* **1995**, *270* (11), 6298–6307. https://doi.org/10.1074/jbc.270.11.6298.
17. Goto, Y.; Katoh, T.; Suga, H. Flexizymes for Genetic Code Reprogramming. *Nat. Protoc.* **2011**, *6* (6), 779–790. https://doi.org/10.1038/nprot.2011.331.
18. Cha, L.; Paris, J. C.; Zanella, B.; Spletzer, M.; Yao, A.; Guo, Y.; Chang, W. Mechanistic Studies of Aziridine Formation Catalyzed by Mononuclear Non-Heme Iron Enzymes. *J. Am. Chem. Soc.* **2023**, *145* (11), 6240–6246. https://doi.org/10.1021/jacs.2c12664.
19. Roskoski, R. [1] Assays of Protein Kinase. In *Methods in Enzymology*; Hormone Action Part F: Protein Kinases; Academic Press, 1983; Vol. 99, pp 3–6. https://doi.org/10.1016/0076-6879(83)99034-1.
20. Franke, J.; Ishida, K.; Ishida-Ito, M.; Hertweck, C. Nitro versus Hydroxamate in Siderophores of Pathogenic Bacteria: Effect of Missing Hydroxylamine Protection in Malleobactin Biosynthesis. *Angew. Chem. Int. Ed.* **2013**, *52* (32), 8271–8275. https://doi.org/10.1002/anie.201303196.
21. Jenul, C.; Sieber, S.; Daeppen, C.; Mathew, A.; Lardi, M.; Pessi, G.; Hoepfner, D.; Neuburger, M.; Linden, A.; Gademann, K.; Eberl, L. Biosynthesis of Fragin Is Controlled by a Novel Quorum Sensing Signal. *Nat. Commun.* **2018**, *9* (1), 1297. https://doi.org/10.1038/s41467-018-03690-2.
22. Ng, T. L.; McCallum, M. E.; Zheng, C. R.; Wang, J. X.; Wu, K. J. Y.; Balskus, E. P. The L-Alanosine Gene Cluster Encodes a Pathway for Diazeniumdiolate Biosynthesis. *Chembiochem Eur. J. Chem. Biol.* **2020**, *21* (8), 1155–1160. https://doi.org/10.1002/cbic.201900565.
23. Wang, M.; Niikura, H.; He, H.-Y.; Daniel-Ivad, P.; Ryan, K. S. Biosynthesis of the N–N-Bond-Containing Compound L-Alanosine. *Angew. Chem. Int. Ed.* **2020**, *59* (10), 3881–3885. https://doi.org/10.1002/anie.201913458.
24. Batova, A.; Diccianni, M. B.; Omura-Minamisawa, M.; Yu, J.; Carrera, C. J.; Bridgeman, L. J.; Kung, F. H.; Pullen, J.; Amylon, M. D.; Yu, A. L. Use of Alanosine as a Methylthioadenosine Phosphorylase-Selective Therapy for T-Cell Acute Lymphoblastic Leukemia in Vitro. *Cancer Res.* **1999**, *59* (7), 1492–1497.
25. Morgan, G. L.; Li, B. In Vitro Reconstitution Reveals a Central Role for the N-Oxygenase PvfB in (Dihydro)Pyrazine-N-Oxide and Valdiazen Biosynthesis. *Angew. Chem. Int. Ed Engl.* **2020**, *59* (48), 21387–21391. https://doi.org/10.1002/anie.202005554.
26. Wang, M.; Ryan, K. S. Reductases Produce Nitric Oxide in an Alternative Pathway to Form the Diazeniumdiolate Group of L-Alanosine. *J. Am. Chem. Soc.* **2023**, *145* (30), 16718–16725. https://doi.org/10.1021/jacs.3c04447.
27. Du, J.; Filipović, M. R.; Wagner, B. A.; Buettner, G. R. Ascorbate Mediates the Non-Enzymatic Reduction of Nitrite to Nitric Oxide. *Adv. Redox Res.* **2023**, *9*, 100079. https://doi.org/10.1016/j.arres.2023.100079.

28. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

29. Horrell, S.; Kekilli, D.; W. Strange, R.; A. Hough, M. Recent Structural Insights into the Function of Copper Nitrite Reductases. *Metallomics* **2017**, *9* (11), 1470–1482. https://doi.org/10.1039/C7MT00146K.

30. Einsle, O.; Messerschmidt, A.; Stach, P.; Bourenkov, G. P.; Bartunik, H. D.; Huber, R.; Kroneck, P. M. Structure of Cytochrome c Nitrite Reductase. *Nature* **1999**, *400* (6743), 476–480. https://doi.org/10.1038/22802.

31. Zhao, G.; Peng, W.; Song, K.; Shi, J.; Lu, X.; Wang, B.; Du, Y.-L. Molecular Basis of Enzymatic Nitrogen-Nitrogen Formation by a Family of Zinc-Binding Cupin Enzymes. *Nat. Commun.* **2021**, *12* (1), 7205. https://doi.org/10.1038/s41467-021-27523-x.

32. Keefer, L. K. Fifty Years of Diazeniumdiolate Research. From Laboratory Curiosity to Broad-Spectrum Biomedical Advances. *ACS Chem. Biol.* **2011**, *6* (11), 1147–1155. https://doi.org/10.1021/cb200274r.

33. Matsuda, K.; Arima, K.; Akiyama, S.; Yamada, Y.; Abe, Y.; Suenaga, H.; Hashimoto, J.; Shin-ya, K.; Nishiyama, M.; Wakimoto, T. A Natural Dihydropyridazinone Scaffold Generated from a Unique Substrate for a Hydrazine-Forming Enzyme. *J. Am. Chem. Soc.* **2022**, *144* (28), 12954–12960. https://doi.org/10.1021/jacs.2c05269.

34. Kuzmishin Nagy, A. B.; Bakhtina, M.; Musier-Forsyth, K. Chapter Four - Trans-Editing by Aminoacyl-tRNA Synthetase-like Editing Domains. In *The Enzymes*; Ribas de Pouplana, L., Kaguni, L. S., Eds.; Biology of Aminoacyl-tRNA Synthetases; Academic Press, 2020; Vol. 48, pp 69–115. https://doi.org/10.1016/bs.enz.2020.07.002.

35. Nguyen, D. T.; Mitchell, D. A.; van der Donk, W. A. Genome Mining for New Enzyme Chemistry. *ACS Catal.* **2024**, *14* (7), 4536–4553. https://doi.org/10.1021/acscatal.3c06322.

36. Burkhart, B. J.; Hudson, G. A.; Dunbar, K. L.; Mitchell, D. A. A Prevalent Peptide-Binding Domain Guides Ribosomal Natural Product Biosynthesis. *Nat. Chem. Biol.* **2015**, *11* (8), 564–570. https://doi.org/10.1038/nchembio.1856.

37. Kretsch, A. M.; Gadgil, M. G.; DiCaprio, A. J.; Barrett, S. E.; Kille, B. L.; Si, Y.; Zhu, L.; Mitchell, D. A. Peptidase Activation by a Leader Peptide-Bound RiPP Recognition Element. *Biochemistry* **2023**, *62* (4), 956–967. https://doi.org/10.1021/acs.biochem.2c00700.

38. Kloosterman, A. M.; Shelton, K. E.; van Wezel, G. P.; Medema, M. H.; Mitchell, D. A. RRE-Finder: A Genome-Mining Tool for Class-Independent RiPP Discovery. *mSystems* **2020**, *5* (5), 10.1128/msystems.00267-20. https://doi.org/10.1128/msystems.00267-20.

39. Zhong, G.; Zhao, Q.; Zhang, Q.; Liu, W. 4-Alkyl-L-(Dehydro)Proline Biosynthesis in Actinobacteria Involves N-Terminal Nucleophile-Hydrolase Activity of γ-Glutamyltranspeptidase Homolog for C-C Bond Cleavage. *Nat. Commun.* **2017**, *8*, 16109. https://doi.org/10.1038/ncomms16109.

40. Patteson, J. B.; Dunn, Z. D.; Li, B. In Vitro Biosynthesis of the Nonproteinogenic Amino Acid Methoxyvinylglycine. *Angew. Chem. Int. Ed.* **2018**, *57* (23), 6780–6785. https://doi.org/10.1002/anie.201713419.

41. Truong, F.; Yoo, T. H.; Lampo, T. J.; Tirrell, D. A. Two-Strain, Cell-Selective Protein Labeling in Mixed Bacterial Cultures. *J. Am. Chem. Soc.* **2012**, *134* (20), 8551–8556. https://doi.org/10.1021/ja3004667.

42. Wang, L.; Schultz, P. G. A General Approach for the Generation of Orthogonal tRNAs. *Chem. Biol.* **2001**, *8* (9), 883–890. https://doi.org/10.1016/s1074-5521(01)00063-1.

43. Wan, W.; Tharp, J. M.; Liu, W. R. Pyrrolysyl-tRNA Synthetase: An Ordinary Enzyme but an Outstanding Genetic Code Expansion Tool. *Biochim. Biophys. Acta* **2014**, *1844* (6), 1059–1070. https://doi.org/10.1016/j.bbapap.2014.03.002.

44. Ding, W.; Zhao, H.; Chen, Y.; Zhang, B.; Yang, Y.; Zang, J.; Wu, J.; Lin, S. Chimeric Design of Pyrrolysyl-tRNA Synthetase/tRNA Pairs and Canonical Synthetase/tRNA Pairs for Genetic Code Expansion. *Nat. Commun.* **2020**, *11* (1), 3154. https://doi.org/10.1038/s41467-020-16898-y.

45. Guo, M.; Chong, Y. E.; Shapiro, R.; Beebe, K.; Yang, X.-L.; Schimmel, P. Paradox of Mistranslation of Serine for Alanine Caused by AlaRS Recognition Dilemma. *Nature* **2009**, *462* (7274), 808–812. https://doi.org/10.1038/nature08612.

46. McMurry, J. L.; Chang, M. C. Y. Fluorothreonyl-tRNA Deacylase Prevents Mistranslation in the Organofluorine Producer *Streptomyces Cattleya*. *Proc. Natl. Acad. Sci.* **2017**, *114* (45), 11920–11925. https://doi.org/10.1073/pnas.1711482114.

47. Thuronyi, B. W.; Privalsky, T. M.; Chang, M. C. Y. Engineered Fluorine Metabolism and Fluoropolymer Production in Living Cells. *Angew. Chem. Int. Ed.* **2017**, *56* (44), 13637–13640. https://doi.org/10.1002/anie.201706696.

48. Ko, H.-J.; Park, E.; Song, J.; Yang, T. H.; Lee, H. J.; Kim, K. H.; Choi, I.-G. Functional Cell Surface Display and Controlled Secretion of Diverse Agarolytic Enzymes by Escherichia Coli with a Novel Ligation-Independent Cloning Vector Based on the Autotransporter YfaL. *Appl. Environ. Microbiol.* **2012**, *78* (9), 3051–3058. https://doi.org/10.1128/AEM.07004-11.

Appendix I: *Plasmids, Oligonucleotides, and Gene Sequences*

Table A1.1 Strains, plasmids, oligonucleotides, and gene sequences for Chapter 2. (A) Strains and plasmids. (B) Oligonucleotides used for plasmid construction and sequencing (for T7 tRNAs, **bold** indicates tRNA sequence and non-bold indicates T7 promoter). (C) Sequences of synthetic genes.

A1.1A. Strains and plasmids

Strain	Description	Source
E. coli DH10b-T1^R	*F-mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara-leu)7697 galU galK λrpsL(StrR) nupG tonA*	Thermo-Fisher
E. coli BL21 Star (DE3)	*F-ompT hsdSB (rB-, mB-) gal dcm rne131* (DE3)	Thermo-Fisher
E. coli GM272	*F- dam-3 dcm-6 hsdS2l metb1 lacYl or Z4 galK2 ga/T22 mtl-2 tonA2or A31 tsx-1 or -78 supE44 (thi-1)*	ATCC
S. cattleya	NRRL 8057 (ATCC 35852)	ATCC
S. cattleya ΔbesG	ΔbesG::Am^R-OriT	This study
Plasmid	**Description**	**Source**
pIJ773	*lacI*, Cb^R, Am^R, OriT, ColE1	JIC
pIJ773-BesG	besG flanking regions (2 kB US/DS)-Am^R-OriT, *lacI*, Cb^R, ColE1	This study
pUZ8002	tra, *neo*, Km^R, RP4	*1*
pET16b-His$_{10}$-Ecoli-ThrRS	His$_{10}$-ThrRS (from E. coli, T7), *lacI*, Cb^R, ColE1	This study
pET16b-His$_{10}$-Scat-ThrRS	His$_{10}$-ThrRS (from S. cattleya, T7), *lacI*, Cb^R, ColE1	This study
pSV272.1-His$_6$-MBP-BesG	His$_6$-MBP-BesG (T7), *lacI*, Km^R, f1, ColE1	This study
PepC-METRSKNFL	METRSKNFL peptide (T7), Cb^R, ColE1	This study
pET16b-His$_{10}$-FusA	His$_{10}$-FusA (T7), *lacI*, Cb^R, ColE1	This study
pET16b-His$_{10}$-FusB	His$_{10}$-FusB (T7), *lacI*, Cb^R, ColE1	This study
pET16b-His$_{10}$-FusC	His$_{10}$-FusC (T7), *lacI*, Cb^R, ColE1	This study
pET16b-His$_{10}$-FusD	His$_{10}$-FusD (T7), *lacI*, Cb^R, ColE1	This study
pET16b-His$_{10}$-FusE	His$_{10}$-FusE (T7), *lacI*, Cb^R, ColE1	This study
pET16b-His$_{10}$-FusF	His$_{10}$-FusF (T7), *lacI*, Cb^R, ColE1	This study
pET16b-His$_{10}$-FusG	His$_{10}$-FusG (T7), *lacI*, Cb^R, ColE1	This study
pETDuet-1	*lacI*, Cb^R, ColE1	Novagen
pETDuet-FusA-FusC	FusA (T7) and FusC (T7), *lacI*, Cb^R, ColE1	This study
pACYCDuet-1	*lacI*, Cm^R, p15a	Novagen
pACYCDuet-FusD-FusF	FusD (T7) and FusF (T7), *lacI*, Cm^R, p15a	This study
pACYCDuet-FusD-FusE-FusF	FusD (T7), FusE (RBS), and FusF (T7), *lacI*, Cm^R, p15a	This study
pRSFDuet-1	*lacI*, Km^R, pRSF	Novagen
pRSFDuet-FusB-FusG	FusB (T7) and FusG (T7), *lacI*, Km^R, pRSF	This study

A1.1B. Oligonucleotide sequences

Name	Sequence
psv272.1-ScatBesG-Fwd	GGAAAACCTGTATTTTCAGGGCATGACCCAGCAGCGGCAGATCTATCTGACG
psv272.1-ScatBesG-Rev	TCGACGGAGCTCGAATTTCAGCCGTCCACCGGGACGTGTTCGGCGGTGTAACCCA CCTTCAGCCGG
PepC-Fwd	ATGGAAACCCGTTCTAAAAACTTTCTGTAATGAGGATCCCGGGAATTCTCGAGTA
PepC-Rev	GTTTTTAGAACGGGTTTCCATATGTATATCTCCTTCTTAAAGTTAAACAAAATTA TTTCTAGACCCTATAGTG

pIJ773-BesG-A-Fwd	CCGCGGTGGCGGCCGCCGGGGCACCGGCCCCACCCCGCT
pIJ773-BesG-A-Rev	CCGGGGGATCCACTAGTCCGGAGGCCGCCTCACCGGGGCG
pIJ773-BesG-B-Fwd	CCCGGAATATCAAGCTTAGAGTTCACCCCAAAAGTCGCGTCACGCG
pIJ773-BesG-B-Rev	GAACAAAAGCTGGTACCAGACGCTGCTGCGGACCCGGTCCG
SeqConfirm-BesG-KO-Fwd	CAGGTTACATCGGTGCAGATGTATCTG
SeqConfirm-BesG-KO-Rev	GAACAACTGCCATAATGCGCACG
pET16b-ThrRS-Ecoli-F	GGCCATCTAGAAGTGCTTTTTCAGGGCCCGCATATGCCTGTTATAACTCTTCCTG ATGGC
pET16b-ThrRS-Ecoli-R	CAGCTTCCTTTCGGGCTTTGTTAGCAGCCGGATCCTTATTCCTCCAATTGTTTAA GACTGCGG
pET16b-ThrRS-Scat-F	GGCCATCTAGAAGTGCTTTTTCAGGGCCCGCATATGTCTCAGATCCGTGTGGTCA TCA
pET16b-ThrRS-Scat-R	CAGCTTCCTTTCGGGCTTTGTTAGCAGCCGGATCCTCACACCTGGACGCGGCGCT C
ThrtRNA-Fwd	CCCCGAAAAGTGCTAGTGGTGC
ThrtRNA-Scat-UCCA-Rev	TGGAGCCCCAATACGGAATCGAAC
ThrtRNA-Ecoli-ACCA-Rev	TGGTGCTGATAGGCAGATTCGAACT
ThrtRNA-Ecoli-UCCA-Rev	TGGAGCTGATAGGCAGATTCGAACT
T7-ThrtRNA-Scat	AAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCTAGTGGTGCT AGCCCCGCGAAATTAATACGACTCACTATAG**CCCCAATAGCTCAGTCGGCAGAGC GTCTCCATGGTAAGGAGAAGGTCTACGGTTCGATTCCGTATTGGGGCTCCA**
T7-ThrtRNA-Ecoli-ACCA	AAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCTAGTGGTGCT AGCCCCGCGAAATTAATACGACTCACTATA**GCTGATATAGCTCAGTTGGTAGAGC GCACCCTTGGTAAGGGTGAGGTCGGCAGTTCGAATCTGCCTATCAGCACCA**
T7-ThrtRNA-Ecoli-UCCA	AAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCTAGTGGTGCT AGCCCCGCGAAATTAATACGACTCACTATA**GCTGATATAGCTCAGTTGGTAGAGC GCACCCTTGGTAAGGGTGAGGTCGGCAGTTCGAATCTGCCTATCAGCTCCA**
pET16b-AlfC-Fwd	TCTAGAAGTGCTTTTTCAGGGCCCGCATATGTACCTTGAAACTGAGCGC
pET16b-AlfC-Rev	TCCTTTCGGGCTTTGTTAGCAGCCGGATCCTTAGGACCAGAAGGTTAAAGGC
pET16b-AlfB-Fwd	TCTAGAAGTGCTTTTTCAGGGCCCGCATATGCAGGACAAGTTATATCTGACC
pET16b-AlfB-Rev	TCCTTTCGGGCTTTGTTAGCAGCCGGATCCTTAGTCCATCATCTCATAGTGGACA
pET16b-Af-MetRS-Opt-Fwd	TCTAGAAGTGCTTTTTCAGGGCCCGCATATGTCACGCACTATCTTCGTAACTACG G
pET16b-Af-MetRS-Opt-Rev	TCCTTTCGGGCTTTGTTAGCAGCCGGATCCTTAATTAATGCGCATCCCAGGCTGC
Af-MettRNA-Fwd	AATTCCTGCAGTAATACGACTCACTATA**GCGCGGGGTGGAGCAGCATGGTAGCTC GTCGG**
Af-MettRNA-Rev	**TmGGTTGCGGGGGCAGGATTTGAACCTACGACCTTCGGGTTATGAGCCCGACGAG CTACCATG**
pET16b-FusA-Fwd	TCTAGAAGTGCTTTTTCAGGGCCCGCATATGGTAATTGGCATTGATCTTGTCC
pET16b-FusA-Rev	TCCTTTCGGGCTTTGTTAGCAGCCGGATCCTTAAACTGCAAAAGCCGTGTCG
pET16b-FusB-Fwd	GTGCTTTTTCAGGGCCCGCATATGTTTGATCCGGTTGGTTCG
pET16b-FusB-Rev	TCGGGCTTTGTTAGCAGCCGGATCCTTACGGCAGGACAGCCAGATC
pET16b-FusC-Fwd	TCTAGAAGTGCTTTTTCAGGGCCCGCATATGCCTTTCGCAGCTCG
pET16b-FusC-Rev	TCCTTTCGGGCTTTGTTAGCAGCCGGATCCTTAGGTGTTGTAACGGCGGTATC
pET16b-FusD-Fwd	TCTAGAAGTGCTTTTTCAGGGCCCGCATATGTCGCTGCCGCGT
pET16b-FusD-Rev	TCCTTTCGGGCTTTGTTAGCAGCCGGATCCTTAACCCGTCGTACCCCAC
pET16b-FusE-Fwd	GTGCTTTTTCAGGGCCCGCATATGGGCCACGACCGTATG
pET16b-FusE-Rev	TCGGGCTTTGTTAGCAGCCGGATCCTTAACTGCGTCCGCTTCCC
pET16b-FusF-Fwd	TCTAGAAGTGCTTTTTCAGGGCCCGCATATGAGCACGCGCTCAAC

pET16b-FusF-Rev	TCCTTTCGGGCTTTGTTAGCAGCCGGATCCTTAGGCAAGAGCATAACGGGAG
pET16b-FusG-Fwd	GTGCTTTTTCAGGGCCCGCATATGCCCGACCGTCCC
pET16b-FusG-Rev	TCGGGCTTTGTTAGCAGCCGGATCCTTAAGGAGTATGTCCGGCAGG

A1.1C. Gene sequences

NCBI Accession: AEW98792.1; Uniprot ID: F8JJ98 (*Streptomyces cattleya* βes-tRNA^{Thr} Deacylase **BesG**, native sequence (capitalized) with N-terminal MBP tag (lower-case))

atgaaaatcgaagaaggtaaactggtaatctggattaacggcgataaaggctataacggtctcgctgaagtcggtaa
gaaattcgagaaagataccggaattaaagtcaccgttgagcatccggataaactggaagagaaattcccacaggttg
cggcaactggcgatggccctgacattatcttctgggcacacgaccgctttggtggctacgctcaatctggcctgttg
gctgaaatcaccccggacaaagcgttccaggacaagctgtatccgtttacctgggatgccgtacgttacaacggcaa
gctgattgcttacccgatcgctgttgaagcgttatcgctgatttataacaaagatctgctgccgaacccgccaaaaa
cctgggaagagatcccggcgctggataaagaactgaaagcgaaaggtaagagcgcgctgatgttcaacctgcaagaa
ccgtacttcacctggccgctgattgctgctgacgggggttatgcgttcaagtatgaaaacggcaagtacgacattaa
agacgtgggcgtggataacgctggcgcgaaagcgggtctgaccttcctggttgacctgattaaaaacaaacacatga
atgcagacaccgattactccatcgcagaagctgcctttaataaaggcgaaacagcgatgaccatcaacggcccgtgg
gcatggtccaacatcgacaccagcaaagtgaattatggtgtaacggtactgccgaccttcaagggtcaaccatccaa
accgttcgttggcgtgctgagcgcaggtattaacgccgccagtccgaacaaagagctggcaaaagagttcctcgaaa
actatctgctgactgatgaaggtctggaagcggttaataaagacaaaccgctgggtgccgtagcgctgaagtcttac
gaggaagagttggcgaaagatccacgtattgccgccactatggaaaacgcccagaaaggtgaaatcatgccgaacat
cccgcagatgtccgctttctggtatgccgtgcgtactgcggtgatcaacgccgccagcggtcgtcagactgtcgatg
aagccctgaaagacgcgcagactaattcgagctcgaacaacaacaacaataacaataacaacaacctcgggatcgag
gaaaacctgtattttcagggcATGACCCAGCAGCGGCAGATCTATCTGACGGACACCTACGCCTACGAGGCGTCGAC
GCGGGTGGTGAGCGCCGAACGCGGACCGGACACGGTACGCATCGCGCTGGCCGACAACATCTTCCACCCGCAGGGCG
GGGGGCAGCCCGACGACCGGGGCTGGGTCGACCAGGTCGAGGTCCGCCCGGTCCGCGACACCGGCCCCGGGCTGGTC
CACCTGACCTGCCCGGCGGATGCCGCGTGGCCGCGGGAGGCGGTGGCTGTCGGCACCACGGTGACCTCCCGGATCGA
CCCCGCGCTGCGCCGCCTCCACGCCGCGCTGCACACCGCGGGGCACCTGGTGGACGCCCTCATCGACGAGTGGGGCT
ACCGTCACGTCGGCAGCAACCACTTCCCCGGCCAGGCCCGCGTCGAGTACGACCTCGGCGGCCTCGACTGCGACAAG
GACGCCCTCGCCGAAGGGCTCACCGAACGCCTCGGCAAACCGCTCGCCGAAGCGCTCCCGGTGATCCCCGGGGAACG
CGACGGACGCCGCACGATCACCATCGAGGGGTACGCCACCGAGTTCTGCGCCGGCACCCACGTCCCCGACCTGTCCC
TGCTGACCGGGGTCGCCATCCGCTCGGTGAAGGTCAAGGGCGGCCGGCTGAAGGTGGGTTACACCGCCGAACACGTC
CCGGTGGACGGCTGA

NCBI Accession: WP_014145994.1; Uniprot ID: F8K3W0 (*Streptomyces cattleya* Threonyl-tRNA synthetase, native sequence)

ATGTCTCAGATCCGTGTGGTCATCAAACGCGATTCCGTTGAGCGGGAAGAGCGGGTGGTCACGACGGGTACGACGGC
CGCGGACCTCTTCGGCGGCGAGCGCACCGTGGTCGCCGCCCGGGTGGCCGGCGAGCTGAAGGACCTGGCCCACAAGG
TCGCCGAGGGCGACGAGATCGAGCCGGTCGAGATCGGCAGCGAGGACGGCCTCGCCATCCTGCGCCACTCCACCGCG
CACGTGATGGCCCAGGCCGTGCAGGAGCTCCACCCCGAGGCCAAGCTGGGCATCGGTCCGCCCATCAAGGACGGCTT
CTACTACGACTTCGACGTCGACAAGCCGTTCCACCCCGATGACATCAAGGCCATCGAGAAGAAGATGCAGGAGATCG
TCAAGCGCGGCCAGCGCTTCTCCCGCCGCGCGGTCAGCGACGAGCAGGCCCGCGAGGAGCTGGCCGGCGAGCCCTAC
AAGCTGGAGCTGATCGGCCTCAAGGGCTCCGCCGCCGACGCCGCGGAGGGCGCCTCCGCCGAGGTCGGCGCCGGCGA
GCTGACCATCTACGACAACCTGGACGCCAAGACCGGCGAGCTGTGCTGGAAGGACCTGTGCCGCGGTCCCCACCTGC
CCACCACCCGGCTGATCCCGGCCTTCAAGCTGATGCGCTCGGCCGCCGCCTACTGGCGCGGCAGCGAGAAGAACCCG
CAGCTCCAGCGGATCTACGGCACCGCCTGGCCCTCCAAGGACGAGCTCAAGGCCCACCTCGACTTCCTCGCCGAGGC
CGAGAAGCGCGACCACCGCAAGCTCGGCAGCGATCTCGACCTGTTCTCCATCCCCGAGGAGATCGGCTCCGGCCTCG
CCGTCTTCCACCCCAAGGGCGGCATCGTGCGCCGCGCCATGGAGGACTACTCGCGCCGCCGCCACGAGGAGGCGGGG
TACGAGTTCGTCTACACCCCGCACGCCACCAAGGGGAAGCTGTTCGAGGTCTCGGGCCACCTGGACTGGTACGCCGA
CGGCATGTACCCGCCCATGCAGCTCGACGAGGGCATCGACTACTACCTCAAGCCGATGAACTGCCCCATGCACAACC
TGATCTTCCGTGCCCGGGGCCGCTCCTACCGCGAACTCCCGCTGCGCCTCTTCGAGTTCGGCACCGTCTACCGCTAC
GAGAAGTCCGGCGTGGTGCACGGCCTCACCCGCGCCCGCGGCTTCACCCAGGACGACGCGCACATCTACTGCACCAA
GGAGCAGATGGCCGACGAGCTCGACTCGCTGCTCACCTTCGTGCTGGGCCTGCTGCGCGACTACGGCCTGGAGGACT
TCTACCTGGAGCTGTCCACCAAGGACCCGGAGAAGTACATCGGGACGGACGAGAACTGGGAGGAGGCCACCGCCACG

CTGCGCAAGGCGGCCGAGAAGCAGGGCTTGGAGCTCGTGCTCGACCCCGGCGGCGCCGCGTTCTACGGCCCCAAGAT
CTCCGTCCAGGCCCGCGACGCCATCGGCCGCACCTGGCAGATGTCCACGATCCAGGTGGACTTCAACCTCCCCGAGC
GCTTCGACCTGGAATACACCGCGGCCGACGGCTCCCGGCAGCGCCCGGTGATGATCCACCGCGCCCTCTTCGGCTCC
ATCGAGCGCTTCTTCGCGGTGCTGCTGGAGCACTACGCGGGCGCCTTCCCGGCGTGGCTCGCCCCGGTCCAGGCGGT
CGGCATCCCGATCGGCGACGCCCACGTGCCGTACCTGGCCGAGTTCGCCGCCAAGGCGAAGGCCAAGGGGGTGCGCA
TGGAGGTGGACTCCTCGGCGGACCGCATGCAGAAGAAGATCCGCAACGCGCAGAAGTCCAAGGTCCCGTTCATGGTC
ATCGCCGGTGATGACGACATGGCGGCCGGCTCGGTGAGCTTCCGCTACCGGGACGGGTCGCAGAAGAACGGCATCCC
GTGTGACCAGGCCATCTCCGAGATCGTTGATGTGGTGGAGCGCCGCGTCCAGGTGTGA

NCBI Accession: WP_001144202.1; Uniprot ID: B1XGI1 (*E. coli* Threonyl-tRNA synthetase, native sequence)

ATGCCTGTTATAACTCTTCCTGATGGCAGCCAACGCCATTACGATCACGCTGTAAGCCCCATGGATGTTGCGCTGGA
CATTGGTCCAGGTCTGGCGAAAGCCTGTATCGCAGGGCGCGTTAATGGCGAACTGGTTGATGCTTGCGATCTGATTG
AAAACGACGCACAACTGTCGATCATTACCGCCAAAGACGAAGAAGGTCTGGAGATCATTCGTCACTCCTGTGCGCAC
CTGTTAGGGCACGCGATTAAACAACTTTGGCCGCATACCAAAATGGCAATCGGCCCGGTTATTGACAACGGTTTTTA
TTACGACGTTGATCTTGACCGCACGTTAACCCAGGAAGATGTCGAAGCACTCGAGAAGCGGATGCATGAGCTTGCTG
AGAAAAACTACGACGTCATTAAGAAGAAAGTCAGCTGGCACGAAGCGCGTGAAACTTTCGCCAACCGTGGGGAGAGC
TACAAAGTCTCCATTCTTGACGAAAACATCGCCCATGATGACAAGCCAGGTCTGTACTTCCATGAAGAATATGTCGA
TATGTGCCGCGGTCCGCACGTACCGAACATGCGTTTCTGCCATCATTTCAAACTAATGAAAACGGCAGGGGCTTACT
GGCGTGGCGACAGCAACAACAAAATGTTGCAACGTATTTACGGTACGGCGTGGGCAGACAAAAAAGCACTTAACGCT
TACCTGCAGCGCCTGGAAGAAGCCGCGAAACGCGACCACCGTAAAATCGGTAAACAGCTCGACCTGTACCATATGCA
GGAAGAAGCGCCGGGTATGGTATTCTGGCACAACGACGGCTGGACCATCTTCCGTGAACTGGAAGTGTTTGTTCGTT
CTAAACTGAAAGAGTACCAGTATCAGGAAGTTAAAGGTCCGTTCATGATGGACCGTGTCCTGTGGGAAAAAACCGGT
CACTGGGACAACTACAAAGATGCAATGTTCACCACATCTTCTGAGAACCGTGAATACTGCATTAAGCCGATGAACTG
CCCGGGTCACGTACAAATTTTCAACCAGGGGCTGAAGTCTTATCGCGATCTGCCGCTGCGTATGGCCGAGTTTGGTA
GCTGCCACCGTAACGAGCCGTCAGGTTCGCTGCATGGCCTGATGCGCGTGCGTGGATTTACCCAGGATGACGCGCAT
ATCTTCTGTACTGAAGAACAAATTCGCGATGAAGTTAACGGATGTATCCGTTTAGTCTATGATATGTACAGCACTTT
TGGCTTCGAGAAGATCGTCGTCAAACTCTCCACTCGTCCTGAAAAACGTATTGGCAGCGACGAAATGTGGGATCGTG
CTGAGGCGGACCTGGCGGTTGCGCTGGAAGAAAACAACATCCCGTTTGAATATCAACTGGGTGAAGGCGCTTTCTAC
GGTCCGAAAATTGAATTTACCCTGTATGACTGCCTCGATCGTGCATGGCAGTGCGGTACAGTACAGCTGGACTTCTC
TTTGCCGTCTCGTCTGAGCGCTTCTTATGTAGGCGAAGACAATGAACGTAAAGTACCGGTAATGATTCACCGCGCAA
TTCTGGGGTCGATGGAACGTTTCATCGGTATCCTGACCGAAGAGTTCGCTGGTTTCTTCCCGACCTGGCTTGCGCCG
GTTCAGGTTGTTATCATGAATATTACCGATTCACAGTCTGAATACGTTAACGAATTGACGCAAAAACTATCAAATGC
GGGCATTCGTGTTAAAGCAGACTTGAGAAATGAGAAGATTGGCTTTAAAATCCGCGAGCACACTTTGCGTCGCGTCC
CATATATGCTGGTCTGTGGTGATAAAGAGGTGGAATCAGGCAAAGTTGCCGTTCGCACCCGCCGTGGTAAAGACCTG
GGAAGCATGGACGTAAATGAAGTGATCGAGAAGCTGCAACAAGAGATTCGCAGCCGCAGTCTTAAACAATTGGAGGA
ATAA

NCBI Accession: WP_109088880.1; (*Alcaligenes faecalis* Methionine-tRNA synthetase, *E. coli* codon optimized)

ATGTCACGCACTATCTTCGTAACTACGGCCTTACCATACGCGAATGGCAGCTTTCATATTGGGCACATTATGGAGTA
TATTCAGGCCGACATCTGGGTGCGTTCTATGCGTATGTCCGGGCACACTGTCCATTTTGTCTGCGCCGATGACGCTC
ACGGGGCGCCGATTATGCTGAAGGCCGAAAGTGCGGGTATCACGCCAGCACAGCTGGTCGACAAAATTGCCGCCGAA
CGCCCAACCTACCTTAACGGTTTTAATGTTAAATTCGATCACTGGCACCGTACTGATTCACCTGATAACGTGGAGCT
TGCGCAGGATATCTATCGTACACTGAAGAGTGCAGGTTTTATTGACACCCGTACCATTGAGCAGTTTTATGATCCAG
TAAAAGGGATGTTCTTGCCGGATCGCTACATCAAGGGTGAATGCCCCAAATGTCACGCTAAAGATCAATACGGCGAC
TCCTGTGAGGTATGTTCAGCTGTATATGCACCCACCGAGCTGATCGAGCCGTACTCCACCTTAACCAACGCTCGTCC
TGTGTTAAAGACGTCGGAACACTTCTTCTTTCGTTTAAGCGACCCGCGCTGTGTCGCGTTTCTGCAAGAATGGACCA
CGGGAAAGAATGCTCAGGGAAAGCCACGCTTGCAGTCCGAAGTCTTAGGCAAAACGCGCGAGTGGTTAGGCACCGGA
GAGGGGGCTGAAGCGTCGTTAAACGATTGGGACATTTCTCGTGATGAACCATATTATGGGATCCCTATTCCAGATGC
CCCCGGGAAATACTTTTACGTCTGGCTTGACGCGCCGGTGGGTTACTTGGCCAGCTTAAAGGCATATTGTGCAAAGG
CGGGGCTGGACTTTGATGCGTTACTTGATCCTGAAGGTAGCACCGAGCAGGTGCATTTTATTGGCAAAGACATCGTA
TATTTTCATGCATTGTTCTGGCCCGCCATGTTAAAATTCTCAGGGCGTAAAGTACCTGACGCTCTTAACGTCCATGG
CTTCATTACTGTGTCTGGCGAGAAAATGAGCAAGTCGCGTGGAACTGGTATCTCGCCACTTCGCTACTTGAAACTTG
GCATGGATGCTGAGTGGATGCGCTATTATATGGCGGCCAAACTGAATTCGCATGTCGAGGACATGGACTTTAATCCG

GACGATTTTATCGCACGTGTGAACTCAGACTTAATTGGTAAATATGTAAATATTGCGTCGCGTGCGGCCAACTTTAT
CTCTAAACACTTTGACGGTAAATTAGCTTATCAAGGAGACACTAGCGAATTGCAGAACCAACTGAAAGAGGTGGCCG
AGAAGGTTCGCGCTGACCTTGAGAGCCGTGAGTATGGTCGCGCTGTACGTCAGATCATGGCTCAGGCCGATATTATT
AATCAGGCGTTCGATACAGCTCAACCCTGGGTGATGGCGAAGGGCATTGCAACAGCCGAGCAGGCCCAGAAAGATGC
GCTTCAGGACATTTGCAGCCGTACGCTTGCAGGTTTCAAGGGGTTAAGTGTTATGCTTACAGCGATTCTTCCAACAT
TGACAGACCGTGTCGCGCGCGAACTGTTTGGACTGGATCGCTCCTTTGTCTGGGATGACGTGGCCGCCTTGCCCGAC
CACATTGCGCCCTTTAAACACTTAATGCAGCGCGTCGATTCTGCCATGGTTGACGAGTTACTTGCGCCCCCCCCAGC
TCCAGTAGTATTACCTGGCGGCGAGGCGATCGCCGATACTATTGATATCAAGGACTTCATTAAAGTCGACCTTCGTA
TCGCTAAGATCGTCAGTTGCGAAGCAGTAGAAGGATCGGACAAATTATTGCGTTTGTCTCTTGATGCCGGCGAGGGA
CGCTTACGCCAAGTATTTTCCGGTATCAAAAGTGCTTATCAGCCGGACGATTTGATTGGTAAACTTACGGTCTTGGT
GGCGAACCTGGCCCCACGCAAGATGCGCTTCGGGGTATCTGAGGGGATGGTACTGGCTGCAAGCCATGCAGACGATA
GTGTGGACCAGGGCATTTACATTCTTGAACCATTTCCTGGTGCGCAGCCTGGGATGCGCATTAATTAA

NCBI Accession: WP_045930716.1; Uniprot ID: A0A2U2BLV1 (*Alcaligenes faecalis* Methionine 3-hydroxylase AlfC, *E. coli* codon optimized)

ATGTACCTTGAAACTGAGCGCCTGCGCACAGAACACATGCAATCCTTGCTTAACGGCGAATTACTTGCAATTCGTTT
GTCGCAATTTATTGACGCGCCAACGAGTCAGAAATTAGGCAAACAAATTATCGACAAAGGTTTCGACCGCTACGTCA
ATGCACCTTCGATTGGTCGCATTGGCATGGCTTTCTACGAAGCTGAAAATCAGCCCGAACGTATGGCGTCTTACTTT
GAAGGCGTTTTCGACAACATCGACGAATTACGCCGTCGCTGTGCTCCCTATTTGTCCCCAATTGATGTTCTTCGCTG
TACGTTAGATGAGGTCTGGCCTGCTGGCGCCACCCTGGAGACGCTGTACGGCAAAAAAATGTACGTGGGTTTGTCCC
GCGTAGTTCAGCCCGACGTTCACTTTCTGGCGCATCACGACATCTTGGCAAAGGACGCTCCCGGCAACTACCAGGCG
ATCTCCTTGCAAGGTCAGTTGGCAGCCAATGTATACTTAAAAATGCCAGAGCAGGGGGGAGAATTGCAAATCTGGCA
GACTGAGCTTTCGCCTGAGGAATTCGATCAAATGCGCGGGGATTCTTATGGTATCCAACCGGATCTGTTAGGCGCCC
CGGAGGTTCAAATTCGCCCAGGCCCCGGTGATTTGATTATTTTTAACAGTCGCAAGATGCATGCCGTAACACCTGGC
AAGGGCGACCTGCGTCTGTCTCTTTCGTGCTTCGTGGGCTATCGCGGCCCGGCAATGCCTTTAACCTTCTGGTCCTA
A

NCBI Accession: WP_045930717.1; Uniprot ID: A0A2U2BLV6 (*Alcaligenes faecalis* 3-hydroxy-methionine $tRNA^{Met}$ Deacylase AlfB, *E. coli* codon optimized)

ATGCAGGACAAGTTATATCTGACCTCTAACTCGTTAACCTGTATCGCCGAGATCCTTGAATGTTACCATAGTGAGAA
CGAGCGCTACGCTGTCCGTTTGGATAAAACTCTGTTTCACCCTCAAGGGGGCGGGCAGCCGTGTGACAAGGGCACAA
TCAACGGGCATGAAGTGCTGCACGTAGCCCAAAACCGTGACGGTCACGTCTTGCATTATTTACACGTCCGTTTGCCA
TTAGGACCCGCTGAATTACACATTGACCGTGACCATCGCTTAGAGCATGCTCGTCTTCACACGGCGGGACATATGAT
TGGACACTTAGTGGAGCGCATCGGATGGCAAGCTGTGAAAGCACATCACTGGCGTGGCGAAAGTCGTGTTGTATTCC
GTAATCCAAAAAACGCTTTGCCCCCGTTCCTGGACGACCTGAGTAATCGCTGTGCGGAGCAAATCCAGTTGGCACAG
CCCATTTCTATCAGTTTTCAGGAAGGGCGTCGTCAGGTTGCAATTGGTGGGTTGGGCCAGTATGCATGTGGAGGCAC
GCATGTTGAGAACACCTCGGAGTTAAACGGACTGAAGATTGTCCGCATTAAAAATAAGAAAGATGAGTTAAGTGTCC
ACTATGAGATGATGGACTAA

NCBI Accession: WP_138460581.1; Uniprot ID: A0A5R8Z9Y2 (*Microbispora fusca* Arginine 3-hydroxylase FuzA, *E. coli* codon optimized)

ATGGTAATTGGCATTGATCTTGTCCACGCTGCTGTCTGGCGTTTTGACATTCAACGTAAGTCTGGTGCAGGTCGTGC
GGCCGACTTTTCAGAGGTCTCAGTCGTTGCCGATTTGGGTGAACTTGCCGTGGATTTACTTGGTGGAACAGTAGCGG
CACTGGAAAACGGCTTTGAGGGTGACGCTCTGAAAGATTTAGGTCTGGTTTTGGTACCAGCACCTGGGGTCGATGGC
GCCGTGTTCACACCCGGGGAATGCGAAGCACTGGACCGCTATGTTCAGGAAGGCGGGGCTGTACTTATTTTGGGGGA
TGGCGAATGGCCGCAGCCACCACAAGGACCCGCACCGTTTGGCTTTGCGTCAGGGGACCTTGCCATCGGTGACCCGG
CAGACCCAGCCCAACCCCATGTACGTCCCTATCAATTCCCGGCAGAGGTGACGGGTGCCCATCCGGCACTTGAGGGT
GTTACACGCGCATATTTGTACCGCCCCAAGCCGGTCAGCGCGCGTTTTCCTTTAACAGGTTTAGTCGAAAAGGATGG
TCAGATTTTCTTAGGGGTCGCACGCCACGGCGCTGGCCGTGTCGCTGTAGTTGGAAATGCCGAAATGTTCGCCCGTC
CTTTTTTAGGCCGTGAAGATAACGCACGTCTTTTATTAAATTTACTGAGCTGGTTACTGGATGGACGCCATCAACGC
TCTGCCGCTGATCTTGCGGCCGAGGTCACTGGCAGCTATCGTTTCTCCGGGCGTGATTTTAGCCCCACGGAAAACCT
TAAAAACGAGCCAGGGCCGCATGTTGTTGACGCCGGAGTATACGCGCACTTATTCACCGGGCTGGCCGATGGCTTCT
TACGCGATTCGGAAACAGACGAGGAGGCTTTTCTGCGTGAGGCTGAGCTTCGCTTTCACGAAATGCCACGCGTTGTT
CGCCAGGCCGTCTCCCGTTTTCGTCAGCACGGTACAGATTCTGGAGTCCTTTTGGTCAAAGGTCTGCCTGTTGATCC
CGCCTTGCCACCTACGCCCGCTGACCCAAAGGCACGCGTACGTTTCACCTCACCCTTATCCGAACTGTGGTTGGCGG

CTTTCGCGCAGGCCCTTGGGGAACCCATTGGATATCGCCAAGAGAAAAATGGCCTGCTTTTCCAAAATGTTGTGCCA
ACGCCGCACAATGCGGCAAAGCTGTCAAGCGAGTCATCGTCTATCCTTCTTGACTATCACACCGAAACAGCATTTCA
TCCCTTCATGCCGGACTACGTATTGTTGCACTGTCTTCGCCCGGATCACGATGGTGCAGCCAAGACGATTTCCGCTT
CGTTACGCATGGCTTTACCCCGTCTGTCTCTTCGTGAACGCGCTATCTTGTCGGAGCCGCTTTATCGCACTGGTATT
GATTATTCTTTTGGAAGTTCACATGGAACTAAGGGCAATGGACCCCTGCTGCGCGTTCTGTACGGTGACCCCTTTGA
TCCTTACCTGACAGTGGACCCGGACCTTATGATGGGGATTGTTCCTGAAGCCGATGCTGCATTGAAACGCCTTCATG
CTGTGCTGCACGAGGTCCGTCGTTGGGTTTGCCTGGACACGGGCGACCTGTTGATTGTAGATAACCGCCGTGCAGTA
CATGGCCGTTCAGAATTCACCCCTCGCTACGATGGGCGCGATCGCTGGCTTCAGCGTGTATGTGTAGTCCGTGACTT
GGTCCCCTCGGCAGGGCATCGCCGCCTGCGTGGTCGTATTATCGACACGGCTTTTGCAGTTTAA

NCBI Accession: WP_138460582.1; Uniprot ID: A0A5R8Z9P6 (*Microbispora fusca* tRNA-Deacylase FuzB, *E. coli* codon optimized)

ATGGGGGAAACTACGTCCGCTCCCTCCACAGTTGTCTTATTCCCTCAGGGTGTTAGTGTTTGCGAGGCTACCGTAGT
TTTGGCCGAAGCTGCCGGGGATGAAGGCCCTGTTGTAATCGTAACCGACCGCACTCCATTCCACCCGCTTGACCACA
CTTGGCCTGATCAACCGGGGGACCGTGGCACGATCACCTTAGGACGCCGCGAGCACGCTGTTGTCGACGCCGTCACC
ATGGCAGTTGGTCCAGATGGTACATTGCACCGTGGAGAGGAAATCCCTGCTCGTCGTACCGATCCGGGCTGGTCTTT
CCTTGTCGGACACGTGCTTGAGACGCCCCCTGATGGTAGTCCCGCCGATTTGGTCGGAGAGCAAGTCGGTCTGGCAG
TAGACGTTGCTCGCCGTAGCTCTCTTAGTGCTGCGCACACAGCTTGTCATTTGATGTCCTTGGCGTTGAATATGGAA
TTACGTGATCTTTGGCGCAAAGTTCCGCGCACGGATTGTTTGGGTAGTCCAGATTTTGATCAAGCTGCTATCACGCG
CAGCTCGATTACACCCTGGGAAAGTGTGGACCGTTATCGCTTAGGGAAAAGTTTACGCAAAGCAGGATTCGATGCAG
AAGCCCTTCCTGGACGTTTGGATGCGGTGCGCGAAGGTGTTACAGCCCGCTTGAGTGAGTGGATTGCTGCCCGTGGC
CGCATTGCCATTGACACTGATGGCCCATGGATTAACAGTCGTCGCTGGTGGGGTTGTGAACTGCCGCCTGGGCCTGC
GCGTCTGCCTTGCGGTGGTACACATTTAGGCAACTTGGGAGAATTAGCCTCCATCGATGTGCGTATCGTAATCGGGG
AGGGAGAACTGGTTGTAAACACCACTGCAAAGCAATCGTAA

NCBI Accession: WP_138460583.1; Uniprot ID: A0A5R8Z9N5 (*Microbispora fusca* L-Ala-L-Arg ATP-grasp FuzC, *E. coli* codon optimized)

ATGTTTGATCCGGTTGGTTCGTCCGTTCTTGTCGTTGAACCTGTGAGTGGTGGAGCAGACGTGGCTCGCCTTGCACA
TCACCTTGGTATGACTGTTGTAATCGCAAGTCAAGACGTCGGAGATTGTGAAATCCCAGCCGATTTGCGTAAAATTG
CCTCGGTTGTGTTAACCGTGGATACCAATGATGAGAAGACGCTGTTTGACGTAGTCGCTGAGCACCACGAACGTGAG
CCTTTAGCCGCAGTTCTGCCCGGCTGCGATTTTTACGTGGCCACTACCGCCCGCCTTGCTGCGCGCTTGGGCTTGCC
CGGCTTACCTGTCGAAACCGTCGACGCAGTACGTAATAAAGCGCGTATGCGCGAGCGTGTGGCAGCGGCCGGATTGC
GTACGCCACGCTTTGCAGAAGCGTCAAGCCCAGAGGCCCTGCGCACAGCCGCTGAGATTGTAGGGTTCCCTAACGTG
ATGAAACCTGTGGAGAGCAGTGGCTCAATCCATGTTAGTCGCGCAGATGACTGGGCGCAACTTGCAGCAGCCTATGC
CGCATTAACTGGCGACCGCGAACTGGACCTGGGACGCCCGATGGGTCATACTGTCGTAGTAGAACAGTACGTTGCTG
GAGAAGAGGTGAGCGCAGATGGCTACGTCGAAAACGGTCGTGTAACGGTCGTTGCCATCACTCGTAAGGTCTTAGGC
GCTGAGCCGCGCTTTGTAGAACTGGGTCACTTAACACCAGCAGATCTGCCTGACGATGTGACGAATGAGGTCGTCGC
GTATACTGAACAAGTCGTACGTGCGCTTAATGTAACTGCCGGTCCGTTTCACTGCGAGATGCGTCTTTCAGACGAGG
GGCCTGTTCTTATGGAGTTGGGTGCGCGTCTTCCGGGAGATGGTATTCCAGAGCTTATTCGTTTAGTAACCGGCACC
GACTTGGCTGCCGTGATGTTGGCGGCAGGATTAGGCGTCGACGCGGCCGAATTGCCGGCGTTCGGACCTCCTGCGGC
GCCTTATGCTGCAATGCGCTTCTTCGTAGCTGATGACTTGGACTCACACAGTGGGATCTTAGGCTGGGATGAACTTG
ATGACCGCCCTGATGTGATCGGTAAGCAATTATTTATTCCCCCGGGAACACCCATTCCCCAAACAGGAGACTGCCGC
AGTCGCTTGGCGTGCGTGCGCTACACGGCCGAGAGTTACGAGGCTGTTCAAGCTTTTTGGAAATCCCTTGACGATCT
GGCTGTCCTGCCGTAA

NCBI Accession: RGA06498.1; Uniprot ID: A0A5R8ZAJ0 (*Microbispora fusca* L-lysine N6-monooxygenase FuzD, *E. coli* codon optimized)

ATGCCTTTCGCAGCTCGCCTTCGCGCACTTCATCGTGGGGTAGTGCGCGGACGCCCTGGAGTTTTGGAAGAGCCTGG
TCGTCCAGGGGGTGCCCCTGTATCAGACGTGCTGGATTATGCCGCTGTAGGAGTGGGCCCAGCTAACTTATCCCTGG
CTGCCCTGGGAGCTGACGTGGATGATTTGCGCTCGGTGCATTTCGAGGCCCGCCCCGATTTTAGCTGGCACCCTGGC
TTATTGCTTCCAGAGGCGACCCTGCAGGTATCTCCCCTTAAAGACCTTGTAACATTGGTAGACCCAACCTCTCGCTA
CTCATTCCTTGCTTTTTTAGCGAATGAGGGACGCTTGTACCAGTTTCTTACAGCTAAATTCCCTGCTGTTTTGCGCG
CGGAATTCGATCAATATCTGCAGTGGGCATCACGCAGTATCCCTAGTATCGAGTTTGGTACCGAAGTCCGTGGAGTC
CGCGTACGTGATGGACATTTTGTCCTGGATACTGGTCACGGCGAAGTGACAGCCCGTAACGTTGTTCTTGGCACGGG
CCAGACCCCGTTTGTCCCGGGGTTTCTGCGCCCACATCTGTGCGCGGACGTCTTTCATGCAGAGGAATTTCTTAGCC

GCCCACGCGAACTTCCCGACCGCGCCGTAACCGTCATCGGAGGTGGACAATCCGGCGCAGAGATTGTACGCCATCTT
ATGCGCTTGGACGTGCCGCCGCGCAGTATTACGTGGATTACGCGTCGCCCAGGCTTCTTACCAATGGACGATTCTCC
GTTCACGAACGAGTACTTTTTCCCGGCGTACGCCCGTGAGTTTCATGGGCTTCCGCAGACCCACAAGCAAGAACTTT
TGGCCGAGCAGAAATTAGCATCCGACGGCATCGACCTTGCCACGTTGCAGGATGTTTATCGTGAACGTTATCGTCAT
GAGTGCTTGTTAGGGCGTAAAGACTGGTTGCGCCTGATGCCGCGTAGCGAGGTAACTGGCTGCGCACGTTTGTCTGA
CGGCTGGGAATTAGAGGTAAGTGGGCCCCACGCCTCAGTGGTACGCAGCGATGTAGTCGTATTAGCGACGGGATATG
AGTATCGCTTACCGGATTGTATGTCGGAGCTTGCACCCCATCTGGCACGTGCCCCGGAGGGTTTGGTAGTCGGGGAT
GACTACGCATTAAGCCTTGATCTGCGTACTAGCGGACGTGTTTTTTTGCAGAACGGAGCCCGCCACACTTGGGGTAT
CGCCGATCCCAACTTGTCTTTACTTGCCTGGCGCTCTGCAGTGATTCTTAACGCAATTTTAGGATACCGCCGTTACA
ACACCTAA

NCBI Accession: WP_111697749.1; Uniprot ID: A0A5R8ZB56 (*Microbispora fusca* Cupin Protein FuzE, *E. coli* codon optimized)

ATGTCGCTGCCGCGTTTGGGTTTTGAGGCGTTACCACGCGAGTTTGAGAAGTTAGATGAATGGGCTGCTTTGTTGGG
ACCATTTGATGGCGACCGCCCATTTAGTATGGCTCGCTTCACAGTCGCGGCGGGAGGGGCCACCCCCCCTGATTCTC
ATGCAGTAAAGGAGATTTGGTTAATTCTGGATGGAGCTGGTACCGTGCGCTATGAAGACCGTCCATATGAGGTGGGA
AAAGGGGATTTATTGGGGTTCGCCCCATTCGCAGAGCACGTCGCAATGGCAGACCGTGGTGTTCCGCTTTCAGTATT
GAGCATCTGGTGGGGTACGACGGGTTAA

NCBI Accession: WP_138460584.1; Uniprot ID: A0A5R8ZCB0 (*Microbispora fusca* Flavin Reductase FuzF, *E. coli* codon optimized)

ATGGGCCACGACCGTATGACGGGAGCTATCCCGGACTTTGATTTGGCTGTAGTCGGTGCTGGGGTCGTGGGAGCAAT
GACTGGACATCTTGCTCTGGAACAGGATCCTGGGCTGCGTCTGCTTACCTTGGATCAACTGGAACGTTTGGACCAGG
ACGTACCGAATGCGACAGGTCTTTCAGCGGGATTAGATGTACCGACTGGGCGCGATGCTCGTCAACGCGAATTGGCT
GCTCGTAGTACCGCTCTGTACCGTCGCTTACGCGCGGAGCGCCCTGGCTTGCCGGTGCGTTCTTTAGATACGTTTTG
GTTATTGGCACGCGGAAATGTTGGTCGCTTACGCGAAACCATGGCTGGTCCAGTATTGGAGGAAGCTGATCCAGATG
CGCTGGACGGCGTCTTCCCTGGATTACGTCCCAGCGGCGACCAGACATTGTTACGCTCTGATGGGTCATGGCGCGCA
GACGTCGCTGGATTAACGCGTGCGTTATTGGATGAACAGCGTGGCCGTCCCGGATGCGCGGTTTGGGAAGACGAAGA
GGTTACTTCTTTACGTCCGTACGGGGACGGATTCTTGTTGGCCGGCGAAGGTGGACGTCCACTTGCGCGTTGCCGCG
CTGCTGTGGTTGCAACGGGCCCTTGGGCTCTGCATGGCCCGTTTGGCGAACGCATGCGCGCGGCGGGACTGCGCATC
AAGAAAGTCGCTGCCTTCCATTTAGACCTGCGTCCAGCGGAAGGAGCCCCCGTCCTTATCTTTGAGGATGATGATGC
TTTTTTACTTCCACAGCCCGAATCGGGACGTTGGCTTTTCTCGTTTACTAGCCAGGAATGGGATGTTTCGCCAGTTC
TGGGCCCTCCGGCCGCTTTATCAGGTGGAGACTATTCGGAGGCTGTTCGTGTCTTGCGTCGTTATGCACCCGCATTA
GCGGGCGCCCTTAGCGGCGGTCGCGCGTTCTACGATGGCTATACCCGCGATCGTTTACCGCTTGTGGAACCTGTGCC
AGACCACCCTGGGGTTGTTATTGCTGTCGGTGGGAATGGAAGTGGCTTTCGTCTGGCACCCGCAATCGCAGAACGTG
CGTTGGGATTGCTTACTCCGGGGCATTTGAATCCACGCAGTTTAGGACCTCAGTCCTTGGGAAGCGGACGCAGTTAA

NCBI Accession: WP_138460585.1; Uniprot ID: A0A5R8ZBS2 (*Microbispora fusca* Hydrazine Synthetase/Methionine-tRNA ligase-like Protein FuzG, *E. coli* codon optimized)

ATGAGCACGCGCTCAACGGGCCTGATCCGTAATGCTGGGACGTTACCTCCTCGCCGTTACCTTCTTGTTCCCCCGGA
GCCTACTCCTAATGGACGTTTGCACTTGGGTCATATCGCAGGGCCATACCTGCGCTCAGATATGTTGAGTCGCTTCG
CACGCATGCGTGGAGACTTACCTCTTACAATTACTGGCATTGACGTTTATGAGCCGTATATGACTCTTAAAAGCGCA
GTAGATGGGCGCCCAGAGGGCGAGATTGCCGCCGAATATCACGCACTGATCGCTGAAGATCTGGCCGCCATGGATAT
CGGACTTGATGCTTTCATTAATCCCGCTGAGGAACCTTGGCGTACTGCCTTCGAGGATGAGGTCGCAGGATGTCTTC
GTCGCCTGCGTGAACGCGGCGCCGTAACGACCTTACGCGAGCGCGTCCCGTACTCTCCAGCCGCAGGTCGCTTCGCA
GTAGGACCGTGGCTGTCAGGTCGCTGTCCTGATTGCGGAGCTGATGCTGGTTCGTACTTCTGTGAGAGCTGCGGTGG
GCACTTCCGTCCTGAGAATCTTCGTGATCCAGAAGGTACCAATGGCGTGGGGCCACTGCAGTGGCGTGAGATTACCT
CGCTTTTCCTTCGTCCAGCTGATCCGCACGACTTGGGTGCTCGTATGGCAGGAATCGGAGTGCCGGCTCGTTACCAG
GAAACCGTTAAACGCTATCTGTCGCGCGAGGGCGGGATGGTACGCGTTTCCGCGCCTGAGAGTTGGGGTATCCCGCT
GGACCAAGTGGATCCGGGGGTTCCCCGCACTCTTTTTCCATATGCCGGCATTTTTATGTTCGCCCGTCTGACAGGCG
CAGTTCATGGTTCGCTTAGTGGGACAGGTGTTAACGCGTTCGATCCAGGTAGTGACGTTGTAACTGTGACCTCATTG
GGCGTCGATAATGTGGTCCCTACATTAGTTTGCATTGTGGGGACGTCGTTGCTTCATGGGGATATGAAGACCTATGA
TCGTACGTTAATTAACGAGTTTTACGGACTTGAAGGCGAAAAGTTCTCGACCTCCCGCGGGCACGTCATTGGAGCGG
CCGACATTGCCGGGACACCTAATGTAGACGTAGACGCCGTGCGCTATTTTTTAGCAGGGACGCGTTTAGAAGAGCGC
CGTGAGTCGATGGAGCCTGAGTCCTTTGTGGAGACGGTAAATCGCCGTCTGGCTGACGGACTTACTCCGTTATTAGA

AAAGGCACGTCGTACTTTAGCCGGTGTAAGTGTTCCTGAAACTATCTCAGATGGCCAACTTGCACGTCTTGAGGCAT
TGCTGGAACGTCAGCAATCTGCCTTAGACGGCCCTGTTGTCTCAACATCGCAGGCCCTGGAGGCATTTGACGCCTGG
TGGCGTTCAGTAGCGGACCTTGAAGAGTCGGCGTACTGGTGGTTGAAGGGAGTATCCGTGTTGGGATTTCCCATTAT
GCCTCGTATGGGGCGTCGTATTTGGGAGGACCTTGGAGGGGTGGGTGAACCACGCCTTGAAGAATTCTTGGCGCCCA
CGCCACCGCGCGAAGCAATGGCACCTTTAGGTTTTGGCGCCATCTCTATTGGCGACCTTGTGCCGGTCTTGCCGGCT
ACCTCCCGTTATGCTCTTGCCTAA

NCBI Accession: ; Uniprot ID: A0A5R8Z9K7 (*Microbispora fusca* L-Pipecolic Acid-Amino Acid ligase ATP-grasp FuzH, *E. coli* codon optimized)

ATGCCCGACCGTCCCCATTACTTAGTCGTTAACCGTTACGACGATTCTGATGCGGAGTATTTTCGCTACTGCGCAGA
TCAACCGTGTGATTTAAGTTATCTTACCCTGCCAATGCGTACGACTGCCTTGGGTCCCGTGGAACCGGGGCGCGTCG
GAACCGTACCCGATCTGCGTTTAGCCACGTTGGTTCCTGCGGCACGCGACTTGGCGGAACGCTATGGTCGTTTTGAC
GGTGTGGCAGCGTTGTCAGAGTTCGACGTTTTGACAGCTGCTTTGTTGCGCGAGGAGTTCGGTGCCCCTGGACCAGG
TGCAGATTTAGTTCGTGCGTTCAAGGACAAGATCGTTATGAAAGAGCGTCTGGCGAAGCGTGGACTGCCAGTCCCCC
GCTTCATGGAACTGGGGGCGTCAACCACTATTGCCGAGATCACGGCGGGTATTGGCTTCCCGCTGGTCCTTAAGCCG
CGCGCAGAGGCTGGAAGTGTAGGTGTTCGTATCATCCGTGACGAGGAGGAGTTGGCCGCCGAGTTAGACGGATTAGA
CGTTACTGCTTTTGAGTGCGAAGAATTCGTTAGTACGGAAATCTACCATGCCGACGGAGTATTGTCGGGCGGCTTGC
CGCATTTTTTATCCGTGGCCGGATACGTGAACACTCCCTTGGATTTTTTGCGCGGTCGCCCGTTAGGTTCTGCAACA
CTTGATCCAGGTCCGTTACGTGACGAGATTCGCGAGTTCGCCGTTGCATGTCTTCGCGCCTTGGGTTTAGAGACAGG
GTGTTTCCATTTGGAGGTCATCCGTCGCCCGGATGGCGGGTTGGTATTTTTGGAAGCGGGGATGCGTGCGGGAGGGG
CAGAAGTGCCTTACGTCCACCGCGAATTACACGGTATTGATTTAGTAACTGAAAGCTTCCGCACCGTCTTAGGTATC
CCTTTGGAATCAGCGGGGGCTCGTTTTGGTGACGCACCCGCAGGTTGGGTCGTGACCCCCGAACCGGGACTGCGTCC
CGGCCGTGTCGTTCGTCGTACCGCTATGACCGGTACAGTTCCGCAAGTCTACGCGGAAGTGTTACCTGAACCTGGTG
ACGTGTTTGATGGTACCGGCGGTTATTTCCACGTGCCCGGACGTTTCCGTTTAACGGGACCCGACGAACGCTCCGTT
CGTGCAGCTGCACGTCGTGTGATGCGCGAGTACGTAGTAGAGTTTGAACCTGCCGGACATACTCCT

Table A1.2. Strains, plasmids, oligonucleotides, and gene sequences for Chapter 3. (A) Strains and plasmids. (B) Oligonucleotides used for plasmid construction and sequencing. (C) Sequences of genes.

A1.2A. Strains and plasmids

Strain	Description	Source
E. coli DH10b-T1^R	*F-mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara-leu)7697 galU galK λrpsL(StrR) nupG tonA*	Thermo-Fisher
E. coli BL21 Star (DE3)	*F-ompT hsdSB (rB-, mB-) gal dcm rne131* (DE3)	Thermo-Fisher
Pseudomonas fluorescens R124	WT Strain	Prof. Hazel Barton

Plasmid	Description	Source
pET16b-His$_{10}$-PfrmB	His$_{10}$-PfrmB (T7), *lacI*, Cb^R, ColE1	This study
pET16b-His$_{10}$-PfrmC	His$_{10}$-PfrmC (T7), *lacI*, Cb^R, ColE1	This study
pET16b-His$_{10}$-PfrmD	His$_{10}$-PfrmD (T7), *lacI*, Cb^R, ColE1	This study
pET16b-His$_{10}$-PfrmG	His$_{10}$-PfrmG (T7), *lacI*, Cb^R, ColE1	This study
pET16b-His$_{10}$-PfrmH	His$_{10}$-PfrmH (T7), *lacI*, Cb^R, ColE1	This study
pET28a.pTet	*ptet,* tet^R, Km^R, ColE1	*2*
pET28a.pTet-PfrmBGC	PfrmB-H from PfR124, *ptet,* tet^R, Km^R, ColE1	This study

A1.2B. Oligonucleotide sequences

Name	Sequence
pET16b-PfrmB-Fwd	GTGCTTTTTCAGGGCCCGCATATGACGGGCTGTGATGCAC
pET16b-PfrmB-Rev	TCGGGCTTTGTTAGCAGCCGGATCCTCATCCCTGAGCTACGTGTAC
pET16b-PfrmC-Fwd	GTGCTTTTTCAGGGCCCGCATATGACGATTCAGGAAAAGAACTTCAC
pET16b-PfrmC-Rev	TCGGGCTTTGTTAGCAGCCGGATCCTCATCGGCTGTGTCCTCC
pET16b-PfrmD-Fwd	GTGCTTTTTCAGGGCCCGCATATGAGTTCGTCTGGAAAGAGCAATC
pET16b-PfrmD-Rev	TCGGGCTTTGTTAGCAGCCGGATCCTCAGGCGAGCTCCATGG
pET16b-PfrmG-Fwd	GTGCTTTTTCAGGGCCCGCATATGAATGCTCCAGTCGTAGAGAAACTCC
pET16b-PfrmG-Rev	TCGGGCTTTGTTAGCAGCCGGATCCTTACCGAGCATCAACAGGCAGCG
pET16b-PfrmH-Fwd	GTGCTTTTTCAGGGCCCGCATATGCGCCATATTTTGATAGTGGGTATTAACAAGG
pET16b-PfrmH-Rev	TCGGGCTTTGTTAGCAGCCGGATCCTCATTTGAAGACCGGGCTTCTAGGG
pET28a-Pfrm-Frag1-Fwd	CGAGGCCCTTTCGTCTTCAGGATCCTCTAGAAGAAGGAGATATACATATGACTGTGACGGGCTGTG
pET28a-Pfrm-Frag1-Rev	CTCATCGGCTGTGTCCTCC
pET28a-Pfrm-Frag2-Fwd	GGAGGACACAGCCGATGAG
pET28a-Pfrm-Frag2-Rev	CTGTGGCTGTTCACTTCTCGG
pET28a-Pfrm-Frag3-Fwd	CCGAGAAGTGAACAGCCACAG
pET28a-Pfrm-Frag3-Rev	CCGGATCTCAGTGGTGGTGGTGGTGGTGCTCGAGTCATTTGAAGACCGGGCTTCTAG

PfrmB-S73A-Fwd	GTGCTTTTTCAGGGCCCGCATATGGGCCACGACCGTATG
PfrmB-S73A-S74A-Fwd	TCGGGCTTTGTTAGCAGCCGGATCCTTAACTGCGTCCGCTTCCC
PfrmB-StoAs-Rev	CACCGACGTAATCGCGAGTATGAAAC
pET16b-BB-Fwd	CCGGATATCCCGCAAGAGGC
pET16b-BB-Rev	GCCTCTTGCGGGATATCCGG
PfrmB-Y69F-S73A-Fwd	GTTTCATACTCGCGATTTCGTCGGTG
PfrmB-Y69F-S73A-Rev	CACCGACGAAATCGCGAGTATGAAAC
PfrmB-Y103F-Fwd	CAGGAAATGGCCTTCGTAAAACATCCC
PfrmB-Y103F-Rev	GGGATGTTTTACGAAGGCCATTTCCTG
PfrmB-Y103A-Fwd	CAGGAAATGGCCGCGGTAAAACATCCC
PfrmB-Y103A-Rev	GGGATGTTTTACCGCGGCCATTTCCTG
PfrmB-Q226A-Fwd	GGTGCAATGCGGCGCACTTC
PfrmB-Q226A-Rev	GAAGTGCGCCGCATTGCACC

A1.2C. Gene sequences

NCBI Accession: EJZ60183.1; Uniprot ID: A0A7U9CRC7 (*Pseudomonas fluorescens* R124 PfrmB, native sequence)

```
ATGACGGGCTGTGATGCACAAGGGATATTCGATTTTGCGGTCGGCACCCATGCCGATCCGGCAGGCTATTTTGCAGC
GGCGGCAGACGACATCCGGGCACAGCTGCTGAGTGAAGGCAAGGTGCTCATTCGTGGCGCCCATCCGTGCAGGGATC
CTGAAGTGTTTGAGCGGGCCATCGGCGTGCTGGGGTTTCATACTCGCGATTACGTCGGTGGCAGCTCTCCACGCTCG
ACGATCCAGGGCAAGGTCATGGAAGCGACGCGCACCCCGCCCGACTGGTCGGTGATCCTTCATCAGGAAATGGCCTA
TGTAAAACATCCCCCCGAGATCATCGCCTTCGTCTGTGTCGAGCCTGCGGGGCGAGGCGGAGAATCCGTGGTCGGGG
ACATGCGCAAACTGGAGCAGTTGATTGATCGCTCGACCTTGAAGCAACTGACCGATCGCGGGCTTAAGTTGCGCCGG
ACGCTGCCCGGCGAGGCGCGGGTGAATCTCAAGCCCGGGGTGAAAAAAAGCTGGCAGGAGACGTTCTCCACTTCCAG
CACCGCTGAGGCAGAAGTGCTGTGCCGGGCGCGGGGGTGGGACTTCGAATGGAGTGGCGGCGACCTGATTCTCTGGC
AGGACTGTATCTCCCCGATGCGCCAGCACCCGCGTAAAGCGGCAGAGATCTGGTGCAATCAAGCGCACTTCTGGGGA
GCGGCCGCGATGATCGAGTGGGCGCGTATCGATGGTCGCGAACAGGATGCCAGTGAACTGGTGAAGGCGCAGCGCAA
CACTCCACAACTGTTGGAGGCCATGTGTTTCGGTGACGGCGAACCACTGCCGGAAGACCTTACCCTCGAGCTGTTCC
ACACCGTCAGAAGTGTTGAACAGGATGTCGATCTGCGGGCGGGCGATATTTTACTGCTCGACAATTTCCAGTATGCG
CACGGTCGGCGCGCATTTTCGGGCAACAGGACAATCAACGTGATCATCGCCAATTGGGCGGATGATCAGGTACACGT
AGCTCAGGGATGA
```

NCBI Accession: EJZ60182.1; Uniprot ID: A0A7U9GV55 (*Pseudomonas fluorescens* R124 PfrmC, native sequence)

```
ATGACGATTCAGGAAAAGAACTTCACGTCGGATAGCGCATTTGTACTGAACCCCGGCGTCGATGTCAGTCTCAAGGA
CTTGATCAGGTATTCCCGGCCCGGGGCTCAGGTGGTCGCCTCGCAAGCCACCCGGCAACGGGTTCAGGCATCGCGAG
AAGCCTTGGAAAACCTCGTCGCGCAGGGTGCGGTGATTTATGGCGTGAACACCGGAATGGGTGGTTTCGTTGATCAT
CTGGTGCCCATCGAGAAAGCCCACGAGCTTCAATCGAACCTGATTCAGGGCGTCTCGACCAACGTTGCCGAGCGGTT
CCCGGACAATGTGGCGCGAGCCACCATGTTTGCGCGAATCGTCTCCCTGTCCCGTGGCAACTCGGCACTCTCGCTGG
AGAATTTCGACCGCTACATCGCCATCTACAACGCGGGCCTGATTCCGGAAATTCCGAGCAAGGGCTCTTTGGGGACG
AGTGGCGACCTGGGGCCTCTGGCGGCGGTTGCGCGGATGCTGACCGGGGACGGTTACGCCTATCTTGACGGTGTGCG
GATTCCGGCGGCGCAAGCGCTTGAGCAATTGGGCCTGGCGCCGCTCCAGCTGAGTTACAAGGAAGGGCTGGCGCTGA
TCAATGGCACCTCCTGCATGGTCGCAGTCGCGGCGATGAATGTGGTCGAAGCCCGGGAATTGCTCGAGCAATACACA
TCGATCTCCGCCTTCTCTTTTGAAGCGTTGCAGGCTCGAGTGCGTCCTTTCCATCCGGATGTGCACCAGTTGAAACC
GCATCTGGGCCAGCAAAAAATTGCCGAAGCCATCTGGGCCGATCTTCAAGGCTCGGAACTGGCGATTGACGATCTCG
AGCTTTCTGCGCAATTGAGCGCCCTGCGCACCGACCAGGTACGCCAGGAAGACATGCCCATCGAAGATGCCTATTCC
ATTCGCTGCACGCCGCAAATCCTCGGGCCGGTGCTGGAAACCATCGAGTTTGTCGAGCGGGTGGTCGGCAACGAGCT
GAACTCCTCCAACGACAACCCGTTGATCACCCCTGAAAACGGCCAGGTTTTTCACAACGGGCACTTTCATGGGCAGT
ACATCGCTTCGGCCATGGACTACCTGACGATCTCCATCATCACCCTGTGCAACCTTTCGGATCGGCGGATCGATCGC
TTCCTGACCAAAGCCAACAGCAATGGTCTGCCGTCCTTCCTGTGCGCGGAAAACGGAGGTCTGCGTTTCGGCCTCAT
```

GGGCGGGCAGTTCATGTCGGCGTCGCTGACGGCGGAAAACCGTTCGCTGGCCACCCCTGTCTCGATCCAGACCCTGA
CCACCACGGGCGATTTCCAGGACGTGGTTTCCTTCGGCCTCGTCGCCGCACGCCGTGTGACCGAAGTGCTGCAGAAC
ACCCGTTACGTTCTCGCGTTCGAACTGATCTGTGCGGCCCAGGCCGCCGATATTCGCGGCAGCGCCAAACTCGGGCC
GAAGGGCCGCGCCTGGCATGCGCGGGTGCGTGAACATATTCCGTATCTGGATCGGGATGTGGCATTGACCGGTTACC
TGGAAGCGATCGTTGCCCAGGTACTGGGAGGACACAGCCGATGA

NCBI Accession: EJZ60181.1; Uniprot ID: A0A7U9CU77 (*Pseudomonas fluorescens* R124 PfrmD, native sequence)

ATGAGTTCGTCTGGAAAGAGCAATCTGTTCGCAACGGCGGCCATCGACCGTGTCCGGCTGCATGCCAACGAAAATCC
GTTGGGTGCATCGGGTCTGGCCCTGCATCTGTGCGCCTCGATGCCGACGCTGCTCAGCGAATATCCTGAAGTCACTT
ACAAGTCTCTTCGCGTGGCAATCGGCGAGCGCTGGCGGGTTGATCCCGAGCGCGTAGTGGTGGGCGCGGGCTCCGAT
GAACTGATCGATATCGCATTGCGCAGCCTGTTGAAGCCCGGCAGTCGTATCGTGACGCAGCAGCACAGCTTCAAGAT
GTACAAAATCCGCGCCGCGCTGCTGGGGGCCGACTATAGCGAGATAGCCCCGGACGCCAACGGCGACTATGAGGCCG
CATTGAAACGTGCGCTGGAGAAAAAGCCGGACGTGGTGGTGCTGGTCAATCCATCCAATCCTTTGGGCGACTATCTG
TTGCCCTCGCAGCTGTTATCGCTGGTGGCACTGATTCCGCCGACCACGACCGTCATCATCGATGAAGCGTACGCAGA
GTTTGCGCCGGATCGGGAAAACACTGCCGGCATTCAGATCGTCAACAGCCGGGAAAACGCGATATGCCTGCGCACGT
TTTCCAAGGCCTATGGTATTGCGGGTCTGCGCCTTGGCTGGGCTTACGCGAGCGCCGATCAGGTGCCGGCGCTGCAA
GCGCGGCGTGCTCCGTACAACGTCAGTTCGTTTTCCGCCCTGGCAGGCGAATACGTCCTGCAGGACTGCGATCATCT
GGACAAGACGGCGGCGTATTGTCAGGCCTGGGGGCGGGAACTCGTGAGTCTGTTCAAGGCTCATGGCATTGAAGCCA
GAAGCCGCTACGTCAACTTTGTTTTCGTCGATTGCGTCAGCCAGACACGCTGCCAGGACTTGCATCAGCACCTGGAG
AGTGAGGGTTTCCTGACGTCTCCATTGACCGACTATGGTCTTGATACGTGCCTGCGCATCTCCTTCGGAAACGAGGC
CCAGATGACCCGGTTGTGCACTGTACTGCAGCAAAAACTGCAAACCATGGAGCTCGCCTGA

NCBI Accession: EJZ60180.1; Uniprot ID: A0A7U9CRD7 (*Pseudomonas fluorescens* R124 PfrmE, native sequence)

ATGAGCCAGACGCTGCCTTTGTTTCACGATGAACCCTACCTGAAGCAATTGGACTGCCGAGTCGATCTGCTTGAGCG
CGACGAAGCCGGTCTGCCTTTCGCACTCCTTGGCGAGACCGTGTTCTACCCCCAGGGCGGCGGACAGTTGGGAGATC
GGGGTGTGCTGAACATTGCGCTCGCCGGGGAGGGTGAGGCACTGCGCACGCTGGAGATCGTCGCGACGAAAAAAAGG
GATGACCAGATCAGGCATTACCTGGGCATTGATGACGAAGCGTTCACACAACTGCAGGCGCGGGTGATCGGCCAGCA
GGCGACCGCCACGCTGGATTGGTCATTGCGATATCATCAGATGCGTGTTCATTCGGCGATGCATTTGATTCATTGCA
TGCTTGAACGGACGCTCGGACGAGCGATTGCACATCCCGTACGTTCGCCGCTGAGCGACGAGGGTGGAGAGAATCAA
TACGAGTTTGTCGGTGAGTTCGACGAAGACAGTCTGAGGCGGGCGACAGACGCCTTGAATGAGTTTTGCTCTGCAAG
TCATGAGATTCGTACGGAAGCGGATACCAGCCAGGGACAGGGATTTCGTTGGTGGCAATGCGAAGACTGGTCAATCC
CCTGTGGGGGCGTTCATGTGCGTAATACGCAAGAAATAGGCGTGGTAACTTCCAGTATGAAGATCAAGAAAAATACG
ACCAGAGTGGTTGTCACTCTGTCGGCAGAGGCTGAAGAGCTTGGATAA

NCBI Accession: EJZ60179.1; Uniprot ID: A0A7U9CSG0 (*Pseudomonas fluorescens* R124 PfrmF, native sequence)

ATGGATAGCAAACGAGCGTTTGGCGCATTTCTCATGCAGTTTCGAATCGACGCGGGGCATCTGAAGATTTCCACGTT
TCTCGACGAGGTCAATATTCCCTTTTCGGCCAATTACTACCGCGATGTCGAGGCCGGAAGGAAGTATCTGAGCATCG
ATGCCGCCGTCGAACTGCATGATCACCTTGGAATCGATGCGGCCAGCAAAGCCAGTTTTGATTATTTCTGGCACTAC
TTCAGGGACTTGTTGCCGGCTGACGTCCACGAGATGTTGTTCGGCAAAGGCGTCGAGGTCACCGATCTTTCTTCGCG
CCTGGAGCTGCGCGAGCACGACTCCAAAATCTTGCGCCGGGCCCTGAGTCTGAGCCGCTACGAGCGTGAGTTCATTG
TCACCGAAGAGGTGATAGCCGCGTTCGAGGCGCAATTCGAACTGCTGCCATTGATGACGTATCTCTACATGGTCGAC
AGCGCCTCCGTGGCGGAGATTCAACTGGTGTGCCAGCAGCTCGGTCTTGCCTACGACTCAGGCACGGAAGCGTTCCT
GAACCTGATCTCCAAGGCGCGGGCAACGCCGGATGAGCCGTTTGTCCGGCATGCACCCACGTTGAGGATGCCAAGGA
CGGTGGACGCGTTGAAGTTGAAGGATCAGTTCCTGCGCCATGAAGTGGACCGGTCGCTGCTCAAGCCTGAGAAGAGC
GAATACTTCGATGTTTCTGGCTCCTTCAAATACAGCTCGATGGTCACGCTCAAGGACGCCGAAATAGAACAGATTCA
GGACCGTCTGGTGGATCTGTTCTCCGTCATTGAAGTGCACAACAAGTCAGGCCGTCAGCTGGAGGACTCCGGCGCGC
AGCCTTATTTCTTCGGGCTGGTCGTTTCCGGACGTCCGGAGTACGACGGGCGTGCGGCCCGAGCCAGTAAAAAAGCC
GAGAAGTAA

NCBI Accession: EJZ60178.1; Uniprot ID: A0A7U9CRC9 (*Pseudomonas fluorescens* R124 PfrmG, native sequence)

ATGAATGCTCCAGTCGTAGAGAAACTCCGTCAGTTGTTCGGCGAGCGTTTAAGGCTCGAACAGGGGATCCGGCAACA
GTTCTGCCAGGCGCCTCCGGGTATCGAGCTGAGCCCTCCCGATGCCGTGGTGACGGTTCACAGCGAAGACGAAATCG

TTCAGGTCGTGAAGCTCTGCCGTGAGTACGGTGTTGCGCTGATCCCGGCGGGCGCGCGTTCATCCCTCGAAGGGCAA
ATCAATGCGCCACAGGGTGGCGTCTGTGTCGACTGCTCGGAAATGAACGCGATTGTCGATGTGAACGTCACCGACAT
GACGGCCACGGTTCAGCCGGGCGTCACTCGCGAACAACTCAACAGCCATTTACGTGACACCGGGCTGTTCTTTTCCG
TCGACCCCGGCGCCAACGCTTCCATGGGCGGGATGGCGGCAACCCGTGCTTCGGGGACCAACTCGGTGCGCTACGGG
ACGATGGCCGGGAATGTCAAAGCCGCGCGCTGGATCATGGCGGACGGCTCAGTCGCCGCCATCGGTTCCCGCGCGCC
GAAATCGGCCATGGGCCTGGATCTGTTGGCGGTTGCCGTAGGTTCAGAGGGCACGCTGGGCATCTTCAGTGAGTTGC
AAGTCCGGCTGCATCCGGTGCCCGAGGCCATCGTTGCCGGTGTCTGTTCGTTCGACTCGGTGGACAGCGCGGTACAG
ACCGTGGCCAGCGCCCTGGGCCAAGGCCTTTCGATTGCCCGGATCGAGCTGCTGGATGTCACGGCTGTGCGTGCCTG
CAATGCCTATTCCCCGGCATTGGGTCTGGTTGAAACGCCGCATCTTTTTCTGGAATTCCATGGCGACCGGTCCACGG
TCGCCAGCAGTGTCGAGCTGTTCAAGGATTGCGCCGGTGAGTACGGCGGGCAGGTCAGGATTTCCGAGCGTCTTGAA
GAACGCACGGCCCTGTGGAAAGCCCGACATGACGCCTGGTGGGCATTTCACGCGTTCTACAAGGGACGCAAGGGGCT
GGCCACCGATGTCTGTGTTCCGTTGTCTTGCCTGAGTGCCTGCATCAGCGCCACTGCGGCCGACATTGCCGCCTCCG
GTCTTGACGCGCCGCTCATCGGCCACGTCGGTGACGGTAACTTTCACCTGCTGATCATGCTCGATGAAGGTGAGGAC
GAAAGTGCGGAAATTCAGGCGTTTCTCAGTCGACTGTCGAGTCGCGCGTTGTCATTCGGCGGCACCGTCAGCGGAGA
ACATGGGGTCGGTCAGGGCAAGTTGAAGTGGATGTCGGCCCAACACGGGGCGGCGCTGCAGTGCATGGCGGCCATCA
AATCGGCACTGGACCCGGACAACCTTTTCAATCCCGGCAAGATGTATCACGCGCTGCCTGTTGATGCTCGGTAA

NCBI Accession: EJZ60177.1; Uniprot ID: A0A7U9GU49 (*Pseudomonas fluorescens* R124 PfrmG, native sequence)

ATGCGCCATATTTTGATAGTGGGTATTAACAAGGTCTCCATCAAGTACGTGTGTGAGGCGATCCGCAGTGCAGGCTT
TGAACCGATACTGTCCGTTGACCTTTCGGGTTTTCACGGCGCGGCGAAAAACGTATTTGAAGGCGTCACCTGCATTC
CGGCCTTTAGCGACAAGGACGCGCTGTCCGGTTACTTGAAGTCCGAGCCGGCGCTGGTCGGAATGATCCATAGCATC
ACGACCTTCTTTGACGAGTTGTTCCCGGTTGTCTGCGCGATTGCCCAAGAGTTCGGCTTCAACGCGCCACCCGCGGT
GTTTGCGCAGCTGTCATCCAAGGAATACGTGGGCAGCCTGGTTCCCGAGCACGTGCCAGCCGAGTCTTGCGTGTCGC
TGGCCGATGGCGACTTCACCTTGCCCTGGCTGACACCCGGTCAGCACAACGATGTGGTGCTCAAACCTTCGGTCGGC
TCCGGTGCGGTGGCAACCACGCTGCTCTCGATTGCCCCGGACGCTGACCCTGCCGCTGTGATCCGCTCGGCCATCCT
CGACAGCGGAATCGATGACCCGCAAGGCATTGCCTGGATTGTCCAGGCGCATTGCGTGGGTGATCTCGTCAGCCTGG
AGGGCTTCGCCCAGGATGGCCGAGTGGTTTTCCTTGGCCTGTCACGGCGTGATCGTGTGGGATTCACGGAAGTCGCC
AACCAATTCCCCGCCGACGCCCGGATTGCCTCGCCCGTGCGTTCGAAAATCGAGACAGCGGTCCGTGATCTGATTGC
CCGCTCGGCCTTTGACAATGGCTTCTTCCATTGCGAGTTCATCGCGAACGAATCCACGGCTTATCTGATTGACGCCA
ACATGGGGCGTCTGGGCGGCGCGAGCATCGTCGAGCAGATTGCGCTGGCCTACGATCTGGCACCCGCGACGGTTCTG
CAGCACGTGGCGCTGCTGCCACTGGGCTTGCTGGAAGCTCCGCTTCAATACCAGCCTGTCGAGTCCTGTCGCCAGAC
CTTGAGCTACTGGTATTGCCTGGATCACGAGGCCATCGTGCATGACTGGGAGATTCCGCCAATGGCCTCCATTCATA
CGCCCATCGCCAACGTCGGCAGCGTGATCCAGCCCATCGGCACGTCCGACTACGCCTGGATCGGCATGCTCACCGGC
TATGCCGATGTGGTTGTGCGGGAGATTGGCCAGGTACGTGTCCTGACCGATCGTGGCCCTAGAAGCCCGGTCTTCAA
ATGA

Table A1.3. Strains, plasmids, oligonucleotides, and gene sequences for Chapter 4. (A) Strains and plasmids. (B) Oligonucleotides used for plasmid construction and sequencing. (for T7 tRNAs, **bold** indicates tRNA sequence and non-bold indicates T7 promoter). (C) Sequences of genes.

A1.3A. Strains and plasmids

Strain	Description	Source
E. coli DH10b-T1^R	*F-mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara-leu)7697 galU galK λrpsL(StrR) nupG tonA*	Thermo-Fisher
E. coli BL21 Star (DE3)	*F-ompT hsdSB (rB-, mB-) gal dcm rne131* (DE3)	Thermo-Fisher
Pantoea sp. GM01	*WT Strain*	Dr. Dale Pelletier
Plasmid	**Description**	**Source**
pET28a.pTet	*ptet,* tet^R, Km^R, ColE1	*2*
pET28a.pTet-PfrmBGC	PfrmB-H from *PfR124*, *ptet,* tet^R, Km^R, ColE1	This study
pET16b-His$_{10}$-PhalA	His$_{10}$-PhalA (T7), *lacI*, Cb^R, ColE1	This study
pET16b-His$_{10}$-PhalD	His$_{10}$-PhalD (T7), *lacI*, Cb^R, ColE1	This study
pET16b-His$_{10}$-PhalF	His$_{10}$-PhalF (T7), *lacI*, Cb^R, ColE1	This study
pET16b-His$_{10}$-PantB	His$_{10}$-PantB (T7), *lacI*, Cb^R, ColE1	This study
pET16b-His$_{10}$-PantC	His$_{10}$-PantC (T7), *lacI*, Cb^R, ColE1	This study
pET16b-His$_{10}$-PantD	His$_{10}$-PantD (T7), *lacI*, Cb^R, ColE1	This study
pET16b-His$_{10}$-PantE	His$_{10}$-PantE (T7), *lacI*, Cb^R, ColE1	This study
pACYCDuet-PhalF-only	PhalF (T7), *lacI*, Cm^R, p15a	This study
pACYCDuet-PhalF-PhalA	PhalF (T7) and PhalA (T7), *lacI*, Cm^R, p15a	This study
pET28a.pTet-PantBGC	PantB-F from *Pantoea sp. GM01*, *ptet,* tet^R, Km^R, ColE1	This study
pET28a-AlnA	AlnA-His$_6$ (T7), *lacI*, Cb^R, ColE1	This study
pET16b-HamB	His$_{10}$-HamB (T7), *lacI*, Cb^R, ColE1	This study
pCDF-chAlaRS-proK-tRNA^{Ala}	chAlaRS (TAC), *lacI*, Sp^R, CloDF13	This study
pATLIC	*araBAD*, Cb^R, pBR322	*3*
pATLIC-mCherry (WT)	mCherry (arabinose), *araBAD*, Cb^R, pBR322	This study
pATLIC-mCherry (TAG)	mCherry-2TAG (arabinose), *araBAD*, Cb^R, pBR322	This study
pET16b-His$_{10}$-BesB	BesB (T7), *lacI*, Cb^R, ColE1	This study

A1.3B. Oligonucleotide sequences

Name	Sequence
PantB-Fwd	TCTAGAAGTGCTTTTTCAGGGCCCGCATATGAATAATCTGTCCACCATGAT
PantB-Rev	TCCTTTCGGGCTTTGTTAGCAGCCGGATCCTTAACGCCCGGAGATC
PantC-Fwd	TCTAGAAGTGCTTTTTCAGGGCCCGCATATGAGAAAGATTTTCCTGCAGATCG
PantC-Rev	TCCTTTCGGGCTTTGTTAGCAGCCGGATCCTTATTCTGCACATAGTAAATCCCT CCTG
PantD-Fwd	TCTAGAAGTGCTTTTTCAGGGCCCGCATATGTGCAGAATGATTGTGGCG
PantD-Rev	TCCTTTCGGGCTTTGTTAGCAGCCGGATCCTTACTCCCCTAAAGGAAATTCATG AAGT
PantE-Fwd	TCTAGAAGTGCTTTTTCAGGGCCCGCATATGGCTATTTTCAATCAGATACGCAT
PantE-Rev	TCCTTTCGGGCTTTGTTAGCAGCCGGATCCTTAGCTCCAGGCGATGGCG

PantBGC-Frag1-Fwd	CGAGGCCCTTTCGTCTTCAGGATCCTCTAGAAGAAGGAGATATACATATGAATA ATCTGTCCACCATGAT
PantBGC-Frag2-Fwd	ATGTGCAGAATGATTGTGGCG
PantBGC-Frag1-Rev	CGCCACAATCATTCTGCACAT
PantBGC-Frag2-Rev	CCGGATCTCAGTGGTGGTGGTGGTGGTGCTCGAGTCAGGCGGTTGCC
PhalA-Fwd	TCTAGAAGTGCTTTTTCAGGGCCCGCATATGAGTTTAAGTAATTTAGCAACCGC
PhalA-Rev	TCCTTTCGGGCTTTGTTAGCAGCCGGATCCTTAAGACCAGATTTGTAATGCATG GG
PhalD-Fwd	TCTAGAAGTGCTTTTTCAGGGCCCGCATATGACCAATATTAAAAATATCGCTATT ATTGGC
PhalD-Rev	TCCTTTCGGGCTTTGTTAGCAGCCGGATCCTTAGGCATAGTATGAAAGTAACCC TTTTTTG
PhalF-Fwd	TCTAGAAGTGCTTTTTCAGGGCCCGCATATGACTACTGCGATTTTATTCTTGAG C
PhalF-Rev	TCCTTTCGGGCTTTGTTAGCAGCCGGATCCTTAACTCGCGTATTGCACTGG
AlnA-Fwd	
AlnA-Rev	

pCDF-chAlaRS-Fwd	GATCGGATCCCAAAGGAGGTGCGGCCGCATGTCCAAAAGTACAGCGGAAATT C
pCDF-chAlaRS-Rev	AATGATGATGATGATGATGGTCGACTTAAGTTGGCACCGGTGAATTTG
mCherry-WT-Fwd	CTAACGGTGTCGCGCCCATGGTGAGCAAGGGCGA
mCherry-TAG-Fwd	CTAACGGTGTCGCGCCCATGTAGAGCAAGGGCGAGG
mCherry-Rev	GTTTTCGGTCGTTGGCCCCTTGTACAGCTCGTCCATGC
T7-tRNA-chAlaRS-Fwd	AATTCCTGCAGTAATACGACTCACTATA**GGAGCTATGATCATGTAGATCGAACG GACTCT**
T7-tRNA-chAlaRS-Rev	**TmGGTGGAGCTACCGGGAATCTAACCCGGCTGAACGGATTTAGAGTCCGTT CGATCTACATGATCATAGCTCC**

A1.3C. Gene sequences

NCBI Accession: WP_046219999.1; Uniprot ID: A0A0F5VEF2 (*Photobacterium halotolerans* PhalA, codon optimized sequence)

```
ATGAGTTTAAGTAATTTAGCAACCGCATTGAGCCAACAAAGCGAGAATATTATCCTGACGAATGAGTTGGAAAAGTG
GCATCTGGATGGGTTATTTAATCGCACCATTGATGTTGTGCGCGTACCAGGATACTTTAGCACTCGTAGTGCTACCG
CCATTGCCAATAAAATCAAGCAATCTGCTTTTTTCGGAAGCTACGTTAATGCTCCAAAGATCGGTCGTATCGGTCAG
GCTTTCTTCGAATGCCAAAACGACGAGGTAAGCTTATCCCGCTATCGCGAATTTGCCAAGGTATGGATTAAGGAGAT
GCGTAAAGAAATCTCGCCCTATCTGTCCCCCATTGATCGTTTGCGCCTTGAACTTGATGAAGTGTGGCCATCGCATT
GTTCGTTGGCGGAAATCGGGGGTCATAAGGCGTTTGCTGGATTAGTTCGTGAATTTAAAGCGGGAGCTTATGCGGAG
CCTCATAATGATGTCCTTAGCTGGGACTTGGTTAATGAGATGAACACAGATATCACGAACCAGTTAGCGGCTAACGT
ATATCTGGAAACCTCAGAAATTGGAGGCGAATTGAAACTGTGGCATGAGTGGCCAGCGACAAAGGAGGAGTATGATT
CCATCCGTATTGAAGGTAGCTACGGAGTTAAAGAGGAGCAGCTTAATGATTCGTGTTTTGAAATCACCCCGGAAACG
GGTGAGTTAATTTTGTTTAATCCCATGCGTATTCATTCCGTGAACGAAATCCTTCAGGGCTCGCGTCTGACCTGGAG
TTGTTTCATCGGTTTGGAGAAAAAGACCCATGCATTACAAATCTGGTCT
```

NCBI Accession: WP_046220002.1; Uniprot ID: A0A0F5VE59 (*Photobacterium halotolerans* PhalD, codon optimized sequence)

```
MTNIKNIAIIGAGPSGISAYLQLVKELHNDLTSITIFDPIGIAESFSFNTDLASSLNNTSVGVSSLYSDDKLDYFKW
LQDKKSDLNVSSSDFISRYYFKEYCQDRFWQSKDYADAFGCSTFVENSEVHKLAFKNGRMIVTSQNDIDFFFDAVIL
ATGVTTTTPGGLDNSCTNLITNPYPESLFLQRVNNDSKVLILGSKLSAIDVSVAIHNKYPNAQMTMVSRSGELPSVR
SSLLIKSTVPQDYNILSRLSASSCADKINAEYKLEQDIACCKSSLNHWEDIVGQFIEEFNEQIPTMTPDRQLRLMSK
YQYFIKHYVSSFPLANAEIVLQAIKSGNLTILKKDISDYIVVKDKHIFGKERDRLKNFDVVINASGISHKPIDDTLI
```

HSLSQLSVTKNKAGGLKVSSSSMRVLSENDSVPLYVIGGPAAGEVPITNYVRSSVTQAKKIVSDIVRLTRCSKKGLL
SYYA

NCBI Accession: WP_046220004.1; Uniprot ID: A0A0F5VEF6 (*Photobacterium halotolerans* PhalF, codon optimized sequence)

ATGACTACTGCGATTTTATTCTTGAGCACACGTCAACTGACCCAGGTGGACTTCTCACTTCCATTGCTGGGTGAACA
CGATTTGATCCTTGTAGCATCGAAACAGGAATTAGATCGTCTGCCTGACAACATTTCTTCCCACTTCACACTTCAGT
TCCCAGTCAGCATCTCAGCCGACGACGGGGTTATCATTGAGTACGTTTACAATGAGATCGTAGAGATTGTTGCTTCA
ATCCGTCAGTCTTACTCTACAGTTGATTTCGTCTGTTTCGACGAAGGCAACGTGGAGTTGGCTGATTCTTTGCGTCA
AAATTCGAAATCTAACCAGAACTTAGAGCGTTTTCGTGATAAGGTGGAGATGAAGAAGTGGCTTAAAAAGCGTAATA
TTCCTACCCCTTTCTTCATCGAGACAATCAGTGAAAATCAACGTTTCGACGAAATTTCCTCAAAGGTTGGAACCCCC
TTCGTAGTAAAACCAAAACGTTCCGCTGGATCCAATAATATCTATATCATTCGTAGTAAGGATGATTTTCTGGAAGT
GCAGAATACTGCGGCGGGATCACTTTCGAATTATGAGGCCGAAGAATACATCAATGGGGACCTGTACCATTGTGATA
TTGCGTTATGGATGGGGGAGTCTATTTTTGCTGAATGTACTGAATATCTGTGCCCAACTATCGACTTTCAATCAGGT
TCGCCTTTGGGGGGTCGCGTGATGGACACTAACTCCCCACTGCGTGAGCAACTTATTGAATTTGCCAAGGTTAGCTT
AAAGGCTTTGGAGAGTTCTGATGGTGTTTATCATATGGAGATCTTTGTGCGTAATAACGATGCTAAAGATCTTGTCT
TCTTAGAAGTTGGCGCTCGCCCCCCGGGAATGCTGGTAACAACGATGTATGAGAAAGCAACTGGGGTCAATCTGTTA
AATCTGGACATTATGATCCAAACGCGTTCAGTTCCCAACGATTTTTCATTCAGCCGCAAGAAACACGCTTTCTACCT
TGTATATCCCAAGCGTAACGGCCGTGTGCAGGCTTTGAACACGCCGTCAAATCACTGCGATATTGAGATGAACTTTG
TCATCCATTCAGTCATTGGTGAGCAGCACCAAGGTTGCTTCTCGAATCTTGATTTTTGCGCGTTCATCACTTGTAGC
GCGGTTGATAAGGCGGAGATCGACTCCGCCTATCAGTATATGTCTGAGTTCAAGCCAGTGCAATACGCGAGT

NCBI Accession: WP_007893500.1; Uniprot ID: J3CZW3 (*Pantoea sp. GM01* PantB, native sequence)

ATGAATAATCTGTCCACCATGATGAGAGACATTGAAGGATATCTCACCCCCACGATTATCGATTCAGCTCAGCACTG
CGCACGCCGTTTACTTGCCGCCCTCCCCGATGCCGCACAGCCGGAAAATAACCGCGTGCTGCTGGCCTACGGCGGCG
GCAAAGACAGCAGCTACATGGTGGCATATGTCCGCTACATTCAGGGACTGATCTGGCAACAGCACGGCAGCACCTTT
CAACTGCGCATCAGCACTAACCGCCATGCGGCGATGAACGACAGCGTGATGGAGAATATCGATCGCGTGTATCGCGC
GTTGCATATCTTTGAAGATGATTTCACCGAATGTCTGGTGACTGATGGCCTGCAAATCCGCCACTTCGATCGCCATC
TACCCATGCCGGACAGCGTGCGCCAGCAGAATCGCAACGATGTGCTGATGAATGGCCACCGCTTCCGCGCCGATGCC
CGCTCCACCTTTTGTAATGCCTGTAACCTCAGCATGGTGAACTCGTTTGGCCTGGCGGCGCATTACGACGGCGGCGT
TGATGTGATCATCACCGGCGATTCGCCGCAGGAACAAACCGCTTATCTCGCCTGGGCGCGGCATCTTTCGCGTCTGT
TTGACGCGCCGGTGGCGGATAAATCGCGCGGCTTTAGCGGCTTTTTGCAGACGCTGGATGGCGTGGCCGAGCGCTAC
TTCGCCGATATTTACGGCGCCGATCAGGTGACGCCCGCGCACCGCATTACCTGGCAACTGGCACGCGATCCACTGTT
CTTCAACATCTATCAGGACACCGCCTATGAAGCCGGTGCGCACTGGGCGCTGCTGACTGAATTTATGCATTTCCGCT
TTGATGAGCTGATGTTCAGCTTTTCCGAATCGGACTGCGGTAACCCAACGTTAATGGCACATATTCGCGGTCTGCGT
GCGGAGACGCTGCTGCAGCGTAACTACGCCGAAGGCATTCGCGAATATGTGCGCTTTGCGCTTGGGCTGATGGTGAA
AAAGGATTTCCCGCCGCAGCTGATCGATGAGATGGCGGCACGTTATCAGGACGATGAGGCGATTAAGCGCGCCCGCC
AGCGCGCGGCGCAGTTCGCCGAGCAGGCATTTGGTCTGAATGAAACCCAGCTGATTTGCATGATCTATTCGCCGTTT
ATCGCCGAAGGACAGCAGCTGAAAAACTGGCTGCAGCAGTTGGTCGATGCGCCGGATGAAGCTGCGGTGCGCGCGCT
GTTGCGCGGTGAAAGCGACGATCGGTTACTGGCACAGCAGCTTGAGCAACTTTCTGGTTTGACCGTCGCGCAGATGC
GCCACTGCTGGCGCAGCGCACAGGTGAGTTCACTGCTGGACGATCGGCGCGGCGATGCGCTCTCACGCGTGCTGCGT
TTTGATCCGCACAAAGCAGTGATTCAGGTGAATAACGGCGTGATGCACGACGTCGTCACTGAAGTGATCTCCGGGCG
TTAA

NCBI Accession: WP_007893502.1; Uniprot ID: J2TSH6 (*Pantoea sp. GM01* PantC, native sequence)

ATGAGAAAGATTTTCCTGCAGATCGGCGCCACCCGCGACGGACAAAATCCGTACTGTGCGGTGGCCAAACGCGACGG
CTATTTTACCGTGCTGGCGGAAATGGGCGATTTCGTCGATTACCAATCGCTAAGCATTGCGTTGCCGTTCGATCTGA
TTGTGCGGCTCGATCGCCCGGAAAATCCGTGGGATGTGCTGCAAGCCTGTTTGCATGCCGGGCTGATGGAACATCCT
CATGTGGTGCTGGCGGGTTTTGAAGCTTACAACGCCAGCGCCGGTCGCGTACGCGAGCTGCTGGCGGGACCCGATGC
CGGGGCGAGCTTTATTCCACCGGATAAGTACGCGCAGCGCATGGCGTTGCGCAAAGCCTGGCAACGCTTTCCACAGC
CGGAGTTTCGTTTCTTCGCTTCGCCGCTCAGCATCCGCGACGCGCGTGATGCCTTGCTTTATCCCTGCGTGATCAAG
CCGGTTGACGGCGGCGGCGGTCTGGGTATCTGGCTGGTCAAAGAAGCCCGCCAGCTGGATGTGGTGATTGGCAAGCT
GGTCGATACCATGAACTACGGCGGGCGCGGCTTCAGCGGCTTTATCGTCGAGAGTTGGCTGGCCGGCGATGAGTATT
CGCTGCAAGGCGTGGTGGCGGACGGCCATGCGCTGGCGCTGACCTGCTGCCAAAAGGTGATCGAACATCACACCGAT

GCCGACGGCTGCGTCAGTTTTTACGAGTCCGGCCACGTGGCGGTCGCGGCGCAGCAGCTGCCTGCCAGCTTCTCCAG
CCTGATGAGTTTCTGCTGCGAAACCTTTGGCTATCGTCAGGGCGCATTTCACATCGACTTTATTGTAGTGAATGGCG
AACCGCACTTTCTCGAAATGGGCTTCCGCCTTTCGGGCATGGGCGTGGTGAATCTGGTGCAGGAAGTGACAGGCATC
AATTGGGCTGAAGTGGCGTTTCACATCGAAGCCGGTCACGGCATGCCGCCGCTGCATTTCCCGCTGCAGCCGCTCGC
GGTGGGCCGGGTACGCCTGCGCCAGCGCCTGCAGTATGACGCGGCAGAACGGTGGATCGCGCAGCATCAGCGCGGTC
AGCTGTTGCCGCCGCTTAATCTGCCGACGCTGGCCGTTTCACCGCGCTCTTCTCTCTATGCAGATTTAACCCGTCAC
GCCGGAATTCTGGCGACCTTACGCCTTGCAGCCGCGAACCGCGACGAGATTCTGGCGGTGTTCCATCAGATTCTTGA
GGTCTCATCGCTCACCACCAGGAGGGATTTACTATGTGCAGAATAA

NCBI Accession: WP_007893504.1; Uniprot ID: J2L912 (*Pantoea sp. GM01* PantD, native sequence)

ATGTGCAGAATGATTGTGGCGCACGGCCAATTTGACGCCGCCGCAGTGCTGGATGCAGCCCGCGCCATGAGCTGTGG
AGAAACCGCCACCCATGAAGGGCCGATAAAACAGCATCCCAACGGCTGGGGTTGTTTATGGCTGGAGCAAGGACACA
TTCGCACGCTGCACGGCACCGGCACCTTTGCCGATGCGCTGCCGACCATCGATGTTGACCGCCTGCGTGGCCGCTTT
CTGGCGGTGCATGTGCGCCACGCTACGCTGAGCAAAAATCACGGTATCGAATTTTCCCACCCGCTGTGGCGCACTAG
CGGCGCAACCCAGTGGTACATGATGCATAACGGGTTTTTGCCCACCGTCTACCAGCAGCTGGGCATGGCGGAGTCGC
ACTTTGATTCTGCCGAATACCTCGAATACATGGTTGATCGCGCCACGCCCGCGGATTTTACCCGCGATTATCTGCGC
AACAAGATGGCGCAGGTGGCGCCGGGCGGCAGCTCTGCCAACGCGATTTTTGTCACCCGCGATAAAGCCTGGGCATG
GCAATGGCATCCCGACGATACGCCCTATCCAAACTATTTCACGCTGCACCGTTTCCAGAATGACAACACGACGTTTA
TCTCTTCTGAACCCATTTCAACACTAGGCACGGCGTCTGAATGGCGTCGCATGAGCAACCATGAACTTCATGAATTT
CCTTTAGGGGAGTAA

NCBI Accession: WP_007893506.1; Uniprot ID: J3CZI2 (*Pantoea sp. GM01* PantE, native sequence)

ATGGCTATTTTCAATCAGATACGCATTATCAATCCTGACATGTTTGTGGAAAGCACCTGGCTGGATGACATCGTGAC
GGCGCGCGTGCCGCTGCTGGTTTTACGCAACTTTCTGACGCCACAGGTGCGGGAAGCCGTGGTCGACGATTTGCAGC
AGTGCCGCGAACAGGTCCGCGTATCGCACTATCCAAACGGTGCGTTGACCACGCTCGGCCCCTATCTGGCAAAGCAT
ACTGCCGCCCCGGAAAACTACTTCACCGAACTGCGCGATATTCAGTCGGCGCTGCCGCCGTCGCTGTACTGCCTGCG
CGATCAAATCTATAACTGGGTGAAGCACGCGCTGCGCCTGGAGAGTCTGCAAACCGCCAGCGAACCCGATCTGGGCG
ATTACGCCGGTTCGATTGTGCGCTTCCACGCCAATGGCGTCGCCAATCCTCTGCATAACGACAACATCATGCGCGAT
GCCGCCGGCAGCGGATTGGCGGTGACTAACATTTTGCATCAGCTGAGCTGCGTGGTGTGTTTGCAGGAGTGCAATGC
AGGTGGCGAGCTGCGCATCTTCAATAAGCAGTGGCAGCCGGTAGATGAACGTTTCAAAACGCCGGGCGAACTGGGTT
ATCAAAGCGGCGTCATCGCGCACAGCGAGGTGTGTGAGTTCTCACCGCGCAGTGGCGATATCTATCTGTTTAACCCC
GCTTTTTATCACGAAATTGATCGCGTGGAAGGCGACACGCGCATCACCATGGGCTTCTTCTTTGGCCTGACCGACAA
AAAAATGAAACACGCCATCGCCTGGAGCTAA

NCBI Accession: WP_055594934.1; Uniprot ID: A0A0N0S5E7 (*Streptomyces hirsutus* AlnA, codon optimized sequence) (capitalized portion is the Cupin domain that was cloned from the entire AlnA gene)

ATGACAAAGGTCCGCCATACTCCTGTCGCACCAACACGCACTCAACGTTTGGCTTCGGGTGCAGAGATCGATGCACA
CCGCCACGACGACCATCAAATTGTTTACGCTGGGCGTGGCGTCTTGAGCGTGACGACCTCAGCGGGTTCATGGGTGG
CACCAGCTACCCGTGCGATTTGGGTTCCGGCTAACACCGTTCACGCCCATCAAGCACACGGGGACTTGGAGTTGCAC
TTAGTTGGCTTGCCCGCCTGGGAGAATCCTTTACGCTTAGACGCACCCACAGTCTTGTCTGTGAGCTTGCTTCTGCG
TGAGCTTATTATCGCTTATACTCGTACAGAGGACGACGATAGTCCGGAACGCATGCGCCTTCATGCGGTGCTTCTTG
ATCAGCTTCGTATTTCTGTGCAACAGCCAATGCATCTGCCTACGCCTAGTGACCCCTTATTAcgcaccgtctgtgac
ctgcttcgcgcggacccctcagacaatcgtactcttgccgaattagggcgcgaagtaggcgcgtccgaccgtacatt
atcgcgcttgttcaacaaagatttgggaatgacgttcccccagtggcgcactcagttgcgtttacaccgtgccttag
tgttgttggccgaacagactcctgttactaaggtggcgaatcagtgtggatggagcagtaccagcgcatttatcgat
gttttccgccgtgcattcggccataccccaggaacgcattcccgt

NCBI Accession: WP_006490841.1; Uniprot ID: A0A1V2W1T2 (*Burkholderia* HamB, codon optimized sequence)

ATGAACGCTCGCGGACTTTCGGAACGTATCGTAGATTTCGCGGCTTTGGAATGGGAGCACCCCGCGGCCTTTGATGA
CGGTATTGATCAGGTCGTCCTGGTCGATACCTTAGATGAACACGCAAAGCGTGGCATCCGTACTCGCATTGTGCGTT
TCGATCCGGGGGCAGGCACTCGTGTACCATTCCTTCACGATTATCACGAAGAAGTCCATTTGATTGACGGCGATCAG
TCATTGGTCGATGTTGCGACCCGCGAGTACCGCGCAACTTATCGTGAGGGTTCGTACTTCGTTCGTCCTGCTGGTAC

ACCACACGGCCCATTTCGTTCTGATGCAGGGTGCCTTTTACTTGAGATTCATTACTACGGTAGTCCAGAGGATACAC
GC

chAlaRS (WT Sequence)

ATGTCCAAAAGTACAGCGGAAATTCGCCAAGCCTTTTTGGACTTTTTCCACTCTAAGGGTCATCAGGTGGTCGCATC
CTCCAGTTTGGTCCCCCATAATGATCCTACTCTGCTTTTTACGAACGCGGGCATGAATCAGTTTAAGGATGTCTTCT
TGGGTCTTGACAAGCGTAATTATTCCCGCGCCACTACTTCTCAACGCTGCGTTCGTGCCGGCGGAAAACATAACGAT
CTGGAGAACGTCGGCTATACAGCTCGCCATCACACTTTTTTCGAGATGTTAGGCAATTTTAGTTTTGGAGACTACTT
TAAACATGACGCGATTCAATTTGCTTGGGAGTTGCTGACCTCGGAGAAGTGGTTCGCACTTCCCAAAGAGCGCCTGT
GGGTTACTGTCTATGAATCAGACGACGAGGCATATGAAATCTGGGAAAAGGAAGTAGGCATTCCCCGCGAGCGTATC
ATCCGCATTGGAGATAACAAAGGGGCTCCATACGCCAGTGACAACTTTTGGCAGATGGGAGATACAGGACCATGCGG
CCCCTGCACCGAAATCTTTTACGATCATGGCGACCATATCTGGGGAGGTCCGCCTGGTAGCCCAGAAGAGGACGGTG
ACCGTTACATTGAGATCTGGAACATTGTTTTTATGCAGTTCAATCGTCAGGCGGATGGAACGATGGAGCCGCTGCCC
AAACCTTCGGTTGACACTGGCATGGGGTTAGAGCGCATTGCAGCCGTATTACAACACGTTAATTCTAATTATGATAT
TGATTTATTTCGCACCCTTATTCAGGCGGTGGCGAAGGTAACGGGAGCAACTGATCTTAGCAATAAGAGCTTACGCG
TCATCGCAGATCACATCCGTAGCTGCGCGTTCTTGATCGCGGACGGGGTGATGCCGTCCAATGAAAACCGTGGCTAT
GTTTTGCGCCGCATCATTCGTCGTGCAGTGCGTCACGGTAATATGTTGGGTGCCAAGGAGACTTTTTTCTATAAGTT
AGTAGGTCCATTAATCGACGTGATGGGTAGCGCAGGCGAAGACCTTAAACGTCAACAGGCCCAAGTCGAGCAGGTAT
TGAAAACCGAAGAAGAGCAATTCGCCCGCACGCTTGAACGTGGGTTGGCTCTGCTGGACGAAGAGCTTGCTGGTGGT
GGCAGCGGTGGTGGCTCCGGAGGAGGGTCGGGCGGTGGTTCAGGAGGAGGCTCAGGAGGTGGAAGCATGGACAAAAA
GCCTCTGGACGTATTAATTTCCGCAACTGGGTTATGGATGTCACGCACGGGGACTTTGCACAAGATCAAACATTACG
AGATTTCCCGTAGCAAGATCTATATCGAAATGGCTTGTGGGGACCATTTGGTAGTAAACAACTCACGCTCATGCCGC
CCAGCGCGCGCATTCCGTTATCATAAGTACCGTAAGACGTGCAAGCGCTGCCGCGTTAGTGACGAGGACATTAACAA
CTTTCTGACTCGCTCGACTGAAGGGAAAACTTCCGTGAAGGTAAAGGTCGTGAGCGAACCCAAAGTCAAGAAGGCGA
TGCCCAAGTCAGTATCACGCGCCCCAAAACCGTTGGAGAATCCTGTTAGTGCGAAAGCATCGACCGATACCTCGCGT
AGTGTACCTTCTCCCGCAAAGAGCACCCCAAATTCACCGGTGCCAACT

mCherry (WT Sequence)

ATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTC
CGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGA
AGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTAC
GTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAA
CTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGC
TGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGG
ATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGC
TGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACA
TCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATG
GACGAGCTGTACAAG

mCherry (N-term TAG Sequence). Amber codon is in **bold**.

ATG**TAG**AGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTC
CGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGA
AGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTAC
GTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAA
CTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGC
TGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGG
ATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGC
TGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACA
TCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATG
GACGAGCTGTACAAG

NCBI Accession: WP_030410683.1; (*Streptomyces sp. NRRL 1448* BesB, codon-optimized sequence)

ATGACCACTGAGGCCAGTGGGGAATTGGGAGCCGCCTGTTCTTTGCGTCATACGGCAGTCGGTCGTCCGGTACCGGG
TTCGCTGCACTCCGTAAGCGTCTCCATCCCAGATGTAGCAGCCGTAATTGGATACGAATCCGGGGACGCAGCGACCC
GTTCCTGCATCTCTTGGGGGTACCCCCGCTTCCGTACTCACCCTTACGTGACGCGTGTCGCAGGATTACTTGCGCAA
GATGGTATTGGACCTGCCAATGGACTGCTGTTGACCCGTTCTGCCCGCGCTGCGGGCGCAGCGGCGACCTATGCTGG

ACTTACTCCAGAAGCAGTTTTCGAGCGTAACGGTTTACACGGAGTGCGTTTACCGGCGGACGGTCCTTCGGCCGCTC
GCGCACGCGCTTATGTTCAACACACAGGAAGCCATCTTAGTTCACGTGAGGCTGAAGACATTCTGTTGGATGCGGGA
CTTATTGAAAACCGTCAGGCCGAAGCCGCAGTGGAAGAGGCTCCAGCAGAAGCTGTGCGTGCCGTCCTTGCCCAAGC
CTATGGAGTTCGCGATGCTGCAGATGTTTCCTTGCACAACTCTGGAATGAACGCTGTTGCTGCCGCGCTGGCAGCGG
TCTCAGGCATTCAGCGCGAGAATGGCCGCCGTCGTTGGTTACAGTTAGGCTGGATCTTTTTTGACACGATGTCCTTA
CTTGAGAAGCGTGTGATTGACGTGGAGCATACGACCGTGCCAGACCCGTTCGACCTGAGTGAAATCGCTCGCGTGGC
AGACGCACACGCAGGCGAATTAGCTGGTATCATCGCGGAGGTTCCTTCCAATCCCAGCCTTCGCACACCTGATATTC
CAGCGCTGCGCGAAATTGCCACACGTGCAGGCTGTGCTTTAGTTGTAGATGCCACCATTGCAACACCATATAACGTA
GAAGTGCTGCCGTACGCCGATGTTGTCTGTGAAAGTCTTACTAAATATGCGACTGGCTCGGCAGATGTCCTGATGGG
TGCGGCTGTCGTCAATCCCGCGAGTTCCTTTGGACCTGACTTGCGCACGGAGTTGCGTCGTCACGGTGATGAACCAT
ACCATCGTGATACTGCACGCGTGGCCGCGCGCATCCGCGGTTATGCCGAACGTATGGACCGTGTAAACGCTAACGCT
CTGGCCTTAGCTGCTTGTTTAGCTCGTCACGATAATGTGGTCCGTGGTGTCGGATGGGCATACGATGCAGCCTCGCG
CGGCAATTACCGCAAGGTAGAGCGTCGCCCCGACGCCCCTGGTGGACTGTTGATGGTGGACTTGAAAGTGCCTCTGG
AGCAAGTCTATGATCGTCTGGCAGTTGCTAAGGGCCCTTCCTTTGGGGCCGAATTCACTATGGCGAGTCCCCAAATT
TTCATTGCACACTACGACTTGCTGTCTACACCGGAGGGTCGCGCAGCACTTCGTGCTCGCGGTTTGCACCGTGATAT
GCTGCGTATCTCTGTTGGAACTGAGGAGCCCGAACGTATTGTTGAAACGTTTGAGCAAGCTCTGCGTCCCTTGTGA

pCDF-chAlaRS-proKtRNA^{Ala} (whole plasmid) (chAlaRS and tRNA are in bold)

cctcaggaagatcgcactccagccagctttccggcaccgcttctggtgccggaaaccaggcaaagcgccattcgcca
ttcacgctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatg
tgctgcaaggcgattaagttgggtaacgccagggttttcccagtcacgacgttgtaaaacgacggccagtgaatccg
taatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcat
aaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagt
cgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccag
ggtggtttttcttttcaccagtgagacgggcaacagctgattgcccttcaccgcctggccctgagagagttgcagca
agcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttgacggcgggatataacatgagctg
tcttcggtatcgtcgtatcccactaccgagatatccgcaccaacgcgcagcccggactcggtaatggcgcgcattgc
gcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcattcagcatttgcatggtttgtt
gaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatatttatgc
cagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgc
gaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagaga
catcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagtta
atgatcagcccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctac
catcgacaccaccacgctggcacccagttgatcggcgcgagatttaatcgccgcgacaatttgcgacggcgcgtgca
gggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgccacgcggttgggaatg
taattcagctccgccatcgccgcttccactttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgcg
ggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttcacattcaccaccctga
attgactctcttccgggcgctatcatgccataccgcgaaaggttttgcaccattcgatggtgtcctggcacgacagg
tttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtgagttagctcactcattaggcaccccaggcttt
acactttatgcttaaacaggatcgatccatcgatgagcttactccccatcccccctgttgacaattaatcatcggctc
gtataatgtgtggaattgtgagcggataacaatttcacacaggaaacaggatcggatcccaaaggaggtgcggccgc
ATGTCCAAAAGTACAGCGGAAATTCGCCAAGCCTTTTTGGACTTTTTCCACTCTAAGGGTCATCAGGTGGTCGCATC
CTCCAGTTTGGTCCCCCATAATGATCCTACTCTGCTTTTTACGAACGCGGGCATGAATCAGTTTAAGGATGTCTTCT
TGGGTCTTGACAAGCGTAATTATTCCCGCGCCACTACTTCTCAACGCTGCGTTCGTGCCGGCGGAAAACATAACGAT
CTGGAGAACGTCGGCTATACAGCTCGCCATCACACTTTTTTCGAGATGTTAGGCAATTTTAGTTTTGGAGACTACTT
TAAACATGACGCGATTCAATTTGCTTGGGAGTTGCTGACCTCGGAGAAGTGGTTCGCACTTCCCAAAGAGCGCCTGT
GGGTTACTGTCTATGAATCAGACGACGAGGCATATGAAATCTGGGAAAAGGAAGTAGGCATTCCCCGCGAGCGTATC
ATCCGCATTGGAGATAACAAAGGGGCTCCATACGCCAGTGACAACTTTTGGCAGATGGGAGATACAGGACCATGCGG
CCCCTGCACCGAAATCTTTTACGATCATGGCGACCATATCTGGGGAGGTCCGCCTGGTAGCCCAGAAGAGGACGGTG
ACCGTTACATTGAGATCTGGAACATTGTTTTTATGCAGTTCAATCGTCAGGCGGATGGAACGATGGAGCCGCTGCCC
AAACCTTCGGTTGACACTGGCATGGGGTTAGAGCGCATTGCAGCCGTATTACAACACGTTAATTCTAATTATGATAT
TGATTTATTTCGCACCCTTATTCAGGCGGTGGCGAAGGTAACGGGAGCAACTGATCTTAGCAATAAGAGCTTACGCG
TCATCGCAGATCACATCCGTAGCTGCGCGTTCTTGATCGCGGACGGGGTGATGCCGTCCAATGAAAACCGTGGCTAT
GTTTTGCGCCGCATCATTCGTCGTGCAGTGCGTCACGGTAATATGTTGGGTGCCAAGGAGACTTTTTTCTATAAGTT
AGTAGGTCCATTAATCGACGTGATGGGTAGCGCAGGCGAAGACCTTAAACGTCAACAGGCCCAAGTCGAGCAGGTAT
TGAAAACCGAAGAAGAGCAATTCGCCCGCACGCTTGAACGTGGGTTGGCTCTGCTGGACGAAGAGCTTGCTGGTGGT
GGCAGCGGTGGTGGCTCCGGAGGAGGGTCGGGCGGTGGTTCAGGAGGAGGCTCAGGAGGTGGAAGCATGGACAAAAA

GCCTCTGGACGTATTAATTTCCGCAACTGGGTTATGGATGTCACGCACGGGGACTTTGCACAAGATCAAACATTACG
AGATTTCCCGTAGCAAGATCTATATCGAAATGGCTTGTGGGGACCATTTGGTAGTAAACAACTCACGCTCATGCCGC
CCAGCGCGCGCATTCCGTTATCATAAGTACCGTAAGACGTGCAAGCGCTGCCGCGTTAGTGACGAGGACATTAACAA
CTTTCTGACTCGCTCGACTGAAGGGAAAACTTCCGTGAAGGTAAAGGTCGTGAGCGAACCCAAAGTCAAGAAGGCGA
TGCCCAAGTCAGTATCACGCGCCCCAAAACCGTTGGAGAATCCTGTTAGTGCGAAAGCATCGACCGATACCTCGCGT
AGTGTACCTTCTCCCGCAAAGAGCACCCCAAATTCACCGGTGCCAACTTAAgtcgaccatcatcatcatcattg
agtttaaacggtctccagcttggctgttttggcggatgagagaagattttcagcctgatacagattaaatcagaacg
cagaagcggtctgataaaacagaatttgcctggcggcagtagcgcggtggtcccacctgaccccatgccgaactcag
aagtgaaacgccgtagcgccgatggtagtgtggggtctccccatgcgagagtagggaactgccaggcatcaaataaa
acgaaaggctcagtcgaaagactgggccttgtttgtgagctcccggtcatcaatcatccccataatccttgttagaa
gcttttatttgccgactaccttggtgatctcgcctttcacgtagtggacaaattcttccaactgatctgcgcgcgag
gccaagcgatcttcttcttgtccaagataagcctgtctagcttcaagtatgacgggctgatactgggccggcaggcg
ctccattgcccagtcggcagcgacatccttcggcgcgattttgccggttactgcgctgtaccaaatgcgggacaacg
taagcactacatttcgctcatcgccagcccagtcgggcggcgagttccatagcgttaaggtttcatttagcgcctca
aatagatcctgttcaggaaccggatcaaagagttcctccgccgctggacctaccaaggcaacgctatgttctcttgc
ttttgtcagcaagatagccagatcaatgtcgatcgtggctggctcgaagatacctgcaagaatgtcattgcgctgcc
attctccaaattgcagttcgcgcttagctggataacgccacggaatgatgtcgtcgtgcacaacaatggtgacttct
acagcgcggagaatctcgctctctccaggggaagccgaagtttccaaaaggtcgttgatcaaagctcgccgcgttgt
ttcatcaagccttacggtcaccgtaaccagcaaatcaatatcactgtgtggcttcaggccgccatccactgcggagc
cgtacaaatgtacggccagcaacgtcggttcgagatggcgctcgatgacgccaactacctctgatagttgagtcgat
acttcggcgatcaccgcttccctcatactcttccttttcaatattattgaagcatttatcagggttattgtctcat
gagcggatacatatttgaatgtatttagaaaaataaacaaatagctgctttttcagcgcccgcataattccgcttcg
caacatgtgagcaccggtttattgactaccggaagcagtgtgaccgtgtgcttctcaaatgcctgaggccagtttgc
tcaggctctccccgtggaggtaataattgacgatatgatcagtgcacggctaactaagcggcctgctgactttctcg
ccgatcaaaaggcattttgctattaagggattgacgagggcgtatctgcgcagtaagatgcgccccgcatt**GGAGCT**
ATGATCATGTAGATCGAACGGACTCTAAATCCGTTCAGCCGGGTTAGATTCCCGGTAGCTCCACCAaattcgaaaag
cctgctcaacgagcaggctttttgcatgctcgagcagctcagggtcgaatttgcggatggctaagtatccggctgc
tctcactcggtcgctacgctccgggcgtgagactgcggcgggcgctgcggacacatacaaagttacccacagattcc
gtggataagcaggggactaacatgtgaggcaaaacagcagggccgcgccggtggcgtttttccataggctccgccct
cctgccagagttcacataaacagacgcttttccggtgcatctgtgggagccgtgaggctcaaccatgaatctgacag
tacgggcgaaacccgacaggacttaaagatccccaccgtttccggcgggtcgctccctcttgcgctctcctgttccg
accctgccgtttaccggatacctgttccgcctttctcccttacgggaagtgtggcgctttctcatagctcacacact
ggtatctcggctcggtgtaggtcgttcgctccaagctgggctgtaagcaagaactccccgttcagcccgactgctgc
gccttatccggtaactgttcacttgagtccaacccggaaaagcacggtaaaacgccactggcagcagccattggtaa
ctgggagttcgcagaggatttgtttagctaaacacgcggttgctcttgaagtgtgcgccaaagtccggctacactgg
aaggacagatttggttgctgtgctctgcgaaagccagttaccacggttaagcagtttcccaactgacttaaccttcg
atcaaaccacctccccaggtggttttttcgtttacagggcaaaagattacgcgcagaaaaaaaggatctcaagaaga
tcctttgatc

Appendix I References

1. Gust, B.; Challis, G. L.; Fowler, K.; Kieser, T.; Chater, K. F. PCR-Targeted *Streptomyces* Gene Replacement Identifies a Protein Domain Needed for Biosynthesis of the Sesquiterpene Soil Odor Geosmin. *Proc. Natl. Acad. Sci.* 2003, *100* (4), 1541–1546. https://doi.org/10.1073/pnas.0337542100.
2. Thuronyi, B. W.; Privalsky, T. M.; Chang, M. C. Y. Engineered Fluorine Metabolism and Fluoropolymer Production in Living Cells. *Angew. Chem. Int. Ed.* 2017, *56* (44), 13637–13640. https://doi.org/10.1002/anie.201706696.
3. Ko, H.-J.; Park, E.; Song, J.; Yang, T. H.; Lee, H. J.; Kim, K. H.; Choi, I.-G. Functional Cell Surface Display and Controlled Secretion of Diverse Agarolytic Enzymes by *Escherichia Coli* with a Novel Ligation-Independent Cloning Vector Based on the Autotransporter YfaL. *Appl. Environ. Microbiol.* 2012, *78* (9), 3051–3058. https://doi.org/10.1128/AEM.07004-11.

Appendix II: *Supplementary Tables*

Table A2.1. Minimal media amino acid supplement for comparative metabolomics experiments. The recipe for 1L of 5X amino acid supplement to M9 media for metabolomics experiments, taken from the *E. coli* genome project, which is sufficient for 5L of M9 media. Amino acids were dissolved as individual stock solutions, mixed according to this recipe, and kept frozen until use.

Amino Acid	FW (g/mol)	Weight (g)	Vol (ml)	Concentration (μM)	Notes
alanine (free)	89.09	1.78	25	799	
arginine (HCl)	210.7	8.44	100	401	
asparagine (free)	132.1	0.66	100	50	
aspartic acid (K salt)	171.2	1.71	25	400	
cysteine (HCl H_2O)	175.6	0.09	50	10	should be kept frozen
glutamic acid (K salt)	185.3	2.78	25	600	
glutamine (free)	146.2	1.8	100	123	
glycine (free)	75.07	1.5	25	799	
histidine (HCl H_2O)	209.6	1.05	25	200	
isoleucine (free)	131.2	0.65	25	198	
leucine (free)	131.2	0.66	100	50	
lysine	182.7	1.83	25	401	
methionine (free)	149.2	0.75	25	201	
phenylalanine (free)	165.2	0.83	100	50	
proline (free)	115.1	1.15	25	400	
serine (free)	105.1	4.2	100	400	
threonine (free)	119.1	1.19	25	400	
tryptophan (free)	204.2	0.26	25	51	
tyrosine (free)	181.2	0.045	25	10	dissolve in 0.01 M KOH
valine (free)	117.2	0.88	25	300	

Table A2.2 Proteomic analysis of BesG knockout strain. Tryptic peptides detected by MSFragger to contain at least one βes+2 mass incorporation at a Thr site in the BesG knockout strain.

Protein ID	Protein	Peptide Sequence
F8K3L8	1D-myo-inositol 2-acetamido-2-deoxy-alpha-D-glucopyranoside deacetylase	TGELAAAMAELGVTDHR
F8JUX7	3-demethylubiquinone-9 3-methyltransferase	YGEAGPGTPGTVMTVVFELAGQR
G8WUC8	3-hydroxy-acyl-CoA dehydrogenase	AATVPATGLAELAGAGLVIEAIAER
F8JU50	3-oxoacyl-ACP reductase	MSVSDWDTVLGVHLR
F8JJ25	4-chloro-allylglycine synthase	TDLNTPESTSKPVWEHFDHVEPGIR
F8JJ25	4-chloro-allylglycine synthase	HLYTETSINK
G8WX65	50S ribosomal protein L6	TLVANMITGVTAGYSK
F8JZN8	56kDa selenium binding protein	HPHPSAPATGLTGGPQMVEVSR
G8WUP1	ABC transporter protein	LSDTIMQNDVPAVPFAYDK
F8JZM5	Acetamidase/Formamidase	DYGLSELDAYQLVTQAVESPLANVCDANYTVVAK
F8K1P4	Acetyl-coenzyme A synthetase	ALGDVTTLTDSTVMELIQAK
F8K1P4	Acetyl-coenzyme A synthetase	VDDVMLVSGHNISTTEVESALVSHPK
F8JWN6	Aconitate hydratase	ATIGNMSPEFGSTAAIFPIDVETINYLK
F8JPI9	Acyl-CoA dehydrogenase	FMLADMATAIEAGR
F8JX35	Acyl-CoA dehydrogenase	HATNLESVLTYEGTVEMHQLVLGK
F8JX35	Acyl-CoA dehydrogenase	MDAGTLRPEQISFGK
F8JVA4	Adenosylhomocysteinase	EITLAEHEMPGLMAIR
F8JSA0	Agmatinase	LMTLGGDHTIALPLLR
G8WXQ8	alanine--glyoxylate transaminase	VFFTSSGTEATDTALLLATAHR
F8K2J3	Aminoglycoside phosphotransferase	HLTDVLHLLEI
F8K2J3	Aminoglycoside phosphotransferase	TTNPSAELLDNLLTMAGQTTAVR
F8JTB5	Aminopeptidase N	TGEVIDWESSIFGR
G8WQK8	Anti-sigma factor antagonist	IFRITGLTKVFPIHDSVEAAVSATD
G8WPM0	ArsC	SQMAAAWLTHLAGDR
F8JR89	Aspartate--tRNA ligase	VMTELFEEFGGGR
F8JQ89	ATP synthase subunit alpha	MSDDTIQALDEAVETFK
F8JPD2	ATP/GTP binding protein	MDYVTEADREPMVDTVQMHR
G8WYT4	beta-glucosidase	FEQTLVSEVR
G8X1X8	Beta-ketoacyl-[acyl-carrier-protein] synthase III	LGTGNPPAFDISAGCAGFTYGLTLAKGMVVDGTAK
F8JR58	Bifunctional protein (Histidine kinase and regulator)	ESETTVMPMPAEPARGTR
F8JQA9	Cellulose-binding protein	SDTSSPFGFELVR
F8JVH3	Chaperone protein DnaK	LGMFELTGLPPAPR
F8JVH3	Chaperone protein DnaK	MTVTGGSALPKDDIDR
G8WS40	Chaperonin GroEL	EQIASTASISAADTQIGELIAEAMDK
F8JZD2	Chaperonin GroEL	AMLGDMAVLTGATVVAEEVGLK
G8WS40	Chaperonin GroEL	KAMLGDIAILTGGTVISEEVGLK
F8JIZ7	chitinase	FTTMWSDLAALQR

G8WRR8	Citrate synthase	MTFAVPAEEYELDPVVVSALDK
F8JZ84	citrate synthase (unknown stereospecificity)	VIASTGADVAAALSGAVGAMSGPLHGGAPSR
F8JZD1	Co-chaperonin GroES	VVVQPLEAEQTTASGLVIPDTAK
F8JR06	Cupin 4 family protein	HYLAHTPPAATFTLAPGDVLWLPR
F8JR06	Cupin 4 family protein	DYLIGALLSLDPETAVK
F8K0X5	Cytochrome bc1 complex cytochrome b subunit	TNENVVGMPLLPVYMAK
F8K0X7	Cytochrome bc1 complex cytochrome c subunit	AIYNQAEIDQLAAYIASLGPGPATPTK
F8K394	Cytochrome P450 hydroxylase	NAITHGMHAFLTHPEQWELFK
F8JSH4	D-3-phosphoglycerate dehydrogenase	AGVMVVNAPTSNIVTAAELACGLLVSTAR
F8K302	D-alanyl-D-alanine carboxypeptidase	VMTALVVLDRR
G8XF20	Dihydrodipicolinate synthase	GIMTATALPLR
F8K306	Dihydrolipoyl dehydrogenase	HGELLGAHLIGPEVTELLPELTLAQK
F8JYJ2	Dihydropyrimidinase	AAAWGGTTTIVDFAIQSVGGSLR
G8WPQ5	DNA-binding protein	AAGLRREELAMLAGISVDYVTR
F8JYD2	DNA-directed RNA polymerase subunit beta	MCPIETPEGPNIGLIGSLASYGR
F8JYD3	DNA-directed RNA polymerase subunit beta'	LVDIGEAVGIIAAQSIGEPGTQLTMR
F8JYD3	DNA-directed RNA polymerase subunit beta'	LMAGTENPHDVLR
F8JR28	DSBA-like thioredoxin domain-containing protein	TVADFWFDPVCPWAWMTSR
F8JPR4	DUF885 domain-containing protein	HSGRPTEFLDSELIR
G8WZ01	Electron transfer flavoprotein	GVHVEDDALHGTDVMGTSLVLAK
F8JYD6	Elongation factor G	LGATPIVMQLPIGTEADFK
F8JTP9	Elongation factor Ts	HTAATKPADLAALLASEIESGK
F8JYD7	Elongation factor Tu	LLGLMHTIDEAIPTPQR
F8JYD7	Elongation factor Tu	TTDVTGVVTLPEGTEMVMPGDNSSMTVSLIQPVA MEEGLR
F8JQT8	Extracellular solute-binding protein family 1	PTMGIQFVDIPEFSDLGTK
F8JPG4	Fluorinase	RPIIAFMSDLGTTDDSVAQCK
F8JX40	Fluoroacetaldehyde dehydrogenase	FDGYDDAISIANDTLYGLGAGVWTR
F8JX40	Fluoroacetaldehyde dehydrogenase	VDLGGSLSGGYYVAPTIFEGDNR
F8K1J1	FMN-dependent monooxygenase	TPAMTAMTAATLDSLSGGR
F8JVI8	Fructose-bisphosphate aldolase	PIATPEIYNEMLDR
F8JVI8	Fructose-bisphosphate aldolase	DMVTGAVALAEFAHVVADKYPVHVALHTDHCPK
F8JYJ6	Gamma-glutamyltranspeptidase	LQMTWLEPGLPATLTPGR
F8JP94	GatB/YqeY domain-containing protein	QLTDEELGALVDEAVAEAAASGTEGPK
F8K1Z7	Glutamine synthetase	FCDLPGVMQHFTVPVQAFDPSEELAFDGSSIR
F8JYI5	Glutathione hydrolase proenzyme	DLYGMAPSSSGGTTVGEALNILSHTDLGK
F8JYZ8	glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)	GLMTTVHAYTNDQR
G8XGC1	Glycogen debranching enzyme	YWVTEMHVDGFR
G8WQF8	GPP34 family phosphoprotein	PTGDPTLDSALELLR
F8JUQ7	histidine kinase	NTVNTMVDQLSSFASEVTR
F8JQ26	histidine kinase	LMSTLGEELISNER

F8JUQ7	histidine kinase	NTVNTMVDQLSSFASEVTR
F8JUQ7	histidine kinase	TTINTMVGQLSAFGDEVTR
F8JX19	Homogentisate 1,2-dioxygenase	WSPPPAPEGPTDFVDGLVTVGGNGDVLQR
F8JZV3	Imidazole glycerol phosphate synthase subunit HisF	GTGIDAVEWAHRAAELGAGEILLNSMDADGTK
F8K2G9	Integral membrane protein	HIATNYAYDAHGADELSLR
G8X2W0	Lactococcin 972 family bacteriocin	HSATATIANAVDK
F8JYD0	Large ribosomal subunit protein uL10	IAANEAGISSLDDLLNGSTALAFVTGDPVEAAK
G8WX50	Large ribosomal subunit protein uL3	TSDASEYTLGQEIAADVFEAGVK
F8K234	Lipoprotein	SVTLVDLAEQVVR
F8JPL3	Lysine--tRNA ligase	TDTLGQVLAR
F8K090	Malate dehydrogenase	AAEGTAMELDDCAFPLLR
F8JM60	Maltokinase N-terminal cap domain-containing protein	GAPLDGADHALVGTMEHGVLGR
G8WXU0	Mannosyl-oligosaccharide alpha-1,2-mannosidase	TGAVSGSSVPLAEVGTNVMEFGLLTR
G8WXU0	Mannosyl-oligosaccharide alpha-1,2-mannosidase	LGDLTPAYWFAENPK
F8JTR7	methylmalonate-semialdehyde dehydrogenase (CoA acylating)	VTGQVDFASIAEVDQAVGAAVTAFR
F8K0D6	methylmalonyl-CoA mutase	MTDELEEAAVALMR
F8JML4	Methyltransferase domain-containing protein	EPTPETIEILGDFADLATTVDR
F8K4P1	Molybdenum ABC transporter, periplasmic molybdate-binding protein	TMDGLREETGTPTVIAK
F8JX81	Molybdopterin molybdenumtransferase	TLMGLEPVR
F8JZV2	Mycothiol-dependent maleylpyruvate isomerase metal-binding domain-containing protein	LLLADLLESAGPDAPTLCEGWTTR
F8JP11	Neutral metalloproteinase	HVLTTRPASAAAEGVAPQATAADHPER
F8JP11	Neutral metalloproteinase	HADPEVHRPAQYTLEVDAGHR
G8WX46	NlmB	HFAVPLPTTVICEMLGVPEADR
G8WS69	Nucleoid-associated protein SCATT_29380	AVDPEDTETLADLVLAAVR
F8JKN8	Oxidase	HLPFPGWTLPGVVAAGGAQAMLK
F8JW22	Oxidoreductase	YAAGGADVTGDIVDLLDPAATR
F8JSV6	Oxidoreductase	DLGAFMAAVTPDVMLR
F8JW22	Oxidoreductase	TDALDGAVIAVAGAGGPAGR
F8K1I6	Peptidase M20 dimerisation domain-containing protein	ALETSFDGALVEAMQTALR
G8XES1	Peptidase M55 D-aminopeptidase	NPDLADMATWITGVTR
F8JWB6	Phosphoenolpyruvate carboxylase	HVFDTIEAEHDLTVR
F8JJE0	phospholipase C	HGHGWYDLTVTCDTDAR
F8JVB0	Phosphomannomutase/phosphoglucomutase	QEMADTGAIFGGEHSAHYYFR
F8JP14	PknH-like extracellular domain-containing protein	PAAGMITAADLPGGGWVR
F8K3Q1	Polyketide cyclase/dehydrase	LPVLGAPLTR
F8K1S4	Prephenate dehydrogenase	TAPAAETPAAADKASDSAAEASAAEAPAR
F8K1Y0	Probable cytosol aminopeptidase	LGLPVNVTGWLALAENMPSGTATR
F8K1Y0	Probable cytosol aminopeptidase	TFGIMANDDAFR
F8JRC3	Probable enoyl-CoA hydratase EchA17	MVTAEEALAIGLVDR

F8JKP0	Proline racemase	HVFHAVDSHTEGMPTR
F8K2M4	Proteasome component	LETDPMSLAGELDWVAK
F8K2M4	Proteasome component	TLTAVQLQMEYFELAR
F8JL33	Putative acyl-CoA thioesterase	HLATVVQEGMIR
G8WZJ6	Putative ATP-dependent Clp protease	VSITDAALVAAATLADR
F8JNE1	Putative beta-lactam synthetase	HHELVFEPEQLTALLPEMIR
F8JNE1	Putative beta-lactam synthetase	FAGVTAVPAGTAVTLAVAGFDTEPLLTR
F8JNE1	Putative beta-lactam synthetase	GADDTLLVATSAGALLPLLGTAPVLGEGDR
F8JWN7	Putative extracellular solute-binding receptor	TWADFLSLGGEIK
F8JKY7	Putative FAD-binding dehydrogenase	MDADVIVVGAGLAGLVAAHELTSR
F8JZ46	Putative hydroxylase	DLGTMLVAADPGGAVFGVWQGGAHQGFGAK
F8JND9	Putative methyltransferase	HRFEAAEVTSPAEPVLTEQR
F8JNE4	Putative methyltransferase	HPNALTVLGGVGASPIAR
F8JNE3	Putative oxidoreductase	HDPFTPVLLTEALYEAGLPEHAVSVVHGDAPALR
G8XDE6	Putative oxigenase	GEPMTSFPLVGDLVR
F8JLE8	Putative phage tail sheath protein	PSYLTPGVYVEEVQSGAR
F8K0L4	Putative phytoene dehydrogenase (Phytoene desaturase)	LGAWLTEAGWPAPPVR
G8XGW0	Putative polyketide oxygenase/hydroxylase	TVLTEWFGAPLDD
G8WMZ9	Putative secreted tripeptidyl aminopeptidase	LLAHMTTLDTAR
F8K4N7	Putative sulfurylase	NEFLMGTALGDLPQDK
F8JXU1	Putative TerD-family protein	TINDLMGSIVPHL
F8K499	Pyruvate dehydrogenase E1 component	NMVEGMTDDQILHLGR
F8JNZ0	Pyruvate, phosphate dikinase	IMETLETHYR
F8JNZ0	Pyruvate, phosphate dikinase	GANLAEMTNLGLPVPPGFTITTEACK
F8JUQ4	Ribonuclease J	TPAGMVVHTGDFK
F8JYR8	RNA polymerase sigma factor	GLTMLDLVQEGNLGLIR
F8JPA9	RNA polymerase sigma factor	HRPVPHIDATDALEQADPMER
G8WNG6	S1 motif domain-containing protein	KATAPVGPPAVSEEPAAEAPAAEEAPAR
F8JJ36	Secreted protein	EAVTAIEGAVGAALVDYTSGMALGTIGGDK
F8K0C6	Serine hydroxymethyltransferase	MIPSENYVSQAVLEATGTVLTNK
F8JP36	Serine protease	IGGTSLAAPVISAIQALAQQAR
F8JYD4	Small ribosomal subunit protein uS12	PTIQQLVR
F8JTP8	Small ribosomal subunit protein uS2	WLGGMLTNFSTVYK
F8JYF6	Small ribosomal subunit protein uS3	NPETDAQLVAQAVAEQLASR
F8JYG3	Small ribosomal subunit protein uS8	VLGGLGVAIISTSQGLLTGQQAAK
F8JYG3	Small ribosomal subunit protein uS8	TMTDPIADMLTR
F8JPG5	S-methyl-5'-thioadenosine phosphorylase	TVAEVMAANVIAAR
F8JPG5	S-methyl-5'-thioadenosine phosphorylase	AWGMDVIGMTAQPEAK
G8WXY3	Solute binding transport lipoprotein	VMHFPTGGTTPEPDAAK
F8JYJ5	Sterol-binding domain protein	TPDLTLTMAAPELLK
F8JJ80	Streptogrisin C	TQMSATDDVLAIRPASK

F8K3J6	Succinate dehydrogenase flavoprotein subunit	HTYDVVVVGAGGAGLR
F8JRQ9	Sugar ABC transporter substrate-binding protein	VAMVTHGGPGDTFWDLVR
F8K4T8	Sugar transporter sugar binding protein	SLFMTAVLQGQDPAAAAK
F8K173	Thymidine phosphorylase	ELATTMVGLGTDHGVR
F8K173	Thymidine phosphorylase	HVVTAPATGVLTGLDAYAVGVAAWR
F8JNK9	Transglycosylase SLT domain-containing protein	TGASASPSASAPGTPAPTPSGMAAGNQDPSRFAK
F8K034	Trehalose 6-phosphate phosphatase	HTWVETAHIEHK
F8JRL0	Type-5 uracil-DNA glycosylase	PGDGLELLGTR
F8K2J4	Uncharacterized protein	SQLLNVDETTESAVEQLALDFIK
F8JJW6	Uncharacterized protein	VDTTPVMMSANPADISGGIVWAPR
F8JS38	Uncharacterized protein	TVALLASGPWLR
F8JTK6	Uncharacterized protein	HTALTNLVGAVK
F8K091	Uncharacterized protein	VHTQGGMLVLEFVPHVLPPVR
F8K2J4	Uncharacterized protein	TTIDTTAITVELPEAFDPR
F8JPS3	YceI family protein	VTAVIDMASVDSGDGTR

Table A2.3. Bioinformatic analysis of tRD co-occurrence with TauD-like oxygenases. Uniprot IDs of TauD-like proteins that co-occur with AlaX tRDs with **bold** indicating characterized TauD-like family members included in the phylogenetic analysis.

Gene IDs for TauD-like Proteins (Bold = previously characterized)				
A0A068R7I6	A0A354V2Q8	A0A7Y1U8U5	A0A2D9PTE7	**P37610**
A0A0G4MNG9	A0A354V3U1	A0A839VMU6	A0A2E1PCZ6	**P83309**
A0A0P1H0M6	A0A354V546	A0A856MCN9	A0A2E4UYB6	**P83310**
A0A0P6VFR0	A0A367FPC6	A0A8H6A2U1	A0A2E4XYG3	**P9WKZ1**
A0A168SAF5	A0A3M9XZM2	A0A8I2Z9A1	A0A2E8XSD1	**Q9WWU5**
A0A1G0FHP3	A0A4D8PSG2	B5XPQ2	A0A2J8E656	**A0A0E3USC3**
A0A1G0J532	A0A4P8QU60	C9SZ68	A0A2L1KSR9	**A0A4V8H042**
A0A1G6X5I6	A0A520XTR5	D3VCN5	A0A2N3KR91	**E3VWI8**
A0A1H0FNU9	A0A521MA23	Q4ZNC5	A0A2T0RFH8	**Q05581**
A0A1H3JV32	A0A521R3E4	A0A5R8Z9Y2	A0A2T3XXG4	**Q6WZB0**
A0A1I5XWK8	A0A522ZJ96	A0A6B0WKD2	A0A7X4LJB9	**Q96UB1**
A0A1L9R7J7	A0A537UDE8	A0A6B1E9G7	A0A7X5H3M1	**Q9I1L4**
A0A1L9R7L7	A0A538QCZ2	A0A6B3D1L9	A0A329VER5	**Q9XB59**
A0A1M7PQN6	A0A5B0WNY0	A0A6J6CFU3	A0A329X5M0	**Q9Z4Z5**
A0A1Q5TPF5	A0A5D0NGY0	A0A6N9EWV1	A0A7Y1TFG4	**A0A0A2IJP3**
A0A1Q5U120	A0A5E7AQX1	A0A7K1B3F2	**V5TD18**	**C7Q942**
A0A1T4KZS4	A0A5E7UAW5	A0A7K2A4F5	**A0A3B6UEU3**	**A0A0A2J5V5**
A0A1Y2SFF6	A0A5J5KAD2	A0A7K2AT71	**C8VSZ2**	**E2JA28**
A0A2A2PUC6	A0A5N6FZS7	A0A7U9CRC7	**O75936**	**P67756**
A0A2D0L901	A0A5N7BUE4	A0A7X0B2M7	**P10088**	**Q4W9F7**

Table A2.4. Proteins discovered through genome mining from Chapter 4. (A) Sequences identified from genome mining from Chapter 4. Nearest homologs are reported for proteins that have had their Uniprot entry removed from the database. (B) YbaK/INS Uniprot accessions for Chapter 4 Fig. 4.36.

A

Protein Name	Uniprot Accession ID
PhalA	A0A0F5VEF2
PhalD	A0A0F5VE59
PhalE	A0A0F5VEF6
PantB	J3CZW3
PantC	J2TSH6
PantD	J2L912
PantE	J3CZI2
CreE	A0A0J2KL70
HpyO	Q8PDQ6
AlnA	A0A0N0S5E7 (nearest homolog)
HamB	A0A1V2W1T2
DAP N_{β}-monooxygenase	A0A0N0FVR4
DAP N_{α}-monooxygenase	A0A0N1JFP2
DABA *N*-monooxygenase Apy 11	B1G5G9
DAP C_{β}-hydroxylase	A0A2G1XLZ8

B

Uniprot Accession ID of YbaK/INS tRD
A0A2V6SZS3
A0A2D6FRJ2
A0A2D7Z1P3
A0A1Y2NIP6
A0A2C8XM24
A0A1Q5AL21
A0A0M8SN63
A0A2D3E325
A0A5N5W9X4
A0A6N2HTB1
A0A1G8Y1G2
A0A0F5VIG1
A0A4P5RQQ3
A0A257RGP9
A0A0Q9LGF0

Table A2.5. Enriched metabolites from fuscazine BGC metabolomics. Enriched metabolites from overexpression of FuzA and FuzC-H with supplemented $^{13}C_6$,$^{15}N_4$-L-arginine.

RT (min)	*m/z*	Type	Signal/Noise	ANOVA *p*-value	Fold Increase
3.67	604.3123	[M+H]+	161493	0.07414	8099724
3.78	427.1856	[M+H]+	53852	2.45E-07	189.032
3.45	385.1758	[M+H]+	51406	3.94E-06	279.5775
3.78	411.1913	[M+H]+	42160	4.23E-10	332.6115
4.68	130.0856	[M+H]+	35242	7.78E-06	29.33825
3.67	594.3025	[M+H]+	35203	0.107073	875.94
5.29	583.2186	[M+H]+	30766	3.40E-07	13084.26
3.30	385.1753	[M+H]+	28018	0.520004	47.78965
4.12	363.1336	[M+H]+	24802	7.50E-05	1240111
1.91	270.1434	[M+H]+	23861	0.136063	37.3971
3.78	409.1751	[M+H]+	20418	3.62E-07	1074.116
3.69	363.134	[M+H]+	19170	2.79E-06	959629.3
3.60	620.3055	[M+H]+	18858	0.103766	303.8553
3.61	610.297	[M+H]+	15355	0.162168	119.1927
5.23	607.2434	[M+H]+	14625	0.000211	731238.9
3.74	388.1211	[M+H]+	13523	0.051667	199.6118
4.17	348.1201	[M+H]+	10441	0.019672	290.6913
3.63	590.295	[M+H]+	9873	4.40E-05	86.77592
5.02	322.2956	[M+H]+	8733	0.454792	146444.6
4.22	594.3535	[M+H]+	8435	1.82E-09	128.6802
4.16	326.1384	[M+H]+	7829	0.016698	89.04315
4.17	481.1976	[M+H]+	6943	0.008827	68.24628
6.09	282.2784	[M+H]+	6852	0.075827	34.17553
5.19	368.1855	[M+H]+	6216	1.45E-06	645999.6
5.18	146.1169	[M+H]+	5911	5.06E-06	82.14187
3.26	305.0087	[M+H]+	5718	5.78E-05	2941.695
4.17	380.0575	[M+H]+	4642	0.005724	225573.2
4.17	104.0704	[M+H]+	4427	0.019067	46.97865
2.31	177.1376	[M+H]+	4008	0.0002	37.26972
5.88	268.263	[M+H]+	3809	0.011516	336.4062
3.89	325.1556	[M+H]+	3797	0.038861	7.190962
4.20	445.1942	[M+H]+	3629	0.015584	210615
2.20	163.1219	[M+H]+	3544	1.91E-06	60.06951
5.84	342.2991	[M+H]+	2751	0.154932	11.49298
1.98	268.1756	[M+H]+	2687	0.49992	112515.5
3.80	402.2177	[M+H]+	2560	0.003465	7.320316
4.84	525.2378	[M+H]+	2407	0.500596	41643.43
5.89	368.3145	[M+H]+	2338	0.023045	16.33324
4.16	86.06014	[M+H]+	2141	0.019985	57.83809
4.96	385.1766	[M+H]+	2135	0.01209	11.66523

5.95	904.3092	[M+H]+	2130	0.000582	16.66034
3.64	672.336	[M+H]+	2101	0.231484	97.68811
4.71	359.3593	[M+H]+	2087	0.565245	417.3318
3.95	772.3994	[M+H]+	1951	0.0077	178.898
3.59	598.2964	[M+H]+	1881	0.047729	12.58908
4.88	607.2416	[M+H]+	1846	0.045064	2440524
4.71	353.3518	[M+H]+	1841	0.556744	179274.3
3.78	189.1229	[M+H]+	1811	3.35E-09	15.00091
4.16	705.1876	[M+H]+	1808	0.167936	25.25655
3.78	465.1428	[M+H]+	1783	4.52E-06	138066.2
1.93	145.0484	[M+H]+	1770	6.76E-06	15.30141
6.41	462.3017	[M+H]+	1754	0.069668	5.916421
3.57	539.2598	[M+H]+	1669	0.106337	11.0881
5.82	597.2757	[M+H]+	1631	4.28E-05	81556.04
3.63	589.2979	[M+H]+	1620	0.067494	29.36492
4.35	568.2431	[M+H]+	1576	0.520965	26274.65
3.26	276.9949	[M+H]+	1546	8.39E-05	176839.3
3.47	207.0988	[M+H]+	1527	2.06E-05	113354.9
4.55	753.5378	[M+H]+	1519	0.516063	59394.18
1.94	261.0757	[M+H]+	1485	1.02E-06	19.13572
4.85	568.244	[M+H]+	1476	0.469423	24598.24
4.92	666.2907	[M+H]+	1461	0.000303	77084.24
6.01	282.2778	[M+H]+	1414	0.075827	34.17553
4.41	406.1674	[M+H]+	1410	0.519552	27321.87
6.22	382.3308	[M+H]+	1386	0.026682	8.510197
1.93	85.02836	[M+H]+	1384	6.84E-06	17.16813
4.11	349.121	[M+H]+	1384	0.00478	330128
5.77	687.2796	[M+H]+	1354	0.01973	65310.69
6.41	976.4014	[M+H]+	1298	0.613414	34.78185
5.33	240.2308	[M+H]+	1289	0.006632	409.2407
4.52	708.4663	[M+H]+	1261	0.432276	21891.21
5.41	583.1908	[M+H]+	1257	0.021698	146.956
6.06	529.3669	[M+H]+	1253	0.123419	11.47568
1.98	431.2123	[M+H]+	1246	0.494162	36.27711
6.28	533.3721	[M+H]+	1239	0.343816	13.27274
4.17	130.0507	[M+H]+	1230	0.017282	9.009397
4.56	759.5453	[M+H]+	1224	0.520382	118.3114
2.06	182.0804	[M+H]+	1206	0.001272	18.96531
5.01	226.2154	[M+H]+	1179	0.010306	11.48457
5.21	336.1221	[M+H]+	1156	8.96E-10	57813.28
5.20	413.242	[M+H]+	1146	3.80E-07	6.793242
4.59	305.0081	[M+H]+	1143	4.81E-06	1432.564
2.01	182.0802	[M+H]+	1141	0.00134	18.96531
4.33	156.0647	[M+H]+	1136	3.09E-06	32.3845
3.37	377.6981	[M+H]+	1126	0.382309	167.5858

3.74	410.1033	[M+H]+	1121	0.042	56069.7
4.21	311.3037	[M+H]+	1107	0.498328	30985.13
4.17	156.065	[M+H]+	1089	0.011026	9.696875
3.59	596.2815	[M+H]+	1084	0.146073	131.9282
5.19	390.1799	[M+H]+	1077	6.49E-06	53870.35
4.56	345.3435	[M+H]+	1041	0.522978	105577.1
4.17	481.225	[M+H]+	1016	0.01124	22.93781
5.18	465.2413	[M+H]+	1014	0.291679	5.116294
1.95	501.1447	[M+H]+	1011	2.50E-05	57.58204
6.06	525.3655	[M+H]+	1010	0.419486	5.147125
3.92	427.1909	[M+H]+	1010	0.236321	35.81286
5.87	401.3638	[M+H]+	968	0.553466	16131.19
5.77	292.1327	[M+H]+	960	0.018237	57099.57
5.95	904.5573	[M+H]+	942	0.027634	40503.03
1.92	398.2023	[M+H]+	935	0.495199	75.92919
1.92	163.059	[M+H]+	935	1.77E-05	15.84759
0.96	285.1417	[M+H]+	934	0.128617	59.04119
1.93	499.1621	[M+H]+	929	9.75E-07	36.31974
5.68	660.2441	[M+H]+	925	0.000329	43049.84
1.92	531.0995	[M+H]+	918	3.39E-05	153.3218
4.17	282.149	[M+H]+	908	0.024256	6.723564
1.94	203.0731	[M+H]+	904	4.93E-07	15.71814
3.41	721.1819	[M+H]+	904	0.004377	115.4263
2.76	465.2171	[M+H]+	898	0.391943	22.68282
1.95	161.0282	[M+H]+	890	3.85E-06	12.99333
5.41	618.4589	[M+H]+	882	0.554781	22262.04
4.55	438.4012	[M+H]+	881	0.482195	14686.03
1.89	181.0966	[M+H]+	842	0.000106	16.61259
4.60	421.0918	[M+H]+	837	2.33E-05	18.00971
4.70	350.1768	[M+H]+	833	0.000126	6.248376
5.92	708.3874	[M+H]+	818	0.497837	13628.51
5.64	652.2336	[M+H]+	802	1.30E-08	36807.29
5.62	779.3312	[M+H]+	782	0.49323	13232.79
3.87	347.1389	[M+H]+	782	0.038985	34575.11
6.19	529.3922	[M+H]+	776	0.002141	33.34822
4.40	668.0942	[M+H]+	773	0.000117	71701.81
3.82	375.1892	[M+H]+	763	1.20E-05	25.67118
6.27	824.3661	[M+H]+	759	0.007168	203.062
4.87	723.2859	[M+H]+	753	0.025699	27648.54
5.89	527.3217	[M+H]+	751	0.462831	12515.82
3.32	588.3727	[M+H]+	749	0.30554	12.53346
5.76	666.1871	[M+H]+	739	0.032906	45767.11
1.91	185.1261	[M+H]+	737	0.144128	17.70936
5.28	292.1343	[M+H]+	719	1.53E-05	35960.11
3.58	549.2684	[M+H]+	705	0.141078	5.26396

5.56	254.2464	[M+H]+	694	0.016506	20.99645
4.49	397.333	[M+H]+	694	0.462831	11558.93
5.06	429.3605	[M+H]+	693	7.13E-11	34632.92
5.82	640.2372	[M+H]+	683	0.046301	22763.09
5.20	190.1063	[M+H]+	683	3.24E-06	7.487683
1.95	530.0935	[M+H]+	668	2.24E-05	33436.45
4.92	189.0875	[M+H]+	659	0.054321	6.590505
2.82	423.1368	[M+H]+	649	0.299128	12345.46
4.16	689.2164	[M+H]+	647	0.097627	30630.06
4.37	832.3636	[M+H]+	646	0.031329	26995.15
4.93	207.1041	[M+H]+	640	0.0206	26876.4
4.40	670.0914	[M+H]+	638	0.000107	50439.85
1.93	277.0497	[M+H]+	629	8.36E-08	7.965393
5.02	431.1152	[M+H]+	614	1.99E-05	51677.52
3.57	626.2916	[M+H]+	602	0.154112	19.77353
5.01	487.132	[M+H]+	575	0.005595	47906.82
5.02	307.0053	[M+H]+	558	2.52E-05	1069.284
5.81	268.2617	[M+H]+	436	0.011832	68.91762
4.44	811.4006	[M+H]+	423	0.022353	5.75026
1.95	561.1291	[M+H]+	356	9.94E-06	53.16471
1.95	186.9689	[M+H]+	320	9.94E-08	8.009474
4.54	383.3191	[M+H]+	273	0.511713	4557.441
4.55	444.4081	[M+H]+	184	0.503478	5024.336
3.61	611.3417	[M+H]+	176	0.164082	130.2019
4.54	430.4209	[M+H]+	105	0.462831	1749.003
1.95	258.5681	[M+H]+	99	3.05E-05	33.20153
1.97	501.1839	[M+H]+	82	2.84E-05	51.68677
5.02	544.058	[M+H]+	68	0.000538	2746.977
5.81	544.7141	[M+H]+	41	0.000111	6.003357

Table A2.6. Enriched metabolites from Pfrm BGC metabolomics. Enriched metabolites from overexpression of Pfrm BGC with supplemented $^{13}C_6$-L-phenylalanine, with Signal/Noise cutoff above 530.

RT (min)	*m/z*	Type	Signal/Noise	ANOVA *p*-value	Fold Increase
12.41	293.1144	[M-H]-	8364	0.002712	9.701816
4.57	345.5365	[M-H]-	6499	7.85E-05	22.02267
4.70	345.5367	[M-H]-	6493	2.46E-05	20.15867
15.50	311.0757	[M-H]-	5509	7.49E-05	3.424789
10.76	180.0654	[M-H]-	4771	2.92E-05	1246.698
4.86	345.5365	[M-H]-	4459	1.34E-06	17.26018
8.52	218.1029	[M-H]-	3930	0.003879	1.523934
14.14	288.1243	[M-H]-	3921	9.73E-07	15.1232
16.49	236.9048	[M-H]-	3771	0.044611	1.035853
11.33	278.1492	[M-H]-	3144	1.88E-05	1412.304
18.43	166.9734	[M-H]-	3134	0.987383	1.032804
7.27	337.0969	[M-H]-	2832	0.004828	2.711737
18.96	293.9888	[M-H]-	2611	0.057857	4.966346
14.81	306.0805	[M-H]-	2493	0.008777	1.142628
20.14	375.1033	[M-H]-	2324	0.001932	2.758518
9.22	371.0565	[M-H]-	2292	0.004923	119115.5
14.14	234.0421	[M-H]-	2197	0.096943	2.113285
18.93	438.5829	[M-H]-	2120	0.168496	3.1429
10.75	186.085	[M-H]-	1906	1.51E-08	1533.97
10.15	234.0983	[M-H]-	1864	0.000144	3.234355
14.71	382.5496	[M-H]-	1844	0.018055	5.925889
16.06	236.904	[M-H]-	1844	0.550634	1.031654
16.14	292.9218	[M-H]-	1794	0.682677	1.346723
6.64	233.1158	[M-H]-	1741	0.014277	1.77754
23.23	371.6069	[M-H]-	1682	0.092747	4.074765
2.30	209.0932	[M-H]-	1673	0.422386	81.97237
9.77	181.0489	[M-H]-	1659	0.001086	1.740459
6.89	234.0769	[M-H]-	1599	3.90E-05	4.810294
18.06	184.9842	[M-H]-	1562	0.041108	2.056638
11.33	234.1599	[M-H]-	1539	1.42E-05	124696
8.42	151.0243	[M-H]-	1514	0.002715	1.662499
7.41	293.0643	[M-H]-	1452	0.554983	28166.55
14.32	357.14	[M-H]-	1433	0.010171	597.4886
18.80	311.0427	[M-H]-	1410	3.47E-07	71100.3
14.26	337.0375	[M-H]-	1396	0.313303	1.579155
18.33	293.9934	[M-H]-	1390	0.001987	1.109572
15.70	353.0091	[M-H]-	1373	0.058451	1.351156
9.21	352.9863	[M-H]-	1349	0.731678	4.189681
10.45	160.0605	[M-H]-	1339	0.0048	1.829108
12.42	192.98	[M-H]-	1329	0.000569	16.28933

13.35	132.0288	[M-H]-	1298	0.000173	3.811912
14.71	431.0025	[M-H]-	1244	3.31E-05	3.79019
15.14	102.0548	[M-H]-	1197	0.094432	7.811147
8.12	204.0647	[M-H]-	1194	9.25E-05	4.374167
11.97	327.0941	[M-H]-	1131	0.00014	3.319925
14.65	234.0374	[M-H]-	1108	0.001102	98033.35
18.04	226.9403	[M-H]-	1098	0.024618	3.300282
10.70	357.0889	[M-H]-	1050	0.042628	4.203961
14.81	272.0872	[M-H]-	1048	0.005571	1.380719
13.04	491.9461	[M-H]-	1004	2.69E-07	103278.1
2.83	209.0933	[M-H]-	974	0.422704	74.48235
18.50	246.0611	[M-H]-	971	1.72E-05	4.145247
14.11	266.9383	[M-H]-	919	0.539487	1.561036
16.45	434.8135	[M-H]-	909	0.00396	1.077372
7.91	135.0295	[M-H]-	895	0.337393	1.320409
3.32	311.2827	[M-H]-	893	0.126394	2.116759
11.81	128.0339	[M-H]-	861	0.831443	1.083903
18.56	96.96012	[M-H]-	847	2.43E-05	1.227418
8.54	392.0678	[M-H]-	802	0.004164	1.525209
3.20	311.2873	[M-H]-	800	0.353152	1.779898
7.88	233.0948	[M-H]-	788	0.065424	1.599198
9.63	130.0857	[M-H]-	775	0.001148	1.870349
9.21	430.9725	[M-H]-	774	0.52831	3.413557
3.42	311.2825	[M-H]-	771	0.446419	1.779266
18.51	228.0504	[M-H]-	750	3.83E-05	4.499278
18.11	242.0663	[M-H]-	743	0.003062	2.231116
7.83	203.0479	[M-H]-	717	9.54E-05	73.20596
16.50	336.9875	[M-H]-	711	0.45042	1.789311
15.85	274.103	[M-H]-	709	0.000404	2.23285
18.75	293.9897	[M-H]-	708	0.034275	1.145695
23.23	380.612	[M-H]-	703	0.134559	2.228034
13.71	234.0381	[M-H]-	703	0.017613	34.58146
9.67	353.1114	[M-H]-	697	0.000519	9.786666
9.54	164.0707	[M-H]-	695	1.11E-11	46.12691
7.51	233.1012	[M-H]-	692	0.094988	1.139025
10.77	491.9753	[M-H]-	677	3.85E-06	84634.35
10.60	357.0871	[M-H]-	639	0.261738	1.893711
15.69	357.1029	[M-H]-	634	1.27E-05	4.614871
11.14	371.0852	[M-H]-	629	0.003216	31465.08
2.99	311.2813	[M-H]-	624	0.543147	1.523913
7.63	413.0456	[M-H]-	602	2.77E-07	46760.64
17.79	337.0619	[M-H]-	598	0.030845	27129.72
12.42	134.9742	[M-H]-	590	0.001832	7.930513
14.06	346.0542	[M-H]-	579	0.337923	1.428452
18.25	371.6151	[M-H]-	575	0.018256	1.914232

19.01	402.994	[M-H]-	567	0.04091	2.267038
13.08	321.0476	[M-H]-	562	0.003721	7.079821
4.59	165.0547	[M-H]-	554	7.86E-06	158.3563
4.52	346.0379	[M-H]-	554	0.000141	16.63755
13.70	311.097	[M-H]-	551	0.022589	16219.67
8.52	233.0887	[M-H]-	546	0.098609	15791.78
11.32	284.1691	[M-H]-	532	1.04E-05	492.1818

Table A2.7. Enriched metabolites from *Pha*l BGC metabolomics. Enriched metabolites from overexpression of *Pha*l BGC with supplemented $^{13}C_5$,^{15}N valine; $^{13}C_6$ leucine; or $^{13}C_6$,^{15}N isoleucine, with Signal/Noise cutoff above 230.

RT (min)	*m/z*	Type	Signal/Noise	ANOVA *p*-value	Fold Increase
10.40	76.09444	[M+H]+	5351	0.667443	62.87611
8.91	91.11343	[M+H]+	5317	0.79209	119.563
8.52	235.1646	[M+H]+	5165	0.29207	2.123797
3.01	198.1554	[M+H]+	5002	0.588573	498.8376
8.52	118.0862	[M+H]+	4919	0.271012	1.112275
6.37	252.1668	[M+H]+	4489	0.560522	329.2127
9.30	91.11337	[M+H]+	3783	0.928815	2.775917
12.19	217.129	[M+H]+	3681	0.058142	8.943001
10.40	77.09174	[M+H]+	3275	0.805723	30.08019
17.54	248.0756	[M+H]+	2867	0.094987	11.99298
7.28	231.1339	[M+H]+	2826	0.143246	402.9583
8.87	120.0806	[M+H]+	2666	0.266761	1.367995
10.16	154.1167	[M+H]+	2471	0.570155	508.5817
10.28	254.1609	[M+H]+	2164	0.006104	2.668573
10.16	76.03104	[M+H]+	2040	0.568547	847.1406
8.91	138.1218	[M+H]+	1925	0.791703	103.0483
2.39	228.1942	[M+H]+	1877	0.705672	228.5206
15.11	397.105	[M+H]+	1874	5.23E-06	215167.3
19.33	175.1188	[M+H]+	1868	0.000352	7.729481
16.65	397.1054	[M+H]+	1829	1.13E-07	188067.9
7.29	237.1478	[M+H]+	1735	0.565539	118.9976
8.00	192.1588	[M+H]+	1683	0.420665	34.14632
17.54	230.0648	[M+H]+	1678	0.102449	9.966813
7.01	166.1102	[M+H]+	1473	0.573752	13.64759
6.81	254.1819	[M+H]+	1347	0.342306	55.28849
8.46	119.0891	[M+H]+	1344	0.000241	1.826373
6.30	268.1037	[M+H]+	1334	0.012146	4.383598
6.66	254.1817	[M+H]+	1315	0.343453	50.22222
10.04	216.1703	[M+H]+	1274	0.001512	5.526651
10.16	89.09766	[M+H]+	1254	0.574052	72.91909
15.76	226.9508	[M+H]+	1243	0.146547	1.750913
8.53	58.06596	[M+H]+	1158	0.461646	1.137476
19.43	175.1187	[M+H]+	1129	0.000573	7.483103
15.86	226.9505	[M+H]+	1035	0.069255	1.933679
10.17	107.108	[M+H]+	1027	0.573192	159.696
10.16	136.1059	[M+H]+	1020	0.574087	133.2413
8.87	166.0855	[M+H]+	1003	0.412425	1.412562
8.64	119.089	[M+H]+	997	0.135666	1.311591
7.88	235.1182	[M+H]+	981	0.073174	66.95696

9.63	107.1079	[M+H]+	964	0.579982	146.1577
8.49	187.6105	[M+H]+	960	0.039655	45246.01
13.95	308.0898	[M+H]+	947	0.076512	3.913489
3.34	86.06035	[M+H]+	931	0.011434	172.0641
15.59	164.9293	[M+H]+	893	0.322171	1.915009
12.82	203.1136	[M+H]+	889	1.50E-05	85.17293
4.99	180.1318	[M+H]+	869	0.634999	52.38423
9.60	62.05173	[M+H]+	831	0.574384	111.9521
17.64	248.0758	[M+H]+	821	0.291002	5.136249
13.31	348.0688	[M+H]+	813	0.381889	3.56513
3.78	228.17	[M+H]+	790	0.00445	37.03149
6.16	136.0614	[M+H]+	779	0.695288	1.963268
2.37	227.1954	[M+H]+	777	0.641104	59.51311
7.68	235.1181	[M+H]+	742	0.338365	27.10761
2.98	192.1413	[M+H]+	738	0.5789	105.4541
8.54	257.1472	[M+H]+	717	0.003035	9.921052
15.33	397.1054	[M+H]+	707	0.009617	61607.52
8.81	235.1649	[M+H]+	705	0.188558	9.393742
6.48	235.1187	[M+H]+	702	0.000892	128.5721
9.60	89.09743	[M+H]+	681	0.577271	21.17023
7.28	236.1472	[M+H]+	677	0.595397	26.72344
6.87	251.1694	[M+H]+	676	0.208543	50.56455
5.00	181.1303	[M+H]+	673	0.674058	35.00855
9.30	74.0869	[M+H]+	668	0.988752	1.778374
5.88	180.1321	[M+H]+	646	0.45327	32.63597
3.01	194.1416	[M+H]+	614	0.590157	78.82815
14.91	259.0913	[M+H]+	587	0.002777	27.03246
11.77	395.1665	[M+H]+	585	0.089844	5.549832
10.48	186.1596	[M+H]+	583	0.261417	2.743488
5.52	275.2163	[M+H]+	571	0.458948	210.2722
8.13	192.1585	[M+H]+	571	0.42382	44.20298
6.05	180.1322	[M+H]+	568	0.480369	34.98561
10.16	434.6816	[M+H]+	564	0.421073	246.687
10.34	265.1111	[M+H]+	554	4.58E-05	4.133431
3.29	228.1703	[M+H]+	552	0.538639	5.64768
2.23	257.0242	[M+H]+	542	0.000501	323728.8
6.76	61.04049	[M+H]+	528	0.168723	1.651212
7.60	220.1174	[M+H]+	525	0.086428	22.62345
14.66	275.1339	[M+H]+	522	0.027496	7.833322
10.17	498.2819	[M+H]+	516	0.512649	34.40117
10.31	186.159	[M+H]+	515	0.462121	46.00642
8.92	86.09663	[M+H]+	504	0.050085	19.32401
12.21	146.0917	[M+H]+	497	0.842766	1.295365
2.16	264.1006	[M+H]+	485	0.177099	1.436011
5.53	268.1978	[M+H]+	484	0.462669	49.78882

10.16	74.02391	[M+H]+	471	1.10E-06	87.11577
11.69	245.1602	[M+H]+	471	0.1475	6.409393
5.49	274.2173	[M+H]+	470	0.466213	228.5679
9.70	230.1852	[M+H]+	466	0.499222	1.282456
10.16	364.1467	[M+H]+	465	0.455399	43.10622
15.20	98.98367	[M+H]+	462	0.000368	2.799801
6.38	245.1493	[M+H]+	457	0.111378	72.10612
8.53	195.599	[M+H]+	451	0.00967	68.39832
5.51	273.2167	[M+H]+	444	0.473168	129.8731
2.85	227.1959	[M+H]+	435	0.441192	46.95399
2.91	198.1548	[M+H]+	415	0.594083	485.9052
11.08	344.8738	[M+H]+	414	0.00278	386.9384
10.46	59.06789	[M+H]+	411	0.636159	1.395048
10.42	281.634	[M+H]+	403	0.468588	228.2646
9.28	86.09657	[M+H]+	403	0.536107	2.61779
6.85	61.04027	[M+H]+	402	0.001991	1.94602
10.42	364.6612	[M+H]+	400	0.377328	125.8284
10.28	238.0689	[M+H]+	396	0.283372	658.1019
5.51	92.11035	[M+H]+	395	0.458158	987.71
9.27	434.3192	[M+H]+	394	0.434714	62.71141
2.53	208.1628	[M+H]+	393	0.526236	43.1305
9.60	76.09453	[M+H]+	392	0.577911	132.6508
6.99	235.1185	[M+H]+	391	0.989766	1.441814
7.07	137.0454	[M+H]+	391	0.424524	1.847413
13.96	346.0452	[M+H]+	387	0.243871	2.839099
19.33	116.0701	[M+H]+	382	0.000497	7.046805
10.43	131.1174	[M+H]+	380	0.005696	3.674468
19.36	70.06553	[M+H]+	374	0.000114	6.970295
10.42	446.1792	[M+H]+	373	0.485161	11568.44
13.34	291.1313	[M+H]+	372	0.779332	4.756989
8.87	103.0536	[M+H]+	367	0.364506	1.452729
15.67	254.8985	[M+H]+	365	0.224233	1.959452
2.53	209.1608	[M+H]+	362	0.610407	14.83068
10.15	148.0963	[M+H]+	361	1.41E-06	172.2733
10.35	144.047	[M+H]+	353	0.000104	3.507297
8.48	304.6891	[M+H]+	350	0.049614	31990.52
3.53	221.0841	[M+H]+	347	8.35E-05	15.92715
6.66	235.1184	[M+H]+	347	0.004985	19.01326
15.08	259.0912	[M+H]+	343	0.017254	19.50369
17.60	270.0573	[M+H]+	342	0.06592	26.02479
8.92	132.1013	[M+H]+	342	0.047875	17.90121
10.28	131.1171	[M+H]+	342	0.441792	12.33373
15.08	276.119	[M+H]+	342	0.018186	13.53235
10.55	291.0067	[M+H]+	341	0.446277	110.4986
3.63	228.17	[M+H]+	340	1.48E-05	30.71315

6.71	166.1094	[M+H]+	337	0.674091	9.492727
10.34	122.0706	[M+H]+	337	0.000117	3.638674
14.87	233.1129	[M+H]+	333	0.016394	13.56385
11.16	162.1593	[M+H]+	332	0.031841	28.41789
13.31	312.1029	[M+H]+	330	0.536851	5.781041
11.17	174.1595	[M+H]+	329	0.000493	9.465472
10.17	492.2619	[M+H]+	327	0.50982	20.06357
18.11	476.1853	[M+H]+	323	0.420597	24206.49
10.17	269.1019	[M+H]+	323	0.588285	9.640928
2.97	197.1567	[M+H]+	312	0.587802	27.70944
10.71	86.09644	[M+H]+	311	0.041497	142.5892
8.89	132.0514	[M+H]+	309	0.057335	20.81749
12.72	87.0443	[M+H]+	305	0.090976	22.75127
8.89	428.2988	[M+H]+	303	0.454767	17.16426
9.40	152.0561	[M+H]+	303	0.13371	39.82918
12.39	497.1975	[M+H]+	303	0.001356	11.23551
9.67	353.1154	[M+H]+	303	1.46E-05	181133.2
8.92	434.3195	[M+H]+	299	0.469577	77810.09
3.53	261.0765	[M+H]+	296	0.000199	14.96968
10.42	266.1175	[M+H]+	295	0.000113	4.220295
19.32	349.2296	[M+H]+	294	0.001963	87.82513
7.79	192.1588	[M+H]+	292	0.274444	3.858959
9.31	310.6717	[M+H]+	289	0.470095	223.1853
17.54	148.0597	[M+H]+	287	0.101355	10.16533
8.32	226.1508	[M+H]+	287	0.454393	86.01878
10.41	284.1417	[M+H]+	283	0.529546	36.30194
14.87	271.0683	[M+H]+	279	0.066662	11.53252
14.35	122.9237	[M+H]+	272	0.000766	4.898169
8.57	140.0677	[M+H]+	270	0.001009	6.191748
9.27	454.7769	[M+H]+	267	0.453147	101.3744
10.43	114.0909	[M+H]+	266	0.012282	3.176147
8.52	312.678	[M+H]+	265	0.029433	432.0743
7.85	220.1175	[M+H]+	265	0.815054	1.783607
11.08	214.9168	[M+H]+	260	0.001458	382.2624
15.06	233.1129	[M+H]+	258	0.010732	13.5047
15.77	90.97648	[M+H]+	257	0.180299	1.595924
15.51	210.9218	[M+H]+	251	0.119439	5701.341
10.42	222.596	[M+H]+	251	0.463232	314.9649
8.91	376.7474	[M+H]+	250	0.454189	9412.188
12.63	439.1556	[M+H]+	249	0.001382	40.31219
18.86	84.08087	[M+H]+	246	0.000101	23.64171
7.33	70.0656	[M+H]+	245	0.000995	148.8932
7.41	192.1587	[M+H]+	245	0.157284	11.85336
6.81	245.149	[M+H]+	244	0.123372	46.23321
7.79	225.1338	[M+H]+	244	0.604782	8.019804

10.16	173.4861	[M+H]+	244	0.430886	3544.225
8.52	352.2435	[M+H]+	244	0.602075	3.144829
7.24	70.0655	[M+H]+	243	0.002016	147.1997
6.58	253.1768	[M+H]+	243	0.158155	13.43626
11.60	114.1018	[M+H]+	241	0.013945	17.17513
7.74	220.1175	[M+H]+	239	0.394374	2.43002
10.16	102.091	[M+H]+	237	4.40E-06	17.12372
2.16	299.1365	[M+H]+	235	0.28881	1.317248
10.42	300.6072	[M+H]+	233	0.427216	165.1156
15.47	142.9472	[M+H]+	233	0.244519	2.020303
10.44	441.674	[M+H]+	232	0.363364	197.3549
7.48	245.1765	[M+H]+	230	0.09338	13.74575

Table A2.8. Enriched metabolites from *Pant* BGC metabolomics. Enriched metabolites from overexpression of *Pant* BGC, with Signal/Noise cutoff above 67.

RT (min)	*m/z*	Type	Signal/Noise	ANOVA *p*-value	Fold Increase
8.53	235.1652	[M+H]+	8615	0.7598	1.059469
12.57	191.1021	[M+H]+	6572	0.000679	106.6489
8.53	118.0859	[M+H]+	5110	0.110199	1.062555
12.06	477.2054	[M+H]+	4875	0.508141	1.497171
9.00	120.0794	[M+H]+	4321	0.018357	1.588732
12.07	277.061	[M+H]+	3409	0.294816	1.214954
15.36	259.0918	[M+H]+	3197	0.000298	42.58605
17.87	230.0651	[M+H]+	2994	0.001964	2.552661
12.32	217.1292	[M+H]+	2861	0.007945	2.286002
17.87	248.0766	[M+H]+	2750	0.002354	2.363548
12.05	499.1845	[M+H]+	2680	0.266798	1.620891
14.15	346.0468	[M+H]+	2457	0.013077	8.349243
15.16	314.0742	[M+H]+	2270	0.003013	188.3582
12.03	261.0901	[M+H]+	2093	0.146771	1.201583
12.62	279.1004	[M+H]+	1967	0.2863	24.36271
19.70	175.1183	[M+H]+	1896	0.016444	1.742033
8.33	192.1593	[M+H]+	1800	0.114093	8.350397
15.18	259.0916	[M+H]+	1768	8.12E-06	65.76051
10.26	216.1706	[M+H]+	1739	0.01356	1.408588
14.15	308.0909	[M+H]+	1612	0.002424	6.974936
11.87	295.1289	[M+H]+	1485	0.014527	39.15179
14.87	275.1346	[M+H]+	1451	0.000127	4.104786
2.18	415.2127	[M+H]+	1417	0.060009	1.681895
10.67	188.1752	[M+H]+	1374	0.404553	2.369835
12.92	161.0914	[M+H]+	1359	0.02328	2.295537
17.93	270.0573	[M+H]+	1265	0.012434	4.925559
8.09	192.1588	[M+H]+	1220	0.085575	2.315274
7.51	220.1176	[M+H]+	1063	1.15E-06	4.701174
9.00	166.0857	[M+H]+	1059	0.017416	1.539327
17.47	399.144	[M+H]+	1040	0.020654	4.77599
16.05	226.9506	[M+H]+	1015	0.737625	1.062614
11.86	232.096	[M+H]+	970	0.015153	22.90928
7.94	192.1584	[M+H]+	960	0.906524	1.078057
8.53	58.06527	[M+H]+	949	0.19346	1.091866
15.09	271.0686	[M+H]+	922	0.031202	2.296407
3.38	298.0966	[M+H]+	918	0.002285	3.448403
11.87	278.1014	[M+H]+	913	0.01748	5.791392
15.09	233.1131	[M+H]+	908	0.010961	2.345125
15.97	226.9501	[M+H]+	896	0.737625	1.062614
7.56	192.159	[M+H]+	876	0.063456	29.11738

17.79	305.0974	[M+H]+	841	0.005391	210.1242
13.50	322.1055	[M+H]+	824	0.027631	3.057415
18.43	218.0649	[M+H]+	823	0.016596	2.388118
13.18	147.0751	[M+H]+	774	0.051541	2.205125
11.80	261.144	[M+H]+	731	0.017116	763948
9.92	230.1861	[M+H]+	722	0.009635	2.235882
6.45	268.1038	[M+H]+	694	0.17692	3.126201
11.76	277.0613	[M+H]+	670	0.058117	1.360862
10.49	254.1604	[M+H]+	664	0.273944	1.669816
11.83	173.0911	[M+H]+	648	0.001144	68.84696
14.15	179.0474	[M+H]+	648	0.002142	7.49809
11.77	261.0861	[M+H]+	626	0.955539	1.03122
10.50	72.08033	[M+H]+	619	0.120954	1.23911
15.34	275.1349	[M+H]+	619	0.00355	2.793897
11.46	204.1697	[M+H]+	613	0.456995	5.145366
13.48	348.0692	[M+H]+	600	0.189987	1.705326
13.29	322.1068	[M+H]+	590	0.001105	88.8037
19.75	176.1203	[M+H]+	575	0.0167	1.747318
13.00	203.1135	[M+H]+	565	0.021975	3.216018
11.85	333.0798	[M+H]+	563	0.00148	58.913
15.37	241.0825	[M+H]+	562	0.000431	9.868783
9.39	119.1168	[M+H]+	560	0.500228	1.066009
10.37	90.09039	[M+H]+	545	0.005122	1.078128
12.07	239.1058	[M+H]+	536	0.384067	1.331792
18.42	200.0543	[M+H]+	535	0.022225	2.349673
11.87	373.1063	[M+H]+	533	0.00726	55.6659
7.61	220.1174	[M+H]+	532	0.000103	2.660001
9.02	86.09579	[M+H]+	530	0.718654	1.069837
15.10	407.1712	[M+H]+	518	0.011024	33.55019
17.40	399.147	[M+H]+	516	0.026321	4.687641
13.42	265.0851	[M+H]+	515	0.053706	26.31618
11.93	395.1742	[M+H]+	506	0.082796	2.527537
17.45	284.0733	[M+H]+	497	0.049969	4.177599
4.47	183.0955	[M+H]+	496	0.445413	1.284139
14.19	384.0009	[M+H]+	495	0.00041	2.397731
9.02	132.1004	[M+H]+	494	0.968727	1.005884
11.87	120.0793	[M+H]+	481	0.014985	33.88662
9.13	118.0853	[M+H]+	474	0.175225	1.457393
9.32	322.1062	[M+H]+	472	0.043147	15.74096
6.63	235.1187	[M+H]+	471	0.035952	17.60066
14.11	397.1149	[M+H]+	451	0.098042	57897.31
12.82	247.1284	[M+H]+	435	0.035533	2.983989
14.22	421.9564	[M+H]+	429	0.018899	1.637699
15.32	233.1125	[M+H]+	420	0.478664	1.218529
10.53	265.1114	[M+H]+	415	0.4966	1.137083

19.73	70.0648	[M+H]+	392	0.01233	1.946944
9.00	103.0527	[M+H]+	379	0.016322	1.577058
13.54	290.1339	[M+H]+	377	0.592364	1.172238
9.38	132.1005	[M+H]+	375	0.021775	1.958102
10.58	59.06059	[M+H]+	370	0.685317	1.018036
12.91	410.0513	[M+H]+	364	0.134711	2.748345
12.36	146.0911	[M+H]+	354	0.067888	1.333382
4.80	183.0964	[M+H]+	347	0.166695	1.765561
2.75	427.1505	[M+H]+	343	0.019985	6.149094
12.57	62.05975	[M+H]+	342	0.001007	48.04217
12.79	410.0514	[M+H]+	334	0.232053	2.572036
12.11	370.1778	[M+H]+	333	0.386581	3.071435
12.65	439.1713	[M+H]+	330	0.033924	2.624128
17.87	184.0593	[M+H]+	326	0.001829	2.542679
6.53	377.151	[M+H]+	325	0.833217	1.048093
15.18	298.1002	[M+H]+	322	0.001675	34.13953
13.50	309.0755	[M+H]+	321	0.013565	45912.45
7.22	137.0451	[M+H]+	315	0.120039	1.268858
10.66	171.1482	[M+H]+	313	0.327096	2.922269
9.39	86.09576	[M+H]+	304	0.032316	2.430545
14.15	233.0585	[M+H]+	304	0.002286	5.796685
12.20	406.1628	[M+H]+	301	0.120602	26549.74
11.87	166.0855	[M+H]+	297	0.01446	32.81614
13.73	277.0616	[M+H]+	294	0.565428	1.106738
9.52	118.084	[M+H]+	292	0.495306	1.090299
15.18	241.0835	[M+H]+	285	2.95E-05	7.703447
3.25	228.1705	[M+H]+	279	0.230302	3.063816
9.33	129.1006	[M+H]+	276	0.00302	1.105401
12.67	189.1226	[M+H]+	272	0.470355	1.153116
6.99	235.1187	[M+H]+	270	0.992081	1.007559
17.77	395.6771	[M+H]+	267	0.630409	1.246898
12.55	497.2061	[M+H]+	266	0.014617	3.702505
19.24	147.1116	[M+H]+	260	0.009729	6.649824
17.87	202.0702	[M+H]+	259	0.001797	2.49858
2.19	368.4253	[M+H]+	259	0.880096	1.048103
10.60	117.1008	[M+H]+	256	0.656448	1.099386
6.84	61.03942	[M+H]+	255	0.055015	1.36772
3.32	466.293	[M+H]+	255	0.003485	1.648337
11.57	162.1584	[M+H]+	252	0.585032	2.311798
12.88	366.2226	[M+H]+	251	0.25932	1.623439
9.34	174.1592	[M+H]+	249	0.006386	1.101185
19.70	398.2062	[M+H]+	249	0.006939	2.477427
12.54	277.061	[M+H]+	248	0.189471	1.156331
11.73	204.1694	[M+H]+	247	0.732276	1.647192
14.90	247.1282	[M+H]+	247	0.190222	1.305702

14.11	277.0618	[M+H]+	244	0.072469	2.171808
13.66	365.1113	[M+H]+	243	0.373901	12203.73
13.75	435.1876	[M+H]+	242	0.103447	1.123617
16.68	378.1842	[M+H]+	239	0.034614	6.939834
13.69	381.0816	[M+H]+	238	0.245475	45184.35
8.53	352.2436	[M+H]+	236	0.79092	1.099065
9.43	118.0841	[M+H]+	236	0.968066	1.002846
12.98	344.0999	[M+H]+	234	0.001925	6.90365
15.33	298.1003	[M+H]+	233	0.000831	14.8713
19.71	116.069	[M+H]+	232	0.020044	1.634508
8.96	132.0503	[M+H]+	231	0.866522	1.121171
11.23	214.9171	[M+H]+	230	0.066623	2.014222
15.22	475.205	[M+H]+	229	0.901768	1.008927
11.92	358.2008	[M+H]+	228	0.006847	9.793694
15.22	443.2087	[M+H]+	227	0.572726	1.028161
13.79	475.2055	[M+H]+	224	0.186374	1.066432
10.41	245.1236	[M+H]+	224	0.011496	29.67584
12.91	183.0798	[M+H]+	221	0.144664	1.455054
15.16	413.1793	[M+H]+	221	0.110268	1.188144
3.96	228.1709	[M+H]+	219	0.188033	9.185374
17.58	457.2131	[M+H]+	219	0.130684	1.162702
12.84	239.106	[M+H]+	217	0.949307	1.014724
9.36	210.0812	[M+H]+	217	0.215004	2.042183
13.79	473.1937	[M+H]+	216	0.890565	1.006452
13.84	413.1799	[M+H]+	215	0.122485	1.08836
14.86	313.0911	[M+H]+	213	0.002793	2.928293
4.28	183.0982	[M+H]+	212	0.103236	1.450403
15.09	471.2119	[M+H]+	211	0.027733	1.811311
8.03	235.1197	[M+H]+	209	0.350425	8.222692
12.83	365.1066	[M+H]+	209	0.262967	1.390984
13.18	84.04363	[M+H]+	209	0.048492	2.036557
10.53	70.06456	[M+H]+	207	0.536084	1.131677
9.48	221.0952	[M+H]+	206	0.015458	10.49102
8.57	120.0894	[M+H]+	206	0.062103	1.081783
16.01	158.9624	[M+H]+	201	0.843767	1.038306
13.78	431.1816	[M+H]+	200	0.104079	1.091335
10.59	186.1592	[M+H]+	197	0.013129	2.304037
10.53	144.0463	[M+H]+	193	0.523635	1.13022
17.44	261.0895	[M+H]+	191	0.33611	1.251227
3.23	494.3239	[M+H]+	191	0.006778	1.632806
13.17	130.0487	[M+H]+	190	0.05511	2.011267
7.62	242.1005	[M+H]+	188	0.372593	1.619273
2.84	427.1507	[M+H]+	185	0.019307	5.328991
13.82	299.1016	[M+H]+	181	0.02614	1.196357
15.20	431.1822	[M+H]+	181	0.181275	1.14019

15.05	417.1681	[M+H]+	179	0.01281	1.252752
13.82	417.1764	[M+H]+	179	0.042681	1.109612
12.91	332.0746	[M+H]+	175	0.065857	3.694897
13.84	459.205	[M+H]+	175	0.054979	1.051999
15.95	164.9283	[M+H]+	175	0.806672	1.048056
12.21	468.1745	[M+H]+	175	0.210865	5.237749
9.93	200.1747	[M+H]+	175	0.002781	3.166217
2.19	284.3306	[M+H]+	171	0.167864	1.802204
15.97	254.898	[M+H]+	170	0.806903	1.047819
12.20	435.1923	[M+H]+	168	0.145641	1.537178
9.75	228.0832	[M+H]+	166	0.02918	26308.15
10.95	60.05549	[M+H]+	165	0.601799	1.086429
15.05	459.2025	[M+H]+	164	0.033526	1.113368
14.15	291.0646	[M+H]+	164	0.001772	6.232829
19.70	414.2063	[M+H]+	163	0.004951	1.716005
12.22	367.1746	[M+H]+	162	0.380586	1.14046
16.83	261.0894	[M+H]+	162	0.104676	1.580151
13.84	443.2012	[M+H]+	161	0.023989	1.071256
10.47	116.069	[M+H]+	160	0.939888	1.015466
3.32	480.308	[M+H]+	159	4.45E-07	8.984313
10.59	357.3389	[M+H]+	158	0.235369	1.306059
10.14	86.09534	[M+H]+	155	0.076946	1.016289
15.28	299.1083	[M+H]+	155	0.038986	1.355313
11.42	211.0751	[M+H]+	155	0.050642	9699.852
11.76	162.1587	[M+H]+	152	0.816919	1.336593
16.05	90.97528	[M+H]+	152	0.866488	1.031501
13.70	389.1806	[M+H]+	151	0.996777	1.000203
9.31	193.0663	[M+H]+	151	0.164403	6867.626
11.94	262.0925	[M+H]+	150	0.008486	1.581942
8.51	257.1463	[M+H]+	150	0.3657	1.161884
16.27	381.195	[M+H]+	150	0.807448	1.08209
15.08	457.2056	[M+H]+	149	0.066366	1.119223
12.81	439.1916	[M+H]+	147	0.071989	1.671382
5.39	202.0903	[M+H]+	147	0.181491	1.991622
3.29	494.3232	[M+H]+	146	0.006778	1.632806
15.44	416.8582	[M+H]+	145	0.252209	1.372664
8.71	192.1586	[M+H]+	145	0.137793	11.59907
19.69	386.2057	[M+H]+	145	0.018813	1.750629
9.80	118.0837	[M+H]+	144	0.255247	1.282398
14.70	443.2045	[M+H]+	143	0.460578	1.068206
14.78	459.2079	[M+H]+	143	0.000768	1.316363
13.53	233.1122	[M+H]+	143	0.096422	2.436755
2.93	120.0793	[M+H]+	141	0.217535	3.10957
13.85	399.1729	[M+H]+	141	0.050691	1.094859
8.50	118.7062	[M+H]+	140	0.319342	1.549698

12.18	457.1979	[M+H]+	140	0.327427	1.083351
15.31	389.1926	[M+H]+	140	0.260503	1.223109
2.19	388.3935	[M+H]+	139	0.764045	1.078046
14.84	351.0482	[M+H]+	137	0.039992	2.942554
14.46	277.0641	[M+H]+	137	0.030935	2.092728
17.85	395.6764	[M+H]+	137	0.630409	1.246898
11.52	431.2108	[M+H]+	137	0.027928	1.279355
9.31	344.0922	[M+H]+	136	0.156786	34.71981
2.18	432.2408	[M+H]+	135	0.05571	1.686319
15.01	433.1888	[M+H]+	134	0.098799	1.119995
12.88	399.1955	[M+H]+	132	0.174011	1.454506
12.52	173.0912	[M+H]+	132	0.420282	1.134624
10.22	185.0987	[M+H]+	132	0.076435	1.701206
10.62	156.1481	[M+H]+	132	0.029684	2.02398
14.32	122.9227	[M+H]+	131	0.04528	2.648911
12.38	427.1503	[M+H]+	130	0.00709	3.683539
11.09	367.203	[M+H]+	130	0.00834	3.507349
14.29	388.1402	[M+H]+	130	0.003438	13.0565
14.69	457.2118	[M+H]+	127	0.045037	1.187402
9.00	77.03766	[M+H]+	127	0.020056	1.549186
15.21	397.201	[M+H]+	127	0.416108	1.039849
15.09	487.2117	[M+H]+	125	0.085452	1.174024
11.86	186.0906	[M+H]+	124	0.027983	5.798069
12.07	450.1973	[M+H]+	124	0.558043	1.16192
14.47	84.04354	[M+H]+	123	0.006636	3.013044
15.68	382.8637	[M+H]+	123	0.182926	1.102789
15.11	433.1867	[M+H]+	119	0.082955	1.149487
15.10	197.0918	[M+H]+	119	0.006275	2.659406
10.57	114.0898	[M+H]+	118	0.018799	1.93293
15.30	433.2071	[M+H]+	118	0.993942	1.001625
12.81	332.0749	[M+H]+	118	0.139211	3.173695
17.87	148.0589	[M+H]+	117	0.001688	2.519702
11.94	245.1595	[M+H]+	117	0.017342	2.541152
11.63	270.1446	[M+H]+	117	0.003239	18.91348
13.11	277.0611	[M+H]+	117	0.530094	1.101112
13.91	287.0915	[M+H]+	116	0.010088	1.19552
10.56	357.3553	[M+H]+	115	0.087679	1.62694
15.09	255.0961	[M+H]+	115	0.004964	1.707521
11.79	406.1848	[M+H]+	114	0.06746	7528.039
3.25	448.1504	[M+H]+	114	0.096566	8608.301
12.25	129.0648	[M+H]+	114	0.570518	1.135277
14.16	250.0504	[M+H]+	114	0.457149	1.062175
11.71	403.1993	[M+H]+	113	0.152026	1.061895
12.01	472.1837	[M+H]+	113	0.014387	1.586844
11.75	299.1	[M+H]+	112	0.101068	1.626013

15.41	262.1025	[M+H]+	112	0.069539	2.019587
14.46	250.0499	[M+H]+	111	0.050577	1.25081
3.08	228.1703	[M+H]+	111	0.56982	1.351474
12.91	174.0976	[M+H]+	110	0.070399	1.254187
15.08	443.2002	[M+H]+	110	0.007193	1.218887
15.07	250.0502	[M+H]+	110	0.217498	1.219159
14.03	459.2052	[M+H]+	108	0.015412	1.291073
16.94	277.0952	[M+H]+	107	0.961988	1.012166
8.94	129.1004	[M+H]+	107	0.000604	1.351306
12.07	496.1822	[M+H]+	107	0.602928	1.553763
11.87	317.11	[M+H]+	107	0.028786	15.50411
19.74	127.0963	[M+H]+	107	0.957112	1.002502
12.86	87.04337	[M+H]+	106	0.084475	1.34856
15.97	322.8861	[M+H]+	106	0.754644	1.061133
12.69	386.131	[M+H]+	106	0.070948	14323.11
19.27	101.105	[M+H]+	105	0.017718	1.423965
14.58	443.2067	[M+H]+	105	0.231201	1.140405
11.63	397.2092	[M+H]+	104	0.002858	1.519506
13.53	161.0915	[M+H]+	104	0.273697	1.278357
15.98	430.9113	[M+H]+	104	0.742965	1.055167
9.01	103.1166	[M+H]+	104	0.000224	1.454302
15.07	431.1765	[M+H]+	103	0.081221	1.165737
10.57	131.1163	[M+H]+	102	0.025706	1.877552
13.79	373.1437	[M+H]+	102	0.115761	1.11502
11.71	431.2014	[M+H]+	101	0.681077	1.013588
11.86	434.2018	[M+H]+	101	0.18243	5256.905
10.57	373.3319	[M+H]+	100	0.012371	2.090109
10.62	100.0743	[M+H]+	100	0.310678	2.19848
10.57	343.3675	[M+H]+	100	0.011311	2.019139
6.83	202.0926	[M+H]+	100	0.40332	1.572556
17.48	298.097	[M+H]+	100	0.019105	4.289637
13.56	328.0883	[M+H]+	100	0.660273	1.080522
12.79	488.0282	[M+H]+	99	0.338098	1.607245
19.25	358.2023	[M+H]+	99	0.018165	5.594928
14.86	216.086	[M+H]+	98	0.000309	3.681408
15.93	402.1763	[M+H]+	98	0.002131	1.701852
13.97	244.0971	[M+H]+	97	0.004422	3.807069
12.85	74.05924	[M+H]+	97	0.164639	1.367774
14.31	457.2123	[M+H]+	97	0.02165	1.227929
17.87	130.0485	[M+H]+	96	0.002094	2.532216
3.12	427.1536	[M+H]+	96	0.021299	3.606383
19.70	158.0918	[M+H]+	96	0.014588	1.644936
15.36	159.0757	[M+H]+	96	0.000283	10.72035
14.44	261.927	[M+H]+	96	0.297764	1.238861
13.80	457.2074	[M+H]+	95	0.143495	1.071173

15.25	381.2057	[M+H]+	94	0.583728	1.020884
14.76	387.1909	[M+H]+	94	0.008533	2.307497
15.19	429.2031	[M+H]+	94	0.343762	1.073454
8.83	162.0571	[M+H]+	93	0.02241	60449.45
15.25	287.0945	[M+H]+	91	0.008741	1.373845
17.66	307.0869	[M+H]+	91	0.039029	36.75595
14.76	473.2087	[M+H]+	90	0.004458	1.548957
12.05	176.0912	[M+H]+	90	0.049612	8.301586
14.94	459.2055	[M+H]+	89	0.727412	1.017022
15.66	262.905	[M+H]+	89	0.52099	1.050221
9.32	103.1221	[M+H]+	88	0.008342	1.106596
14.46	459.2052	[M+H]+	88	0.042142	1.441081
9.16	58.06502	[M+H]+	88	0.208975	1.335085
16.06	362.9256	[M+H]+	88	0.595323	1.098982
8.61	118.7107	[M+H]+	88	0.356633	1.415616
19.91	127.0963	[M+H]+	87	0.008469	1.326139
17.60	367.2117	[M+H]+	87	0.085919	1.188914
10.77	89.10616	[M+H]+	87	0.046766	22.232
10.50	158.0917	[M+H]+	87	0.56836	1.102381
15.21	411.2117	[M+H]+	87	0.14795	1.062657
2.20	494.3234	[M+H]+	87	0.006987	1.758663
14.33	443.205	[M+H]+	87	0.035734	1.271204
14.01	417.1701	[M+H]+	86	0.08418	1.17879
17.51	400.2019	[M+H]+	86	0.373901	2892.316
10.22	134.1152	[M+H]+	86	0.312448	1.506747
17.26	397.209	[M+H]+	86	0.500216	1.057056
19.71	349.2296	[M+H]+	85	0.051453	12.85728
2.20	181.0266	[M+H]+	85	0.527683	1.522347
15.75	142.9465	[M+H]+	85	0.473789	1.162598
11.23	198.0874	[M+H]+	85	0.124657	1.519835
10.41	387.3935	[M+H]+	84	0.06943	2.796323
10.09	200.1747	[M+H]+	84	0.459824	1.148771
13.66	388.0583	[M+H]+	84	0.021809	2.738182
14.59	421.9576	[M+H]+	84	0.000467	3.831325
14.55	122.9228	[M+H]+	84	0.032645	5.069591
3.71	120.079	[M+H]+	84	0.322134	5.221838
8.53	297.0893	[M+H]+	84	0.52849	1.130552
16.06	430.9092	[M+H]+	84	0.476035	1.122733
7.20	235.1203	[M+H]+	84	0.567864	1.469976
15.04	475.2005	[M+H]+	83	0.148645	1.101796
14.89	285.094	[M+H]+	83	0.131536	1.397858
12.57	128.0695	[M+H]+	83	0.000564	19.98656
15.97	362.9258	[M+H]+	83	0.963199	1.009409
13.48	342.165	[M+H]+	82	0.003816	14.15211
2.98	368.4238	[M+H]+	82	0.770519	1.046639

17.46	277.0725	[M+H]+	82	0.690712	1.25176
15.43	454.8145	[M+H]+	81	0.154788	1.467024
15.07	415.1768	[M+H]+	81	0.012199	1.245163
14.32	206.8841	[M+H]+	81	0.044273	2.927184
15.87	457.2081	[M+H]+	81	0.348673	1.060363
10.81	188.175	[M+H]+	80	0.381648	2.329021
15.14	389.1688	[M+H]+	80	0.156615	1.175574
14.45	263.9262	[M+H]+	80	0.379447	1.184551
5.96	146.116	[M+H]+	80	0.367859	4.611162
12.92	488.0286	[M+H]+	80	0.119216	1.845415
15.56	190.0933	[M+H]+	79	0.006958	1.543476
19.24	84.07996	[M+H]+	79	0.010572	5.424913
15.28	367.1913	[M+H]+	79	0.52397	1.044037
8.53	195.5982	[M+H]+	79	0.392321	1.192739
11.88	103.0526	[M+H]+	79	0.015642	25.81133
15.90	234.0764	[M+H]+	79	0.733901	1.031443
12.32	158.0829	[M+H]+	78	0.010615	2.159422
12.18	90.05521	[M+H]+	78	0.544642	1.484716
13.73	447.209	[M+H]+	78	0.415638	1.0744
10.50	95.05959	[M+H]+	77	4.09E-05	13.76527
10.67	116.1028	[M+H]+	77	0.869908	1.050631
11.94	395.1986	[M+H]+	77	0.082796	2.527537
3.17	368.4252	[M+H]+	77	0.404873	1.276032
10.53	122.0703	[M+H]+	77	0.517147	1.123803
14.76	299.1012	[M+H]+	76	0.391211	1.078996
6.64	297.0419	[M+H]+	76	0.033461	407.9211
11.24	62.05967	[M+H]+	76	0.01823	1.89591
7.54	114.0674	[M+H]+	76	0.375266	2.685594
11.40	182.0806	[M+H]+	75	0.05932	6.093566
9.03	210.0916	[M+H]+	75	0.070672	1.095708
2.72	410.1245	[M+H]+	75	0.009237	17.06905
14.15	379.1352	[M+H]+	75	0.000867	13798.48
19.72	130.0959	[M+H]+	75	0.007961	1.997942
8.50	118.8952	[M+H]+	75	0.564802	1.308725
15.21	441.2122	[M+H]+	74	0.118087	1.124
15.35	197.0935	[M+H]+	74	0.004851	2.492353
17.87	84.04328	[M+H]+	74	0.001874	2.406324
12.16	132.0757	[M+H]+	74	0.376499	10.68702
9.00	387.0847	[M+H]+	74	0.003332	2.442276
9.98	74.09434	[M+H]+	74	0.047087	1.245049
8.59	118.7987	[M+H]+	73	0.607665	1.366666
14.20	457.2145	[M+H]+	73	0.005088	1.610836
8.31	156.1362	[M+H]+	73	0.121748	2.871753
9.97	431.4188	[M+H]+	73	0.078393	3.656358
14.93	373.1444	[M+H]+	73	0.596228	1.054411

17.43	250.0923	[M+H]+	73	0.019592	5.087869
16.99	261.0895	[M+H]+	72	0.182525	1.451419
10.63	189.132	[M+H]+	72	0.413909	1.972502
10.57	325.3369	[M+H]+	72	0.1239	1.195601
14.58	441.2072	[M+H]+	72	0.02791	1.200054
13.46	426.0464	[M+H]+	72	0.015498	3.492613
14.16	441.2093	[M+H]+	72	0.000558	1.668774
2.34	284.3306	[M+H]+	72	0.388099	1.816878
9.76	286.1756	[M+H]+	71	0.026064	9.329539
2.21	466.2933	[M+H]+	71	0.003954	1.719301
15.80	459.2206	[M+H]+	71	0.17976	1.372596
14.32	290.8452	[M+H]+	71	0.060978	2.675766
17.52	250.0924	[M+H]+	70	0.014586	5.106285
15.31	353.1737	[M+H]+	70	0.166999	1.11775
14.05	299.0996	[M+H]+	70	0.000235	1.483436
12.83	360.1686	[M+H]+	69	0.36937	1.168016
12.33	200.1026	[M+H]+	69	0.009981	1.987728
14.81	461.2135	[M+H]+	69	0.097299	1.279834
12.10	256.1048	[M+H]+	69	0.172813	1.6686
5.57	202.0927	[M+H]+	69	0.157933	1.905086
15.77	284.8872	[M+H]+	69	0.453204	1.132929
9.78	162.0753	[M+H]+	69	0.090035	2.049528
14.11	443.2048	[M+H]+	69	0.007152	1.687734
12.78	277.0642	[M+H]+	69	0.228652	1.233242
14.15	76.02072	[M+H]+	68	0.0022	7.752041
5.08	234.1346	[M+H]+	68	0.016099	7.877453
10.05	101.1085	[M+H]+	68	0.103712	1.098636
19.70	382.2124	[M+H]+	67	0.010403	2.431317
11.25	198.2067	[M+H]+	67	0.130199	1.350576
12.57	174.0772	[M+H]+	67	0.001213	6.514649

Appendix III: *Supplementary Figures*

Figure A3.1. MS/MS spectra of βes+2 *m/z* incorporation into peptides. High resolution MS/MS spectra showing incorporation of a βes+2 *m/z* at a Thr position in the BesG KO strain. (A) Peptide from the aminoglycoside transferase (F8K2J3). (B) Peptide from Elongation Factor Tu (F8JYD7). (C) Peptide from the thymidine phosphorylase (F8K173).

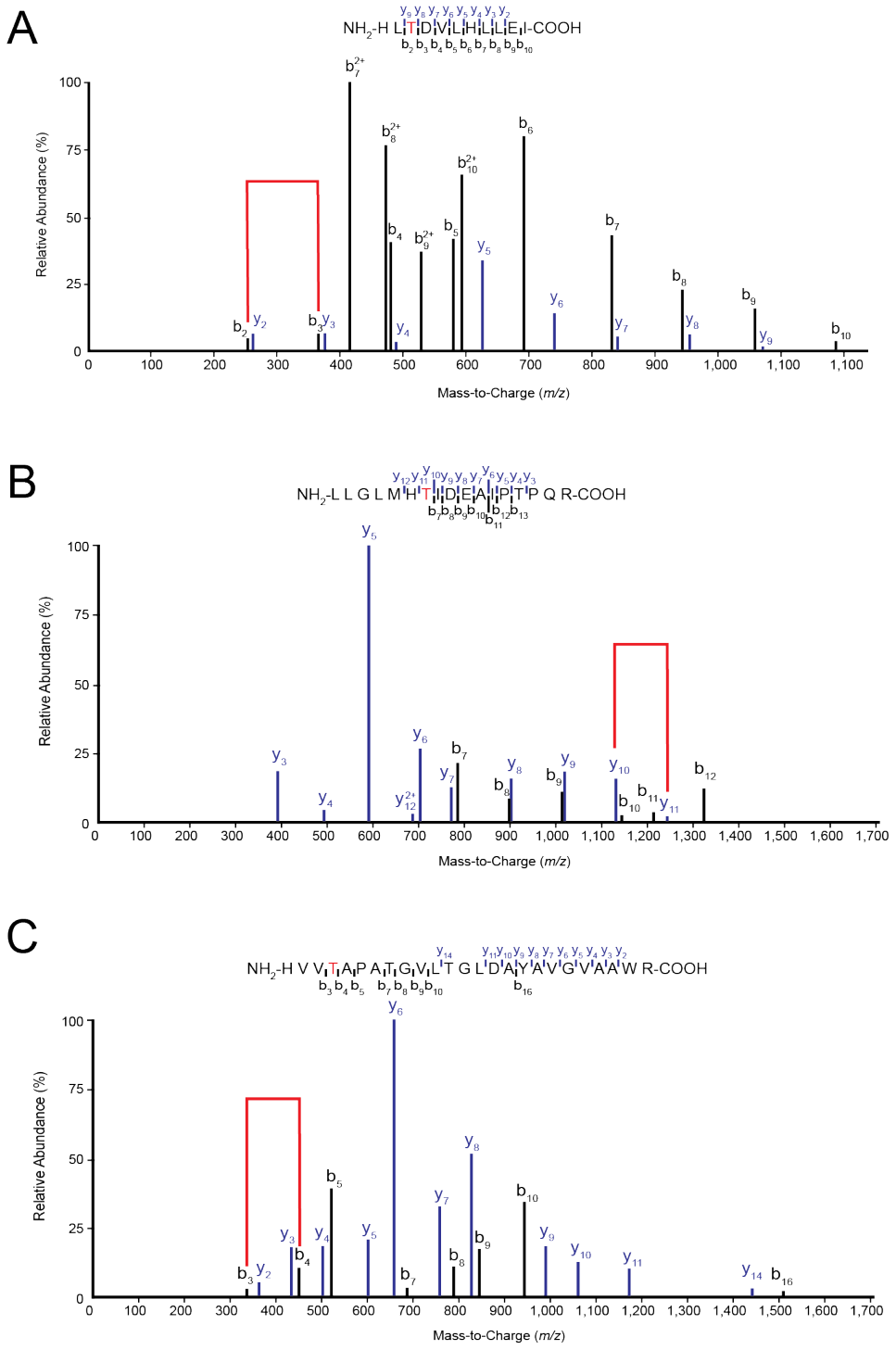


Figure A3.2. Putative biosynthetic gene clusters with AlaX tRNA-Deacylases. (A) Additional biosynthetic gene clusters (BGCs) identified through tR3D to potentially produce secondary metabolites. (B) Gene accession IDs for the co-clustered AlaX tRNA-deacylases in these BGCs.

A



B

SSN Cluster (No.)	Gene ID
1	A0A349BZI0
1	A0A352WGH3
1	A0A353VAT0
1	A0A395YFU0
1	A0A1I7IUL1
1	A0A1K1QS26
1	A0A0B6X1R7
1	A0A0C5VSR9
1	A0A0E0TC32
1	A0A0E2V1R8
1	A0A0G2Z9J1
1	A0A2D5QB74
1	A0A0J0V3S2
1	A0A1F5BIE5
1	A0A1F8KNP3
1	A0A1Q7Y246
1	A0A1V5WFB0
1	A0A1V5WKX4
1	A0A1Y4H8W5
1	A0A2D8TTC1
1	A0A2H9MB76
1	A0A521YGW0
1	A0A6G9J6F2
1	A0A6N2E527
5	A0A1D8ITP8
5	A0A1E2VAM5
5	A0A0F7HGQ0
5	A0A1H3HCS3
5	A0A1R7QG13
5	A0A241W1K3
5	A0A2X5B6Y9
5	A0A373B7I8
5	R7KJA3
5	E3BHA2
6	A0A0F4TWK6
7	A0A1H4GTI8
7	A0A3M2GHW3
7	A0A7L6AQ55
7	Q4ULE5
8	A0A016AAV5
8	A0A069S1R8
8	A0A3D1KHZ0
8	A0A413IT53
8	U6RLZ9
9	A0A2A2XUT3
9	A0A2D4SJW2
10	A0A2K2V6S8
10	A0A1Q8A5W8
10	A0A3G1A851
10	A0A497F5K3
10	A0A7C5VTJ9
10	H9ZZH5
11	A0A0F9YCM3
11	A0A0V8HTG7
11	A0A2A3XZ84
12	A0A1I0MDA7
12	A0A1S8AVJ3
12	A0A365TCI0
12	A0A4D6HLX3
13	A0A0C1WC14
14	A0A1U7MBY9
14	A0A3D0WN32
15	A0A0Q6BZ07
15	A0A1C4TCR7
15	A0A1H3JVZ4
15	A0A4D8PSJ1
15	A0A4P6S1V7
15	A0A4Y6IEL6
15	A0A5R8Z9P6
15	A0A6H1W567
15	A0A6J7I4Z3
15	A0A7W5TLA6
15	A0A841PUY5
15	A0A212A9C6
15	A0A856MBM9
15	C7QJ43
15	C8X9H3
16	A0A073K9G7
16	A0A235B870
17	A0A3P1S676
18	A0A0F9PA48
18	A0A7K4CVU1
19	A0A2V4LHS9
19	A0A7X1U745
21	A0A1H5BNC4
21	K9UDQ2
21	A0A2T1FSD1
25	A0A0R1PCV5
25	A0A380L106
25	D2MMJ8
25	K1LP04
26	A0A1N5VNL2
28	A0A0P7JQA8
28	A0A7U5CYZ1
28	A0A857C2Y4
29	A0A550H3Z7
30	A0A3N7EWS7
30	A0A5B9YD17
30	U7UUU5
30	A0A848BTW6
31	D0I6N6
31	A0A1Z2SF43
33	A0A7Y8A1K1
34	A0A1G2U5T6
34	A0A858AQQ9
34	S5MJX2
34	W6A7U5
37	R4U3D3
41	A0A5P8JSU5
42	A0A3N2DGX1
44	A0A8J7X8D4
45	U2HUE4
46	A0A7Y3PLL7
53	L9Z963
58	A0A7K4FKM2
60	A0A846Z3J6
60	A0A7H8M2W2
65	A0A1F0FTP0
66	D2EF82
67	A0A0P6VYU1
69	A0A7G7HSS4
70	A0A2S9IT69
70	K2NLA6
72	V6MJU4
73	Q3MBR7
74	M4Z4W0
75	K7W230
76	C5A2G3
77	B5YA28
79	A0A833JB23
84	A0A7V1AGG7
85	A0A7J3WPL0
86	A0A7X1GSJ5
88	A0A7H0HZI1
93	A0A3M4LSE5
105	A0A1I5XWL6
108	W0BER3
132	A0A447TJC4

Figure A3.3. Purification results of BesG, ThrRSs, *alf* cluster, *Af* MetRS and *fuz* cluster. (A) Purification of ThrRSs (Ec = *E. coli*, *Sc* = *S. cattleya*) and MBP-BesG for tRNA acylation experiments, with arrow for AlfC for clarity. (B) Purification of genes from the Fuz cluster with His_{10} N-terminal tags.

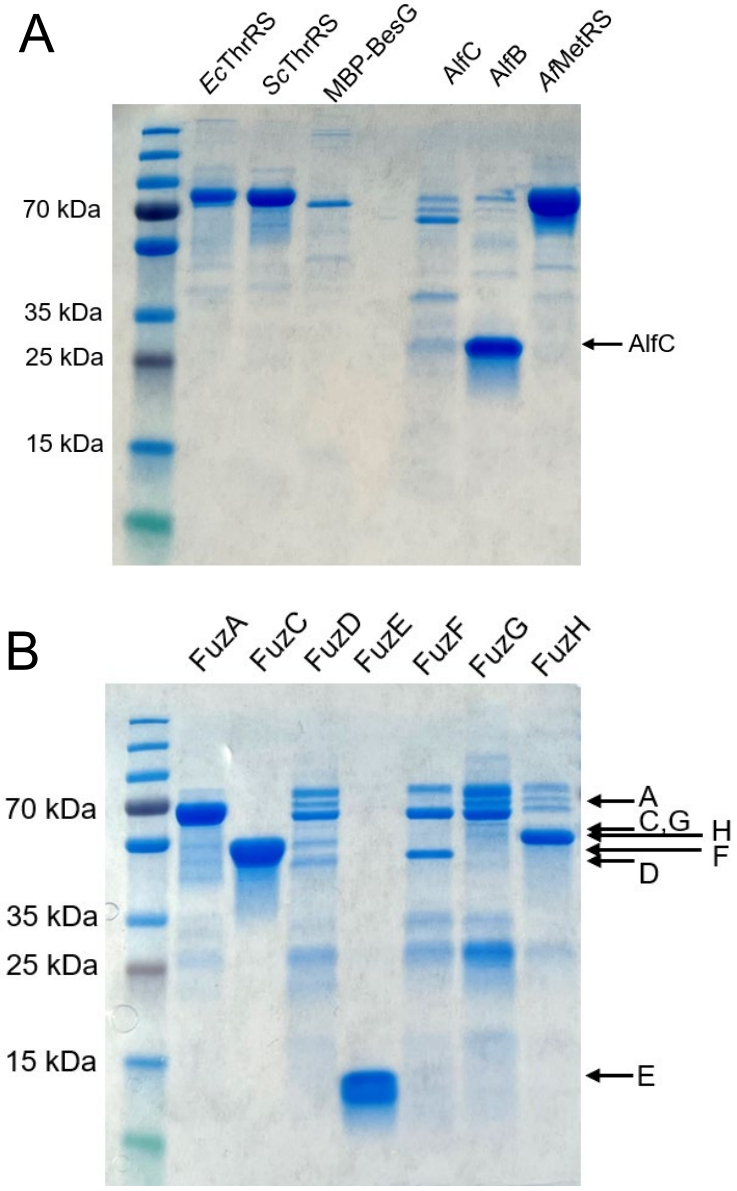


Figure A3.4. Genome mining to identify Pfrm biosynthetic gene cluster in *Pseudomonas fluorescens R124*. (A) TauD-like protein Uniprot ID and Taxon ID for the nine organisms that contain the Pfrm BGC. (B) Genomic neighborhoods of other TauD-like proteins present in clade G.

A

Organism	Gene ID (PfrmB Homologs)	NCBI Taxon ID
Pseudomonas koreensis	A0A1G6X5I6	198620
Pseudomonas moraviensis	A0A2A2PUC6	321662
Klebsiella pneumoniae	A0A2L1KSR9	573
Brenneria rubrifaciens	A0A4P8QU60	55213
Pseudomonas fluorescens	A0A5E7UAW5	294
Pseudomonas fluorescens R124	**A0A7U9CRC7**	743713
Klebsiella pneumoniae (strain 342)	B5XPQ2	507522
Gammaproteobacteria bacterium RIFCSPHIGHO2_12_38_15	A0A1G0FHP3	1798273
Gammaproteobacteria bacterium RIFCSPLOWO2_02_FULL_38_11	A0A1G0J532	1798293

B

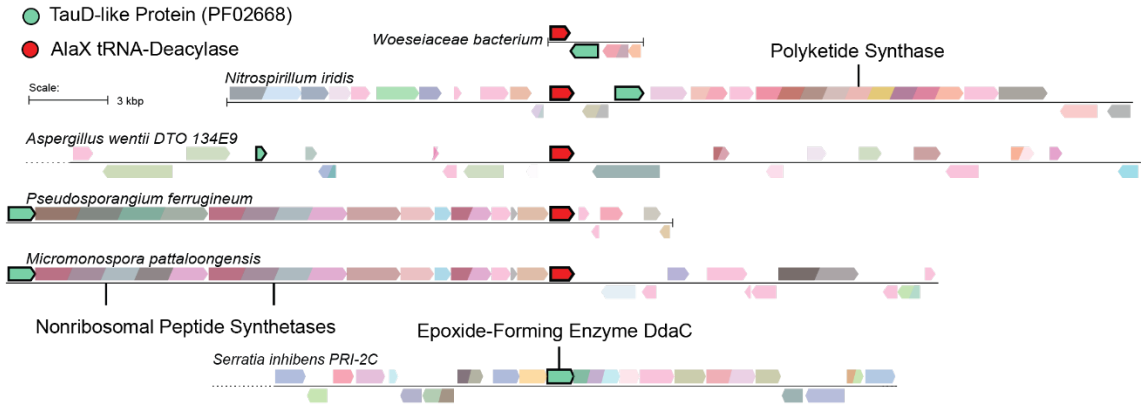


Figure A3.5. Purification of genes from Pfrm BGC. Purification of genes from Pfrm BGC with N-terminal His_{10} tags via heterologous expression in *E. coli*.

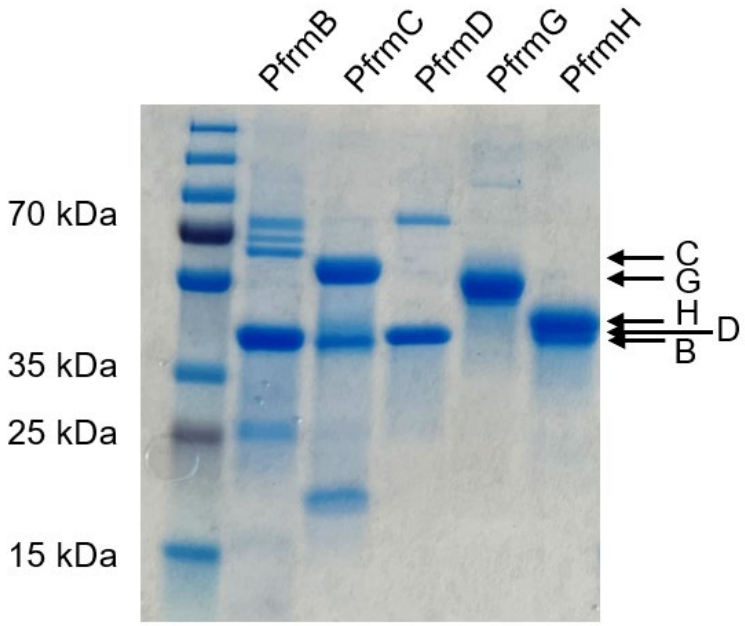


Figure A3.6. Purification of genes from Phal and Pant BGCs. Purification of genes from (A) Phal and (B) Pant BGCs with N-terminal His_{10} tags via heterologous expression in *E. coli*.

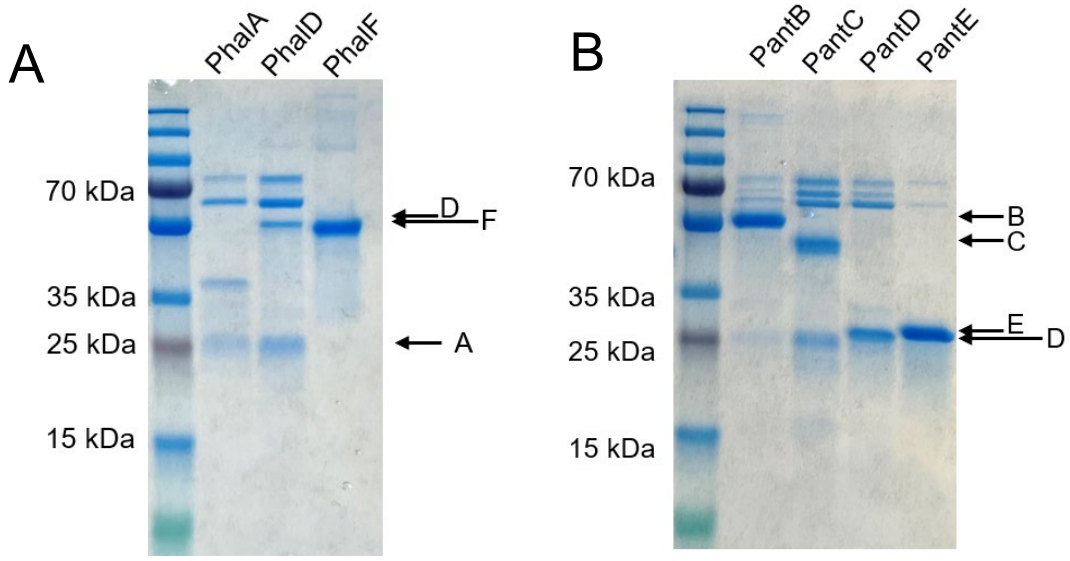
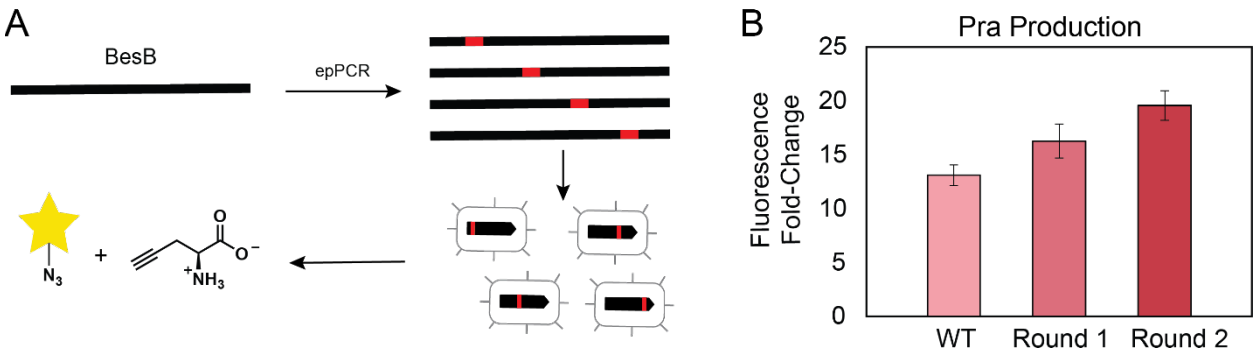


Figure A3.7. Engineering of BesB. (A) Error-prone PCR (epPCR) was used to randomly mutate BesB, where production was measured with an azide-fluorophore. (B) After 2 rounds, Pra production improved roughly 50%.



BesB-improving mutations
S56P
A172P
D185E
G208R
A257V
I267I (ATC to ATA)
A478T

We performed some engineering of the Pra biosynthetic enzyme BesB to improve titers *in vivo*. This enzyme suffers from insolubility, so error-prone PCR was used to install changes randomly throughout the protein to improve Pra production. After 2 rounds of engineering, Pra production increased about 50% with 7 mutations. We anticipate future rounds of engineering could provide a means to both biosynthesize and incorporate Pra *in vivo*.