

ELUCIDATION AND CONTROL OF SUBSTRATE RECOGNITION DURING RIPP
BIOSYNTHESIS

BY

BRANDON JAY BURKHART

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemistry
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:

Associate Professor Douglas A. Mitchell, Chair
Assistant Professor Kami L. Hull
Professor William W. Metcalf
Professor Wilfred A. van der Donk

Abstract

Ribosomally synthesized and posttranslationally modified peptides (RiPPs) are a rapidly growing class of natural products. RiPP precursor peptides can undergo extensive enzymatic tailoring to yield structurally and functionally diverse products. Cyclodehydratases are a type of RiPP modifying enzyme that catalyze phosphorylation of the peptide backbone and subsequent nucleophilic attack by the sidechains of Cys, Ser, or Thr to form azoline heterocycles (or azoles upon oxidation). The catalytic unit of the cyclodehydratase is a YcaO-family protein which is often accompanied by a partner protein from the E1-like superfamily (depending on the type of cyclodehydratase). Although primarily known for azoline formation, recent work suggests that YcaO proteins can use different nucleophiles and partner proteins to generate thioamide, macroamidine, and possibly other peptide posttranslational modifications. In Chapter 1, I comprehensively review the biosynthetic gene clusters (BGCs), natural products, functions, mechanisms, and applications of YcaO proteins and outline future directions for this protein superfamily.

In Chapter 2, I report my investigations into the substrate recognition of canonical cyclodehydratases. Recent work suggested that unrelated RiPP modifying enzymes contain structurally similar precursor peptide binding domains. Using profile hidden Markov model comparisons, I discovered related and previously unrecognized peptide binding domains in proteins spanning the majority of known prokaryotic RiPP classes. This conserved domain was designated the RiPP precursor peptide recognition element (RRE). Through binding studies, I verified the RRE's role for three distinct RiPP classes: linear azole-containing peptides, thiopeptides, and lasso peptides. Because numerous RiPP biosynthetic enzymes act on peptide substrates, these findings have powerful predictive value as to which protein(s) drive substrate binding, thereby laying a foundation for further characterization of RiPP biosynthetic pathways and the rational engineering of new peptide binding activities.

In Chapter 3, I use knowledge gained from precursor peptide binding studies to engineer a peptide that can be recognized and modified by two biosynthetic enzymes from different pathways. Combining enzymes from multiple pathways is an attractive approach for producing molecules with desired structural features, but this strategy thus far has been hampered by limited substrate tolerance of enzymes from unrelated pathways. Because RiPP biosynthetic enzymes modify their substrates by binding motifs located usually in the N-terminal leader region of precursor peptides, RiPP biosynthetic systems are highly amenable to the engineering of new compounds. I exploit this by designing chimeric leader peptides that can be bound and processed by multiple enzymes from unrelated RiPP pathways. Using this broadly applicable strategy, a cyclodehydratase was combined with enzymes from the sactipeptide and lanthipeptide RiPP classes to create new-to-nature hybrid RiPPs. These hybrids provide insight into biosynthetic timing and enzyme compatibility and establish a general platform for the engineering of additional hybrid RiPPs.

To my family for their unwavering support, encouragement, and prayers

Acknowledgements

First and foremost I thank God (Father, Son, and Holy Spirit) for enabling me to complete my doctoral work and for His countless blessings in my life. He formed the universe with all its complexity and appointed humans to rule over it. I join in humanity's God-given purpose by studying the intricacies of His creation. Despite the errors of man, God has redeemed the world through Jesus so that His plan for creation will not fail. Accordingly, I am motivated to do my best every day as I partner with and serve God in my work and other activities. I know that I can never earn His favor and that the blessing of a successful graduate career has come not because I was faithful to Him, but because God was faithful to me.

I am thankful for the many communities and people that I have built strong relationships with during graduate school. My church family at Twin City Bible Church has been a welcoming environment that pushed me to grow in faith and maturity as I engaged with people from all walks of life. Special thanks to Jeff Roesler for his mentorship over the last year. Graduate Christian fellowship (GCF) provided a similar source of encouragement. After a difficult first year in graduate school, GCF provided the opportunity to create intentional and authentic Christian relationships and to build a strong emotional and spiritual foundation. I am grateful for the wisdom and leadership of David Suryk, and the many books he introduced me to. I also want to thank the Wednesday night Bible study at Stephen and Heidi Mayhew's house for their weekly prayers and encouragement. Additionally, I acknowledge my former roommate Ian Robertson for his friendship and reliability through four years of living together. Had his wife not moved in, we would likely have lived together until completion of our degrees. Nonetheless, I truly appreciate living in Centennial House now, a house named for its many years as living and meeting space for GCF members. I thank housemates Ben Chidester, Ben Fulan, Stephen Tillman, Morgan Hammer, and Joshua Whitman for creating a positive environment and for fellowship through meals and conversations.

Another central part of my graduate school experience was rock climbing and getting to know people like Selina Kempton, Alex James, and Alex Dragas who helped me go from the "super strong guy that nobody knows" to a "regular" at the UIUC rock wall. Alex Dragas is a special climbing partner who got me on real rock and was good company during the 45-minute drive to Upper Limits in Bloomington, Illinois. Nothing cleared my mind and relieved stress like leaving town for climbing.

I also owe a huge debt of gratitude to my family for being a source of consistent encouragement through the years during Skype conversations and phone calls. They were always proud of me and my accomplishments even when I had not achieved much in the lab for weeks. Going home to spend time with Dad, Mom, Ryan, Hannah, and Megan was the greatest treat during my PhD work. I am truly grateful that my parents raised me and my siblings with strong Christian values and the countless hours they sacrificed to build a cohesive family. They always encouraged me to pursue my passions and develop my God-given talents so that I could become the man that I am today. Similarly, I am grateful to my grandparents for their

support through the years, and I want to especially acknowledge Grandma Burkhart for her steadfast prayers. I also know that my late grandfather “Opie” Westerbeek would be proud of this accomplishment.

Of course, I wouldn’t be anywhere without the incredible members of the Mitchell Laboratory and collaborators from other labs (Nair and van der Donk). Lab members who worked in the IGB (Joel Melby, Courtney Cox, Katie Molohon Hess, and Caitlin Deane) absorbed hundreds of questions when I first started and helped me as I developed into an independent scientist. Subgroup meetings with Kyle Dunbar and Graham Hudson also shaped my research career, and their fervent work accelerated my first publication. Chris Schwalen was crucial to completing my massive review paper. I additionally thank all members of the MMG theme for their insights and friendship. Abraham Wang in particular has been a great source of knowledge and conversation, both scientific and otherwise, and I’ve had a blast with him outside of the lab. I also appreciate more recent additions to the IGB crew (Adam DiCaprio, Andi Liu, Arash Firouzbakht, Nilkamal Mahanta). Playing the occasional chess game with Arash has also been a highlight (when I win).

The National Science Foundation and the Robert C. and Carolyn J. Springborn Graduate Fellowship provided another boost to my scientific career and eliminated financial concerns during graduate school. Being a part of the Chemistry Biology Interface training program, which has been headed by Wilfred van der Donk and Paul Hergenrother, has been a wonderful experience and privilege that broadened my scientific knowledge. My PhD studies have also benefitted from the input of my thesis committee. I have been a coauthor with Professor van der Donk and Professor Nair, and their feedback during MMG theme meetings along with other MMG faculty helped shape the direction of my work. I thank Professor Lu for providing a unique perspective from outside of my immediate research field.

Lastly, I whole-heartedly thank my advisor Doug Mitchell. Doug has provided excellent mentorship throughout my career and has encouraged me to be the best scientist that I can be. He has devoted many hours to reading and revising multiple drafts of different manuscripts and fellowship applications. Doug is an excellent advisor who is quick to do whatever is needed for his students. I am grateful for the opportunities that he gave me to investigate RiPPs and to pursue projects that interested me which ultimately resulted in this dissertation.

Table of Contents

| | |
|---|-----|
| Chapter 1: YcaO-Dependent Posttranslational Amide Activation: Biosynthesis, Structure, and Function.. | 1 |
| 1.1 Introduction | 1 |
| 1.2 Linear azole-containing peptides (LAPs) | 6 |
| 1.3 Thiopeptides | 17 |
| 1.4 Cyanobactins | 34 |
| 1.5 Miscellaneous azol(in)e-containing RiPPs | 40 |
| 1.6 Characterization of RiPP cyclodehydratases | 43 |
| 1.7 Non-canonical YcaOs | 61 |
| 1.8 Biotechnology | 67 |
| 1.9 Summary and outlook | 70 |
| 1.10 Acknowledgements | 71 |
| 1.11 Figures | 72 |
| 1.12 References | 112 |
| Chapter 2: A Prevalent Peptide-Binding Domain Guides Ribosomal Natural Product Biosynthesis..... | 167 |
| 2.1 Introduction | 167 |
| 2.2 Results | 168 |
| 2.3 Discussion | 174 |
| 2.4 Methods | 176 |
| 2.5 Acknowledgements | 179 |
| 2.6 Figures | 179 |
| 2.7 Tables | 191 |
| 2.8 References | 200 |
| Chapter 3: Chimeric Leader Peptides for the Generation of Non-Natural Hybrid RiPP Products | 206 |
| 3.1 Introduction | 206 |
| 3.2 Results | 207 |
| 3.3 Discussion | 213 |
| 3.4 Methods | 214 |
| 3.5 Acknowledgements | 219 |
| 3.6 Figures | 220 |
| 3.7 Tables | 252 |
| 3.8 References | 254 |
| Appendix A: PDF Reprints of other coauthored publications | 261 |

| | |
|--|-----|
| A.1 Structural and functional insight into an unexpectedly selective <i>N</i> -methyltransferase involved in plantazolicin biosynthesis..... | 261 |
| A.2 Discovery of a new ATP-binding motif involved in peptidic azoline biosynthesis | 268 |
| A.3 Identification of an auxiliary leader peptide-binding protein required for azoline formation in ribosomal natural products | 276 |
| A.4 <i>In vitro</i> biosynthesis and substrate tolerance of the plantazolicin family of natural products | 283 |

Chapter 1: YcaO-Dependent Posttranslational Amide Activation: Biosynthesis, Structure, and Function¹

1.1 Introduction

With the advent of high throughput genome sequencing technology, the number of available protein sequences has grown at a staggering rate. To organize all of this data, public databases collate protein sequences into superfamilies based on amino acid sequence similarity and conserved domains. Computationally, these groupings are determined by hidden Markov models (HMMs),^{1,2} which use pattern recognition algorithms that predict whether one protein sequence matches an established group. Once a protein is matched with a superfamily, it is presumed to be functionally similar with other members of that superfamily. Alternatively, function can often be understood by analyzing a protein's constituent domain(s).^{3,4} This function inference process is essential for automating the annotation of whole genomes, and fittingly, has been employed to predict the functions for millions of proteins without them having been experimentally validated.

Unsurprisingly, these homology-based approaches to protein function annotation are not without errors or inaccuracies.^{5,6} Although useful in certain situations, simply matching a superfamily does not necessarily imply one specific function because many protein superfamilies consist of multiple different iso-functional families, which may only share a partial mechanistic step or certain sequence motifs.⁷⁻¹⁰ Moreover, there are many proteins, ~20% of all classified protein domains, that consist entirely of domains of unknown function (DUFs) and possess no characterized homologs.^{11,12}

Thus, much can be learned by investigating proteins that are too divergent to be confidently assigned a function by homology or contain DUFs with no assigned function. This is especially true for such proteins found in natural product biosynthetic pathways. The enzymes encoded by these pathways catalyze the chemical reactions that build complex molecules and exploration of their reactions often reveals unexpected or entirely new reaction mechanisms.¹³⁻¹⁶ The topic of this review, the YcaO superfamily, exemplifies many of the challenges, triumphs, and future benefits in studying these uncharacterized biosynthetic proteins.

Originally defined as DUF181, YcaO proteins were studied through genetic and biochemical methods for over two decades until sufficient evidence could be gathered to suggest a common function for this superfamily in ATP utilization.^{17,18} However, what ATP is being used for is unknown for much of the superfamily. The term “YcaO” itself alludes to the enigmatic nature of this superfamily as it originates from

¹ This chapter has been reproduced with permission from Burkhart, B.J.; Schwalen, C.J.; Mann, G.; Naismith, J.H.; Mitchell, D.A. YcaO-dependent posttranslational amide activation: Biosynthesis, structure, and function. *Chem. Rev.* **2017**, *117*, 5389-5456. Copyright 2017 American Chemical Society. B.J.B wrote this work with contributions from C.J.S. for figure generation and section 1.7 and from G.M. and J.H.N. for sections 1.4, 1.6, and 1.8. All coauthors reviewed and edited the manuscript.

a gene-naming rubric established in *Escherichia coli* (*E. coli*). If a gene was of unknown function, it was given a provisional, four-letter name beginning with “y” and ending with three letters that indicated its genomic location.¹⁹ Hence, the *ycaO* gene of *E. coli* is located between the 21- and 30-min region of the genome, after *ycaN* but before *ycaP*. Many of these “y” proteins remain uncharacterized to this day, even for the most studied model organism, which underscores the difficulty in defining roles for genes when no clear phenotype is evident upon deletion, overexpression, or other functional genomics methods.²⁰⁻²²

Despite the mystery that still surrounds many YcaO proteins, the currently available data suggest four functions within the YcaO superfamily: (1) azoline formation, (2) macroamidine formation, (3) thioamide formation, and (4) potentiation of RimO-dependent methylthiolation. Little rationale has been provided for YcaO involvement in methylthiolation beyond the associated effect from gene deletion studies (section 1.7.1).²³⁻²⁵ However, the other functions are better understood and come from work on YcaO proteins encoded within the biosynthetic gene clusters (BGCs) of ribosomally synthesized and posttranslationally modified peptide (RiPPs).²⁶

RiPPs comprise a structurally and functionally diverse class of natural products and are categorized into subclasses based on common features installed by conserved biosynthetic enzymes.²⁶ With rare exception, RiPP biosynthesis begins with posttranslational modification of an inactive precursor peptide, composed of N-terminal leader and C-terminal core regions (Figure 1.1). Within the leader region are specific motifs that are recognized and bound by key biosynthetic proteins. To govern this substrate-binding event, several RiPP modifying enzymes rely on a structurally conserved domain, known as the RiPP precursor peptide recognition element (RRE),²⁷ whereas other RiPP biosynthetic protein bind their cognate leaders peptides in structurally unique ways.^{28,29} Once engaged with the substrate, the biosynthetic enzymes (e.g. YcaO) then install residue-specific modifications to the core region. After these modifications are complete, the leader region is removed while other tailoring enzymes, if present, further modify peptide. After enzymatic processing is completed, the mature compound is usually exported by ATP-binding cassette (ABC) transporters.²⁶ In some cases, mature RiPPs have been so heavily modified that they no longer resemble ribosomal peptides and had been postulated to derive from non-ribosomal pathways.³⁰⁻³² Additionally, because of their biosynthetic logic, RiPPs have many alluring biotechnological applications in engineering new natural products (further discussed in section 1.8).

Within the field of RiPP biosynthesis, the first conclusive role of a YcaO protein was elucidated by studying the biosynthesis of the antibiotic microcin B17 (MccB17). In 1986, the MccB17 *ycaO* gene was linked to an unknown posttranslational modification,³³ which was later identified as peptide backbone azole heterocycles.³⁴⁻³⁶ A decade later in 1996, azole biosynthesis was reconstituted *in vitro* and demonstrated to depend on ATP and a trimeric enzyme complex of the YcaO protein, a dehydrogenase, and a member of the E1-ubiquitin activating (E1-like) superfamily.³⁷ At the time, it was clear that the FMN-dependent

dehydrogenase (encoded by a gene adjacent to the YcaO) would oxidize azolines to the corresponding azoles, but it was unclear what roles ATP, the YcaO, and E1-like protein played in azoline heterocycle synthesis. In 2012, it was determined again through *in vitro* reconstitution that the YcaO and E1-like protein functioned together as a cyclodehydratase wherein the YcaO protein performs the ATP-dependent cyclodehydration reaction by catalyzing the attack of a β -nucleophilic side chain (Cys, Ser, or Thr) into the preceding amide bond to form a presumed hemiorthoamide intermediate (Figure 1.1c).¹⁷ This species then undergoes an ATP-driven *O*-phosphorylation, followed by *N*-deprotonation and elimination of phosphate to yield the azoline heterocycle. In the absence of the E1-like protein, the catalytic activity of the YcaO was minimal; the E1-like protein potentiates cyclodehydratase activity by nearly three orders and binds the precursor peptide through what is now known to be its N-terminal RRE domain.^{17, 18, 27} This finding resolved the historical confusion over the roles of the two components of the cyclodehydratase. In the past, the E1-like and YcaO proteins have been wrongly referred to as the cyclodehydratase and docking protein, respectively. Although these were the originally postulated functions, data now establishes the YcaO as the cyclodehydratase while the E1-like protein brings the substrate into proximity of the catalytic center of the YcaO. Since most YcaO cyclodehydratases are all but inactive without their cognate E1-like protein, the cyclodehydratase is most properly defined as the combination of both proteins. More detail into the cyclodehydratase reaction mechanism and substrate processing is covered in section 1.6.

As YcaO/DUF181 members became functionally linked with azoline heterocycles in RiPPs, intriguing counterexamples surfaced which suggested a YcaO-mediated role in other peptide backbone modifications, namely amidine and thioamide posttranslational modifications. For example, the BGC for the antibiotic natural product bottromycin encodes two YcaO proteins yet only one azoline heterocycle is present.³⁸⁻⁴⁰ However, bottromycin contains a second amide backbone modification in the form of a macroamidine (covered in section 1.7.2). Surprisingly, the two YcaOs within this gene cluster also do not have an E1-like partner protein suggesting that fully functional, standalone YcaOs are a possible division within the YcaO superfamily. Another case for additional YcaO functional capacity comes from thioviridamide biosynthesis, as it is devoid of azol(in)e heterocycles. Instead, it has several thioamides (sulfur in place of the oxygen in the peptide backbone) and an uncharacterized protein (TfuA) seems to occur immediately adjacent to the YcaO gene in many divergent YcaO-containing gene clusters, indicating that this may be another distinguishable division (section 1.7.3). While the role of the YcaO proteins during bottromycin and thioviridamide have yet to be experimentally established, these posttranslational modifications could share a mechanism, proceeding through a common phosphorylated amide intermediate, with the only variable being the identity of the nucleophile.

With the varied activities associated with the YcaO superfamily, categories have been introduced to help distinguish different types of YcaO (Figure 1.2). The most well-studied are azoline-forming YcaOs

that cyclodehydrate Cys, Ser, and Thr through collaboration with a member of the E1-like superfamily. These azoline-forming YcaOs are further described in the next section. Related to this type are standalone azoline-forming YcaOs, like in bottromycin, that appear to install azoline heterocycles without any partner protein. Also found in the bottromycin BGC is an amidine-forming YcaO. Another emerging type are YcaOs that co-occur with a TfuA gene. The only investigated example to date is found in the thioviridiamide gene cluster and may be a thioamide-forming YcaO (also referred to as TfuA-associated). Lastly, because some YcaOs lack the experimental data to be confidently assigned a function, the remaining YcaOs are generally referred to as enigmatic YcaOs. These would include members like the YcaO associated with RimO-dependent thiomethylation and other YcaOs with no predicted function. This review covers the current knowledge for each of these groups, with emphasis on the extensively explored azoline-forming YcaOs, and outlines future research directions and applications.

1.1.1 Azoline-forming YcaOs

By far the most explored group of the YcaO superfamily are the azoline-forming YcaOs. These are associated with thiazoline and (methyl)oxazoline biosynthesis and have been found exclusively within RiPP biosynthetic pathways. Among RiPPs, backbone azole and azoline heterocycles are found within several formally defined subclasses, including the linear azol(in)e-containing peptides (LAPs), azol(in)e-containing cyanobactins, thiopeptides, and bottromycin.^{26, 41} In each RiPP subclass, the YcaO protein is responsible for azoline formation, but additional structural modifications further differentiate them. For instance, LAPs are linear peptides with backbone azol(in)e heterocycles. However, azol(in)e-containing cyanobactins are formed by N- to C-macrocylation of a LAP intermediate,^{42, 43} and it is this macrocylation that defines the cyanobactins as some lack azol(in)e heterocycles and correspondingly do not have a *ycaO*.²⁶ Similarly, during thiopeptide biosynthesis, a LAP intermediate is formed which later undergoes a class-defining [4+2] cycloaddition to give rise to a six-membered, nitrogen-containing heterocycle (often pyridine).^{44, 45} Lastly, the bottromycins bear not only a Cys-derived thiazole, but also a macroamidine. To underscore the biosynthetic commonality of these four RiPP subclasses, we will generally refer to them as azol(in)e-containing RiPPs.⁴¹ Throughout this review, an azoline-forming YcaO is defined as any member of the YcaO superfamily that is involved in heterocycle formation with a RiPP substrate through the cyclodehydration of Cys, Ser, or Thr. Additionally, we will use RiPP cyclodehydratase to refer to the protein complex of an azoline-forming YcaO and any associated partner proteins that catalyze heterocycle formation in RiPP pathways.

The first YcaO to be studied was azoline-forming and involved in the biosynthesis of a LAP named microcin B17 (MccB17).^{37, 46, 47} Owing to their relative chemical simplicity, LAPs have historically provided the most significant insights into the function of YcaO proteins. Moreover, because other more

structurally complex azol(in)e-containing RiPPs proceed through a LAP-like intermediate, this classic work laid a foundation for the more recent comprehension of RiPP cyclodehydratase function *senso lato*. However, it would require the revolution in genome sequencing before the broad distribution of the cyclodehydratase and the diversity of molecular scaffolds that were biosynthesized by azoline-forming YcaOs was fully realized. In the mid-2000s, the next group to be biochemically characterized after MccB17 was the cyanobactins.⁴⁸⁻⁵⁰ The number of BGCs was then greatly expanded as a result of studying the genetic distribution of the cytolytic bacterial virulence factor, streptolysin S (SLS).⁵¹ As sequencing technology continued to advance, the long-known and potent antibacterial thiopeptides were also demonstrated to be RiPPs that employed azoline-forming YcaO proteins.⁵²⁻⁵⁶ A 2015 survey found a total of ~1,500 azoline-forming YcaO gene clusters distributed among numerous bacterial phyla and even some archaea.⁵⁷

The ability of RiPP biosynthetic pathways to incorporate thiazole(in)e and oxazol(in)e heterocycles through a YcaO protein greatly expands their structural variability and bioactive potential. From a structural perspective, azol(in)e heterocycles rigidify and enforce specific orientations in peptidic natural products so they presumably can interact with their intended targets more efficiently. Consequently, these heterocycles are also found in the products of non-ribosomal peptide synthetases (NRPS) or hybrid polyketide synthase (PKS)/NRPS pathways.^{58, 59} Bleomycin, a PKS/NRPS approved anticancer agent, has a crucial bisheterocycle that drives DNA intercalation.⁶⁰ Another anticancerazole-containing PKS/NRPS is epothilone and has a synthetic macrolactam analog, ixabepilone, which has also been approved for human use (Figure 1.3).⁶¹ In addition to their appearance in natural products, thiazoles are a popular heterocycle for medicinal chemists and thus they can be found in a number of approved drugs with activities spanning many therapeutic categories.⁶²⁻⁶⁴ For example, meloxicam, a nonsteroidal anti-inflammatory drug, and ritonavir, an HIV protease inhibitor, both possess thiazoles (Figure 1.3). However, no azoline-containing RiPP has been approved for use in humans because they tend to have inadequate pharmacological properties. Their peptidic nature lowers stability *in vivo*, and more highly modified azol(in)e-containing RiPPs, which are not susceptible to proteases are often limited by poor solubility, absorption, and bioavailability. Nonetheless, several are used in animals for various purposes with two examples being growth promotion in livestock (nosiheptide) and treatment of skin infections for both companion animals and livestock (thiostrepton).⁶⁵⁻⁶⁷ Additionally, azol(in)e-containing RiPPs often represent enticing, but challenging, drug leads for treatment of human disease. For instance a semi-synthetic derivative of GE2270A named LFF571 has enhanced solubility (12 mg/mL compared to <0.001 mg/mL) and is showing promise in clinical trials for treating *Clostridium difficile* infections (Figure 1.4).⁶⁸⁻⁷⁰

Based on the bioactivities of azol(in)e-containing RiPPs, there has been much interest in identifying gene clusters that have potential to encode structurally and functionally unique compounds. This has

generated interest in identifying YcaO-containing BGCs, but as new YcaO sequences are added to public databases regularly, there is a need to bioinformatically distinguish which may catalyze azoline formation. Fortunately, azoline-forming YcaOs are in many cases readily differentiated from other YcaOs based on whether the a partner (E1-like) protein is locally encoded or if one is fused to the YcaO to yield a single polypeptide.^{17, 18, 57} Identifying standalone azoline-forming YcaOs is more difficult and currently relies on homology to proteins encoded by the bottromycin and trifolitoxin BGCs (see sections 1.7.1 and 1.7.2, respectively).

From the known or putative azoline-forming YcaOs, a powerful insight into the diversity of their products has been revealed by plotting all proteins on a sequence similarity network (a method of clustering homologous sequences in two dimensional space).⁵⁷ Multiple distinct clusters of related YcaO proteins are formed as a result and can be viewed as a “family” given their genetic relatedness. An analogy can be drawn to human genealogy: identical twins look more alike than non-identical twins, parents and children also tend to more closely resemble each other in appearance as opposed to distant cousins or non-family members. Thus, if two YcaO proteins have high sequence similarity, they will be part of the same cluster and family, and correspondingly their BGCs and the structure of their mature natural products will generally be similar as well. However, more divergent YcaOs will be part of other discrete clusters and are likely to make a structurally different product. By comparing and analyzing the distribution of sequence clusters, an overview of the different potential molecular scaffolds encoded by azol(in)e-containing RiPPs is revealed, and distinct families are identified. Interestingly, many groups emerged from this analysis that fall within established subclasses (e.g. LAPs, cyanobactins, thiopeptides) but others had no characterized members, which implies new opportunities for the discovery of biologically active azol(in)e-containing RiPPs.⁵⁷ Another very recent bioinformatic survey went further and predicted the products of putative LAP, cyanobactin, and thiopeptide BGCs based on the encoded genes/peptides.⁷¹ The result of this analysis again underscored the great diversity of potential azol(in)e-containing RiPPs. In the next section, we review the subset of BGCs found through these bioinformatic surveys that have been experimentally investigated.

1.2 Linear azole-containing peptides (LAPs)

The LAPs are a heterogeneous subclass of RiPPs, with YcaO proteins being the primary element of their BGCs.²⁶ Their unifying feature is azole or azoline heterocycles within a linear ribosomal peptide. Accordingly, their BGCs can be as minimal as a precursor peptide and a cyclodehydratase. However, the cyclodehydratase is often accompanied by a flavin-dependent dehydrogenase which oxidizes azoline to azole heterocycles (Figure 1.5).^{37, 72} Additional tailoring enzymes, where present, further modify the peptide and are responsible for the chemical diversity amongst LAPs. For example, the selective anti-*Bacillus anthracis* compound named plantazolicin (PZN) has a dimethylated N-terminus whereas the secondary

metabolism inducer goadsporin undergoes N-terminal acetylation and contains dehydroamino acids (see sections 1.2.3 and 1.2.7, respectively). Meanwhile, other notable members of the LAP subclass possess no additional modifications (other than leader peptidolysis) such as MccB17 and the cytolytic virulence factor SLS.

With the unification of the subclass,²⁶ LAP biosynthetic genes and their products share a common nomenclature with A (precursor peptide), B (dehydrogenase), C (E1-like protein), and D (YcaO) although genes in some older BGCs are designated contrary to this convention based on historical precedent (i.e. MccB17). Also, the C and D proteins are found as a single fusion protein in roughly half of all known pathways.⁵⁷ The other modifying enzymes are pathway specific and are usually named sequentially starting at E (following from D).

1.2.1 Microcin B17

The secretion of a growth-suppressive metabolite is an effective strategy for bacteria to compete within their niche, where nutrients are a precious commodity. In cases where the growth-suppressive metabolite is of ribosomal origin, the bioactive species are referred to as bacteriocins. These can vary widely in molecular weight, structural modifications, and mode of action, which has convoluted their classification.⁷³⁻⁷⁷ The lower molecular weight bacteriocins are commonly referred to as microcins⁷⁸ to distinguish them from the higher molecular weight protein toxins, which will not be discussed further in this review. It is important to note, despite the incorrect terminology propagated, “microcins” need not come from Enterobacteriaceae. Indeed, the first description of microcins include examples from *Pseudomonas*, which are taxonomically distinct and non-enteric.⁷⁸

A well-studied microcin is microcin B17 (MccB17), a narrow-spectrum bactericidal peptide that targets DNA gyrase.⁷⁹⁻⁸¹ MccB17 was isolated from strain 17 during a screen of antibacterial compounds produced by intestinal bacteria of infants.⁷⁸ Soon the name MccB17 was established, based on the observation that it targeted DNA replication (class B).⁷⁴ Interest into MccB17 biosynthesis grew as the plasmid encoding the genes necessary for its production was isolated from *E. coli*.^{82, 83} Early experiments with this plasmid indicated that transcription of the MccB17 BGC was under the control of the extensively studied outer membrane protein regulator (OmpR) and increased greatly as cultures approached the stationary phase.⁸⁴⁻⁸⁶ After further genetic studies, the entire BGC was sequenced to reveal seven genes: *mcbABCDEFG* (Figure 1.5).^{87, 88}

While most of the MccB17 biosynthetic genes would be functionally analyzed by *in vitro* reconstitution or genetic deletion, the nucleotide sequence of *mcbA* immediately indicated that it encoded a precursor peptide as it matched the structure of MccB17, which was obtained from Edman degradation.³³ From these initial structural data, it was evident that a number of residues were posttranslationally modified,

including all four Cys and four of the six Ser.^{46, 47} Structural characterization by NMR showed these modifications to beazole heterocycles,^{34, 35} and they are formed by the action of a trimeric synthetase comprised of McbB, McbC, and McbD (E1-like, dehydrogenase, and YcaO proteins, respectively).^{36, 37} Afterazole formation on the core region of McbA, the leader region is removed by a protease (*tldD/tldE*) encoded on the chromosome of *E. coli*.^{89, 90} After complete processing, a two-component ABC transporter (McbE/McbF) presumably exports mature MccB17 while at the same time providing a level of self-protection. Previous reports have also designated McbG as an immunity protein, although it has no experimentally verified immunity mechanism.⁹¹ McbG is however homologous to pentapeptide repeat proteins which are known to structurally mimic DNA, interact with DNA gyrases, and protect against the quinolone antibiotics which represent another class of gyrase inhibitors.⁹²

Like most classically-defined microcins, MccB17 has a narrow spectrum of activity affecting only other Enterobacteriaceae and some *Pseudomonas* species. The narrow spectrum activity is largely due to the cellular uptake machinery that mediates the entrance of MccB17 into cells. For example, deletion or alteration of the C-terminal residues of McbA results in MccB17 analogs that retain inhibitory activity against DNA gyrase *in vitro* but are significantly less active towards live bacteria.^{93, 94} Similarly, MccB17 variants from *Pseudomonas syringae* inhibited both *Pseudomonas* sp. and *E. coli*, but their effectiveness varied depending on the precursor peptide sequence despite having similar gyrase inhibitory activity *in vitro*. Interestingly, when a Gly₃ sequence from the *P. syringae* MccB17 variant was substituted into *E. coli* McbA, this chimeric structure exhibited a broader spectrum of activity that included additional *Pseudomonas* strains. Together, these data demonstrate that multiple segments of MccB17 contribute to its uptake and bioactivity and that sequence modification is a viable option for controlling the activity spectrum of microcins.⁹⁵

Two proteins, OmpF and SmbA, are largely responsible for governing the entry of MccB17 into cells, and thus define its spectrum of activity. The outer membrane porin OmpF (and partially OmpC) first allows entry into the periplasm.⁹⁶ Subsequently, SmbA, an inner membrane protein involved in the uptake of many different peptide antibiotics, enables entrance to the cytoplasm.⁹⁶⁻⁹⁹ Once inside, MccB17 inhibits DNA gyrase and triggers double strand breaks in DNA. The initial DNA damage occurs through exploitation of the enzymatic mechanism.^{79, 81} DNA gyrase introduces negative supercoils, which are required for DNA replication, by wrapping DNA around itself, cleaving the two strands of the DNA, passing bound DNA through this newly formed gap, and finally re-ligating the DNA.^{100, 101} The active enzyme is a heterotetramer, A₂B₂, where the A subunit (GyrA) is responsible for DNA wrapping, cleavage, and re-ligation while the B subunit (GyrB) has an N-terminal ATPase domain and interacts with DNA and GyrA through its C-terminus. Although the exact binding site is unknown, a resistance mutant (W751R) suggests that MccB17 interacts directly with the C-terminal domain of GyrB.^{79, 102} MccB17 has little effect

on the rate of supercoiling by DNA gyrase *in vitro* (~2-fold reduction), but it does appear to stabilize a transient enzyme intermediate during DNA strand passage that results in release of bound DNA from the enzyme before it can be re-ligated (known as the cleavage complex).¹⁰³⁻¹⁰⁷ Since the damaged DNA cannot be replicated, inhibited cells undergo rapid DNA degradation that is prevented by resistance genes encoded in the MccB17 BGC.^{80, 108} In addition to MccB17, the proteinaceous DNA gyrase inhibitor GyrI also protects against MccB17 because it blocks the enzyme's interaction with DNA, thereby preventing DNA cleavage, and ultimately allows cell growth despite its inhibition of DNA gyrase.¹⁰⁹⁻¹¹¹

It has been reported that MccB17 isolated from natural sources actually replaces a backbone amide with one ester linkage resulting from an N → O acyl shift at residue 52.¹¹² This does not affect activity, as both the amide and ester forms exhibit activity against *E. coli* (minimum inhibitory concentration, MIC, of 60 nM).⁹⁴ The thiazole and oxazole heterocycles are critically important and cannot be substituted for each other without loss of activity, and the antibacterial potency correlates with the total number of heterocycles present in the peptide.^{107, 113} Of these azoles, the second bisheterocycle (B-site) is the most important for antibiotic activity and may intercalate DNA, like the bisthiazole of bleomycin (Figure 1.3), although this has yet to be proven.⁶⁰ Notably, MccB17 does not directly bind naked DNA but likely interacts with DNA once bound to gyrase.¹⁰⁵ Conceivably, many new MccB17 derivatives could be explored through solid phase peptide synthesis, but in practice, the Gly-rich region is synthetically troublesome and limits the usefulness of this approach.^{114, 115} Overall, MccB17 has a complex structure-activity relationship that remains poorly understood, since changes can affect export, import as well as inhibition. However, it is the study of MccB17 that has laid much of the foundation for our current understanding of RiPP cyclodehydratases.¹¹⁶ Later in this review, we will return to the topic of MccB17 biosynthesis for its insights into cyclodehydratase function (section 1.6).

1.2.2 Cytolysins

Certain human and animal pathogens employ cytolytic toxins that disrupt host cell membranes to enhance virulence during infection.¹¹⁷⁻¹²⁰ *Streptococcus pyogenes*, a bacterial cause of pharyngitis and more severe infections such as necrotizing fasciitis,¹²¹ produces a cytolytic LAP named SLS.¹²² Though its identity and nature were unknown at the time, the cytolytic activity of SLS was first reported in 1895 when certain Streptococci were observed to produce a hemolytic phenotype when grown on blood agar media.¹²³ Variations of this phenotype include the γ -, α - β -, and hemolytic phenotypes and describe no hemolysis, partial hemolysis, or complete lysis of erythrocytes, respectively. SLS causes β -hemolysis and results in complete clearing of blood agar around producing bacteria.¹²⁴ Early efforts to isolate and characterize SLS were limited by its poor physicochemical properties, and its need for stabilizing carrier molecules (e.g. RNA, serum, detergents) to retain activity.^{125, 126} Despite these challenges, analysis of crude preparations indicated

that the active component of SLS was a peptide.¹²⁷⁻¹³⁰ However, it was not until the year 2000 that transposon mutagenesis and DNA sequencing studies^{131, 132} revealed that SLS was a RiPP.^{133, 134} These genomic studies demonstrated that nine genes, *sagABCDEFGHI*, comprised the SLS operon (Figure 1.6).¹³⁵

As is standard for LAP BGCs, a precursor peptide (SagA) and azole synthesizing machinery are present (SagB, SagC, and SagD).^{51, 136} SagE is presumably the leader peptidase,^{137, 138} as it is the only gene within the operon with homology to a protease, in this case, the type II CaaX proteases¹³⁹⁻¹⁴¹ and some homologs are also involved in bacteriocin maturation pathways.¹⁴² SagGHI comprise a multi-component ABC transporter and thus are presumed to export mature SLS. Another predicted membrane protein, SagF, currently has an unknown function despite being conserved in multiple SLS-like gene clusters.^{51, 135, 137}

Although the gene cluster for SLS is known and should aid in determining its structure, the mature structure of SLS remains elusive due to difficulties in working with the molecule. From bioinformatics and gel filtration experiments,^{130, 143} the SLS core is predicted to begin just before the Cys-rich region at the C-terminus of SagA. Most Cys and Ser residues are expected to be heterocyclized based on the ability of heterologously expressed and purified SagBCD to accept MccA (the MccB17 precursor) as a substrate.^{51, 136} Partial MS support for conversion of the final two Ser in SagA to oxazoles has also been obtained.¹⁴⁴ The only other inference into which SagA residues are modified has come through mutagenesis of heterocyclizable residues (Cys, Ser, or Thr) to Ala or Val.¹³⁶ Assuming that a residue is normally heterocyclized into an azole, then the introduction of a free side chain at such a position should be detrimental to the resulting mutant's lytic activity. By assessing SLS mutants in this way, many of the residues in the N-terminal poly-heterocyclizable (NPH) region of the core, CCCCTTCCFS, were identified as important for blood cell lysis while only three of the remaining Ser outside of the NPH contribute to its effect.¹³⁶ Interestingly, some SLS-like gene clusters from *Borrelia* were predicted to encode a short cytotoxin, possessing only this short NPH, and indeed the minimal cytolytic unit of SLS is the NPH region (the first 11 residues of SLS).¹⁴⁵ This indicates that the extended and variable C-terminal region of SLS could play a role in governing cellular specificity.

Various types of membrane bound compartments are sensitive to SLS including leukocytes, platelets, subcellular organelles, and even some bacteria lacking a cell wall.¹⁴⁶⁻¹⁵⁰ SLS appears to interact with membranes and causes temperature dependent osmotic lysis by forming pores in the membrane.^{151, 152} Recently, more mechanistic insights were gained when SLS was shown to trigger influx of Cl⁻ by interacting with anion exchanger 1 (band 3), resulting in lysis, but the precise binding site and mechanism of ion dysregulation are not known.¹⁵³ Given the broad range of membranes affected by SLS, homologous anion transporters are likely also targeted in other eukaryotic cell types; however, sub-lethal amounts of SLS activate programmed cell death and inflammation in epithelial cells so SLS seems capable of affecting cells in multiple ways.¹⁵⁴

Because of the enhanced virulence of SLS-producing streptococci,^{137, 155} it is not surprising that other pathogens similarly employ SLS-like toxins. Other *Streptococcus* species that have been found to produce SLS-like cytotoxins include the following *Streptococcus* species: *S. dysgalactiae* subsp. *equisimilis*,¹⁵⁶ *S. iniae*,¹⁵⁷ *S. equi*,¹⁵⁸ and *S. anginosus*.¹⁵⁹ Additional Firmicutes encode SLS-like BGCs, including select strains of *Clostridium*, *Listeria*, *Staphylococcus*, and even certain Spirochete species from *Borrelia* and *Brachyspira*. These putative toxins have correspondingly been named clostridiolysin S (CLS),¹⁴⁴ lysteriolysin S (LLS),¹⁶⁰⁻¹⁶² stapholysin S,⁵¹ and borreliolysin S.¹⁴⁵ While only CLS and LLS have been isolated from their native organisms,^{144, 160} fusing the core sequence for any one of these toxins to the SagA leader peptide followed by treatment with SagBCD resulted in the production of a cytolytic agent (presumably upon the installation of the heterocycles).^{136, 145} The BGC for clostridiolysin S has the same architecture as SLS, exemplified by *C. botulinum*, and its cyclodehydratase (ClosCD) has been reconstituted *in vitro* with SagB as its dehydrogenase (the cognate ClosB was inactive).¹⁴⁴ In addition to these, other putative SLS-like gene clusters have also been bioinformatically identified among *Firmicutes* and *Actinobacteria* but require additional work to demonstrate their predicted activity.^{57, 137, 145, 163}

1.2.3 Plantazolicins

Plantazolicin (PZN) is a highly specific, bactericidal antibiotic that targets *Bacillus anthracis*, the causative agent of anthrax (Figure 1.7).¹⁶⁴ PZN is part of a growing trend where new molecules are first identified through genomics rather than traditional bioactivity-guided screening as a consequence of the increased availability of organism genomes.¹⁶⁵ The BGC of PZN was discovered in 2008 in the genome of *Bacillus velezensis* FZB42 (formerly *Bacillus amyloliquefaciens*) through its similarity to the SLS gene cluster.⁵¹ Three years later when PZN was isolated from its native host and structurally characterized, much of its structure had been correctly predicted based on its gene cluster and the biosynthetic logic of azoline-forming YcaOs. Indeed, with the movement toward genome-guided natural product discovery, genome-guided structure elucidation has become increasingly powerful.¹⁶⁶ Based on the PZN precursor peptide (BamA), and presumed similarity to SLS, all Cys, Ser, and Thr residues were predicted to be converted to azol(in)es by the cyclodehydratase/dehydrogenase (BamBCD). By homology to SagE, the previously described type II CaaX-like protease, BamE was hypothesized to remove the leader.¹³⁸ Additionally, methylation was anticipated based on the presence of a *S*-adenosylmethionine (SAM)-dependent methyltransferase (BamL). This methyltransferase was later shown to selectively dimethylate the newly formed N-terminus of PZN through a narrow substrate tunnel which limited access to the active site.¹⁶⁷⁻¹⁷⁰ Although the exact position and selectivity of these modifying enzymes could not be predicted *a priori*, the high resolution mass of purified PZN was consistent with 10 cyclodehydrations, 9 dehydrogenations, 2 methylations, and leader

peptidolysis between Ala-Arg (Figure 1.7). With this information, NMR was used primarily to confirm the structure rather than elucidate it outright.^{164, 171, 172}

Mature PZN is presumably exported by an ABC transporter comprised of BamG and BamH. Of the remaining proteins encoded in the BGC, BamK is a predicted transcriptional regulator, but BamF, BamJ, and BamI have unclear functions. Deletion studies showed BamF and BamI were dispensable for PZN production whereas deletion of BamJ abolished PZN production.¹⁷¹ Notably, BamJ is present in every PZN-like gene cluster,^{164, 173} but it is clearly not required for the *in vitro* formation of PZN, given that the biosynthetic pathway has been reconstituted for two PZN family members and neither required the addition of BamJ.¹⁷³

Since the identification of the first PZN gene cluster, 14 additional PZN-like BGCs have been identified through conservation of the genes *bamJBCE* and similarity of motifs in the precursor peptides (two discrete regions of 4 to 6 heterocyclizable). The product from *Bacillus pumilus* (Bpum) has been confirmed to be identical to PZN as indicated by its core peptide,¹⁶⁴ and one other close PZN variant from *Bacillus badius* (Bbad) named “badiazolicin” (BZN) has also been isolated and shown to be similarly active against *B. anthracis*.¹⁷³ Importantly, work with proteins from the Bpum BGC allowed reconstitution of a PZN azole-synthetase by combining BpumC with BamB and BamD (the cognate BamC expressed poorly and was unstable). With an active azole-forming synthetase (BamB/BpumC/BamD) *in vitro*, numerous precursor peptides could be tested as substrates and partially processed intermediates could be analyzed to determine the rules governing substrate processing (further discussed in section 1.6.3). Encouraged by this success, the azole synthetase (CurBCD) from *Corynebacterium urealyticum* was also reconstituted and demonstrated to install ten azoles on CurA. Then, taking this decazole intermediate forward, its leader peptide was removed, and it was subjected to chemical N-terminal dimethylation to yield the putative natural product “coryneazolicin” (CZN).¹⁷³

PZN is only known to target *B. anthracis*, with the exception of one *B. anthracis* strain that is non-susceptible and one *B. cereus* strain that is susceptible.^{164, 174} Given this ultra-narrow spectrum of activity, there has been much investigation into the mode of action of PZN.¹⁷⁴ Nonetheless, it is still not clear what mediates PZN susceptibility or selectivity. PZN resistance mutants, which could indicate the target, upregulated multidrug resistance efflux pumps or had missense mutations in two-component regulatory gene directly upstream of a cardiolipin synthase.¹⁷⁴ As efflux pumps provide general resistance, this result suggested that cardiolipin may be involved in the PZN mode of action, and in support of this, fluorescently-labeled PZN co-localized with dye that approximates the location of cardiolipin in *B. anthracis* membranes. Cardiolipin, a diphosphatidylglycerol lipid associated with regions of membrane curvature,¹⁷⁵⁻¹⁷⁷ has an unusual localization in *B. anthracis* relative to other bacteria.¹⁷⁴ Accordingly, it was hypothesized that PZN

interacts with these unique regions of cardiolipin to destabilize and lyse the cell membrane of only *B. anthracis*, but the molecular details of this process remain to be elucidated.

Given the narrow spectrum of PZN, there has been interest in obtaining new variants to explore structure-activity relationships and identify the key features for bioactivity. Deletion of the methyltransferase from *B. velezensis*, produces PZN with an unmodified N-terminus (desmethyl-PZN), but this variant had dramatically reduced potency.¹⁶⁴ Full heterocyclization was also important as acid hydrolysis of the methyloxazoline in PZN showed a 10-fold loss of activity (Figure 1.7)¹⁶⁴ In another experiment designed to generate and test novel PZN derivatives, the PZN gene cluster was heterologously expressed in *E. coli* with different precursor peptides.¹⁷⁸ However, the heterologous biosynthetic pathway was essentially intolerant to substitution at positions normally converted to azoles. Ten variants at other accepted positions were purified from *E. coli* but were found to be less active.¹⁷⁸ Overall the data suggest that few PZN structural features can be changed without reducing activity against *B. anthracis*.

Additional information about the structure-activity relationships of PZN comes from chemically synthesized analogs. PZN has been the subject of several total syntheses¹⁷⁹⁻¹⁸¹ and the N-terminal half of PZN (Me₂-Arg-Az₅) has also been prepared synthetically as a methylester (Figure 1.8). Unexpectedly, this truncated pentazole analog retained nearly all of its antibiotic potency towards *B. anthracis* but additionally killed *S. aureus* and other Gram-positive bacteria unaffected by PZN.^{167, 169} In accord with a broader spectrum of activity, the mode of action for Me₂-Arg-Az₅ appears distinct from PZN based on the observation that PZN resistance mutants are sensitive to Me₂-Arg-Az₅, sub-lethal treatments induce different expression signatures captured by RNA-seq, and that Me₂-Arg-Az₅ kills *B. anthracis* without cell lysis.¹⁷⁴ The molecular target of Me₂-Arg-Az₅ is unknown, but these results suggest that the N-terminal half of PZN is the bioactive portion while the C-terminal region could control targeting or selectivity. These data are reminiscent of what was observed for SLS, with the NPH being the minimal portion required for activity.¹⁴⁵

1.2.4 Hakacins

The hakacins represent an example of a RiPP BGC found in *Bacillus* sp. Al Hakam (Balh) for which no known compound has been isolated but is presumed to be a LAP based on the genes present.¹⁸² The Balh gene cluster was first identified in 2008 and contains two precursor peptides (BalhA1 and BalhA2), a cyclodehydratase (BalhC and BalhD), a dehydrogenase (BalhB), and a nearby transporter; no specific protease appears linked with the gene cluster (Figure 1.9).^{51, 182} Other nearby genes have unknown function but also could play a role in modifying the peptide. Highly related BGCs can also be found in closely related *Bacillus* species.^{57, 182} Even though the hakacin natural product has not been identified, the likely location of heterocycles in BalhA1/A2 has been determined from *in vitro* reconstitution reactions.^{182, 183}

Unfortunately, BalhB did not co-purify with its flavin co-factor so it was substituted with a dehydrogenase (BcerB) from one of its homologous BGCs in *B. cereus*, as has been done for other previously discussed cyclodehydratases.^{144, 173} Because the active azole synthetase complex could be formed in this way, the Balh cyclodehydratase has proved valuable for probing YcaO-dependent heterocycle formation (section 1.6).

1.2.5 Heterocycloanthracins

The heterocycloanthracins (HCAs) represent another predicted group of LAP BGCs found through genomic analysis that identifies characteristic features such as a YcaO protein and Cys-rich regions at the C-terminus of their precursor peptides. The HCAs are primarily found in *B. anthracis* and *B. cereus* and possess numerous repeats of Cys-Xxx-Xxx within their precursor peptides, which often are distally encoded relative to the biosynthetic enzymes.¹⁸⁴ Generally, the only other genes that compose an HCA gene cluster are a cyclodehydratase, a partner protein required for cyclodehydratase activity (“ocin-ThiF”), and often a dehydrogenase (Figure 1.10). Other potential tailoring enzymes are often present, such as methyltransferases, succinyltransferases, and 2-oxoglutarate dehydrogenases, but no transporters, immunity proteins, or proteases appear in the local genomic region. Additionally, because most HCA clusters have high sequence identity, it can be more difficult to predict BGC boundaries by conservation analysis.

Insight into HCA biosynthesis has come from studying the cyclodehydratase from *B. Al Hakam*. These studies revealed that the “ocin-ThiF-like” protein, named HcaF, encoded immediately adjacent to the cyclodehydratase HcaD, was required to reconstitute azoline biosynthesis.¹⁸⁵ HcaF is homologous to ThiF, a member of the E1-like superfamily but has a distinct role in the biosynthesis of the thiazole moiety of thiamine.¹⁸⁶⁻¹⁸⁸ Consequently, HcaF is referred to as a cyclodehydratase partner protein and is referred to as the F protein. Ocin-ThiF-like partner proteins are also commonly found in thiopeptide BGCs (section 1.3). Interestingly, azol(in)e-containing RiPP BGCs that encode ocin-ThiF-like proteins (e.g. HCA) lack a canonical E1-like “C” protein, which is found in more standard examples (e.g. hakacin), but part of the canonical LAP C protein appears fused to the YcaO protein. To distinguish these pathways, the term F-protein dependent cyclodehydratase was introduced.¹⁸⁵

Of all the identified HCA BGCs, only one product has been purportedly isolated from *Bacillus sonorensis* MT93. Named sonorensin, it was demonstrated to have potent antibacterial activity against *Listeria monocytogenes* and hence, may be useful as a food preservative.¹⁸⁹⁻¹⁹¹ However, the structure and the biosynthetic origin remain unverified, and the reported low-resolution mass of the isolated antibacterial compound does not match the putative core peptide. Further, the proposed sonorensin leader peptide cleavage site is unlikely given that multiple Cys residues appear before it, and for most other azol(in)e-containing RiPPs, the first Cys is indicative of the beginning of the core. This proposed cleavage site, if

correct, would be highly unusual.¹⁸⁹ Recombinant expression of only the proposed sonorensin core peptide, in the absence of any modifying enzymes, resulted in a peptide with anti-*Listeria* activity, an observation that led to the conclusion there is no modification of the natural product.¹⁸⁹ However, until further studies including high-resolution mass spectrometry, NMR, and biochemical testing, the actual structure of sonorensin and whether it belongs among azol(in)e-containing RiPPs remains uncertain.

1.2.6 Azolemycins

One of the primary differences between LAPs is the variable functional groups installed by the ancillary tailoring enzymes. In this regard, the azolemycins are highly unique LAPs because they feature a rare oxime moiety. A few other natural products possess an oxime moiety, for example caerulomycin A,¹⁹² althiomycin,¹⁹³ collismycin A,¹⁹⁴ and nocardicins A and B,¹⁹⁵ but no RiPP was previously known to display this group. Azolemycin was found in extracts of *Streptomyces* sp. FXJ1.264 based on its mass and predicted molecular formula, which did not match any known compound.¹⁹⁶ Further NMR characterization revealed four derivatives that vary based on the configuration of the oxime and its methylation pattern. The structure was also independently confirmed by total synthesis (Figure 1.11).¹⁹⁷ No definitive bioactivity has been found for the azolemycins with only modest anti-proliferative activity being observed.¹⁹⁷

Based solely on the structure, it was difficult to predict the genetic origin of the azolemycins so the producer, *Streptomyces* sp. FXJ1.264, was sequenced.¹⁹⁶ This revealed a BGC with a precursor peptide (AzmA) containing the sequence VVSTCTI that could give rise to the azolemycins. A fused cyclodehydratase (AzmC/D) and discrete dehydrogenase (AzmB) form the four azoles. The protein responsible for oxime formation (AzmF) has homology to flavin-dependent monooxygenases. Deletion of *azmF* gave products with no oxime and primarily yielded a truncation product lacking the N-terminal Val residue, indicating that the oxime protects against proteolysis. AzmE is an SAM-dependent methyltransferase that installs two methyl groups. Similar to PZN, desmethyl derivatives can be isolated from the Δ *azmE* strain.¹⁹⁶ Notably, there are no locally encoded proteases encoded in the BGC, yet AzmA has both N- and C-terminal regions flanking the core peptide sequence which must be removed during the maturation process. Overall, these investigations confirmed the ribosomal origin of the azolemycins and demonstrated how oxime functional groups can be introduced in RiPPs.

1.2.7 Goadsporin

Since sequencing the genome of *Streptomyces coelicolor* A3(2),¹⁹⁸ it has been evident that many BGCs are cryptic in that they do not produce detectable amounts of product. This is true even for organisms that are known to produce many different secondary metabolites. While these cryptic BGCs can be activated by various genetic methods or in heterologous hosts, altering growth conditions or media is often a fruitful

approach.^{199, 200} In their native environments, bacteria live in complex multi-species communities and communicate with each other through various physical and chemical cues. Mimicking this environment by supplementing growth media with extracts from different bacteria can cause the activation of pathways that are silent under normal laboratory culturing conditions. Goadsporin was initially discovered by screening several hundred actinomycete broths for the ability to stimulate production of a red pigment, actinorhodin, by *Streptomyces lividans* TK23.²⁰¹ Goadsporin was isolated from extracts of *Streptomyces* sp. TP-A0584 and further characterization indicated that it induced secondary metabolism and sporulation in various streptomycetes. Its bioactivity spectrum was broad among *Streptomyces*, affecting 36 of 42 tested strains at a concentration of 1 μ M. Above this concentration, goadsporin displays growth inhibitory activity.

The structure of goadsporin is unusual for a LAP because in addition to itsazole heterocycles it contains two dehydrated amino acids (Figure 1.12).^{202, 203} Genetic deletion studies support a biosynthetic pathway in which the azoles are formed first by a fused cyclodehydratase (GodD) and a dehydrogenase (GodE).²⁰⁴ In the next likely step, dehydroamino acids are formed by what is now termed a split LanB dehydratase that catalyzes the glutamylation of select Ser and Thr (GodG) and subsequent elimination (GodF) to yield the corresponding dehydroalanine and dehydrobutyrine moieties.^{204, 205} Putative membrane-associated proteases (GodB and GodC) cleave the modified precursor peptide (GodA), and an N-acetyltransferase (GodH) modifies the newly formed N-terminus.^{206, 207} Transcription of the goadsporin BGC is controlled by GodR, an activator,²⁰⁸ whose overexpression results in increased production of goadsporin, although the best titers were achieved through heterologous expression in *S. lividans* and co-culturing with *Tsukamurella pulmonis*.^{208, 209} In order to prevent inhibition of its own growth, the BGC encodes the immunity protein GodI (see below).

The dehydrated residues of goadsporin are noteworthy because the split LanB dehydratase that forms them is so named owing to homologous proteins being involved in lanthipeptide biosynthesis. Lanthipeptides comprise their own RiPP class and are beyond the scope of this review; however, they have been extensively reviewed elsewhere^{210, 211} and unique aspects of their unusual enzymology are found in an accompanying review from van der Donk, Nair, and co-workers.²¹² Genes encoding a split LanB dehydratase are also found nearby cyclodehydratase genes in the thiopeptide class of RiPPs (section 1.3). Thus, a LanB dehydratase can be found in multiple RiPP classes and can provide increased chemical/structural diversity to a natural product pathway. Indeed, the dehydroamino acids of goadsporin are required for bioactivity.²⁰⁴ Other structural features that have been investigated show flexibility in swapping oxazole and methyloxazole heterocycles,²¹³ but alteration of the C-terminus or to Gly10 compromise its activity.

Insight into the mode of action of goadsporin has come from studying its immunity protein. Because GodI has homology to the 54 kDa “Ffh” subunit of the signal recognition particle (SRP),

goadsporin is believed to bind and inhibit native Ffh. This would interfere with the normal function of the SRP, which transports specific proteins to the plasma membrane. However, GodI is insensitive to goadsporin and presumably replaces the inhibited subunit, allowing normal SRP function.²⁰⁷ The observations that goadsporin induces secondary metabolism but at higher concentrations is growth inhibitory raises questions about what its “true” function may be in vivo. Goadsporin serves as a reminder that not all natural compounds that inhibit the growth of various microorganisms were necessarily intended for niche competition.^{214, 215}

1.3 Thiopeptides

The thiopeptides (originally thiazolyl peptides²¹⁶) are an extensively studied group of azol(in)e-containing RiPPs known for their inhibition of protein synthesis and potency towards various Gram-positive bacteria, especially multi-drug resistant strains, e.g. methicillin-resistant *Staphylococcus aureus* (MRSA), penicillin-resistant *Streptococcus pneumoniae*, and vancomycin-resistant enterococci.^{32, 217, 218} Since the first thiopeptide micrococcin was discovered in 1948,²¹⁹ over 100 variants have been isolated. In addition to their antibiotic activity, other reported bioactivities include anticancer, antimalarial, antifungal, and immunosuppressive activities, as well as inhibition of renin and RNA polymerase (see section 1.3.3).⁶⁵ Beyond these diverse activities, thiopeptides have received attention for their architectural complexity (Figure 1.13). The defining feature of a thiopeptide is six-membered, nitrogenous heterocycle within a macrocyclic framework also decorated with multipleazole heterocycles and dehydroamino acids.^{26, 32, 65} Thiopeptides frequently feature additional tailoring, including methylations, hydroxylations, and in some cases, a second, ester-linked macrocycle derived from Trp or a thioether macrocycle derived from Cys. Due to the molecular complexity, early structure determinations required much effort and relied heavily on chemical degradation. Eventually, mass spectrometry, NMR, X-ray crystallography, and total synthesis became standard methods for structural elucidation and confirmation. These studies and advances in thiopeptide total synthesis have been reviewed elsewhere.^{32, 65, 220-224} Based on their structural complexity, many believed thiopeptides would be produced by non-ribosomal peptide synthetases (NRPSs).^{32, 225} However in 2009 their ribosomal origin was firmly established by multiple independent research groups, and *ycaO* genes were found in all cases, providing a rationale for the azol(in)e heterocycles found in all members of this class.^{52-56, 226, 227} Since this time, much has been learned about the role of individual biosynthetic proteins,^{65, 224, 228, 229} and taking advantage of their biosynthetic logic, many new thiopeptide derivatives have been made easily using different biosynthetic approaches.^{45, 224, 229, 230} We focus on the thiopeptides from a biosynthetic perspective with an emphasis on how their observed structures and activities are encoded by their BGCs.

1.3.1 Common biosynthetic features

Despite the large number of known thiopeptides and their congeners, these compounds share many overall similarities in their structures, biosynthesis, and biological effects. As mentioned above, the six-membered nitrogenous heterocycle is a class-defining feature of the thiopeptides, but it appears in different oxidation states and substitution patterns which has led to the classification of five different thiopeptide series termed *a* through *e* (Figure 1.13).³² The *a* series possess a fully reduced piperidine, as exemplified by some thiopeptins, whereas the *b* series has a 1,2-dehydropiperidine ring with thiostrepton being the archetype member. The *c* series currently has only one member, Sch 40832, which displays a bicyclic piperidine fused to an imidazoline. The *d* series is most populous and harbors a tri-substituted pyridine ring as in micrococcin P1. Lastly, the *e* series exhibit hydroxylated pyridines as illustrated by nosiheptide. In addition to the macrocycle that is formed by the central pyridine/piperidine heterocycle, members of the *a-c* series contain a second macrocycle made from a connection at this central heterocycle to a side chain of the macrocycle through a quinaldic acid moiety. The *e* series also harbor a second macrocycle, but it is smaller and attached to two side chains of the primary macrocycle through an indolic acid moiety.

Early insights into the biosynthesis of thiopeptides came through feeding studies using isotopically labeled amino acids (thoroughly reviewed elsewhere³²), among other metabolic precursors, to trace the chemical origins for various parts of the structure.^{225, 231-241} These studies revealed that all structural components of thiopeptides were derived from amino acids, as implied by their structures and degradation products. The azol(in)e heterocycles came from the cyclodehydration of Cys, Ser, and Thr while the dehydroamino acids derived from dehydration of Ser and Thr. The nitrogenous six-membered heterocycle originated from Ser residues and was first proposed to form through a cycloaddition reaction in 1977 (Figure 1.13).²⁴² The quinaldic acid moiety in the secondary macrocycle of the *a-c* series thiopeptides originated from tryptophan and required a ring expansion transformation, and similarly the indolic acid linked secondary ring of the *e*-series also derived from tryptophan through a unique rearrangement reaction.

The next major advance in understanding thiopeptide biosynthesis required discovery of their BGCs. Because of their peptidic nature, thiopeptide could be made ribosomally (i.e. RiPPs) or by NRPSs.²²⁶ Attempts to identify DNA sequences within producing organisms that might encode a precursor peptide matching the deduced precursor sequence of known compounds were initially fruitless, and chloramphenicol, an antibiotic that inhibits protein translation, did not seem to block production of nosiheptide, which suggested that the ribosome is not needed in thiopeptide biosynthesis.^{243, 244} Other evidence showed that the quinaldic acid and indolic acid moieties found in some thiopeptides had their carboxylic acid groups activated with ATP, consistent with NRPS coupling mechanisms.²⁴⁵ Moreover, a NRPS putatively involved in micrococcin biosynthesis was identified,²⁴⁶ and other heterocycle containing and macrocyclic compounds (e.g. bacitracin) were known to be made by modular NRPSs.^{247, 248} However,

as more RiPPs were studied and discovered in the 2000s, the potential of a ribosomal origin for thiopeptides gained more traction.²²⁶ Dehydratases from the lanthipeptide subclass of RiPPs were known to transform Ser and Thr residues into dehydroamino acids,²¹⁰ and heterocycles derived from the peptide backbone in the azol(in)e-containing RiPPs were also recognized as being much more prevalent.⁴⁹⁻⁵¹ Consequently in 2009, a RiPP biosynthetic origin for thiopeptides was proven nearly simultaneously by several independent research groups.⁵²⁻⁵⁶

At present, BGCs for 14 thiopeptides have been verified using genetic manipulations and/or heterologous expression. Currently, these include thiocillin,^{52, 54, 249, 250} thiostrepton,^{54, 56, 251} siomycin,⁵⁴ nosiheptide,^{53, 252, 253} GE2270,^{55, 254, 255} thiomuracin,^{45, 55} nocathiacin,^{256, 257} cyclothiazomycin,^{258, 259} TP-1161,²⁶⁰ GE37468,²⁶¹ berninamycin,²⁶² lactocillin,²⁶³ nocardithiocin,²⁶⁴ and lactazole.²⁶⁵ Many natural congeners of these compounds are known to be produced by the same BGC. Based on these studies, all necessary biosynthetic genes were found to cluster near a precursor peptide (designated the “A” gene/protein) of typically 50-60 amino acids. Additionally, a primary set of six essential proteins emerged for building the core thiopeptide scaffold and included a split LanB dehydratase (B and C), [4+2] cycloaddition enzyme (D), dehydrogenase (E), cyclodehydratase partner protein (F), and cyclodehydratase (G). Despite the institution of a “standard” nomenclature,²⁶ the labeling of these genes tends to vary between BGCs, and this nomenclature is quite different compared to other subclasses. Usually, thiopeptide genes are clustered, but one counter example, cyclothiazomycin C, has a fragmented BGC with the [4+2] enzyme being distally encoded.²⁶⁶ In the remainder of this section, the biosynthesis of the core thiopeptide scaffold, the formation of the two tryptophan derived secondary rings, and C-terminal amidation are described. These steps have been expertly reviewed,^{65, 228, 229} but here we again cover this material with additional recent findings.

The roles of the aforementioned core biosynthetic enzymes were first assigned through bioinformatic analysis and characterization of partially processed intermediates in genetically manipulated strains.^{44, 204, 255, 262, 267} Although not conclusive, these studies suggested the biosynthetic order of events began with azol(in)e formation, then Ser/Thr dehydration, followed by macrocyclization through a putative [4+2] cycloaddition. The cyclodehydratase (G) was readily identified based on its obvious homology to azoline-forming YcaO proteins,^{17, 18} and similarly, the role of the dehydrogenase (E) in azoline oxidation to their corresponding azole heterocycles was easily recognized via homology with other flavin-dependent dehydrogenases known to dehydrogenate azoline heterocycles in RiPP pathways (Figure 1.5).⁷² However, the role of the cyclodehydratase partner protein (F) was not realized because its genetic deletion abolished thiopeptide production without providing any intermediates.^{53, 258, 265} The F-protein has low similarity to the cyclodehydratase protein (G) but is most clearly homologous to ThiF, an adenylating protein involved in thiamine biosynthesis and part of the E1-like superfamily.^{188, 268} In public genome databases, this protein is

most often annotated as “Ocin-ThiF-like.” Importantly, this ThiF-like partner protein contains the RRE, which binds the precursor peptide, and potentiates YcaO catalyzed cyclodehydration.^{27, 45, 185} Accordingly, the thiopeptide cyclodehydratase is most accurately described by the combined activities of the partner protein and YcaO because both are necessary for azoline biosynthesis.⁴⁵ Notwithstanding these observations, there are rare examples of thiopeptide producers that do not locally encode a ThiF-like protein homolog (e.g. *Streptomyces bernensis*, the producer of berninamycin²⁶²). The next step, dehydration of Ser/Thr residues is performed by a dehydratase composed of two proteins (B and C) that are homologous to the N- and C-terminal domains of class I lanthipeptide dehydratases respectively (referred to as a split LanB). These proteins catalyze dehydration through glutamylation of the hydroxyl group (B protein) using glutamyl-tRNA^{Glu} with subsequent elimination (C protein).^{45, 205, 269} A formal [4+2] cycloaddition of two dehydroalanines forms the six-membered nitrogenous heterocycle (Figure 1.13).^{45, 270} The identity of the [4+2] cycloaddition enzyme was inferred by its absence in goadsporin (section 1.2.7),^{204, 213} which structurally resembles a presumed intermediate during thiopeptide biosynthesis. Deleting this gene in thiopeptide BGCs resulted in the isolation of linear intermediates lacking this 6-membered heterocycle.^{44, 265, 271} Unexpectedly, the [4+2] enzyme has high similarity to the elimination domain of LanB dehydratases and possibly evolved through a duplication event. Whether or not the cycloaddition is concerted (Diels-Alder) or stepwise has yet to be evaluated, but these studies support a cycloaddition product as common biosynthetic intermediate from which subsequent reductions lead to the (dehydro)piperidine in the *b* and *a* series (Figure 1.13c). Also from this intermediate, dehydration and β -elimination of the leader peptide gives rise to the pyridine of the *d* series; for the *e*-series a cytochrome P450 monooxygenase must further hydroxylate this to form the hydroxypyridine.²⁵² Ultimately, the biosynthetic order of events leading to the primary macrocycle were verified by *in vitro* reconstitution.⁴⁵ More detail into the rules governing the heterocycle formation by the YcaO and how this azol(in)e formation interfaces with modification by other enzymes is further discussed in section 1.6.3.

After formation of the distinctive (primary) macrocyclic structure of thiopeptides, another common modification common is a secondary macrocycle connected through a quinaldic or indolic acid moiety (Figure 1.14).²²⁹ As shown in Figure 1.15a, the quinaldic acid derivative found in thiostrepton (Tsr) and other *a-c* series thiopeptides begins with C-2 methylation of tryptophan with overall retention of stereochemistry of the methyl group via a radical *S*-adenosylmethionine (rSAM) methyltransferase (TsrM, also known as TsrT),^{234, 272-274} and then an aminotransferase catalyzes deamination (TsrV, also called TsrA, note: naming conventions varies between research groups).⁵⁶ Next a putative oxidase (TsrQ) forms an intermediate that presumably undergoes ring opening and recyclization that may be enhanced by the action of polyketide-like cyclase (TsrS, also called TsrD).²⁷⁵ After formation of the quinaldic acid moiety, the ketone is reduced stereospecifically to an alcohol (TsrN, also called TsrU),²⁷⁵ and a probable next step is

attachment to a side chain of the primary macrocycle, which has been suggested to be catalyzed by an adenylate forming enzyme (TsrJ).^{239, 245} A putative cytochrome P450 oxygenase (TsrI) forms an epoxide likely after initial attachment to a side chain of the main scaffold.²⁷⁶ Lastly, the leader peptide is removed by an α/β -hydrolase fold protein (TsrB, also called TsrI),^{56, 277, 278} and the newly formed N-terminus attacks the epoxide stereospecifically to complete the macrocycle with the aid of the hydrolase.²⁷⁸ Formation of the quinaldic acid macrocycle seems required for *in vivo* production as no intermediates are isolated when these genes are inactivated.²⁷⁵ In contrast, the indolic acid moiety found in *e* series thiopeptides (e.g. nosiheptide, Nos) begins with a radical-mediated rearrangement to form 3-methylindolic acid (Figure 1.15b).^{53, 228} This step is performed by a rSAM enzyme (NosL) that has been extensively investigated *in vitro* and begins by abstracting hydrogen from the amino group.²⁷⁹⁻²⁸⁴ After fragmentation-recombination, the order of biosynthetic steps is not clear, but the 2-methylindole appears to be linked through its carboxylate to the main macrocycle by an ATP-dependent ligase (NosI?), based on a previously observed adenylated intermediate.^{53, 245, 256} The side ring can be further processed while attached to the peptide with methylation by a rSAM methyltransferase (NosN) and subsequent hydroxylation by an unidentified protein.⁵³ Lastly, the side ring is then likely closed by an acyltransferase (NosK?) to complete its maturation. This final condensation which completes the side ring appears to take place prior to macrocyclization.^{53, 256, 271} Thus, thiopeptide secondary rings have many commonalities in their biosynthesis and assembly,^{228, 229, 245} and attachment of the indole/quinaldic acid is somewhat malleable, exemplified by the observation that exogenously supplied fluorinated indole/quinaldic acids can be incorporated into the mature product without loss of activity.^{275, 282} The only exception to this related biosynthetic pathway for side ring formation is cyclothiazamycin which forms its secondary ring through a thioether linkage between the main macrocycle to its tail via an unknown mechanism (Figure 1.14c).

The last prevalent modification during thiopeptide biosynthesis is amidation of the C-terminal residue in the tail and is accomplished through two distinct strategies (Figure 1.16).^{228, 229} The first pathway, as exemplified during nosiheptide biosynthesis, utilizes an “amidase” (NosA) to catalyze acrylic acid dealkylation of a C-terminal dehydroalanine residue (Figure 1.16a). Notably, the precursor peptide contains an extra Ser that is absent in the mature product.^{53, 253, 285, 286} This mechanism is unlike other peptide amidations, which usually occur through oxidative cleavage of a C-terminal Gly residue.^{287, 288} A second route to C-terminal amide formation is seen in thiostrepton-like pathways and necessarily follows a different mechanism as there is no extra Ser residue in the core peptide (Figure 1.16b). Accordingly, amidation was proposed through transamidation performed by an asparagine synthetase-like protein (TsrT, also called TsrC) and requires a de-esterification process.^{54, 56} The C-terminus is normally in the form of a methylester installed by a methyltransferase (TsrP, also called TsrF), but a carboxylesterase (TsrU, also called TsrB) reforms the C-terminal carboxylate which undergoes subsequent amidation.²⁵¹ The free

carboxylic acid has some reduced potency relative to the amide but the methylester is the most potent derivative, indicating that the C-terminus has an important influence on bioactivity.²⁵¹ C-terminal amidation also enhances potency of other biologically active peptides, particularly hormone/signaling peptides,²⁸⁹ but its effect on activity varies among thiopeptides as it does not markedly affect nosiheptide or thiomuracin potency.^{45, 253} Overall C-terminal amidation appears to be a commonly occurring step in thiopeptide biosynthesis as very few thiopeptides have a free C-terminus. Other unique modifications (hydroxylation, methylation, glycosylation) more specific to individual thiopeptides are discussed later in individual sections.

1.3.2 Thiopeptide BGCs

Although thiopeptides contain common features and biosynthetic steps as described above, each compound is made unique by variable ancillary tailoring steps. Thiopeptide BGCs often possess methyltransferases, cytochrome P450 oxidases, or other enzymes that have been linked to specific modifications, and although the precise timing for many of these modification is not known, they are not required for the formation of the thiopeptide core structure.^{45, 251, 252, 255} Nonetheless, most thiopeptide BGCs produce multiple congeners through these enzymes, so even though there are over 100 known thiopeptides, the number of encoding BGCs is much less. Analysis of the known thiopeptide parent compounds indicate that ~40 precursor peptides are necessary to account for all compounds, as inferred from their structures. The peptides also tend to share a common pattern in the position of their azol(in)es, especially among thiopeptides with the same ring size (Figure 1.17). Accordingly, there is much that can be learned about how thiopeptide scaffolds are formed by comparing their different structures, even if the BGC remains unknown. Here, these distinctive thiopeptide subfamilies are reviewed with a focus on their unique biosynthetic steps and chemical properties.

1.3.2.1 Thiostrepton-like

Since its isolation in 1955, thiostrepton²⁹⁰⁻²⁹³ (also known as bryamycin²⁹⁴⁻²⁹⁷) has been one of the most studied thiopeptides. Several highly related, 26-membered macrocycle compounds have been reported, including siomycin,^{298, 299} Sch 18640,³⁰⁰ thiopeptin,^{301, 302} and Sch 40832.³⁰³ With the discovery of the thiostrepton and siomycin BGCs in 2009 (Figure 1.18),^{54, 56} the biosynthesis of their primary scaffold, secondary macrocycle, and C-terminal amide is known to be performed by the enzymes discussed in section 1.3.1. However, thiostrepton/siomycin are unusual in that they encode an extra dehydratase glutamylatone protein (glutamylatone domain) and possess a doubly hydroxylated Ile, which is likely installed by a cytochrome P450 hydroxylase (TsrR). Thiostrepton also possesses inverted stereochemistry at the thiazoline in position 9, which could originate from D-Cys, but this inversion seems likely to occur

spontaneously after posttranslational modification.³⁰⁴ Additionally, it is not clear which protein is responsible for reducing the presumed dihydropyridine intermediate post-cycloaddition to the final dehydropiperidine.

Reprogramming of the thiostrepton precursor peptide has shown that the presence of only one dehydrated residue in the secondary macrocycle of the mature product is not due to a limitation of the enzymes, as substitution with Ser or Thr at Ala2 or Ala4 also results in modification like in the other members.^{271, 277, 305} Other studies have also sought to increase the water solubility of thiostrepton without reducing target potency by introducing polar residues at positions 2 and 4. These positions were chosen because they point toward bulk solvent in the co-crystal structure of thiostrepton bound to the 50S ribosome and are less likely to affect binding potency. Unfortunately, charged residues at these positions were rejected by the biosynthetic pathway. The biosynthetic machinery were more tolerant to variations at position 4 than position 2, but of the accepted residues, generally hydrophobic residues were the only substitutions that retained activity.^{277, 305} Interestingly, the introduction of dehydrobutyrines in this region by genetic manipulation resulted in decreased activity, and none of the native sequences utilize Thr, perhaps suggesting that nature has already selected against less potent derivatives. Despite these lower potencies against live bacteria, *in vitro* translation inhibition was approximately identical for most of the substitutions, indicating the lowering in whole cell activity comes from other sources such as uptake, export, or stability.^{277, 305}

Other changes to the tryptophan-derived macrocycle have also been investigated. Exogenously supplied fluorinated or methylated quinaldic acid can be incorporated into thiostrepton by the native biosynthetic pathway and results in a slightly more potent derivative.^{277, 306} The quinaldic acid ring can also be contracted if Val or Ile are substituted at position 2 because the leader protease appears to cleave the peptide just before Val/Ile so the traditional position 1 will be removed when position 2 is Val/Ile.²⁷⁷ Additionally, a fourth ring has been formed by introducing a Cys into the side ring that can then adventitiously react with a dehydroalanine in the tail region.³⁰⁵ However, both of these modifications likely disrupt the structure of the quinaldic acid ring, preventing it from interacting with the 23S ribosomal subunit properly, and result in dramatic loss of antibacterial activity. By similar logic, mutation of Thr7 also decreases potency as this conserved residue also interacts with RNA when bound.^{276, 307} Unexpectedly, a C-terminal carboxylate is also detrimental to thiostrepton potency,²⁵¹ but it is not clear why from a structural view and could be due to uptake or other steps outside of target binding. Overall, these data indicate that bioengineering of thiostrepton's side ring tended to abrogate antibacterial activity, but it should also be noted that most of the tested mutants did not significantly change thiostrepton ability to inhibit the proteasome so this latter effect could more promiscuous.^{277, 308}

1.3.2.2 Nosiheptide-like

Closely related to the thiostrepton-like family, the *e* series thiopeptides like nosiheptide have a different secondary ring but identical core macrocycle size (26-membered). Nosiheptide (also known as multithiomycin³⁰⁹⁻³¹¹), was first isolated in 1970³¹² and is the most studied member of this group (Figure 1.19). The discovery of the nosiheptide BGC (*nos*) has provided many insights.^{53, 228, 257} The BGC includes a transcriptional regulator (NosP) and the standard complement of core biosynthetic proteins, indole-derived macrocycle-forming enzymes, and C-terminal amide-forming enzyme (NosA) under the control of a bidirectional promoter.³¹³ The C-terminal amide appears to be installed only after pyridine hydroxylation by a P450 (NosC).²⁵² In addition to these standard enzymes, a cytochrome P450 (NosB) is present that carries out the γ -hydroxylation of Glu.^{228, 252} This hydroxyl is the site of glycosylation in many other *e* series thiopeptides, such as glycothiohexide,³¹⁴⁻³¹⁶ philipimycin,³¹⁷ nocathiacin^{318, 319} MJ347-81³²⁰, thiazomycin,³²¹ and S-54832;³²² however, nosiheptide is naturally an aglycone. Because the primary structural difference between *e* series thiopeptides is the identity of their sugar (their core peptides vary only by Cys to Ser substitution), nosiheptide is representative of a likely common intermediate that would form during the biosynthesis of these other thiopeptides. Indeed, when the nocathiacin BGC (*noc*) was discovered,²⁵⁶ comparison with the nosiheptide BGC provided significant insight. The nocathiacin BGC is organized nearly identically to nosiheptide's but has additional modifying enzymes. Five total P450s (NocB, NocT, NocU, NocR, NocV) which corresponds to three additional enzymes presumed to be responsible for producing the most oxidized congener of nocathiacin. Additional oxygens appear at the dehydrobutyrine, indolic nitrogen, and β -carbon of Glu that is crosslinked to a methyl indole carbon resulting in a cyclic ether (Figure 1.19). After these oxidations, the hydroxybutyrine is likely *O*-methylated by NocQ, and a glycosyltransferase (NocS1) appends the *N*-dimethylated aminodeoxysugar (NosS2-5) to complete the biosynthetic pathways.²⁵⁶

Attachment of the sugar moiety to nocathiacin does not appear to enhance potency as the naturally isolated aglycone (nocathiacin III) was just as active as the more modified nocathiacins.³¹⁹ Despite the sugar moiety, nocathiacin exhibits the characteristically low solubility of other thiopeptides but more soluble variants of nocathiacin have been extensively studied through semi-synthetic modification.³²³⁻³³⁰ However, not as much is known about the importance of the structural features of nosiheptide aside from the previously mentioned minimal effects of converting the C-terminus to a carboxylate or incorporating fluorinated indolic acid moiety. One mutagenesis study has shown that substitution of Ser13 does not particularly affect the biosynthetic pathway except for preventing de-acylation to the C-terminal amide by NosA.³³¹ There has been more interest in optimizing nosiheptide production³³²⁻³³⁵ due to its use as a livestock growth promoter.^{66, 67}

1.3.2.3 Thiocillin-like

Another family of thiopeptides with 26 (atom) membered macrocycles lacks the secondary macrocycle while retaining the same pattern of azol(in)e heterocycles (Figure 1.17). Although micrococcin^{219, 336} was the first discovered member of this group (1948), thiocillin^{337, 338} was the first to have its BGC reported (Figure 1.20).⁵² The thiocillin BGC produces multiple congeners including micrococcin^{219, 336} and YM-266183^{339, 340} (also known as QN3223³⁴¹). Given that these variants escape modification by some of the ancillary tailoring enzymes, they are now generally referred to as thiocillins. More recently, an authentic micrococcin BGC has been identified and heterologously reconstituted in *Bacillus subtilis*, which has provided further insight in the biosynthesis of the thiocillin family.^{249, 250} Aside from the core biosynthetic enzymes, thiocillin BGCs contain two NAD(P)-dependent short chain dehydrogenases which convert a C-terminal Thr to an aminoacetone (TclP) or further to an aminoisopropanol (TclS). Although TP-1161 (section 1.3.2.7) also bears an aminoacetone moiety, it derives from a C-terminal Ser and the responsible enzyme (TpaJ) is an Fe-dependent dehydrogenase. In micrococcin, aminoacetone formation is required for antibacterial activity,²⁴⁹ whereas TP-1161 required the modification for production in vivo. The optional methylation of Thr8 in thiocillin is likely performed by a methyltransferase (TclO), and the Val6 hydroxylation appears to be installed by an Fe(II) oxygenase unlike other thiopeptides that usually employ a cytochrome P450 for hydroxylations. Although not involved in maturation, a putative *N*-acyltransferase may play a role in detoxifying stalled thiopeptide maturation by succinylating Lys residues on linear intermediates.³⁴²

Nocardithiocin³⁴³ and lactocillin are two other thiocillin-like molecules for which BGCs are known. The nocardithiocin pathway follows biosynthetic steps seen in other thiopeptides and contains a putative regulator (NotD). It has an Ile with two hydroxyl groups reminiscent of modifications seen in thioestrepton and thiomuracin that are likely installed by a P450. A second P450 likely hydroxylates a dehydroalanine (Dha4), with a modification also observed in some berninamycin-like thiopeptides. Two methyltransferases (NotC and NotE) are also present and could catalyze methylation of the hydroxydehydroalanine and of the C-terminus. In contrast, lactocillin is a more enigmatic and structurally divergent thiocillin-like molecule. Outside of the core biosynthetic enzymes, its unique modification is the indolyl-S-Cys residue at position 8. This moiety bears strong resemblance to nosiheptide's indole side ring, but the proteins that are likely involved in the thioether linkage, an adenylyltransferase (LclJ), hydrolase (LclK), and thiolation domain (LclI), are only convergently related to the nosiheptide side ring forming enzymes.²⁶³ Additionally, it is very unusual for thiopeptides to have an unmodified C-terminus, but lactocillin still possesses antibacterial activity.

Based on their macrocycle size, promothiocin and JBIR-83 are also included in the thiocillin-like group, but they bear a clear sequence resemblance to berninamycin-like thiopeptides which have a 35-

member macrocycle (i.e. radamycin). This is unexpected as thiopeptides within a given macrocycle size usually are more similar to each other than those of a different size. Currently, it is not known if the conformation of substrate or the cycloaddition enzyme has a stronger effect on controlling macrocycle size so the similarity between these thiopeptides with 26- and 35-member macrocycles may indicate the importance of the cycloaddition enzyme. Indeed, experiments with the thiocillin biosynthetic pathway demonstrate that insertion of Gly residues between Thr3 and Thr4 or deletion of Thr3 allows for the creation of 35 to 23 membered macrocycles, indicating that the enzyme recognizes the residues around the macrocyclization site (Ser1 and Ser10) regardless of the ring size.³⁴⁴ Additionally, thiocillins with residue substitutions that convert one heterocycle to an unmodified residue or vice versa inside the macrocycle can also be macrocyclized in select instances so multiple peptide conformations can be recognized by the [4+2] cycloaddition enzyme.^{267, 342} Introducing additional dehydroalanines (via Ser substitution) adjacent to the default macrocyclization site indicated that the default site is favored (Ser1) as the dienophile, but there is some flexibility with the diene site (Ser11) provided that there are two thiazoles C-terminal to the position,³⁴⁴ which may explain differences in cyclization between the 26- and 35-member macrocycles. However, it should be noted that all engineered thiopeptides with altered ring sizes appeared to be devoid of any antibiotic activity.³⁴⁴ Lastly, promothiocin/JBIR-83 are different from the other thiocillin-like compounds because they do not have antibiotic activity, but this is likely due to mutation of Thr3/4, which are essential for activity in thiocillin.³⁴² Indeed, promothiocin barely interacts with the 23S/L11 complex ($K_d > 10 \mu\text{M}$).³⁴⁵ However, these thiopeptides do still activate TipA demonstrating that TipA stimulation is not equivalent to antibiotic activity.³⁴⁶

Many thiocillin variants have been tested as substrates for the pathway to gain insight into rules governing processing (further discussed in section 1.6.3) and to decipher structure activity relationships. Based on the structure of micrococcin bound to the ribosome, the heterocycles are important for enforcing the geometry necessary for interacting with the ribosome.³⁴⁷ Replacement of anazole with a noncyclized residue significantly alters the structure of thiocillin,²⁶⁷ and accordingly no heterocycle positions within the macrocycle can be substituted without loss of activity. Similarly, the Ser residues that form the pyridine cannot be mutated, but generally the tail region is more accommodating to substitution. Mutation at other positions indicated that solubilizing, charged residues are not accepted or are inactive. Thr3 and Dha4 are both important for activity with Thr3 possibly involved in a hydrogen bond.²⁶⁷ Additionally, the non-canonical amino acids N ϵ -(tert-butoxycarbonyl)-L-lysine (BocK), N ϵ -allyloxycarbonyl-L-lysine, (AlocK) and N ϵ -prop-2-ynylloxycarbonyl-L-lysine (ProcK) have been incorporated into thiocillin with only partial loss of activity at Thr3 and Val8.²³⁰ Notably, using a Thr3ProcK variant, an azirine for crosslinking and biotin purification handle were attached and used to pulldown ribosomal protein L11, the target of

thiocillin.²³⁰ Overall, these analogs again suggest that the highly modified nature of thiopeptides are already optimized for binding to the ribosome as nearly all alterations reduce activity.

1.3.2.4 GE2270-like

A related group of thiopeptides with an expanded macrocycle size of 29-atoms is represented by GE2270³⁴⁸⁻³⁵⁰ but also includes amythiamicin,³⁵¹⁻³⁵⁴ thiomuracin,⁵⁵ GE37468,^{355, 356} and baringolin³⁵⁷ (also known as kocurin^{358, 359}) (Figure 1.17). Comparison of the BGCs for GE2270,^{55, 255} GE37468,²⁶¹ and thiomuracin⁵⁵ has given insight into the differences in the ancillary tailoring of these thiopeptides (Figure 1.21). The GE2270 BGC is normally activated during exponential growth and has been heterologously expressed in multiple organisms to better understand its biosynthesis.^{55, 254, 255, 360} Gene deletion studies indicate that Phe β -hydroxylation, Asn *N*-methylation, and *C*-methylation of thiazole 4 are performed by a cytochrome P450 oxidase (TpdQ), *N*-methyltransferase (TpdT), and radical SAM methyltransferase (TpdU), respectively.^{55, 255} An additional rSAM methyltransferase (TpdL) *C*-methylates the thiazole in the sixth position, which then undergoes hydroxylation by an unidentified protein and *O*-methylation (TpdM) to yield the methoxymethyl-thiazole (Figure 1.21). The C-terminal amide is probably formed as in other thiopeptides through cleavage of a C-terminal dehydroalanine (TpdK). In addition to these distinguishing ancillary tailoring steps, the GE2270 BGC has an unexpected duplication of the YcaO cyclodehydratase (TpdO) and the glutamylation domain of the split LanB dehydratase (TpdP). The reason for this is not fully understood, but the extra YcaO is associated with formation of the sole oxazoline. Other members of the GE2270 group convert this Ser to dehydroalanine.²⁵⁵ Given the similarity to GE2270, amythiamicin probably undergoes the same steps except at the site of Phe to Val substitution.

Thiomuracin biosynthesis is highly analogous to GE2270, however, there is only one rSAM methyltransferase (TpdI) for thiazole *C*-methylation. Further, there are two cytochrome P450 enzymes (TpdJ1 and TpdJ2) with one catalyzing Phe β -hydroxylation as in GE2270 biosynthesis and the other tailoring the side chain of Ile to various oxidation states (Figure 1.21).⁵⁵ Although the C-terminal Ala present in the precursor peptide is not detrimental to antibiotic activity,⁴⁵ it is presumably removed by a putative carboxypeptidase (TpdH) in the mature product. As previously stated, it is uncommon for thiopeptides to have free C-termini. On the other hand, the GE37468 BGC has a putative carboxypeptidase (GetM) for removal of the C-terminal Asn but lacks the methyltransferase responsible for ester formation. GE3748 thus is an exception, as it bears a C-terminal carboxylate. Another notable feature of GE3748 is the cytochrome P450-dependent (GetJ) formation of methylhydroxyproline, which derives from oxidation and cyclization of Ile8 and is likely installed prior to macrocyclization.^{261, 361} Interestingly, baringolin appears to natively encode a Pro at this position; however, replacement of the methylhydroxyproline

modification with unmodified Ile (via GetJ deletion) or Ala (via core sequence substitution) in GE37468 has little effect on bioactivity.²⁶¹

Using precursor peptide replacement, many potential GE37468 variants have been assessed for their ability to be produced and their antibiotic activity.³⁶¹ Despite the overall similarity in the core sequences of GE2270 and GE37468, precursor replacement of the GE37468 core peptide with a quadruple amino acid substituted variant (Thr2Cys, Phe5Val, Tyr7Gly, Ile8Phe) to make it more GE2270-like resulted in no product from the GE37468 biosynthetic pathway. To further investigate tolerance and processing rules of the GE37468 biosynthetic machinery, over 100 additional substitutions were assessed for their ability to be produced and their antibiotic activity.³⁶¹ In total, 29 mutants were tolerated by the GE37468 biosynthetic pathway, of which 12 had detectable bioactivity. Thr2 could only be replaced by other cyclizable residues Ser or Cys, which were converted to azoles, and retained bioactivity. Asn3 accepted mostly smaller residues, but none were bioactive as the native Asn donates a transannular hydrogen bond stabilizing the macrocycle conformation and accepts a hydrogen bond to aid binding of EF-Tu. Cys4 cannot be mutated and is crucial for production. Positions 5, 7, and 8 were the most tolerant in terms of biosynthesis, but the only substitutions which remained bioactive, albeit moderately reduced, were Phe5Tyr and Tyr7Phe. Interestingly, Phe/Tyr are both required at positions 5 and 7 for Ile8 oxidation and cyclization and are likely key recognition determinants for the P450 tailoring enzyme.³⁶¹ Lastly, mutation of Ser13 in the tail tended to produce linear or inactive intermediates, demonstrating that alteration of this position can affect downstream dehydration necessary for cycloaddition. Of all these GE37468 analogs, only a Thr2Cys variant had slightly improved activity over the native GE37468 indicating that evolution may have reached an optimal structure for bioactivity.³⁶¹

The promiscuity of thiomuracin and GE2270 biosynthesis have not yet been tested through precursor peptide replacement, but with the reconstitution of the core thiomuracin biosynthetic pathway from *Thermobispora bispora*,⁴⁵ there is an opportunity to investigate tolerance and create new derivatives *in vitro*. Semi-synthetic derivatives have been investigated as an alternative to improve solubility of these compounds,³⁶²⁻³⁶⁷ with GE2270 optimization eventually leading to LFF571,³⁶⁸⁻³⁷³ a compound that has undergone clinical trials for treatment of *C. difficile*.⁶⁸⁻⁷⁰ Baringolin has also been investigated through alteration of its tail.³⁷⁴ Overall, these studies have shown that semi-synthetic tailoring of the tail region of thiopeptides holds promise for enhancing solubility and potency against gram-positive organisms, but the advantage of the *in vivo* engineering is the ability to rapidly examine the importance of different positions within the macrocycle which are more difficult to test synthetically.²²⁴

1.3.2.5 Cyclothiazomycins

The cyclothiazomycins *d* series thiopeptides have also 29-membered macrocycles (Figure 1.22). Their distinguishing feature is an unusual secondary macrocycle that links the C α of a Ser residue with the thiol group of the C-terminal Cys in the form of a thioether.^{266, 375-378} Cyclothiazomycins also have three thiazolines while most other thiopeptides contain oxidized heterocycles. The BGC for cyclothiazomycin A contains the core thiopeptide genes (cyclodehydratase, dehydrogenase, dehydratase, cycloaddition) under the control of a LuxR-type regulatory gene (*cltH*).^{258, 259} Recent genome mining studies have led to the isolation of cyclothiazomycin B and C.²⁶⁶ However, the genes flanking the core biosynthetic proteins are variable, so it is not clear which genes are responsible for forming the thioether-linked secondary macrocycle. Heterologous expression of cyclothiazomycin A required *cltMN*, but these are absent from the genomes of the other producers and their putative annotations (ether degradation and unknown, respectively) do not suggest a clear role in thioether formation. Thus, the enigmatic nature of the thioether modification combined with the nontraditional bioactivities (RNA polymerase and renin inhibition) reported highlight the distinctiveness of this group.³⁷⁷⁻³⁷⁹

1.3.2.6 Lactazoles

The lactazoles are an unprecedented group of thiopeptides first reported in 2014. They are the only thiopeptide family with a 32-membered macrocycle, and they exhibit an unusual amino acid composition in the core region of the precursor peptide (Figure 1.23).²⁶⁵ The lactazole BGC is by far the most compact (9.8 kb) and contains only the core modifying enzymes (Figure 1.23). Also, this BGC fuses the elimination domain of the split LanB dehydratase to the azoline dehydrogenase.²⁶⁵ In stark contrast to all other thiopeptides, the lactazoles have no known antibiotic activity. It is worth mentioning that thiopeptides have historically been discovered through bioassay-guided isolation; however, the lactazoles were identified through genome mining so having an associated bioactivity was not a prerequisite for discovery. The lactazole biosynthetic pathway was found to tolerate various amino acid substitutions in the core, such as W2S and S11C. The latter substitutes a thiazole for an oxazole in the mature structure. Future genome-guided discoveries may reveal more lactazole-like thiopeptides, but currently this is a rather unusual group.

1.3.2.7 Berninamycin-like

The largest known macrocycles for thiopeptides are 35-membered, and members of this family are closely related in the sequence of their core peptides (Figure 1.17). Being first isolated in 1969, berninamycin and sulfomycin are the oldest members of this group,³⁸⁰⁻³⁸² but with the confirmation of the berninamycin BGC through heterologous expression,²⁶² it is the most suitable representative for understanding this group (Figure 1.24). Berninamycin is not heavily modified outside of the core thiopeptide scaffold, but the BGC

does contain one cytochrome P450 oxidase (BerH) and a hypothetical monooxygenase (BerI) that are putatively responsible for hydroxylvaline formation, which appears to occur after macrocyclization. Its C-terminal amide is installed as in other thiopeptides though removal of the C-terminal dehydroalanine (BerJ). Geninthiocin differs from berninamycin by a single methyl group,³⁸³ due to a Thr to Ser substitution, and likely is biosynthesized in the same way. The only other 35-membered thiopeptide with a known BGC is TP-1161,²⁶⁰ which was verified through heterologous expression. Akin to berninamycin, TP-1161 has only one ancillary modification, a C-terminal aminoacetone reminiscent of thiocillin (section 1.3.2.3). As mentioned above, this group is believed to be installed by an Fe-dependent dehydrogenase (TpaJ) followed by decarboxylation and appears to be required for maturation.²⁶⁰ However, the unique feature of these two BGCs is that their cyclodehydratase and dehydrogenase, responsible for azole formation, are split in nontraditional ways. The berninamycin YcaO “G” protein is split into BerG1/2. There is no “F” protein, but there is an E1-like homolog that is fused to part of a dehydrogenase (BerE2) near another dehydrogenase component (BerE1). Given this genetic structure, it is hard to predict how the active azole forming proteins work with each other. Perhaps not coincidentally, the TP-1161 BGC also has an unusual array of proteins with cyclodehydratase homology. TpaGH appear to be a standard cyclodehydratase (partner protein and YcaO, respectively), but there is a second YcaO (TpaC) next to a truncated E1-like protein (TpaD). As in berninamycin, there appears to be a canonical E1-like protein fused to a dehydrogenase (TpaE) and another smaller dehydrogenase domain (TpaF) nearby. These atypical azole forming proteins may influence the processing of the peptide in such a way that leads to the larger 35-member macrocycle being formed. The spacing of azol(in)e heterocycles in the macrocycle region of the berninamycin-like thiopeptides is nearly identical to the thiocillin-like thiopeptides which have 26-member macrocycles. In particular, promothiocin³⁸⁴ and JBIR-83³⁸⁵ are nearly identical to radamycin^{386, 387} with regard to the sequence around the macrocycle, but a dehydroalanine becomes an oxazole in radamycin and the larger macrocycle results.

Although speculative, future discovery of BGCs for other 35-membered thiopeptides may shed light on the rules that govern macrocycle size and downstream tailoring. It is likely that these BGCs would possess additional cytochrome P450 and methyltransferase genes to install the additional hydroxylations and methylations seen in (methyl)sulfomycin,^{382, 388} promoinducin,³⁸⁹ thiotipin,³⁹⁰ and A10255.^{391, 392} Additionally, it is curious that the cyclodehydratases for thioxamycin³⁹³ and thioactin³⁹⁴ apparently skip over Cys4, which becomes an *S*-methylcysteine in the mature compounds. These unexplored areas may reveal new enzyme chemistry and rationalize the unusual features of this thiopeptide family. Another future avenue for research is to further explore structure-activity relationships through precursor peptide replacement. Berninamycin has had only three altered precursor peptides tested *in vivo*. Thr3Ala was tolerated but inactive while Thr4Ala and Thr5Ala did not give any macrocyclic products.²⁶² Further

bioengineering of berninamycin may also reveal the unique characteristics of the enzymes in these pathways.

1.3.3 Bioactivities

Because of the structural similarity of thiopeptides, they also share biological activities. In particular, thiopeptides are renowned for their potent antibiotic activity towards Gram-positive bacteria through inhibition of protein translation.³² Ribosomes are composed of two RNA-protein complexes known as the small (30S) and large (50S) subunit that can assemble into a full (70S) ribosome.^{395, 396} Translation begins when the 30S subunit binds an mRNA and associates with an initiator tRNA (fMet-tRNA^{fMet}) and a free 50S subunit through the help of initiation factors. This complex places the initiator tRNA in the peptidyl site (P site), and its acceptor site (A site) is empty for incoming aminoacyl-tRNAs. Aminoacyl-tRNAs are delivered by GTP-bound, elongation factor Tu (EF-Tu). After GTP hydrolysis and dissociation of EF-Tu·GDP, a new amide bond is formed between the P site and the A site tRNA, which transfers the growing peptide and extends it by one amino acid. Next, GTP-bound elongation factor G (EF-G) binds the ribosome and triggers translocation, promoted by GTP hydrolysis, that shifts all associated tRNAs downstream one site (bringing the base paired mRNA with it).³⁹⁷ This movement opens the A site for the next aminoacyl-tRNA, and the cycle can repeat after release of EF-G·GDP.

Thiopeptides interfere with the protein translation process in two distinct ways, and the size of the main macrocycle largely determines how inhibition will occur. Those with 29-membered macrocycles including GE2270, GE37468, amythiamicin, and their congeners bind to and inhibit elongation factor Tu (EF-Tu) at nM concentrations.^{55, 351, 355, 365, 369, 398-400} Binding sterically blocks tRNA binding to EF-Tu and stabilizes a protein conformation that disrupts its interaction with ribosomes.^{348, 401-403} Multiple analogs have also been co-crystallized with the target further verifying the binding site of these molecules.^{362, 363, 365, 369} Consistent with this data, resistance to thiopeptides is mediated through EF-Tu variants or mutation,^{254, 351, 399, 404, 405} but the overall mechanism and binding site is unique compared to other antibiotics that target EF-Tu.⁴⁰⁵⁻⁴⁰⁷

In contrast, thiopeptides with 26-membered macrocycles (e.g. thiostrepton, siomycin, nosiheptide, berninamycin, and thiocillin/micrococcin) target the 50S subunit of by interacting with the 23S rRNA and ribosomal protein L11.^{65, 345, 408-410} Binding is mediated primarily by interaction with 23S rRNA and is cooperative with L11.⁴¹¹⁻⁴¹⁴ Resistance to thiopeptides of this variety occurs through mutation of either 23S rRNA or L11.^{345, 415-421} Alternatively, methylation of the 23S rRNA also confers resistance.⁴²²⁻⁴³³ Improving initial models,^{434, 435} more recent structural studies based on NMR, X-ray crystallography, and molecular modeling showed that these 26-membered thiopeptides bind to a cleft between the 23S rRNA and L11,^{347, 436-442} but covalent capture data suggest binding may be more on the surface of the rRNA and not between

23S/L11.⁴⁴³ The interface of the 23S rRNA and L11 is called the “GTPase-associated center” and is crucial for ribosome function given its interaction with many translation factors.⁴⁴⁴⁻⁴⁵⁰ Consequently, the binding of thiopeptides with a 26-member macrocycle in this area can affect all phases of translation,⁴⁵¹⁻⁴⁵⁸ but the most studied effects are on EF-Tu and EF-G during elongation.⁴⁵⁸⁻⁴⁷⁴ Inhibition of EF-Tu and tRNA delivery presumably occurs because thiopeptides partially conflict with the ternary (EF-Tu, GTP, and aminoacyl-tRNA) complex’s binding site near the 23S rRNA/L11.^{447, 448, 473} In contrast, the mechanism of action against EF-G varies depending on the thiopeptide. Specifically, thiopeptides with a secondary quinaldic/indolic ring (i.e. thiostrepton/nosiheptide) interact differently with EF-G than those that have just one macrocycle (i.e. micrococcin) because they bind in slightly different orientations at the 23S rRNA/L11 interface.^{347, 440, 474, 475} Thiostrepton disrupts binding of EF-G to the ribosome, thereby inhibiting GTP hydrolysis,^{458, 463-467, 476, 477} and stabilizes a conformation of 23S rRNA/L11 that prevents the structural transitions necessary for translocation.^{410, 438, 442, 478} This means that even with transient EF-G binding, as suggested by some studies of thiostrepton-bound ribosomes,⁴⁷⁹⁻⁴⁸⁵ translation remains arrested. Conversely, micrococcin enhances EF-G GTPase activity without effecting its binding affinity and blocks the translocation process through a similar rigidifying and restriction of 23S rRNA/L11.^{423, 460, 486, 487}

Similarly, thiopeptides with 35-membered macrocycles (i.e. berninamycin-like thiopeptides) are known to target the 50S subunit and likely bind to the 23S/L11 complex.^{381, 488} The precise binding site and orientation are not known, but these larger thiopeptides almost certainly overlap with the primarily binding site of thiopeptides with a 26-member macrocycle based on partial competition with thiostrepton binding to ribosomes.⁴⁸⁸ Additionally, the thiostrepton resistance RNA methyltransferase, which methylates RNA at thiostrepton’s binding site, confers resistance to berninamycin, and the berninamycin producing organisms has the same RNA methylation activity and is also resistance to thiopeptides that bind the 23S/L11 complex. This is in agreement with the BGCs for berninamycin²⁶² and TP-1161²⁶⁰ which contain genes for a thiostrepton-like resistance methyltransferase. However, the binding orientation will likely be slightly different than thiostrepton/micrococcin based its larger macrocycle size and slightly different effect on the ribosome. Only elongation is effected by berninamycin and EF-G GTP hydrolysis is slightly elevated compared to complete inhibition caused by thiostrepton and significant enhancement by micrococcin.⁴⁸⁸ Future structural studies may more precisely identify the binding site of thiopeptides with 35-membered macrocycles.

Beyond protein translation inhibitors, most thiopeptides are known to activate gene transcription independent of their antibiotic activity.^{386, 489} The primary mechanism for this is activation of the *tipA* gene (thiostrepton-induced protein A), which encodes TipAL and TipAS.^{32, 490, 491} TipAL is a transcriptional activator that upregulates a defined set of genes, including itself, upon thiopeptide binding to its C-terminal domain.⁴⁹² TipAS, which is truncated and only possess the C-terminal binding domain, exists primarily to

sequester thiopeptides and represents a resistance mechanism found in some non-thiopeptide producers.^{493, 494} Thiopeptides of various sizes and topologies are recognized by a common four-ring motif.³⁴⁶ Upon binding, dehydroamino acids often present in thiopeptides can undergo a Michael-like addition with Cys residues of TipA present within the flexible binding pocket.^{346, 495-497} However, thiopeptides with a 29-membered macrocycles lack the conserved four-ring motif. Fittingly, they also do not induce or bind to TipA.³⁴⁶ Additionally, because TipA is highly responsive to thiopeptides, its induction has been applied in thiopeptide discovery.^{383, 384, 390, 394}

Another known activity of thiopeptides targets the malaria parasite, *Plasmodium falciparum*, which have an essential, plastid-like organelle (apicoplast) that contains a 70S ribosome as in bacteria.⁴⁹⁸ Thus, unlike other eukaryotes, *P. falciparum* is sensitive to many thiopeptides such as thiostrepton, micrococcin, and GE2270, and nocathiacin.⁴⁹⁹⁻⁵⁰⁸ Other protein synthesis inhibitors also have demonstrated a similar anti-malarial activity.⁵⁰⁹⁻⁵¹¹ However, it should be noted that protease inhibitors are known to be active against the *Plasmodium* parasite because the proteasome is crucial for development.⁵¹²⁻⁵¹⁴ Given that thiostrepton has been reported to bear proteasome inhibition activity, its antimalarial activity is likely the result of multiple modes of action.⁵¹⁵⁻⁵¹⁷

Another intensively studied activity for thiopeptides is the induction of apoptosis in human cancer cells.⁵¹⁸⁻⁵²¹ This occurs by down regulating the expression of an oncogene FOXM1,⁵²¹⁻⁵²³ which is a transcription factor that activates cell proliferation and cell cycle progression. As an oncogene, FOXM1 is overexpressed in more than 20 types of human cancers and is recognized as a potential anticancer target.^{524, 525} The anticancer activities of certain thiopeptides is also due in part to the inhibition of the proteasome, but this secondary activity appears limited to thiopeptides with a quinaldic acid side ring (i.e. thiostrepton-like).^{515, 517, 526} However, there is also evidence that berninamycin downregulates FOXM1, possibly without proteasome inhibition.⁵²⁷ The proteasome is composed of numerous proteins but is split into two main parts, the core particle (CP) which contains the proteolytic active sites and the regulatory particle (RP) which controls access to the CP and unfolds substrates for degradation.⁵²⁸ Thiostrepton covalently binds the RP via Cys adducts to its dehydroamino acids, which dysregulates proteolysis.⁵²⁹ In an example of overlapping mode of action, proteosomal inhibition by thiostrepton stabilizes a negative regulator of FOXM1 (Hsp70).⁵³⁰ It appears that thiostrepton can also directly bind FOXM1 and blocks its ability to activate gene transcription.⁵³¹ These modes of action are likely most responsible for thiostrepton's anticancer activity, but it remains possible that other means of contributing to apoptosis, such as inhibiting mitochondria protein synthesis, also play a role.^{532, 533}

Other functions for thiopeptides have been observed, however, they are less intensively researched and typically only apply to individual compounds. For instance, siomycin can act as an immunosuppressant and showed efficacy in a murine arthritis model.⁵³⁴ On the other hand, the cyclothiazomycins with their

atypical side ring and relative lack of heterocycles within its primary macrocycle have more unusual activities. They are not particularly effective against gram-positive organisms compared to other thiopeptides.²⁶⁶ Instead, the cyclothiazomycins have other reported activities such as inhibiting RNA polymerase³⁷⁸ and the growth of fungi by binding to chitin in the cell wall,³⁷⁹ although the antifungal activity seems to be narrow spectrum.²⁶⁶ Cyclothiazomycin inhibits renin,³⁷⁷ a protease involved in the renin-angiotensin system that has links to hypertension, cancer, and other diseases.⁵³⁵⁻⁵³⁸ Geninthiocin also has reportedly weak antifungal activity.⁵³⁹ A10255 activates plasminogen, which is involved in fibrinolysis and other cell surface interactions.^{540, 541} Overall, the unique and constrained structure of thiopeptides enables them to effect multiple biological processes, but their poor solubility and large size limits their ability to be used directly as therapeutics. Nonetheless, thiopeptides have recently emerged as drugs leads. More soluble thiopeptide derivatives have been semi-synthetically prepared (e.g. LFF571, Figure 1.4) and are showing promise in human clinical trials for *C. difficile* infections.⁶⁸⁻⁷⁰ Alternatively, the re-engineering of thiopeptides with more hydrophilic sequences that retain bioactivity is also a promising avenue.^{229, 542} Accordingly, there is still much value in investigating new thiopeptides and exploring their potential functions.

1.4 Cyanobactins

The cyanobactin family of RiPPs, produced by both free living and symbiotic cyanobacteria, comprises a family of highly modified peptides of ribosomal origin. As with all RiPPs, the characteristic of the family is that the final natural product derives from a core sequence that is embedded within a longer precursor peptide. Until recently, cyanobactins were synonymous with macrocyclic peptides coming from cyanobacteria⁵⁴³, however linear cyanobactins have now been identified.⁵⁴⁴ Cyanobactins vary in length from as short as a tripeptide for the linear cyanobactins up to around twenty amino acids in some macrocycles.^{544, 545} To date, the vast majority of isolated cyanobactins (and those bioinformatically predicted) are macrocyclic and between 6 and 10 residues in length.⁵⁴⁶ While some cyanobactins contain a high density of posttranslational modifications, the number and types of modifications are determined by the enzymes in the pathway, but larger cyanobactins (more than 15 amino acids) characterized to date appear to have less extensively modified amino acids.^{543, 546, 547} The list of modifications identified so far is extensive and has been reviewed in depth recently.⁵⁴⁶ Here we concentrate on cyclodehydration (heterocyclization) of specific Cys, Ser, and Thr residues that give rise to thiazoline and oxazoline heterocycles, which in turn can be, but are not always, oxidized to thiazoles and more rarely oxazoles.^{548,}

⁵⁴⁹

As a whole, YcaO proteins are quite prevalent among cyanobactin gene clusters; only three families of compounds lack a *ycaO* gene and correspondingly have no azol(in)es: the anacyclamides,

prenylgaramides, and piricyclamides.⁵⁴⁶ This would suggest that cyclodehydration confers an advantageous change in the conformation and chemistry of the peptide backbone of the natural product. Generally speaking, Ser/Thr cyclodehydrations are less common in cyanobactins, but there are notable examples which do contain oxazol(in)e heterocycles such as patellamide⁴⁸ of the *pat* pathway, ulithiacyclamide,⁵⁵⁰ and microcyclamide.⁵⁴⁹ For products of the *tru* pathway, (e.g. trunkamides and patellins) although a YcaO domain is present, it does not readily form oxazolines and Ser and Thr remain uncyclized. Their unmodified hydroxyl groups can instead undergo prenylation.^{49, 551} Prenylation of other amino acids is also known including C-prenylation of tryptophan⁵⁵² and C- and O-prenylation of Tyr.⁵⁵³ In the case of C-prenylated Tyr, the prenyl group is first attached to the phenolic hydroxyl before undergoing a Claisen-type rearrangement (Figure 1.25).⁵⁵³ An additional posttranslational modification observed in several macrocycles, including patellamide A, is epimerization of the amino acids adjacent (N-terminal) to the thiazoline (or thiazole) to D-stereocenters.^{549, 554-556} Rarer modifications include Cys to disulfide bridges found for example in ulithiacyclamide, as well as N- and O-methylation.^{544, 549, 557} The disulfide in ulithiacyclamide hints at the ability to exercise regioselective control of which Cys are modified within the core peptide. Indeed, regioselectivity during cyclodehydration has been experimentally observed as well and is discussed later (section 1.6.3.2).⁵⁴⁸ Variation in amino acid identity, differing combinations of amino acid modifications, and differing lengths of products underscores the vast chemical diversity of the cyanobactins.

In addition to biochemical interest in their biosynthesis, cyanobactins have attracted considerable attention as potential starting points for development of new therapeutics as they possess a diverse range of bioactivities.^{543, 546, 547} Patellamides A, B, C and D display $IC_{50} = 2-4 \mu\text{g mL}^{-1}$ against L1210 murine leukemia cells.⁵⁵⁸ Patellamides B, C and D have each demonstrated activity in reducing multidrug resistance in CCRF-CEM and CEM/VLB100 leukemic cell lines through competitive inhibition of P-glycoprotein (Pg-p), the efflux pump responsible for removing anticancer compounds from tumour cells.^{559, 560} Trunkamide A1, a seven residue cyclic cyanobactin, possesses cytotoxic activity ($IC_{50} = 0.5 \mu\text{g mL}^{-1}$) against HT-29 human colon carcinoma and $IC_{50} = 1 \mu\text{g mL}^{-1}$ against MEL-28 human melanoma cell lines.⁵⁶¹ Ulithiacyclamide is highly cytotoxic with $IC_{50} = 350 \text{ ng mL}^{-1}$ and $IC_{50} = 35 \text{ ng mL}^{-1}$ against L210 murine leukemia and KB cell lines respectively.^{550, 558} Antimalarial activity, but no specific cellular target, has been reported for the aerucyclamides B, C and D displaying $IC_{50} = 1-6 \mu\text{M}$.⁵⁶² The compounds were selective for *Plasmodium falciparum* over L6 rat myoblast cells.⁵⁶² Kawaguchi peptide B presents antibacterial activity with a minimum inhibitory concentration (MIC) = $1 \mu\text{g mL}^{-1}$ against *Staphylococcus aureus*.⁵⁶³

The diversity in biological activity thus mirrors the diversity of the compounds. It is this diversity of cyanobactins that make them attractive starting compounds for fine-tuning to obtain useful clinical candidates. Most natural compounds currently have limitations that preclude their direct use in the clinic.

Nature may have propagated cyanobactins, and other RiPPs, precisely because of their malleable characteristics: simple nucleotide changes can yield new bioactivities. However, there is a large barrier to their use in medicinal chemistry in part because their *de novo* synthesis is problematic. Only a few compounds have been completely synthesized.^{555, 556, 564-570} In most cases, the routes are lengthy, complex and case specific, thus limiting the diversity achievable and so are incompatible with the drug development process. A strike against cyanobactins as drug candidates is that they violate the so-called rule of five for bioavailability⁵⁷¹ derived by Lipinski from an analysis of orally active medicines. Indeed, all cyanobactins that contain more than five amino acids exceed the rule for molecular mass of 500 Da or less. Secondly, their peptidic nature all but ensures they have more than five hydrogen bond donors and ten hydrogen bond acceptors. Indeed, it is precisely because of these considerations that linear peptides are generally regarded as “non-starters” as bioavailable drugs; at best they serve as proof of concept. However, the demonstrated cell permeability and in some cases oral toxicity of macrocyclic peptides indicates a blanket assertion of unsuitability may be wrong.^{572, 573} Cyanobactins, and natural products more generally, are accepted to be exceptions to simple rule of five considerations. The term “beyond rule of five” has been coined to describe such molecules (this field has recently been reviewed in Doak et al.⁵⁷⁴) Given that many peptidic natural products used clinically are macrocyclic, they can adopt conformations that mask polar groups, of which cyclosporine is the most well-known example. Cyclosporine adopts a conformation where internal hydrogen bonds result in a conformation displaying a hydrophobic surface while masking polar regions to facilitate membrane transit.^{575, 576} In the specific case of cyanobactins, macrocyclization also removes the zwitterionic N- and C-terminal charges. Heterocyclization further increases the hydrophobicity of the natural product.^{577, 578} Additional modifications such as prenylation or methylation also decreases polarity as well as the number of hydrogen bond donors.^{572, 578} Although rules are emerging for predicting the bioavailability of macrocycles,^{574, 579} it appears that the conformation of the macrocycle is an important consideration. For chemically complex macrocycles with multiple modifications and epimerizations, predicting the conformation remains challenging. In addition to advantages in cell permeability and decreased hydrogen bonding, the installation of azol(in)e heterocycles and macrocyclization also introduce conformational rigidity when compared to linear precursors. As a consequence of the reduced entropic penalty, rigid molecules that fit a binding site are always more potent than flexible molecules which have to freeze out a specific conformation (or set of conformations) to bind.^{578, 580} This is borne out by many examples where macrocyclization results in more potent compound compared to the linear analogue.^{578, 580}

1.4.1 Cyanobactin biosynthesis

Since chemical synthesis is impractical for generating large libraries of these desirable compounds, attention has turned to protein engineering and biotechnology to produce these molecules at the scale, speed,

and diversity compatible with the drug discovery process. This has spurred investigation of cyanobactin pathways and the discovery of many new BGCs. Overall, 12 families have been identified, and 9 contain azol(in)e heterocycles.⁵⁴⁶ Similar to other RiPPs, the genes associated with cyanobactin biosynthesis are almost always clustered.^{26, 581} As mentioned previously, the primary features encoded within cyanobactin gene clusters are azol(in)e formation, N-to-C macrocyclization, and prenylation although other auxiliary modifications can also be present. Cyanobactin gene clusters mix and match the enzymes responsible for these modifications as well as the core sequences of their peptides to generate diverse products.^{49, 50} The known products from these cyanobactin pathways have recently been reviewed⁵⁴⁶ so in the remainder of this section we provide an overview of the key biosynthetic enzymes and BGC architectures of these pathways.

In cyanobactin BGCs, the nomenclature follows the historic precedent set from the first cluster to be fully sequenced, the patellamide or *pat* cluster.⁴⁸ Accordingly, the biosynthetic proteins are named A through to G in the order of their occurrence in the operon (Figure 1.26). In contrast to other RiPP BGCs, such as that for microcin B17 where the gene encoding the precursor peptide is located at the start of the gene cluster and thus designated “A” (e.g. *mcbA*),^{582, 583} the cyanobactin precursor peptide gene is located in the middle of the cluster and is designated “E” (e.g. *patE*, *truE*)^{48, 584} even where it is not the fifth gene. Other than the precursor peptide, the minimal set of proteins encoded by cyanobactin gene clusters include: an A protein (N-terminal protease), a G protein (macrocyclase / C-terminal protease), and two hypothetical proteins, B and C, which are non-essential for patellamide biosynthesis *in vivo*.⁴⁸ Other genes that often occur are a D protein (cyclodehydratase), oxidase (often fused to the G protein), and an F protein (prenyltransferase). Since not every cyanobactin gene cluster contains a YcaO protein, the class is defined by the presence of a precursor peptide and PatG-like protease which combine to form macrocyclic peptides.

Much has been learned about the main cyanobactin processing enzymes through biochemical characterization and X-ray crystallography.⁵⁸⁵ The A protein is composed of two domains, an N-terminal protease domain which cleaves the precursor peptide immediately N-terminal to the core peptide, and a C-terminal domain of unknown function (DUF).^{586, 587} The G protein is defined by a N-terminal macrocyclase domain and a C-terminal DUF homologous to that found in the A protein.⁵⁸⁸ The macrocyclase domain is also homologous to the protease domain of the A protein, and both belong to the subtilisin protein family.⁵⁸⁷ Compared to subtilisin, however, there is a characteristic insertion that is found in all cyanobactin macrocyclase domains critical for preventing the enzyme from hydrolyzing the acyl enzyme intermediate and producing linear product.⁵⁸⁹ As pointed out above, the principal characteristic feature of cyanobactins is their macrocyclic nature, and there are certain instances where the primary posttranslational modification of the peptide is macrocyclization (due to a lack of additional biosynthetic enzymes) such as in the larger

macrocyclic anacyclamide⁵⁴⁵ and piricyclamide⁵⁹⁰ families (Figure 1.26.) In the compound piricyclamide 7005E4, a disulfide bridge is also present,⁵⁹⁰ but this is likely a spontaneous modification.

Cyanobactin cyclodehydratases catalyze the heterocyclization of Cys and in some cases, Ser and Thr to form azoline heterocycles.⁵⁴⁸ The protein can be broken into three domains. The first domain is homologous to PqqD and serves as the RRE^{27, 591} while the second domain is homologous to the E1-like superfamily. Together, the first two domains are similar to the MccB protein,⁵⁹² an adenylating enzyme in microcin C7 biosynthesis which catalyzes the formation of a five-membered heterocycle.⁵⁹¹ Although similar to MccB, its ATP-binding site is not conserved in these first two domains.¹⁸ Instead, the YcaO (third domain) is responsible for ATP-utilization. In some previously described cyclodehydratase-containing BGCs, the YcaO domain is found as a discrete protein with the LAP C protein containing the first two domains.^{17, 18, 27} The canonical representatives of cyanobactin cyclodehydratases are PatD and TruD, from the patellamide (*pat*) and trunkamide (*tru*) pathways, respectively. These enzymes share 93% sequence similarity so it is surprising that PatD is able to cyclodehydrate Cys, Ser, and Thr whereas TruD only acts on Cys and skips over Ser and Thr residues.⁵⁴⁸ Thus, D proteins can be divided into two types: Cys-specific cyclodehydratases (e.g. TruD, LynD)^{592, 593} and broad spectrum cyclodehydratases (e.g. PatD, McaD, which process Cys, Ser, and Thr).^{549, 585} There is currently no sequenced-based rationale for discerning their reactivities. Given the leader-dependent nature of cyanobactin cyclodehydratases, it must act before macrocyclization and is presumably the first enzyme to act on the peptide.^{594, 595}

The cyanobactin dehydrogenase, which oxidizes YcaO-formed azolines into azoles, is flavin-dependent. This enzyme can be a ‘freestanding’ protein, but frequently, the dehydrogenase domain is fused to the G protein. In the patellamide BGC, PatG is a tridomain protein with an N-terminal dehydrogenase, macrocyclase middle domain, and C-terminal DUF. Sequence analysis of the dehydrogenase reveals a typical FMN-binding site signature. Sequence analysis also indicates that N-terminal to the flavin-binding site is at least one peptide clamp domain or RRE-like domain of the sort found in PatD.²⁷ The dehydrogenase has been experimentally verified to be responsible for azole formation,⁵⁴⁶ and is homologous to the “B” protein in the LAP family of RiPPs.^{27, 72} However, unlike B proteins which are presumed to act quickly after cyclodehydration, the cyanobactin dehydrogenase is thought to act natively after macrocyclization.⁵⁹⁶ Lastly, cyanobactin pathways also have a unique class of prenyltransferase (F protein).⁵⁹⁷ Due to independent naming conventions, the cyanobactin F protein (prenyltransferase) should not be confused with the HCA/thiopeptide F protein (Ocin-ThiF-like protein). In the *pat* pathway, PatF has been shown to be inactive as a prenyl transferase due to mutations at the active site⁵⁹⁸ consistent with the lack of any prenylated patellamides. No other activity has been ascribed to this protein. The prenyltransferase is functional in other gene clusters including trunkamide of the *tru* pathway⁵⁹⁹ (TruF1, a Ser/Thr prenyltransferase),⁵⁸⁹ aestuaramide⁵⁹⁷ (LynF, a Tyr prenyltransferase),^{597, 599} piricyclamide⁵⁹⁰ and

kawaguchipectin.⁵⁵² The structure of the prenyltransferase from the prenylagaramide biosynthetic pathway of *Oscillatoria agardhii* has been determined in complex to linear and cyclic substrate peptides.⁶⁰⁰ The enzyme has a very broad substrate specificity requiring only an N-terminal Tyr followed by an aliphatic or aromatic residue. The structures show that the enzyme only recognizes this two residue motif. Despite this small motif, prenylation has also been shown to occur only after macrocyclization for LynF,⁵⁹⁷ but some must have relaxed substrate specificity given that linear cyanobactins are prenylated and PagF accept linear substrates.^{544, 600}

Depending on the combination and order of the proteins encoded by different operons, cyanobactin biosynthesis has been classified into four genotypes.⁶⁰¹ Genotype I, exemplified by the patellamide pathway, which also includes the trunkamide, teneucyclicamide and microcyclamide pathways, amongst others, is the most common. Genotype I comprises a single operon starting with “A” and ending with “G”⁶⁰¹ (Figure 1.26). Genes encoding both the cyanobactin cyclodehydratase and prenyltransferase are always present in genotype I. Indeed, azol(in)e heterocycles are universal in the final products; however, due to the previously mentioned inactivation of PatF, prenylation is not universal to genotype I cyanobactins. Further, genotype I pathways that encode dehydrogenases nearly invariably do so at the N-terminus of the G protein. When present, thiazole groups are always present in the final product. Notable examples where the dehydrogenase is omitted come from the *tru* pathway; fittingly, these products—the trunkamides and the patellins—contain only thiazolines (Figure 1.26). In some clusters, notably the patellamide pathway, the dehydrogenase is chemoselective for thiazolines since the final product contains thiazoles but not oxazoles. Other clusters that have similar enzymatic reactivity trends include the lissoclinamides and ulithiacyclamides. Conversely, the microcyclamide and teneucyclamide products contain thiazoles and oxazoles (Figure 1.26) suggesting the dehydrogenase has a wider activity.

Genotype II is epitomized by the prenylagaramide and anacyclamide pathways.⁶⁰¹ Gene organization within this genotype is significantly different from that of genotype I with the B and C proteins located at the start of the operon, as well as the addition of multiple hypothetical open reading frames of unknown function. This is the only genotype that uniformly lacks a cyclodehydratase gene and whose final products consequently lack a Cys-, Ser-, and Thr-derived heterocycles (Figure 1.26).^{545, 601} Interestingly, the precursor peptides all contain Pro at the C-terminal position of the core peptide. The macrocyclase domain of PatG was shown to require either a thiazoline or Pro at the C-terminal position of the core peptide in order to carry out the macrocyclization reaction.⁶⁰² The structure of the PatG macrocyclase domain reveals that these backbone conformation-restricting moieties are required for the substrate to bind effectively to the active site of the enzyme.⁵⁸⁹ Since the cyanobactin products of genotype II lack azolines, the biosynthetic clusters unsurprisingly also lack dehydrogenases. Genes encoding the prenyltransferase are always found in genotype II; however, the products are not always prenylated. For example, Leikoski and

co-workers identified the anacyclamides, which besides N-to-C macrocyclization otherwise contain entirely unmodified amino acids, as well as the prenylagramides which contain prenylated amino acids.⁵⁴⁵ Furthermore, the products of genotype II are typically larger (>11 amino acids) than those of genotype I.⁵⁴⁶ The piricyclamide pathway contains the same complement of genes as those found in genotype II, although the organization of genes within the operon is distinct (Figure 1.26).⁵⁹⁰ A cyanobactin gene cluster *osc* was identified from *Oscillatoria sp.* PC6506 and is considered a hybrid of genotype I and II as it contains all of the biosynthetic enzymes necessary to produceazole heterocycles (characteristic of genotype I) but the organization more closely reflects genotype II (Figure 1.26).⁶⁰¹ To date, no products of the *osc* pathway have been identified.

Genotype III contains only the trichamide biosynthetic pathway, and is the only known example of a cyanobactin gene cluster lacking a gene encoding the F protein.⁶⁰¹ Like genotype I, cyclodehydratase and dehydrogenase genes are present, giving rise to thiazoles in the final product (Figure 1.26). In contrast to genotype I, the dehydrogenase is not fused into the G protein, but rather is encoded as a discrete gene. Genotype IV so far contains only the cyanothecamide (*thc*) pathway, which displays a unique gene organization.^{586, 601} Despite the presence of up to nine precursor peptides in a single operon, only three products have so far been identified.⁵⁸⁶ This pathway has a number of other features that are atypical including non-consensus protease signatures in some of its precursor peptides although it can still act on more canonical sequences as well.^{586, 595} Differences in the enzyme active sites between PatA and ThcA may explain the relaxed specificity of ThcA, but this has not been extensively investigated.⁵⁸⁶

Given the vast number of cyanobactin pathways discovered to date it seems likely that more examples of each genotype, and possibly other genotypes, exist. Perhaps the most unusual example to date are the linear cyanobactins. Although N-to-C macrocyclization is the class-defining feature of cyanobactins, viridisamide and the aeruginosamides are linear, seemingly because their N- and/or C-termini are modified and prevent macrocyclization by the PatG macrocyclase domain which is still present in the BGC (Figure 1.27).^{544, 546} These and other emerging pathways are expected give rise to a plethora of diverse compounds suitable for pharmaceutical development, as well as a range of interesting enzymes suitable for biotechnological application.

1.5 Miscellaneous azol(in)e-containing RiPPs

Whereas most known azol(in)e-containing RiPPs clearly belong to one of the established subclasses previously mentioned, some other RiPPs with thiazol(in)e or oxazol(in)e heterocycles contain additional structural features and biosynthetic components that are not fully consistent with these taxonomies. For example, the nitrile hydratase- and Nif11-derived precursor peptides are likely similar to the LAPs, but their precursor peptide sequences connect them to multiple RiPP subclasses. On the other hand, YM-

216391 is clearly similar to the cyanobactins, but because its BGC lacks the class-defining proteases that form its macrocycle and is quite unlike the cyanobactin gene clusters, it does not meet the subclass criteria. Accordingly, these two are currently part of an unassigned group.

1.5.1 Nitrile hydratase- and Nif11-derived precursor peptides

Because of their short length, RiPP precursor peptides are frequently unannotated (not predicted as a gene) in publicly available genomic databases.²⁶ Complicating matters further for gene prediction is that many azol(in)e-containing RiPPs display a large degree of sequence variability so it is difficult to identify new precursors because they do not share sequence similarity to previously identified members.⁶⁰³ However, bioinformatic analysis of presumed azol(in)e-containing RiPP BGCs has revealed a new family with two widely recurring motifs in the putative precursor peptides.⁶⁰⁴ In each case, the leader peptides appeared to derive from other proteins, which presented a novel paradigm for RiPP biosynthesis. The first variety harbored leader peptides closely related to the alpha subunit of nitrile hydratase (NHase).^{605, 606} The nitrile hydratase-like leader peptide (NHLP) sequences are roughly half the size of actual NHases because a ~60 amino acid region including the NHase active site has been lost. Accordingly, these NHLP sequences within putative precursor peptides are not expected to be catalytic but simply serve as part of the peptide substrate for post-translational modification. The second variety had leader peptides strongly resembling the Nif11 protein, a nitrogen-fixing protein from cyanobacteria.^{607, 608} Nif11 remains of unknown function.

Because of the typically longer length of their leader peptides (~75 amino acids) and their conserved sequence motifs, NHLP- and Nif11-derived precursors are less likely to be skipped by automated gene prediction algorithms. For those associated with a cyclodehydratase, the core regions tend to be enriched with heterocyclizable residues (Cys, Ser, and Thr). They also possess a classic Gly-Gly leader peptide cleave site that is useful for predicting their products (Figure 1.28).^{143, 604} In addition to these well-defined precursor peptides, there are usually fused cyclodehydratases and nearby transporters or other putative tailoring enzymes. Phylogenetically, BGCs for this class tend to be found in cyanobacteria, proteobacteria (*Burkholderia* order), and a few anaerobic bacteria, which are not typically known for production of secondary metabolites.^{163, 604} Despite this knowledge, no product from one of these clusters has yet been isolated so this putative azol(in)e-containing RiPP family remains hypothetical.

On the other hand, NHLP/Nif11 derived natural products have been isolated for other RiPP classes including lanthipeptides⁶⁰⁹ and proteusins (i.e. polytheonamide).⁶¹⁰ Additional potential natural products have also been identified through genome mining for these classes.^{163, 604} One notable lanthipeptide example, the prochlorisin BGC from *Prochlorococcus* MIT9313, contains nearly 30 different Nif11 precursor peptides,^{609, 611} Interestingly, most of the Nif11 leader sequence is not required for prochlorisin

biosynthesis.⁶¹² However, it has yet to be investigated whether this is a general trend applicable to cyclodehydratase-containing NHLP/Nif11 pathways as well.

1.5.2. YM-216391

During screening for novel anticancer compounds, *Streptomyces nobilis* was found to produce a compound, YM-216391, with nanomolar activity towards several cell lines.⁶¹³ NMR structure determination revealed YM-216391 to be a cyclic, pentazole-containing peptide, similar to the telomerase inhibitor telomestatin (Figure 1.29).⁶¹⁴ All but one amino acid (Val) of the core peptide have been modified; Marfey's analysis confirmed the presence of D-*allo*-Ile. The complete stereochemistry has also been confirmed through total synthesis.^{615,616} Although the mode of action of YM-216391 is unknown, given its bioactivity and structural similarity to telomestatin, it may mediate cell death through inhibition of telomerase as well.⁶¹⁷

Given the structure of YM-216391, it was difficult to predict its biosynthetic origin. However, in 2012, the BGC was discovered by mining for a peptide sequence (FIVGSSSC) that could give rise to the mature product (Figure 1.29). The BGC was confirmed by heterologous expression in *S. lividans*.⁶¹⁸ The YM-216391 core sequence is flanked by motifs rich in Asp/Glu, reminiscent of cyanobactin precursors which also are macrocyclized peptides (section 1.4). However, YM-216391 is not considered a cyanobactin because the protease/macrocyclase characteristic of the cyanobactins is not present in the BGC, and thus it appears to employ a different strategy for macrocyclization. With the function of very few biosynthetic genes being confidently predicted, the precise mechanism for macrocyclization remains unknown. Indeed, only a few proteins have passable similarity to characterized proteins. The YcaO protein (YmD) is proposed to install the 4 canonicalazole heterocycles with the aid of an unusual fusion of the E1-like protein to the N-terminus of the azoline dehydrogenase (YmBC). However, the E1-like protein appears to be truncated and is missing its N-terminal RRE, and an RRE seems absent from the gene cluster. The biosynthesis of the conjugated benzyloxazole moiety remains unknown. Putatively, β -hydroxylation of Phe by a cytochrome P450 enzyme (YmE) sets the stage by installing a nucleophilic moiety for heterocyclization. After N-to-C macrocyclization (YmF? and YmI?), the hydroxyl could be heterocyclized and dehydrogenated by YmB1/C1, but these proteins do not resemble the YcaO or any other characterized proteins.⁶¹⁸ YmG is homologous to lysine 2,3-aminomutase but lacks the SAM binding motif (CxxxCxxC) so it may cooperate with YmH to epimerize the side chain of L-Ile to D-*allo*-Ile. Heterologous expression studies have indicated that *ymR1* and *ymR2* are likely transcriptional activators. Conversely, YmR3 is a repressor that upon deletion increases YM-216391 production by 20-fold.⁶¹⁸ Lastly, YmR4 is a transmembrane efflux protein that may provide resistance. The overall lack of similarity between the YM-216391 and other cyclodehydratase-containing BGCs suggests that it belongs to a distinctive group of cyclic azol(in)e-containing RiPPs. Interestingly, marthiapeptide is structurally similar to YM-216391 yet its BGC remains

unknown, despite the availability of the whole genome sequence of the producing organism, *Marinactinospora thermotolerans*.⁶¹⁹

1.6 Characterization of RiPP cyclodehydratases

Although the YcaO superfamily is associated with multiple chemical transformations, the best characterized are azoline-forming YcaOs which catalyze cyclodehydration of Cys, Ser, or Thr of a ribosomally synthesized peptide. As exemplified by the many azol(in)e-containing RiPPs, the ability to form these heterocycles through enzymatic modification is crucial for the creation of biologically active compounds. Consequently, there has been much interest in determining how azoline-forming YcaOs heterocyclize Cys, Thr, and Thr residues. As discussed previously, this began with work on McbBCD from MccB17 biosynthesis. The second group to be characterized and which also helped to lay a foundational understanding of cyclodehydration were the cyanobactin cyclodehydratases TruD and PatD. After these, studies of cyclodehydratases rapidly expanded beginning in the late 2000s with the biochemical characterization of cyclodehydratases involved in cytolysin (e.g. SLS pathway) and thiopeptide biosynthesis. Ultimately, the function of the YcaO would be conclusively determined through investigation of BalhD (hakacin pathway) and more recent work continues to advance our understanding of YcaO-mediated cyclodehydration.

From these studies, there are currently four varieties of cyclodehydratases (Figure 1.30). Three of the four groups require a member of the E1-like superfamily as a partner protein for activity. This partner can be encoded as a discrete polypeptide or fused to the YcaO. In addition to an E1-like domain, certain fused cyclodehydratases are notably shorter by ~150 amino acid and only act when present with an “Ocin-ThiF-like” protein. Members of this variety are referred to as F-protein dependent cyclodehydratases. Lastly, in rare cases, a standalone YcaO seems to independently form azoline heterocycles. Little is known about the standalone azoline-forming YcaOs as none have been reconstituted *in vitro*; consequently, this section is focused on insights into how the canonical cyclodehydratases catalyze azoline formation and recognize their substrates. Importantly, despite the differences in domain architecture, these enzymes share common features that apply to all types.

1.6.1 Functional dissection of the cyclodehydratase

Although the function of the YcaO protein and its partner proteins are now well established, the role of individual proteins was ambiguous for many years. Historically, the primary limitations for functionally dissecting the role of each protein in the cyclodehydratase were a lack of robust cyclodehydratases, in part due to the low amount of genomic information available and thus few known pathways, and the inability to characterize YcaO independent of its associated partner proteins. The first-studied cyclodehydratase from

the MccB17 BGC (section 1.2.1) was inactive unless the full trimeric synthetase was formed (i.e. the B, C, and D proteins).^{33, 37, 47, 620} Although the identity of the dehydrogenase was clear due to its yellow appearance from binding an FMN cofactor,⁶²⁰ the individual roles of the E1-like partner protein (McbB) and the YcaO protein (McbD), which composed the azoline-forming cyclodehydratase, were not readily distinguishable.³⁷

The first functional insights into the proteins that compose the cyclodehydratase came from examination of McbB (E1-like) and McbD (YcaO). The protein sequence of McbD revealed supposedly weak similarity to some ATP-binding motifs involved in phosphate or Mg²⁺ binding.⁶²¹ However, McbD did not hydrolyze ATP by itself, and mutation of the ATP-binding residues affected both complex assembly and catalysis, which obscured interpretation of their role.^{620, 621} In another experiment, the McbA leader peptide was synthesized with a biotin purification handle and a photoactivable benzophenone group in order to crosslink and affinity purify the protein responsible for leader peptide binding. Upon exposure to UV light, McbD was identified, and this result suggested that McbD recognized the leader peptide, although crosslinking only occurred when the full McbBCD synthetase was present.⁶²⁰ On the other hand, the McbB protein was found to contain Zn²⁺ and thus was proposed to catalyze cyclodehydration by acting as a Lewis acid;⁶²⁰ however, later work indicated that the Zn, in the form of a tetrathiolate cluster, was only structural.⁶²² Overall, these early studies suggested that McbB (E1-like protein) was the cyclodehydratase and the site of azoline formation while McbD (YcaO) regulated activity through leader peptide-binding and ATP hydrolysis.⁶²⁰ However, McbB was later suggested to be more likely responsible for ATP-binding based on sequence homology to members of the E1-like superfamily, which are also zinc tetrathiolate-containing and ATP-utilizing enzymes.^{48, 188, 594} The YcaO protein then became known as the “docking” protein as it seemed less involved in catalysis, but these previously assigned functions for the azoline-forming proteins are now known to be incorrect.

1.6.1.1 YcaO protein

After the pioneering studies on the microcin B17 azole-forming complex, other cyclodehydratases were investigated *in vitro* to further probe the role of the individual proteins. Work with the cyanobactin cyclodehydratases unequivocally showed that the dehydrogenase was not required for cyclodehydration,^{548, 594} unlike in the previously studied examples (e.g. McbBCD and SagBCD).^{37, 51} However, the fact that the cyanobactin cyclodehydratases are a fusion of the E1-like and YcaO proteins meant their individual roles could not be separated. Upon investigating BalhD (YcaO from the LAP BGC in *B. Al Hakam*), the protein was shown to consume ATP and form azoline heterocycles independent of any other protein.¹⁷ The activity was modest, but this represented the first time a precise function could be ascribed to a YcaO protein.

Later, follow up studies focused on determining the active site and ATP-binding residues of the YcaO protein since it was devoid of established ATP-binding motifs. Working towards this goal, the

structure of the fused cyclodehydratase TruD provided the first insight into the structure of the YcaO domain and its attached E1-like partner protein.⁵⁹² However, no substrates were co-crystallized with the cyclodehydratase, so there was no evidence yet for how the YcaO domain might interact with ATP. The study drew attention to the fact that the canonical ATP-binding site of the E1-like protein was missing multiple key residues. TruD was also shown to produce adenosine monophosphate (AMP) and pyrophosphate (PPi) whereas previously studied cyclodehydratases produced adenosine diphosphate (ADP) and phosphate (Pi).^{17,37,594,621} This difference in mechanism questioned whether the E1-like domain really was involved in adenylation or if the YcaO protein was truly the ATP-utilizing enzyme.⁵⁹² This question was addressed when the structure of the enigmatic YcaO from *E. coli* (Ec-YcaO) was determined with a bound ATP analog.¹⁸ The analog indicates ATP was bound in a novel way and was coordinated by two Mg²⁺ ions (also see section 1.6.3.7). Although Ec-YcaO is not believed to be involved in azoline biosynthesis, the residues which interact with ATP were conserved among the entire YcaO superfamily, regardless of associated function. Furthermore, in the presence of ATP, Ec-YcaO exhibited slow hydrolysis of ATP to AMP and PPi, indicating that ATP can be hydrolyzed by YcaOs even in the absence of recipient substrate. Thus, with this discovery, a more conclusive and universal role for the YcaO protein was established. Later, another fused cyclodehydratase (LynD) was crystallized with an ATP analog, confirming the ATP-binding site for a bona fide azoline-forming YcaO and that the YcaO is the site of cyclodehydration.⁵⁹³ The study noted that the adenylation mechanism suggested from the original TruD study was incompatible with the binding of ATP.

In addition to showing the site of ATP binding, the structure of the fused cyclodehydratase also revealed the interactions between the E1-like and the YcaO domain and the assembly of the overall cyclodehydratase.⁵⁹³ The structure was dimeric, with the two monomers interacting primarily through their E1-like domains (Figure 1.31). The peptide-binding region of the E1-like domain (also known as the RRE and discussed below) also forms dimerization contacts and is responsible for orienting the core region of the precursor peptide toward the active site of the YcaO from its own monomer. It is noteworthy that most azoline-forming YcaO proteins contain a Pro-rich C-terminus, often with the final five residues being a PxPxP motif. This portion of the YcaO points into the ATP-binding pocket and is crucial for organizing the active site residues and enabling the proper interaction of the E1-like and YcaO domains.¹⁸ Replacement of these Pro with Gly, truncation of the PxPxP motif, and shifting of the motif by even two residues by inserting a random sequence upstream all abolish cyclodehydratase activity. The structure of LynD reveals that the PxPxP tail region does not directly interact with the E1-like partner,⁵⁹³ so it likely helps maintain proper structure elsewhere for this critical interaction.

1.6.1.2 E1-like partner protein

As experiments began pointing toward the YcaO protein as the component with cyclodehydratase activity, the function of the E1-like protein needed to be reconsidered. One of the first indications for the role of the E1-like protein was revealed from studies with SagC from SLS biosynthesis (Figure 1.6). When theazole synthetase components were individually tested for binding to an array of leader peptides, only SagC remained bound after extensive washing.¹³⁶ Investigation of BalhC from the hakacin BGC (Figure 1.9) also supported the LAP C protein (E1-like protein) as the site of leader peptide binding. BalhC bound a fluorescently labeled leader peptide whereas BalhD had no significant interaction.¹⁸ At this point, no binding site had been determined, but the structure of TruD demonstrated the similarity of the E1-like partner protein with the protein MccB, which is an adenylyase involved in microcin C7 biosynthesis. During microcin C7 maturation, MccB binds the peptide substrate using its N-terminal, winged helix-turn-helix (wHTH) domain like a peptide clamp. This same wHTH domain was present in LAP C proteins. However, as demonstrated by the leader peptide-bound crystal structure of the cyclodehydratase LynD, a peptide clamp binding mode is not employed during azoline biosynthesis (Figure 1.31). Instead, the leader peptide adds on to an existing three-stranded β -sheet structure as the fourth, antiparallel β -strand in the wHTH domain of the E1-like domain. Although the YcaO portion of the cyclodehydratase is spatially close to the leader peptide, it does not appear to contribute much to substrate binding, which provides a possible explanation for the incorrect conclusions as to the role of individual proteins from the previously-mentioned photolabeling experiments with MccBCD. Importantly, the leader peptide-binding site of the cyclodehydratase was definitively revealed by the LynD structure and a concurrent study further supported the crystallographic data with binding data for different LAP C proteins mutated in this region.²⁷ In addition to controlling and guiding substrate recognition, the E1-like partner protein is also crucial for allosterically activating the YcaO protein.¹⁸ Even in the absence of substrate, the partner protein enhances the background rate of ATP hydrolysis and lowers the K_M for ATP.¹⁸

Unexpectedly, comparison of the leader peptide-binding site on the cyclodehydratase with other protein structures revealed that another RiPP biosynthetic protein, the lanthipeptide dehydratase NisB, has a structurally similar leader peptide-binding domain and the same leader peptide-binding orientation.²⁰⁵⁵⁹³ The structural similarity between these proteins was unanticipated because their protein sequences did not bear any noticeable similarity. However, upon the use of more sensitive profile-HMM comparisons, the leader peptide-binding domains from RiPP cyclodehydratases and lanthipeptide dehydratases were indeed sequence-related to each other, as both were homologous to the PqqD protein from PQQ biosynthesis. Further, these PqqD-like domains were present in many other prokaryotic RiPP biosynthetic pathways.²⁷ The widespread prevalence of this domain in different RiPPs led to its naming as the RiPP precursor peptide recognition element (RRE). Thus, the RRE is a structurally conserved leader peptide-

binding domain that guides RiPP biosynthesis in over half of all identified prokaryotic RiPP classes. While some RiPP pathways and enzymes have alternative mechanisms or protein regions for binding their leader peptides,^{28, 29} it appears that the most common and unified mechanism is the PqqD-like structural motif defined by the RRE as employed by canonical RiPP cyclodehydratases.

1.6.1.3 Ocin-ThiF-like partner protein

Even though most RiPP cyclodehydratases contain an E1-like and YcaO protein, bioinformatic analysis of the putative LAP BGCs from the heterocycloanthracins seemed to suggest yet another domain architecture for the cyclodehydratase. Within the heterocycloanthracins, the YcaO protein is fused to the E1-like protein, but unlike the cyanobactins, where the fusion protein length is about equal to the sum of two individual LAP C and D proteins, the heterocycloanthracins were shorter by ~100-150 amino acids. Additionally, even though a dehydrogenase was occasionally absent from the putative heterocycloanthracin gene clusters, the YcaO protein virtually always co-occurred with an uncharacterized protein.¹⁸⁴ This uncharacterized protein was distantly related to ThiF, a member of the E1 superfamily, and was annotated as Ocin-ThiF-like (or F partner protein).

Indeed, the truncated, fused cyclodehydratases required the F partner protein for azoline formation.¹⁸⁵ Although this protein is distantly related to the E1-like protein, they do not detectably bind Zn and do not have any zinc-tetrathiolate motifs, indicating significant divergence from better known members of the E1 family. This discovery led to the naming of this type of azoline synthetase as an F-dependent cyclodehydratase. Upon closer inspection, the shortened length of the YcaO fusion in heterocycloanthracin BGCs was due to the lack of the PqqD-like N-terminal wHTH domain, which had been implicated in leader peptide binding.^{18, 591} Consistent with this hypothesis, the truncated (fused) cyclodehydratase HcaD did not bind the HcaA precursor whereas the cognate HcaF protein did.¹⁸⁵ Later bioinformatic analysis indicated that the F partner protein contains a RRE at its N-terminus while the truncated cyclodehydratase fusion protein was missing this domain, rationalizing its shorter length and how the F partner protein binds its substrates via the leader peptide of the precursor.²⁷

However, little is known about how the Ocin-ThiF-like protein in HCA gene clusters interacts with the F-dependent cyclodehydratase because of a lack of structural data. Current evidence suggests that the F-protein interacts with the YcaO through the E1-like domain and that the active azoline synthetase is a complex of one F partner protein and one truncated YcaO fusion protein (Figure 1.31).¹⁸⁵ Furthermore, F-dependent cyclodehydratases are also found in thiopeptide BGCs and have been confirmed to be necessary for their biosynthesis.^{45, 53, 258, 265} Lastly, the leader peptide-binding site has been investigated through mutagenesis and binding experiments with all available data consistent with models based on the known structure of LynD.²⁷

1.6.2 Mechanism of azol(in)e formation

Cyclodehydration of Cys, Ser and Thr residues to form azoline heterocycles was known to require ATP and the YcaO domain from early *in vitro* reconstitution experiments with McbBCD (Figure 1.5).^{37, 621, 623} In contrast to McbBCD, which requires the entire heterotrimeric synthetase for azole formation, the cyanobactin cyclodehydratases, which are fusions of the E1-like and YcaO protein (LAP C and D proteins, respectively), are known to install azoline heterocycles in the absence of a dehydrogenase.^{548, 594} Other RiPP cyclodehydratases also do not require a dehydrogenase for activity.^{182, 185} These biosynthetic platforms allowed for separation of function between the cyclization and dehydrogenation events and ruled out a mechanism where dehydration may have preceded cyclodehydration.^{548, 594}

The role of ATP during cyclodehydration was, at this point, still unclear and multiple potential mechanisms for ATP utilization arose (Figure 1.32).^{17, 594} Direct activation of the peptide substrate was proposed to allow for nucleophilic attack on the carbonyl that proceeds through a hemioorthoamide intermediate that eliminates phosphate from the conversion of ATP to ADP. Alternatively, the ATP could have played an indirect role through allosteric activation of the YcaO, designated as the “molecular machine” hypothesis.⁵⁹⁴ Lastly, a mechanism related to intein splicing was proposed as a third possibility by which similar nucleophilic attack on the adjacent amide carbonyl generates a tetrahedral hemioorthoamide intermediate. With subsequent *O*-protonation, water is then eliminated. This intein-like mechanism appeared least likely as cyclodehydration was known to not proceed without ATP.^{51, 136}

1.6.2.1 Amide *O*-(pyro)phosphorylation and elimination

Reconstitution of the Balh cyclodehydratase¹⁸² from *Bacillus* sp. Al Hakam shed critical light on the mechanism of ATP utilization by YcaO proteins. Indeed, it was determined that the YcaO protein, BalhD, and not BalhC (E1-like LAP C protein) was responsible for ATP hydrolysis to ADP and phosphate, disproving the previously proposed function of the YcaO as a “docking scaffold” for assembly of the trimeric azole synthetase. This result was further corroborated by the finding that BalhD alone was necessary and sufficient for cyclodehydration of the cognate precursor peptides, BalhA1 and BalhA2 (Figure 1.9). In contrast to earlier reports regarding the cyanobactin cyclodehydratases,⁵⁹⁴ a stoichiometric consumption of ATP was seen when compared to azoline formation, though notably, reactions with BalhD alone produced deviations from this ratio, suggesting a potential dysregulation that is rectified by the addition of the BalhC partner protein.¹⁷

To differentiate between direct activation and the molecular machine hypotheses, cyclodehydratase reactions were conducted in [¹⁸O]-H₂O. In contrast to the direct activation mechanism, the molecular machine hypothesis involves bulk water in the hydrolysis of the γ -phosphate, and hence, ¹⁸O incorporation

into free phosphate would be expected if this mechanism were operative.¹⁷ Analysis by ³¹P-NMR was consistent with all [¹⁶O]-phosphate (¹⁸O imparts a shielding effect on the ³¹P nucleus), disproving the molecular machine mechanism. A complementary experiment utilized BalhA where the amide oxygen upstream of every cyclized position was replaced with ¹⁸O. The preparation of this key substrate was from a new procedure called AMPL, for Azoline-Mediated Peptide backbone Labeling (discussed further below in Figure 1.34).¹⁸³ Isotopically labeled BalhA1 was then used as substrate for the Balh cyclodehydratase using non-labeled [¹⁶O]-H₂O. This time, mono-[¹⁸O]-phosphate was detected by ³¹P-NMR, providing the strongest support to date that cyclodehydration proceeds via a hemioorthoamide intermediate using the direct activation mechanism (Figure 1.33). Analogous results were observed upon testing the MccB17 enzymes, suggesting more broadly that azoline-forming YcaOs use ATP to activate the amide backbone in the form of a phosphorylated hemioorthoamide, akin to a kinase, for modification during biosynthesis.¹⁷

The crystal structure of the TruD enzyme was the first available for a YcaO domain and it displayed a novel fold.⁵⁹² Alongside the structural study, an NMR-monitored reaction was carried out at 10 °C, which showed ATP consumption to be congruent with the rate of heterocycle formation (1 ATP hydrolyzed for every heterocycle formed). The TruD byproducts detected were AMP, phosphate (Pi) and pyrophosphate (PPi).⁵⁹² This pattern of products is typically seen for adenylating enzymes, with Pi resulting from breakdown of PPi under the assay conditions. This study also looked for the production of ADP, as would be expected from a kinase mechanism, but failed to detect ADP either by NMR or by using an enzyme-coupled assay, suggesting that the peptide backbone was adenylated rather than phosphorylated by TruD.⁵⁹² Studies on discrete cyclodehydratases (i.e. Mcb and Balh), however, detected only ADP, consistent with a kinase mechanism.^{17, 621} Further, AMP and PPi production was observed Ec-YcaO in the absence of substrate,¹⁸ suggesting that different YcaOs may employ either a kinase-like or adenylation mechanism. Structural and biochemical characterization LynD, a close homologue of TruD, showed that this enzyme produced AMP, PPi and Pi in an identical manner to TruD. However, the structures of both LynD⁵⁹³ and Ec-YcaO¹⁸ reveal a binding mode for the ATP that exposes only the γ -phosphate, making adenylation structurally hindered. At the same time, crystal structures only provide a snapshot and can be misleading. Further experimentation with LynD showed that [¹⁸O] can move from the peptide to both Pi and PPi, apparently at odds with a simple phosphorylation mechanism.⁵⁹³ Although multiple lines of evidence points to a kinase mechanism, the production of PPi by some enzymes remains an open mechanistic question, and it is possible that slightly different modes of activation are occurring within the YcaO superfamily.

1.6.2.2 Azoline oxidation

The oxidation (dehydrogenation) of thiazoline to thiazole is common in natural products, especially in RiPPs.^{543, 546} The tendency for RiPP BGCs to install thiazoles may reflect the increased chemical stability

of the aromatic thiazole relative to the non-aromatic thiazoline. Synthetically, thiazolines can readily be oxidized to thiazoles under mild conditions using molecular oxygen.⁶²⁴ In addition to their sensitivity to oxidation, peptidic thiazolines are relatively easily hydrolyzed to Cys by treatment with aqueous acid.⁵⁴⁸ The hydrolytic susceptibility of thiazolines has been exploited to introduce [¹⁸O] labels into peptides for mechanistic studies (AMPL method).¹⁸³ Ring opening of thiazolines in peptides with acidic [¹⁸O]-H₂O results in the incorporation of [¹⁸O] at the preceding carbonyl oxygen position (Figure 1.34). In contrast, thiazoles cannot be hydrolyzed by simple acid or base treatment.

In LAPs that contain Ser- and Thr-derived heterocycles, such as MccB17 and goadsporin, (methyl)oxazoles are more commonly found than (methyl)oxazolines. In contrast, in the cyanobactin family of natural products, (methyl)oxazolines are more common than (methyl)oxazoles. Oxazoles have significantly less aromatic character than thiazoles,⁶²⁵ which is reflected in their ability to readily undergo Diels-Alder cycloaddition reactions. This reactivity has been exploited synthetically, where the oxazole acts as a diene (Figure 1.35).⁶²⁶ Thiazoles can also undergo Diels-Alder chemistry, but they require harsher conditions,⁶²⁷ consistent with their greater aromatic nature. The lower aromaticity of oxazole means the oxidation of oxazolines requires a more electrochemically powerful reagent compared to a thiazoline.

The so-called B protein in the LAP biosynthetic clusters is an FMN-dependent dehydrogenase that oxidizes the azol(in)e heterocycles during LAP biosynthesis.³⁷ In addition to a characteristic FMN-binding site, B proteins contain a conserved Lys-Tyr motif that has been shown to be essential for catalytic activity.^{37, 72} Many B proteins, such as McbC (so-named based on historical precedent) from the MccB17 pathway, dehydrogenate thiazoline and (methyl)oxazoline to the corresponding azole.³⁷ The ability to process methyloxazoline is interesting and may indicate a stereochemical constraint on the mechanism. Starting from L-Thr, the corresponding methyloxazoline has the two hydrogens (one each on C α and C β) that are removed during oxidation. The simplest mechanism would be deprotonation of the more acidic C α proton, followed by hydride transfer from C β atom to the N5 atom of FMN (Figure 1.36). This mechanism is essentially an E2 elimination. If an E2 mechanism is operative for methyloxazoline, then it presumably operates in the same manner for thiazoline and oxazoline. An E2 mechanism would require an L-configured azoline and an anti-configured hydrogen at C β . Such a requirement would mean that L-allo-Thr, D-Thr, D-allo-Thr, D-Ser or D-Cys derived azoline rings would not be substrates for B proteins, as each would have at least one improperly positioned hydrogen. This prediction has yet to be tested experimentally.

The presence of oxazolines alongside thiazoles in many natural products may indicate that not all B proteins are electrochemically capable of forming oxazoles. Indeed, one dehydrogenase from *Bacillus cereus* (BcerB) has been shown experimentally to be only capable of oxidizing thiazolines.⁷² The electrochemical potential of BcerB and McbC (which oxidizes both thiazolines and oxazolines) was measured and found to be identical for both enzymes within error. This finding indicates that the difference

in reactivity of these enzymes cannot be attributed to a simple difference in the electrochemical potential of the FMN cofactor.⁷² Since the potential is equal between these enzymes, it seems unlikely that the different energy of hydride transfer from an oxazoline vs a thiazoline would underpin the difference in reactivity. Alternatively, if an E2 mechanism is operative, it would suggest that the difference may lie in the deprotonation of C α and/or in the larger sulfur atom being required for correct positioning of the ring.

Cyanobactins tend to have oxazolines and thiazoles, suggesting their dehydrogenases share the same preferences as BcerB although this has yet to be established experimentally. There are, however, examples of oxazole-containing cyanobactins, (e.g. tenuocyclamides⁶²⁸) demonstrating that some cyanobactin dehydrogenases are capable of oxidizing oxazolines. In contrast to LAPs, the case of cyanobactins raises the question as to when during biosynthesis does oxidation occur. In theory, the dehydrogenase could act before or after macrocyclization; alternatively, dehydrogenation could be independent of macrocyclization. Using purified enzymes, it has been established that the isolated dehydrogenase domain of ArtG is capable of oxidizing both linear and macrocyclic substrates whereas the dehydrogenase from the *thc* pathway⁵⁸⁶ only oxidizes macrocyclic substrates; however, no information on the relative rates was reported.⁵⁹⁶ LAP dehydrogenases, on the other hand, tend to immediately dehydrogenate azolines given that partially processed intermediates are typically found as azoles.¹⁸² A further complication in many cyanobactins is that the residue N-terminal to the thiazole is epimerized. Simple chemical considerations, such as avoiding the disruption of aromaticity by delocalization of the negative charge, would favor epimerization of the residue when it is adjacent to thiazoline rather than thiazole (i.e. before oxidation, Figure 1.37). Spontaneous epimerization has been observed previously for the thiazoline of lissoclinamide,⁶²⁹ and it likely occurs after macrocyclization.⁵⁹⁶ A related epimerization mechanism may explain the presence of D-Asp next to the thiazole in bottromycin,^{630, 631} but generally these epimerizations remain an enigmatic modification and are not common among RiPPs. Overall, the data are most consistent with epimerization and then oxidation as the final two steps during cyanobactin maturation (Figure 1.38).

The crystal structure of the dehydrogenase domain from the cyanothecamide gene cluster has been determined (Figure 1.39). As expected, the protein has a fold characteristic to FMN-utilizing enzymes and the previously mentioned, conserved Lys-Tyr is adjacent to FMN in the active site. Intriguingly, the structure displays two PqqD-like RRE domains. The role of either RRE in substrate binding or dehydrogenation has yet to be elucidated. Given that the (leaderless) macrocyclic peptide is a substrate for the dehydrogenase,⁵⁹⁶ it may be that the RRE is non-functional, as has been previously observed in a thiopeptide pathway with multiple RRE-containing proteins.⁶³²

1.6.3 Enzymatic processing and promiscuity

Based on the roles of their individual components and the determination of their structure with bound substrates, there is now an established model for the assembly and function of canonical RiPP cyclodehydratases (Figure 1.31). It is striking that YcaO proteins can process multiple positions within a single peptide substrate. Indeed, as highlighted in previous sections, azol(in)e-containing RiPPs with only one backbone heterocycle are rare (e.g. trifolitoxin and bottromycin). However, catalyzing multiple turnovers poses a challenge for these enzymes because each modification changes the substrate for the next turnover. In essence, each site of modification is a different substrate, which implies that the cyclodehydratase is inherently tolerant to a variety of substrates. Much of this ability can be attributed to the leader peptide. Upon a binding to the leader peptide and correct positioning via the RRE, the YcaO can promiscuously process variable core sequences.

Insight into how the peptide substrate is handled by the cyclodehydratase has come from a number of biochemical investigations. YcaO domains utilize ATP directly for catalysis (section 1.6.2.1)^{17, 183} and it has been observed that these enzymes evolved to limit the non-productive hydrolysis of ATP. For example, ATP consumption is stoichiometric (1:1) with heterocycle formation so ATP is efficiently coupled to catalysis.¹⁷ ATP hydrolysis by RiPP cyclodehydratases is also slow in the absence of peptide substrate, although the presence of pseudosubstrates (e.g. modified substrate) seem to stimulate enhanced background hydrolysis rates.^{18, 594} Studies of cyclodehydratases from various BGCs have revealed many common rules for how residues are selected for heterocyclization and suggested an interplay between chemoselectivity, regioselectivity, and local sequence context during processing of multiple positions within a single peptide. These investigations have also demonstrated the importance and role of the leader peptide in allosterically activating the cyclodehydratase and placing the cyclizable residues properly within the active site.

1.6.3.1 Chemoselectivity

Among the standard cyclizable residues, azoline-forming YcaOs have demonstrated a clear preference (30- to >100-fold) for modifying Cys over Ser/Thr at the same position of a peptide, indicative of chemoselectivity for sulfur over oxygen nucleophiles.^{182, 548, 633} Indeed, some characterized cyclodehydratases act primarily on Cys residues,^{45, 182, 185} but several others also process Ser and Thr residues.^{37, 51, 173, 548} Examination of known azol(in)e-containing RiPPs (i.e. those discussed above) reveals that thiazol(in)e heterocycles are more common than (methyl)oxazol(in)e. Given the predicted formation of the hemioorthoamide as a key step in the reaction, the much higher nucleophilicity of the Cys side chain and its lower pK_a compared to Ser or Thr explains this preference.

It has also been shown that PatD will process selenocysteine-containing precursor peptides, which generates seleno analogues of cyanobactins *in vitro*.⁶³⁴ This *in vitro* method represented a new route to

selenazolines circumventing the need for stoichiometric amounts of highly toxic selenium donors, which are normally required for chemical synthesis.⁶³⁴ It has been speculated that since selenium was so easily accommodated, and selenocysteine is a natural compound, that seleno analogues of cyanobactins or LAPs may exist in nature, but these have yet to be discovered. Non-canonical β -nucleophilic amino acids have not been tested, but conceivably other amino acids could be cyclized (e.g. 2,3-diaminopropionic acid to imidazoline). However, amino acids with γ -nucleophiles are unlikely to be heterocyclized into the six-membered ring as homocysteine was rejected by the McbBCD synthetase (Figure 1.5).⁶³³

In thiocillin biosynthesis, the core peptide contains six Cys residues all of which are normally transformed to thiazole (Figure 1.18).²²⁹ When these six Cys residues were each individually replaced with Ser, the major product in each case contained an unmodified Ser residue at the substituted position,²⁶⁷ oxazol(in)e-containing compounds were only observed as minor compounds, and then only at specific positions.²⁶⁷ In this case, a simple model based on chemical reactivity of the substrate is sufficient to explain the difference. However, this difference in chemoselectivity is not significant enough to derail downstream processing, as the GE37468 biosynthetic pathway tolerated Ser/Thr in place of Cys2 although Cys4 could not be substituted.³⁶¹ Additionally, cyanobactins demonstrate a case where, with the same substrate, PatD from the patellamide pathway will produce both thiazolines and (methyl)oxazolines, while the very closely related TruD from the trunkamide pathway (>93% similar to PatD) will only produce thiazolines, skipping over Ser/Thr, which are subsequently prenylated by a prenyltransferase.^{548, 592} Thus the prevalence of thiazol(in)e is not simply a case of substrate reactivity, rather the enzyme itself plays a role.

1.6.3.2 Regioselectivity

Although Cys nucleophiles are favored for heterocyclization, the picture is more complex as the tolerance of these enzymes towards cyclizable residues is also dependent on their position in the sequence. This regioselectivity was first shown for the MccB17-forming McbBCD synthetase (section 1.2.1) where shifting a Cys one position backward in the core sequence reduced the processing rate by 50%.⁶³³ Additionally, although McbBCD modifies Ser, only certain positions are processed.^{113, 635} Similar regioselectivity was seen in experiments with the hakacin synthetase (section 1.2.4). Variants of the hakacin substrate were constructed which had only one Cys per peptide, at each of the native sites, with all other heterocyclizable residues being substituted in order to see if one position was favored for cyclodehydration over the others.¹⁸² Only one variant was rapidly heterocyclized at nearly the rate of the native peptide whereas all others were cyclized ~20-fold slower. Intriguingly, the single favored position was also the first Cys normally cyclized within the native peptide.¹⁸² For cyanobactin cyclodehydratases, the majority seem to process all Cys in the core peptide regardless of position, and Ser/Thr are generally accepted at different positions in the core peptide.⁶³⁶ However, a different study found that Thr/Ser near the N-terminus of the

core tend to be skipped whereas more C-terminal Thr/Ser are more readily cyclodehydrated.⁵⁴⁸ A notable exception is ulithiacyclamide, where a pair of Cys residues remain unprocessed by the cyclodehydratase, forming instead a disulfide bond which thought to be critical for activity.^{557, 558, 637}

While the above examples of regioselectivity reveal the preferences of different cyclodehydratases, the selective control of this enzyme is much more crucial when it is involved in pathways with multiple enzymes that act on the same residues. In these cases, strict regioselective control is usually displayed. For example, in cyanobactin pathways, when a prenyltransferase is present, certain Ser/Thr residues must remain uncyclized in order for it to be able to act on the substrate and may explain why TruD is a Cys-selective cyclodehydratase since *tru* pathways tend to include prenylation.⁵⁴⁸ Similar selectivity is demonstrated by theazole synthetase (GodD/E) from goadsporin biosynthesis (section 1.2.7) which processes all Cys and Thr but only one of four Ser are converted to oxazoles in the mature compound.^{204, 213} Of the other three Ser residues, two become dehydroalanine while the third remains unmodified.^{204, 213} As for thiopeptide biosynthetic pathways (section 1.3), the cyclodehydratase skips over Cys involved in side ring macrocycles (e.g. nosiheptide or lactocillin), but the basis for this site-specificity has not been explored. More investigated is the regioselectivity during Ser/Thr cyclodehydration. Because dehydroalanine formation at certain Ser residues is necessary for macrocyclization, proper thiopeptide maturation requires control of these potentially competing processes. Most thiopeptide BGCs appear to avoid this problem by cyclizing only Cys residues, although there are multiple berninamycin-like thiopeptides where Thr/Ser residues in certain positions are cyclized whereas others are left unmodified or are dehydrated (Figure 1.17).^{260, 262} This control is especially notable given that the cyclodehydratase acts before the dehydratase.⁴⁵ Additionally, the dehydratase responsible for forming dehydroalanine and dehydrobutyrine must also be regioselective as most known thiopeptides have an unmodified Thr immediately next to a dehydrobutyrine.

1.6.3.3 Order and directionality

Closely related to regioselectivity is the directionality of processing. If one position is favored for the first heterocyclization, the second might also have a preferred position, as will each subsequently modified position. The sum of this selectivity results in the order and direction of processing when the cyclodehydratase cyclizes multiple residues within a peptide. In this view, each successive heterocyclization would enhance the processing of the next site. This is indeed the case as indicated by experiments with the hakacin cyclodehydratase and previously mentioned single Cys variants of its substrate which showed that only one position was initially favored for heterocyclization. Notably, when a Pro was introduced at this favored position (as a crude azoline structural mimic) in the substrate with only one Cys at the site of the second modification, processing became 6-fold faster relative to the substrate with

Ala at the first cyclized position. In general, there is no defined order for all cyclodehydratases. The formation of azol(in)e rings on the MccB17 (section 1.2.1) and the PZN precursor (section 1.2.3) were shown proceed in a N- to C-terminal direction^{173, 623} while TruD (section 1.4.1) has been shown to modify peptide peptides with the opposite C- to N-terminal directionality (Figure 1.40).⁵⁹² However, the order of heterocyclization is not always linear as the hakacin synthetase (section 1.2.4) processes BalhA1 in an overall C- to N-terminal direction but doubles back to the most C-terminal position during processing.¹⁸² Similarly, the thiomuracin cyclodehydratase (section 1.3.2.4) has a non-linear preferred order but only for first three cyclodehydrations, as the remaining thiazolines are installed without detectable intermediates or order.⁶³²

The importance of the order of heterocyclization and the presence of previous azol(in)e heterocycles for further modification has also been shown by substituting a cyclizable residue (Cys/Ser/Thr) with a non-cyclizable Ala residue in other examples.^{173, 182, 592} This is best illustrated by the PZN synthetase, whereby substitution of Cys/Ser/Thr with Ala in the PZN precursor peptide BamA, prevents processing of all succeeding cyclizable residues; however, all residues N-terminal to the substitution are processed normally (Figure 1.7).¹⁷³ While substitution of a cyclizable residue with a Pro has been shown to be tolerated by some cyclodehydratases as a steric mimic of an azoline ring,^{136, 182, 589, 592} Pro also blocked processing by the PZN synthetase, indicating its strict requirements. For a cyanobactin example, when the C-terminal core peptide Cys of an engineered patellamide precursor peptide was substituted with Ala, processing of the internal Cys by TruD was impaired; however, substitution of the same Cys residue with Pro enabled complete processing at the internal site,⁵⁹² similar to the previous observations with cyanobactins.⁵⁴⁸ At least for the patellamide system, the reason for this is that the cyclodehydratase binds as tightly to the Ala variant as the Cys-containing native substrate; thus, the enzyme appears locked onto the wrong place. The Pro variants binds much more weakly allowing the enzyme to sample and thus process the other Cys residues.⁵⁹²

The thiopeptide BGCs have also given some insight into how altering the normal number and position of heterocyclized residues affects complete maturation of the peptide product.^{267, 344} The thiocillinazole-synthetase complex seems to be very tolerant in skipping over normally cyclized residues substituted with Ala, with only one thiazole position (Cys11) being required for production of the macrocyclic product (Figure 1.18).²⁶⁷ The requirement for the thiazole moiety at the Cys11 position was noteworthy and mirrors the observation that a C51A PatE mutant aborts further processing.⁵⁹² Two other positions also did not produce a product, Ser1 and Ser10, but have previously been shown to be critical in construction of the central pyridine ring.²³⁶ Therefore, the inability to isolate any product when Ser1 or Ser10 are substituted cannot be attributed to failure of processing by the thiocillin cyclodehydratase and reinforces the importance of cyclodehydratase regioselectivity in biosynthetic pathways. In another example, the biosynthetic

pathway of GE37468 was unable to tolerate substitution of its azole heterocycles with non-cyclizable residues, but again it was not clear if this was due to the cyclodehydratase or downstream processing events.³⁶¹ On the other hand, the thiomuracin cyclodehydratase has no difficulty in processing around altered Cys sites.⁶³² A molecular level understanding of these preferences has yet to be determined.

Because the cyclodehydratase processes the same substrate at multiple locations, there has also been interest in determining whether the enzymes is distributive or processive. A distributive enzyme will process a single site, release the substrate, and then process a second site after reengaging the substrate, resulting in the easy detection of intermediates. In contrast, a processive enzyme will process every site before releasing the substrate. Some enzymes can even be partially processive, in that they will process several sites, release, then process additional sites in a distributive fashion. It has been shown that trunkamide, hakacin, and MccB17 azole synthetases are distributive.^{182, 592, 620, 623} Logically, distributive enzymes have faster dissociation rates for the enzyme-substrate complex than k_{cat} . Processive enzymes are the opposite: they must have a k_{cat} that is faster than the dissociation rate of the enzyme-substrate complex.

1.6.3.4 Tolerance of flanking residues

Given the diversity of azol(in)e-containing RiPPs found in nature, it is difficult to determine a universal set of rules that govern substrate tolerance of RiPP cyclodehydratases. While chemo-/regioselectivity influence their promiscuity, another crucial variable in processing is tolerance, or lack thereof, for the identity of the residues that immediately flank the Cys, Ser or Thr undergoing cyclodehydration. The first insights into the effect of flanking residues came from studies of the MccB17 azole synthetase (section 1.2.1). Substitution of the native Gly in the position immediately before the first cyclizable residue with charged amino acids gave no turnover and actually was an effective inhibitor to processing of the native substrate.⁶³⁸ Thus, the Gly in the -1 position appears to be essential as even substitution with Ala was not accepted by MccBCD. The enzyme was somewhat more tolerant towards variation in the +1 position (immediately following a cyclized residue).⁶³⁸ The effect of flanking heterocycles was also studied with the MccB17 cyclodehydratase because an azole in the -1 position would theoretically decrease the electrophilicity of the proceeding amide bond through conjugation. Experimentally, this could inhibit heterocyclization,^{633, 639} but only at one position.⁶²³ This effect does not carry a significant catalytic consequence in other cases, given the contiguous heterocycles found in PZN (section 1.2.3), and presumably the SLS-like cytolysins (section 1.2.2).^{51, 173}

Similar trends in the tolerance for flanking residues were also observed during investigation of the hakacin azole synthetase (section 1.2.4).¹⁸² Analyzing the sequence of the BalhA1 precursor peptide and homologues revealed that Gly nearly always preceded a Cys whereas hydrophobic amino acids were usually in the succeeding position.¹⁸² To develop an experimental basis for these observations, the Gly in the -1

position (relative to Cys) was substituted with Ala, Asp, or Lys, and these replacements decreased the rate of ATP consumption, and by extension, the heterocyclization reaction.¹⁸² However, since the heterocyclic product was observed, Gly was clearly not an essential feature for the enzyme. Rather, the preference of the enzyme for Gly was predicted to arise from its unique conformation, which enables a greater sampling of Ramachandran space,¹⁸² and thus increases flexibility of the neighboring C α positions for faster processing. Substitutions of the residues in the +1 position (relative to Cys) were widely tolerated, including charged amino acids, but Pro was not accepted.¹⁸² Pro is not tolerated in either flanking position in PZN biosynthesis,¹⁷³ but another cyclodehydratase from *Corynebacterium urealyticum* is able to cyclize Cys with a preceding Pro (see also the thiopeptides GE2270, amythiamicin, and baringolin, Figure 1.17). However, the cyclodehydratases associated with PZN biosynthesis are generally less tolerant than other cyclodehydratases given their requirement for contiguous heterocycles.^{173, 178} As a whole, this highlights that each cyclodehydratase will have a unique set of requirements and tolerance for flanking residues.

In cyanobactin sequences, there appears a general preference for hydrophobic amino acids flanking the heterocyclic moiety.^{543, 546} The cyanobactin cyclodehydratases TruD/PatD process non-native substrates with hydrophobic flanking residues, presumably by mimicking native sites of modification.⁵⁹⁶ However, examples with polar and charged amino acids are known.^{549, 586} Additional work by Goto *et al.* demonstrated that PatD could tolerate charged residues flanking Cys and install multiple contiguous heterocycles *in vitro*.⁶³⁶ In a more extensive study, the promiscuity of trunkamide biosynthesis was examined *in vivo*, with the natural trunkamide (*tru*) biosynthetic pathway.⁵⁸⁴ A library of precursor peptide mutants was constructed using the NNK codon (where N = A/C/G/T and K = G/T),⁶⁴⁰ but Cys was fixed at position 7 of the core peptide since this was required for macrocyclization. Over 300 unique compounds were identified.⁵⁸⁴ It is worth noting that only final macrocyclic products were isolated as a measure for enzyme activity. Since the reaction intermediates were not detected, definitive conclusions about the specific preference of TruD for flanking residues cannot be made. However, Arg, Lys, Asp, Gly, Glu, and Pro were never found in position 6, N-terminal to the cyclizable terminal Cys residue. In other cyanobactins, Pro and Gly residues are found N-terminal to thiazole moieties; however, two sequence motifs “Pro-Cys” and “Gly-Cys” were always located in the middle of the core peptide, and never at the C-terminus.⁵⁴⁶ This implies that the absence of Pro and Gly at position 6 might not be due to a failure of the cyclodehydratase, but due to difficulties in macrocyclization. Overall, these data highlight the combinatorial potential of these enzymes is not limited by substrate specificity, and point to the wider substrate tolerance of the cyclodehydratases, beyond the immediately flanking residues. While certain sequence combinations or individual residues may not be processed, the TruD/PatD cyclodehydratases display a remarkable level of promiscuity. One caveat, however, is that the rate at which substrates were processed was not considered in these studies.

In addition to processing core peptides with a wide range of natural amino acids, both *in vitro* and *in vivo*, the cyanobactin cyclodehydratases have been shown to tolerate peptide substrates containing unnatural amino acids.^{599, 641} Four Phe analogues, *p*-chloro-Phe, *p*-bromo-Phe, *p*-methoxy-Phe and *p*-azido-Phe were incorporated immediately adjacent to the C-terminal Cys within the core peptide and all such substrates were processed by TruD *in vivo*.⁵⁹⁹ *In vitro*, an engineered variant of the aestuaramide cyclodehydratase LynD was shown to process a synthetic peptide substrate containing a derivatizable azidoalanine residue,⁶⁴¹ demonstrating that cyclodehydratases can tolerate unnatural, and unusual functionality only three residues away from the cyclizable Cys. The thiocillin BGC has also been shown to accept unnatural amino acids in the core peptide.²³⁰

Continuing the general trend observed, thiopeptides tend to tolerate small or hydrophobic residues, with charged residues being disfavored.^{267, 305, 361} Substituting Ala next to Cys within the thiocillin precursor peptide only failed to produce a mature product upon replacement of two Ser residues (Ser1 and Ser10) that are directly involved in macrocyclization.²⁶⁷ Multiple positions of GE37468 were also selected for codon randomization and revealed a variety of residues are accepted in positions flanking Cys residues, but no compounds with charged residues were produced.³⁶¹ Substitution of the -1 position of a thiazole in berninamycin further corroborates this pattern for thiopeptide cyclodehydratase.²⁶² Thus, cyclodehydratases generally have broad tolerance for residues in flanking positions except for a tendency to disfavor charged groups, but as previously stated, there are no hard and fast rules; each cyclodehydratase will have its own preference for the local sequence context around a cyclizable residue.

1.6.3.5 Tolerance of core peptide length

While cyclodehydratases have certain preferences as discussed above, they appear to more broadly tolerate expansions or contractions to core regions. Many of the earliest experiments with the MccB17 azole synthetase were performed with a core region truncated by 23 amino acids.^{633, 639} Other LAP cyclodehydratases have been similarly shown to process shorter substrates including the PZN and cytolysin synthetases,^{145, 173} but heterocyclization of longer core regions has yet to be thoroughly investigated for LAPs. Altering the length of thiopeptide core regions also seems to be tolerated,³⁴⁴ but this has not been systematically investigated either. On the other hand, the ability of cyanobactin cyclodehydratases to tolerate variations in the length of the core peptide has been extensively studied and appears to be highly promiscuous. Addition or deletion of one or two amino acids from the core has been accepted in all known cases.^{173, 596, 636} Following from this, much more drastic changes to the length of the substrate were shown to be tolerated as well.⁶³⁶ PatD, for which the core sequence of the substrate naturally contains eight amino acids, is capable of processing much smaller substrates, with just two core peptide residues, and much larger substrates with thirty-six core peptide residue.⁶³⁶ Given that cyanobactin precursors can have multiple core

peptide regions extending from the single leader, it seems that cyanobactin YcaOs may be generally insensitive to increased core peptide length.

However, this general flexibility regarding length has two caveats. The first is that no cyclodehydratase has been shown to cyclize the terminal residue of a peptide. In all known RiPPs, the last azol(in)e always has one or more residues after it (which may be removed by proteolysis as in the case of cyanobactins and bottromycins). This was most clearly demonstrated for PZN because it normally contains a C-terminal methyloxazoline-Phe-COOH, but replacing Phe with a stop codon resulted in an unmodified Thr-COOH.¹⁷³ There are other instances where residues near the C-terminus are also not modified (hakacin and MccB17),^{113, 182} but it is not clear if this is due to a length requirement or the flanking residues.

The other caveat is that the core region must be long enough to reach into the active site of the YcaO protein when the leader peptide is bound. Although the leader peptide is responsible for driving the formation of the enzyme-substrate complex, the binding site of the canonical cyclodehydratases is in a different protein domain from the active site.^{18, 27, 185, 593} This means that spatially, there must be a spacer region between the residues in the leader peptide that interact directly with the RRE and the core residues which are modified. Indeed, shortening the distance between crucial leader peptide residues and the first cyclizable residue has been shown to reduce heterocyclization for the MccB17 synthetase.⁶³⁸ The spacer region in MccB17 contains a Gly₁₀ sequence (Figure 1.5). Deletion of one or three Gly resulted in 5-fold and 13-fold decreased heterocyclization rates, respectively, whereas inserting one additional Gly (Gly₁₁) enhanced heterocycle formation 2-fold.⁶³⁸ Substrates lacking more than three Gly were only partially processed, whereas no heterocycles were detected when the spacer was reduced to three Gly.⁶³⁸ Generally the composition of these spacer regions is not important as demonstrated for cyanobactin cyclodehydratases. Insertions of protease signatures and extensive site directed mutagenesis has no observable effect upon processing.^{592, 636} In hindsight this was not surprising since PatE, the natural substrate of the cyanobactin cyclodehydratase PatD, contains two core peptides within the same precursor peptide, but only a single leader peptide. By definition, the two core peptides have very different numbers and types of intervening residues. Thus, while extensions are generally tolerated, there is a certain minimum requirement.

1.6.3.6 Leader peptide recognition and binding site

The natural hypervariability of RiPP core peptides indicate cyclodehydratases are able to process Cys with few restrictions.^{581, 642} It would appear that the core peptide can be varied almost at will and that a cyclodehydratase can be found to convert Cys within the core sequence.^{113, 173, 182, 592, 595, 636} In contrast to the promiscuity within the core peptide, cyclodehydratases have specific requirements for the leader peptide (this is removed during processing to the final product). It is this specificity that drives binding of the correct

peptide so that other peptides within an organism are not erroneously cyclodehydrated. Although the active site must have some sufficient affinity for the core region to catalyze azoline formation, the core does not appreciably add to the binding affinity of the overall peptide.¹⁷³ It would seem that orientation of the core peptide into the active site by the leader region upon binding is sufficient for catalysis, but the exact mechanics of this process are unknown. Beyond binding and orienting the core, the leader also appears to activate the cyclodehydratase (see below).

The leader peptide has been known to be important for RiPP cyclodehydratases since the study of the MccB17 machinery, in which two key hydrophobic residues were identified as crucial for binding (Phe12 and Leu16, Figure 1.5).⁶⁴³ These residues were later recognized as part of an important binding FXXXB motif (B is any branched chain amino acid) that is found in many LAPs.¹³⁶ Conservation of this leader motif enabled the cytolysin azole synthetase to process non-cognate precursor peptides from other LAP pathways.¹³⁶ However, LAP precursor peptides are usually highly divergent between different families as demonstrated by a PXX(L/V) motif found to be important for binding of the PZN precursor peptide by its synthetase (Figure 1.7).¹⁷³ However, given the promiscuity of processing, as long as the native leader peptide is present, core peptides from different LAP pathways have been shown to create functional chimeric precursor peptides, albeit with different degrees of success.^{136, 173, 182}

The key residues within the leader peptide of cyanobactins have also been mapped out based on bioinformatic, biochemical, and structural data.^{592, 595, 601, 636} These studies have identified the crucial part of the leader is a LAELSEEAL-like sequence, and given the generally high similarity of cyanobactin precursor peptides, most contain this motif and can be moved from one pathway to another.⁶⁴⁴ The thiopeptides also tend to have similar leader sequences, with a region rich in hydrophobic and acid residues, that were recently shown to be important for binding by the cyclodehydratase.⁶³² Given its role in binding, the leader peptide is sensitive to variation of residues within these important recognition regions .

Although the leader peptide is crucial for activity, some cyclodehydratases have shown the ability to act on leaderless substrates, albeit it less efficiently.^{183, 185, 592, 636} Interestingly, for patellamide-like substrates lacking the leader peptide, the cyanobactin azole synthetase could only processed one position: the most C-terminal Cys.^{592, 636} The addition of exogenous leader peptide was shown to accelerate the rate at which the enzyme processed the terminal Cys while also partially restoring heterocyclization elsewhere in the peptide substrate.^{592, 636} These studies suggested that leader peptide binding may allosterically activate the cyclodehydratase as the leader peptide has been shown to enhance processing of leaderless core peptides in other cases.^{183, 185, 593, 636} This has also been demonstrated for other unrelated RiPP subclasses like the lanthipeptides.⁶⁴⁵

Beyond these findings, the most powerful insights came from crystallization of the LynD cyclodehydratase bound to its leader peptide. This structure showed that the highly conserved region of the

cyanobactin leader peptide was bound to the first domain within the cyclodehydratase, with a few minor contacts to the third (YcaO) domain (Figure 1.41).⁵⁹³ This binding mode was first structurally characterized in nisin biosynthesis (NisB protein),²⁰⁵ where the leader peptide adopts a similar β -strand conformation on the exposed edge of the β -sheet region of the conserved, PqqD-like peptide-binding domain of the enzyme.²⁷ Noticeably, binding of the leader seems to induce a change in conformation, possibly providing an explanation for its enhancement of catalysis.⁵⁹³ The domain that binds the leader peptide is conserved in all cyclodehydratases and in many other RiPP modifying enzymes, and is denoted as the RiPP precursor peptide recognition element (RRE).²⁷ For example, the RRE domain is found in the aforementioned NisB and in the PatG protein of the patellamide pathway where it precedes the dehydrogenase domain. The appearance of this conserved RRE domain in very different enzyme reactions, suggests the RRE and the leader peptide evolved distinctly from the cyclodehydratase reaction. What links these enzymes is the physical and spatial separation of recognition from catalysis. The normal intimate connection between recognition and catalysis, shaped by evolution of enzymes and their substrates, often limits the use of enzymes as catalysts since mutations that alter recognition often reduces catalysis. With the use of the RRE domain and disposable leader peptide, nature has evolved an elegant strategy to have enzymes with very broad substrate profiles as previously discussed.

1.6.3.7 ATP Analogs

Although heterocyclization in azol(in)e-containing RiPP pathways was originally shown to be an ATP-dependent reaction,⁶²¹ subsequent investigations have established there is some flexibility in the triphosphate nucleotide that can be used to catalyze the reaction. The LAP cyclodehydratases BalhD, McbBCD and the cyanobactin cyclodehydratase TruD have each been shown to accept GTP.^{17, 592, 621} TruD has been shown to be active with dTTP⁵⁹² as well, but curiously dTTP only supported heterocyclization of the single terminal Cys. Crystal structures of LynD in complex with three different nucleotides, AMP, ADP/P_i and AMPPNP, and the Ec-YcaO in complex with AMP and AMPCPP have been determined.^{18, 593} In all structures, the adenosine ring is bound in essentially the same fashion, predominantly through cation- π stacking and van der Waals interactions (Figure 1.42).⁵⁹³ The analysis of the crystal structures reveal a lack of specific hydrogen bonding between the protein and the adenosine base which rationalized the ability of the YcaO domains to accept other nucleotide triphosphates as co-substrates.^{18, 593} The phosphate groups are, however, specifically coordinated by metal ions which bridge to the protein.

1.7 Non-canonical YcaOs

In contrast to the BGCs discussed above, a significant number of YcaO family members are present in genomic contexts that deviate from the established patterns of characterized YcaOs. Specifically, many

YcaOs in the sequence databases do not occur near or fused to an E1-like^{646, 647} or ThiF-like partner proteins (Figure 1.30),⁶⁴⁸ Given that these partner proteins are essential for azoline-formation, it would suggest these most of these other types of YcaOs are not involved in azoline formation, unless they are able to act as a standalone YcaO. As discussed in the Introduction (section 1.1), not all protein superfamilies are iso-functional so it cannot be assumed that all YcaO proteins participate in thiazoline/oxazoline formation. Thus, to differentiate the different types of YcaOs, we have introduced a nomenclature based on the inferred function and whether there is an identifiable protein that co-occurs with the YcaO. These features are crucial because the function of the active site and whether a protein complex is required for activity will have a profound effect on the conserved motifs present in each type of YcaO. Moreover, establishing these categories is useful when predicting function for large numbers of bioinformatically-identified YcaOs for which only a genome is known because when a function is inferred for one type of YcaO, the similarity of a new YcaO sequence to the different types can help assign its function.⁶⁴⁷

Examples of YcaOs that act without an E1-like or ThiF-like partner have been markedly less well-characterized than their more established azoline-forming YcaO counterparts. However, local genomic neighborhoods and connections to intriguing natural product structures suggest that these less-studied types of YcaOs could harbor distinct functions or biosynthetic capabilities. For example, in addition to the plasmid-born BGC for MccB17 (in section 1.2.1), *E. coli* also encode a YcaO protein (Ec-YcaO) that does not appear to be related to natural product biosynthesis based on the identity of its flanking genes, although it does utilize-ATP (section 1.7.1).¹⁸ Current evidence suggests a role in enhancing RimO-dependent β -thiomethylation of ribosomal protein S12 and is postulated to function as a scaffolding protein.²³ However, no mechanism has been proposed for how ATP might be used so Ec-YcaO remains an enigmatic YcaO. As another example, the bottromycin BGC encodes two YcaO proteins with no obvious candidates for partner proteins. Seeming to correspond with the two YcaOs, the structure of bottromycin includes two amide backbone modifications: a thiazoline and amidine. There is no clear cyclodehydratase partner protein so the azoline-forming YcaO has been classified as “standalone.” The nature of the second, putative amidine-forming YcaO is not clear (section 1.7.2). For a final example of new biosynthetic modifications, the thioviridamide BGC contains a YcaO but is associated with thioamide formation (section 1.7.3). Lastly, this section also covers the molecule trifolitoxin, which does not have a fully elucidated structure, but the partial structure includes a thiazoline. Based on its BGC and the lack of any partner proteins, this appears to be another instance of a standalone azoline-forming YcaO (section 1.7.4).¹⁸

1.7.1 *Escherichia coli* enigmatic YcaO

The YcaO from *E. coli* (Ec-YcaO) has a cryptic function and is thus categorized as an enigmatic YcaO. Unlike those involved in typical azoline formation, this YcaO does not seem to be in a RiPP BGC and lacks

a partner protein.¹⁸ Although the exact function of this YcaO is not known, it has been implicated in a post-translational modification of ribosomal protein S12.

The S12 ribosomal protein, a component of the 30S subunit, contains an Asp88 residue that is universally conserved across all domains of life and known to be crucial for organism viability. In select organisms, such as *Rhodopseudomonas palustris*⁶⁴⁹ and *Thermus thermophilus*,⁶⁵⁰ the Asp88 residue was found to be β -thiomethylated *in vivo*, whereas in eukaryotes this modification is not present.^{651, 652} Studies have shown that RimO, a member of the radical SAM superfamily, is responsible for β -thiomethylation of S12-Asp88.⁶⁵³ RimO is homologous to MiaB, an enzyme involved in the β -thiomethylation of tRNA.⁶⁵⁴

Affinity enrichment experiments using recombinant *E. coli* S12 protein demonstrate the expected association with RimO, but the Ec-YcaO and several other proteins, were found to co-purify as well. Deletion strains lacking RimO and YcaO were then analyzed for the relative amounts of S12 β -thiomethylated Asp88 by quantitative mass spectrometry. The strain lacking RimO contained no detectable level of Asp88 β -thiomethylation while the strain lacking the YcaO showed an 18% reduction.⁶⁵⁵ It is hypothesized that this YcaO functions in a multi-protein complex with S12 and RimO, although the details of the mechanism have not yet been elucidated and will require more effort to firmly establish the role of YcaOs in S12 modification.

Despite being associated β -thiomethylation, the actual function of Ec-YcaO is unknown. It does hydrolyze ATP *in vitro* with K_M comparable to other characterized YcaOs and has had its structure solved by X-ray crystallography (Figure 1.43).¹⁸ This structure was the first conclusive evidence for the site of ATP binding in YcaO proteins and revealed that the region around this binding site was conserved among all members of the YcaO superfamily. For example, even though Ec-YcaO has only 20% identity to the azoline-forming YcaO from *Bacillus* sp. Al Hakam (BalhD), the sequence conservation is centered around the ATP binding site, and mutation of the corresponding residues in BalhD abolish ATP binding and cyclodehydration activity. Furthermore, comparison of the Ec-YcaO structure and the cyanobactin cyclodehydratases TruD⁶⁵⁶ and LynD⁵⁹³ further verify the conservation of structure and ATP-binding site predicted based on sequence homology. These results suggest that insights from one YcaO protein, regardless of any potential involvement in natural product biosynthesis, can inform broadly on other YcaOs.

1.7.2 Bottromycin standalone YcaOs

The bottromycin family of natural products was first reported in 1957 from *Streptomyces bottropensis*.⁶⁵⁷ The proposed structure of bottromycin underwent several revisions. Initially proposed to be a linear peptide,⁶⁵⁸⁻⁶⁶⁰ the major bioactive component bottromycin A₂ was later determined to contain several non-proteinogenic amino acids and a macroamidine.⁶⁶¹ On the basis of ¹⁵N-NMR, the macrocycle vertex was determined to be located in between two *t*-butyl-containing amino acids, consistent with condensation

between the N-terminus and a backbone amide carbonyl.⁶⁶² Minor congeners bottromycin B₂ and C₂ displayed different methylation patterns.⁶⁶³ The structure of bottromycin A₂ was finally confirmed by total synthesis, which also resolved any stereochemical ambiguity.⁶⁶⁴ Several notable structural features include the 12-membered macroamidine, Asp *O*-methylation/epimerization, and a single thiazole that are shared by all currently known members (Figure 1.44). The variation in structure of the analogs comes from differing patterns of Pro, Val, and Phe β-methylation.

The antimicrobial properties of bottromycin have been described since the initial isolation reports, but characterization of its activity has, like its structure, been incremental over decades. Bottromycin A₂, whose biological activity is best characterized, was found to have extremely potent antimicrobial activities against *Mycoplasma gallisepticum*,⁶⁶⁵ methicillin-resistant *Staphylococcus aureus* (MRSA), and vancomycin-resistant *Enterococcus* (VRE)⁶⁶⁴ with MIC of 12 nM, 1.2 μM, and 0.6 μM, respectively. Studies to elucidate the mechanism have indicated a role in protein synthesis inhibition, as demonstrated by synthetic polynucleotide *in vitro* translation systems, which localized bottromycin A₂ binding to the A-site of the 50S subunit.^{666, 667} Moreover, experiments on polysomes showed both a lack of inhibitory activity towards the translocation and peptidyl transferase steps, but released aminoacyl-tRNA.^{668, 669} Hence, bottromycin is believed to interfere with the binding of charged aminoacyl-tRNA to the ribosome, and as such, would act in a manner distinct from other ribosome translation inhibitors such as erythromycin (which does not bind to the A-site), tetracycline, and micrococin (which interact with the A-site, but neither release aminoacyl-tRNA).⁶⁶⁸

Efforts to explore the structure-activity relationship have demonstrated the importance of the Asp-derived methyl ester in biological activity. Amide derivatives of bottromycin B/C outperformed ester analogs in assays against *S. aureus in vivo*, while esters tended to be more active than amides *in vitro*, suggesting a possible difference in pharmacokinetic properties.⁶⁶⁰ Indeed, upon addition of bottromycin A₂ in mouse plasma, rapid degradation of the side chain ester to acid was observed (compared to a chemically hydrolyzed standard), and hypothesized to contribute to the poor antimicrobial efficacy.⁶⁷⁰ Additional amide, thioester, and ketone derivatives of bottromycin A₂ at this position recapitulated these trends: the thioester analogs showed the lowest plasma stability, but most potent MIC values. Conversely, the ketone analogs showed equivalent *in vitro* activity (compared to the original natural product), with better plasma stability compared to the thioesters.

Upon genome sequencing of producing organisms, bottromycin's complex structure was found to originate from a ribosomal precursor peptide, cementing its identity as a RiPP. Whole genome sequencing of *Streptomyces bottropensis*^{38, 671} and *Streptomyces* sp. WMMB272⁶⁷² uncovered the BGC by matching peptides with the deduced sequences for bottromycin A₂/B₂/C₂ (GPVVVFDC) and bottromycin D (GPAVVVFDC). Heterologous expression confirmed that these BGCs were indeed responsible for

bottromycin biosynthesis. For *Streptomyces* sp. WMMB272, mutagenesis via a shuttle vector was employed to mutate the bottromycin D core peptide Ala³ to Val³, resulting in the production of bottromycin A₂ from the native producer. An additional bottromycin BGC was identified in *Streptomyces scabies*, which showed significant sequence similarity to the ones from *S. bottropensis* and *S. sp.* WMMB272.⁶⁷³ Unlike the vast majority of RiPPs, the bottromycin precursor peptide displays an N-terminal core sequence and, instead of an N-terminal leader peptide, the precursor has a C-terminal extension called a “follower” peptide, which presumably contains residues required for binding and subsequent modification by the biosynthetic enzymes.³⁸

The current proposed biosynthetic pathway for bottromycin was deduced using heterologous expression systems and untargeted metabolomics approaches (Figure 1.45). Individual deletions of the three radical SAM methyltransferases halted production of bottromycin A₂ and produced either B₂, D₂ or E₂, allowing the roles of each radical SAM to be assigned.⁶⁷⁴ From mass spectral network analysis of metabolites produced by strains individually lacking each standalone YcaO, among other tailoring genes, a more complete picture of bottromycin biosynthesis was revealed. Thiazoline formation was indeed dependent on the YcaO encoded next to the precursor peptide (Figure 1.44), thus this protein is established as a standalone azoline-forming YcaO. Also confirmed was that all of the β-methylation events occur prior to macroamidation, which is consistent with their functions being follower peptide-dependent.²⁷ Conversely, thiazoline oxidation and O-methylation occur late in the maturation pathway.⁶³¹ The genes required for macroamide formation include the second YcaO and a putative macrocyclase partner with similarity to amidohydrolases; therefore, these proteins were proposed to act together in macroamide formation.¹⁵ The nature of their interaction has not yet been characterized so it is uncertain if the hydrolase should be considered a YcaO partner protein, analogous to the E1-like protein in RiPP cyclodehydrates. The involvement of a YcaO in transforming an N-terminal Gly and internal Val into a macroamide is clearly distinct from the cyclodehydration of Cys, Ser, or Thr, but at a conceptual level, there is a clear similarity; in both cases a nucleophile attacks the peptide bond, formally eliminating water in an ATP-dependent process (Figure 1.46).

1.7.3 Thioviridamide TfuA-associated YcaO

During bioassay-guided isolation for compounds with antitumor activity, a novel natural product was isolated from *Streptomyces olivoviridis*. This compound, thioviridamide, was reported to induce apoptosis selectively against two oncogene-expressing cell lines as well as possessing antibiotic activity.⁶⁷⁵ Structural elucidation revealed thioviridamide contained several unusual functional groups, including a 2-hydroxy-2-methyl-4-oxopentanoyl moiety, five peptide backbone thioamides, β-hydroxylation and N¹, N³-dimethylation of His, and an S-(2-aminovinyl) Cys macrocyclization (Figure 1.47).⁶⁷⁶ The latter modification

has been previously observed in other RiPPs⁶⁷⁷ including gallidermin, mutacin 1140, mersacidin, and cypemycin, suggesting that thioviridamide would also be of ribosomal origin.

Thioviridamide was confirmed to be a RiPP after sequencing the *S. olivoidiridis* genome and identification of the precursor peptide, TvaA, which contained the core sequence VMAAAASIALHC. Heterologous expression in *S. lividans* confirmed the gene cluster *tvaA-O* to be responsible for thioviridamide biosynthesis. Of these genes, *tvaH* belongs to the YcaO superfamily, although thioviridamide contains no azolines. Moreover, no genes in the *tva* cluster are similar to known partner proteins (E1-like or Ocin-ThiF-like) normally associated with azol(in)e biosynthesis, further underscoring a distinct role for this YcaO. The authors suggest that thioamide formation may be catalyzed by *tvaH*, representing the first example of a thioamide-forming YcaO. Interestingly, TvaI which is annotated as a “TfuA-like” protein is often found immediately adjacent to a YcaO protein in many other genomic contexts which has also given rise to this type of YcaO being referred to as TfuA-associated YcaOs.¹⁸

The potential for a thioamide-forming reaction is fascinating, because if it does indeed follow the proposed YcaO-like chemistry (Figure 1.42), then the atom which ultimately replaces the peptide backbone oxygen is not substrate-derived but rather exogenously delivered in a form chemically equivalent to H₂S. If the YcaO domain is confirmed to carry out this transformation, it would significantly enhance the synthetic scope of this enzyme superfamily, since one could imagine other types of nucleophiles being used in similar ways.

1.7.4 Trifolitoxin standalone azol(in)e-forming YcaO

Because plant growth is generally limited by access to mineral nitrogen (NO₂⁻, NO₃⁻, NH₄⁺), many plants form symbiotic relationships with nitrogen-fixing bacteria.^{678, 679} This symbiosis is especially prevalent among legumes because most are able to form special structures in their roots (*i.e.* nodules) that provide bacteria with an exclusive living space and nutrients in exchange for fixed nitrogen, in the form of ammonium or Ala.^{678, 679} Nitrogen-fixing bacteria of this variety are generally referred to as Rhizobia, and the region in or around plant roots is known as the rhizosphere.^{678, 679} Living in root nodules is advantageous due to the supply of nutrients from the host, so *Rhizobium leguminosarum* bv. *trifolii* T24 produces a bacteriostatic, anti-Rhizobial compound named trifolitoxin (TFX) that inhibits the growth of other Rhizobia, which provides a competitive edge in colonizing root nodules.⁶⁸⁰⁻⁶⁸⁶

Investigation into the genetic components responsible for TFX production revealed that *tfxABCDEFG*, were necessary to confer TFX production and resistance,⁶⁸⁷⁻⁶⁸⁹ but an additional gene, *tfxA*, located outside of the BGC, was later also found to be involved (Figure 1.48). Based on sequence similarity, TfxB, TfxC, and TfxF are oxidoreductases.²⁷ TfxD has weak similarity to multidrug and toxic compound extrusion (MATE)-like proteins, but its deletion results in a strain that does not produce TFX but retains

TFX resistance. TfxE is considered a standalone azoline-forming YcaO protein since it is the only candidate for installing the thiazoline ring in TFX. Deletion of TfxE causes sensitivity to TFX,⁶⁹⁰ however, it was not reported if this strain was deficient in TFX production. Intriguingly, no E1-like or Ocin-ThiF-like protein, is present in the TFX BGC, so it is presently unclear if TfxE forms the thiazoline as a standalone protein or if an additional, unidentified partner protein is required. TfxG is related to protein kinases and has an undescribed role in producing TFX isomers.⁶⁹⁰ Some TfxG homologs have been implicated in alkali tolerance in *Sinorhizobium meliloti*, but such organisms lack all other TFX genes so this relationship is likely unrelated to any potential role in TFX biosynthesis.⁶⁹¹ One final distally encoded gene, *tfuA*, also appears related to production of TFX, as insertional mutagenesis of this gene was reported to abolish TFX production, but how it affects the biosynthesis is unknown since it is not required for TFX production in heterologous hosts.⁶⁹² It should be noted that this *R. leguminosarum tfuA* gene is what gave rise to the naming of TfuA-like proteins (*i.e.* TvaI thioviridamide biosynthesis, Figure 1.47); however, the “TfuA-like” protein has no sequence similarity to the *R. leguminosarum* TfuA (GenBank accession AAB17513), which instead has transcriptional activation, TolB, and multiple TPR repeat domains. The two genes downstream of *tfuA* (accession: AAB17513) in *R. leguminosarum* are homologous to TvaH (YcaO; AAB17514) and TvaI (“TfuA-like”; AAB17515) but are completely unrelated to TFX production so we believe an error was made when the TfuA-like genes were originally named. In fact, we predict that in addition to TFX, *R. leguminosarum* may possibly also produce a thioamide-containing RiPP.

Despite the isolation and purification of mature TFX, the chemical structure remains elusive. A partial structure is available, but Arg and Gln are posttranslationally modified to form an unknown chromophore.⁶⁹³ Expression of precursor peptides in a *tfxA* deletion strain has allowed introduction of non-native precursor peptides to investigate the pathway. These experiments showed that substitution of Arg, Gln, and Cys lead to loss of activity. The C-terminal Ala of TFX is also crucial for activity.⁶⁹⁴

1.8 Biotechnology

The biotechnological interest in YcaO proteins has centered on their ability to produce RiPPs with diverse biological activities. This has driven, and to some extent, justified the interest in azol(in)e-containing RiPP biosynthetic pathways and RiPPs more generally. In practical terms, purification of RiPPs from the natural producer in sufficient quantities for detailed pharmacological or medicinal chemistry studies is not practical for most compounds. Total chemical synthesis to make analogues with more desirable properties is also not attractive due to the chemical complexity of these natural products. Since heterocyclization profoundly changes the chemistry and structure of a compound,⁵⁷⁷ such modifications are essential for the biological activity of the natural product. This means that quickly accessing a broad range of azol(in)e-modified peptide variants requires use of the native biosynthetic systems.

For many complex biomolecules, engineering their pathways to produce new variants is well established, but often the diversity is limited because enzymes have coupled catalysis with recognition at the same site. The combinatorial nature of RiPP biosynthesis suggests RiPPs may be particularly suitable for biotechnological exploitation. The separation of substrate binding from catalysis by using the leader peptide, which is removed from the final product, and the presence of the RRE permit variation in the final product, including non-natural elements, while relying on enzyme catalysis. These factors make RiPPs easier to engineer for library generation and the production of new bioactive peptides compared to other enzyme systems, in which substrate binding and catalysis are mediated by the same site.

There are multiple approaches to further developing azol(in)e-containing RiPPs. One approach is genome mining and genetic engineering to combine enzymes and substrates in new ways to make novel molecules, as exemplified by the work of Schmidt and colleagues.^{49, 50, 584, 599, 695, 696} These approaches are flexible and can be combined with amber codon suppression technology to include unnatural amino acids.^{230, 599, 697} The use of unnatural amino acids has even allowed the creation of derivatives which can be labeled with fluorophores or other tags to aid in mode of action studies or visualization in cells.^{230, 599} In general, *in vivo* combinatorial biosynthesis has seen been successful in generating diverse molecules given the wide ranges of compounds detected in culture which could be screened for desired activity. A few individual compounds have been made in quantity, isolated and characterized using these methods,⁵⁹⁹ but the success is limited as an acknowledged drawback is the low yields of most compounds made in this way and the difficulty in their isolation. Thus, efficiently obtaining milligram quantities of a library of pure compounds by this approach remains a major obstacle.

However, a recent breakthrough in the culture conditions for cyanobactin pathways, by augmentation with Cys, increased the cultured yield of patellins by over 150 fold.⁶⁹⁵ Addition of mevalonate further increased the yield of final product. The effect of these compounds was not due to a simple increase in substrate peptide, whose stability and expression level remained unaltered. The study also ruled out changes to redox or energy status in the cell.⁶⁹⁵ The authors concluded from an *in vitro* experiment, that sulfide derived from Cys regulated the activity of one enzyme, and by ensuring it acted at the correct point in the sequence of biosynthetic events, greatly increased the yield of the final desired product.⁶⁹⁵ If this approach is indeed general then it will transform *in vivo* production and the insights into the critical nature of the timing of the reactions have wider implications.

Other azol(in)e-containing RiPPs have also been produced through heterologous expression and have potential for making new bioactive molecules through precursor peptide reprogramming.^{178, 342, 361} Heterologous expression of the LAP BGC of PZN has been carried out. PZN (section 1.2.3) is normally produced by *Bacillus velezensis* and has antibiotic activity against *Bacillus anthracis*. A large array of substrates for the pathway were made by varying the sequence of the precursor peptide in an *E. coli*

heterologous expression system, and 11 compounds were tested for their biological potency.¹⁷⁸ New thiocillin and GE37468 derivatives have also been made and tested in a similar fashion.^{262, 267, 342, 361}

As an alternative to *in vivo* approaches discussed above, the possibility of combining enzymes *in vitro* has been explored.^{596, 644} The patellamide system has attracted considerable attention in this regard.⁴⁸ As a consequence of the structural and biochemical characterization of the enzymes,⁵⁸⁵ it has been possible to replicate the biosynthesis of patellamides. Fourteen patellamide-like precursor peptides were processed *in vitro* by either TruD or PatD, and subsequently other, cyanobactin biosynthetic enzymes. Taking advantage of the different chemoselectivities of the enzymes, the result was 16 new, non-natural patellamide analogues, 13 of which contained different amino acid sequences from the natural compounds.⁵⁹⁶ Chemically synthesized peptides with non-peptidic moieties such as polyketide-like linkers or triazoles have been macrocyclized by the PatG protease.^{602, 698} An early drawback to these *in vitro* approaches was the slow kinetics of the enzymes, most particularly the protease, PatA, which removes the leader peptide. The reason for the inefficient turnover is unclear, although recent studies indicate that PatA is redox sensitive.^{595, 695} A structural study also showed that the catalytic triad was not optimally aligned.⁵⁸⁶ This problem can also be overcome by removing reducing agents from *in vitro* reaction, which are usually added to prevent disulfide bond formation in their Cys-rich precursors. This enabled one-pot synthesis of cyanobactins using purified enzymes and the creation of unnatural derivatives.⁵⁹⁵ Another approach for *in vitro* biosynthesis is to insert an alternative protease site between the core peptide and the leader and swap PatA for another protease (both trypsin and tobacco etch virus have been used),⁵⁹⁶ and thus a non-cognate protease can be used, greatly accelerating the process.

Another improvement to *in vitro* generation of cyanobactins was the design of a biosynthetic scheme that does not require the leader peptide. Normally, the leader peptide is necessary for processing multiple Cys and, due to the cost of synthesizing the longer leader peptide-containing peptides, the number of substrates tested *in vitro* would be limited. However, the discovery that the addition of the leader peptide *in trans* activated the cyclodehydratase, such that it could process substrates without the leader⁶³⁶ stimulated further investigation. The significance of this was that by dispensing with the leader peptide altogether, one could move to much shorter substrate peptides that could be made entirely synthetically, allowing for greater levels of diversification. Inspired by the *trans* activation of the cyclodehydratase and informed by the structure of LynD from *Lyngbya sp.*, this cyclodehydratase was engineered such that a portion of the leader peptide was fused to the N-terminus of the enzyme.⁵⁹³ The fused enzyme was almost as active on leaderless substrates and the native enzyme on full-length substrates (Figure 1.49). Crucially, the fused enzyme was able to completely process the substrates rather than producing mixtures.⁵⁹³ With the need for a leader peptide alleviated, the process of exploring the synthetic utility of the enzymes can now begin for

substrates which can be made in cost-effective manner by solid-phase chemical synthesis and the results could be of high impact.

In addition to the cyanobactin examples discussed above, total *in vitro* biosynthetic reconstitution of other azol(in)e-containing RiPPs has also been achieved. The first example was with a MccB17 analog which could not be produced *in vivo* so the peptide was purified and treated with the azole synthetase and digested with commercial proteases to test the activity of the new compound.¹¹³ Along the same lines, a combination of *in vitro* biosynthesis and synthetic modification has been used to produce a PZN derivative named CZN.¹⁷³ However, a most complicated example is the recent case of the total *in vitro* biosynthesis of a thiomuracin derivative (a thiopeptide). The core scaffold of thiomuracin (termed thiomuracin GZ) was formed entirely through *in vitro* reconstitution of the core thiopeptide biosynthetic genes from *Thermobispora bispora*.⁴⁵ Although thiomuracin GZ lacks many of the additional tailoring steps, it has comparable bioactivity to naturally produced thiomuracins and thus indicated that even more complex products are within the realm of total *in vitro* biosynthesis.⁴⁵ Moreover, *in vitro* approaches allow increased enzyme concentrations and control compared to *in vivo* heterologous systems which can lead to increased range of product generation.^{45, 173}

Although much focus has been put on enzymes from azol(in)e-containing RiPP pathways for their ability to produce new potential natural products, their mechanisms can be taken advantage of *in vitro* to accomplish different ends. This is most clearly demonstrated by the AMPL methodology previously described (section 1.6.2.2). This method takes advantage of the mechanism of azoline-forming YcaOs to install thioamides or [¹⁸O]-labeled amides into peptides for downstream biochemical investigations (Figure 1.34).¹⁸³

1.9 Summary and outlook

Nature has settled on proteins, polymers of amino acids linked by amide bonds, to provide the bulk of the biomaterial that makes life possible. The chemical properties of the amide bond is taught to all undergraduate chemistry majors as an example of how resonance structures stabilize molecules and impart new chemical properties. Unlike ketones and most other carbonyl groups, the amide carbon is not particularly electrophilic, and unlike amine relatives, the amide nitrogen atom is not particularly basic or nucleophilic. Consequently, the amide bond is very stable and a good choice by nature for the framework of proteins that carry out life processes. Indeed, specialized catalysts (proteases) or harsh base or acid treatments are required to hydrolyze the relatively inert amide bond. The ability of the YcaO superfamily to perform chemistry on the peptide backbone is thus of considerable scientific interest. When one thinks of a posttranslational modification, most think of a side chain modification. When YcaOs are at work, it is the main chain that undergoes modification. Moreover, this ability within the functionally diverse

superfamily is unified through a common structural fold that binds ATP for the activation of peptide amide bonds. This role has been studied most extensively in natural product biosynthesis for the creation of azoline heterocycles from the peptide backbone. Although some azoline-forming YcaOs have been extensively characterized, there remain some key questions, such as an unambiguous description of the transition state, a molecular rationale for the order of heterocyclization, and how some YcaOs act without a partner protein or RRE to manage substrate recognition. We have also highlighted the emergence of thioamide-forming and amidine-forming YcaOs. These enzymes utilize different nucleophiles from the better characterized azoline case (Figure 1.46). How these transformations occur, and what other proteins might participate in these reactions will be of considerable future interest. The new insight into these processes will likely reveal unique protein chemistry and set new precedents for biological catalysis.

In addition to studying YcaOs associated with non-canonical functions, further exploring azoline-forming YcaOs in BGCs with no known product also promises to be a fruitful endeavor. The relationship between natural product structure and YcaO similarity, means that identifying the product of a YcaO in one BGC often informs on other sequence similar YcaOs. Current estimates of the abundance of azoline-containing RiPP BGCs indicate that the majority remain to be discovered.^{699, 700} Already YcaOs have shown their versatility in coordinating modifications with other primary modifying enzymes like dehydratases (i.e. thiopeptides) and macrocyclizing proteases (i.e. cyanobactins) so new combinations with other modifying enzymes may be found in nature to produce entirely different molecular structures. With all these exciting opportunities for discovery related to YcaOs, their products, and their mechanism, we anticipate a wealth of new discoveries that will expand our knowledge of this impressive superfamily.

1.10 Acknowledgements

This work was supported in part by the U.S. National Institutes of Health (GM097142 to D.A.M.) and fellowships from the Chemistry-Biology Interface Training Grant Program (2T32 GM070421 to B.J.B. and C.J.S.). Additional financial support came from the David and Lucile Packard Fellowship for Science and Engineering (to D.A.M.) and the University of Illinois at Urbana-Champaign Department of Chemistry (Robert C. and Carolyn J. Springborn Endowment to B.J.B) and a National Science Foundation Graduate Research Fellowship (DGE-1144245 to B.J.B.). JHN acknowledges the European Research Council (339367), UK Biotechnology and Biological Sciences Research Council (K015508/1), and is a Royal Society Wolfson Merit Award Holder and 1000 talent scholar at Sichuan University.

1.11 Figures

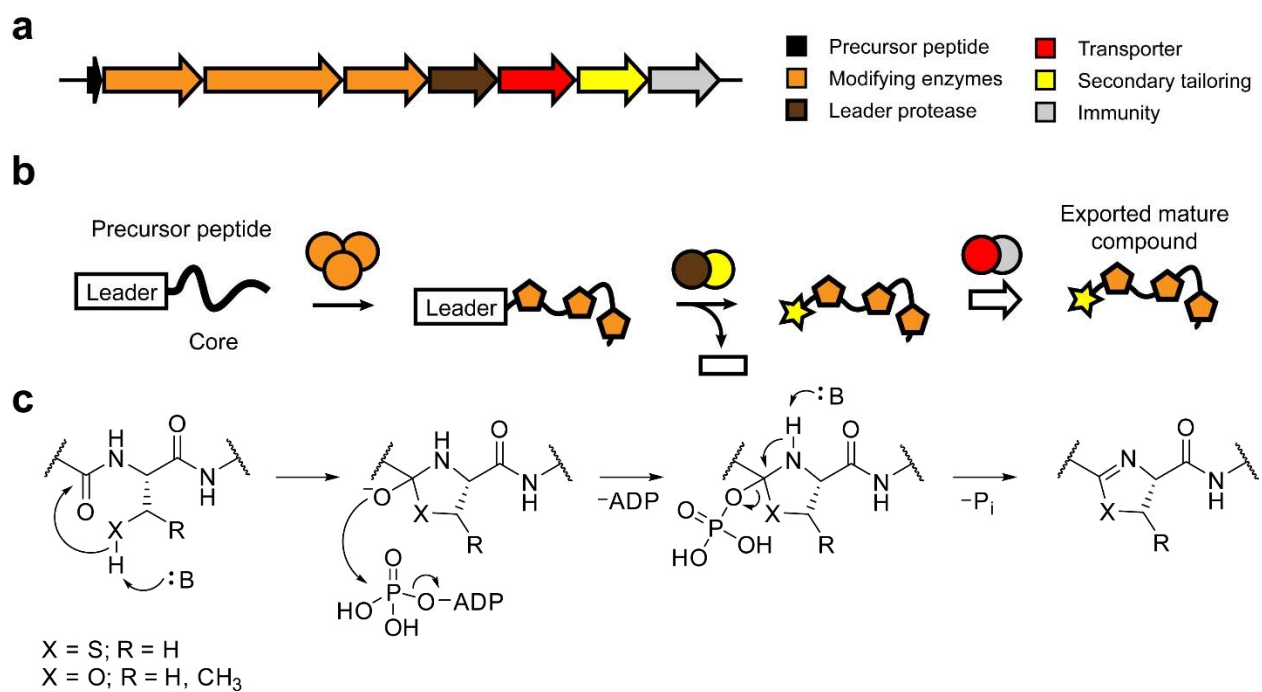


Figure 1.1 (a) Generic RiPP pathway. (b) Overview of common steps in maturation of RiPP natural products. (c) Example of a posttranslational modification within a RiPP biosynthetic pathway. Here, a peptidic azoline is generated by the ATP-dependent cyclodehydration of a Cys, Ser, or Thr residue. This is the best characterized function for a YcaO protein.

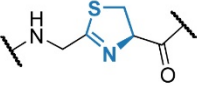
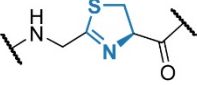
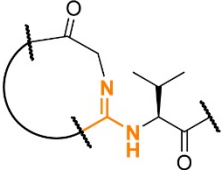
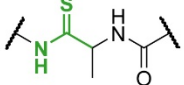
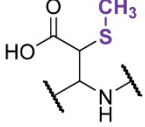
| YcaO superfamily | | | |
|---|--|---------------------|---------------------------------|
| Types | Function | Associated proteins | Representative natural products |
| <i>Azoline-forming</i> |  | E1-like | Microcin B17 |
| <i>Standalone azoline-forming</i> |  | N/A | Bottromycin |
| <i>Amidine-forming*</i> |  | ? | Bottromycin |
| <i>Thioamide-forming*</i> (<i>TfuA-associated</i>) |  | TfuA | Thioviridamide |
| <i>Enigmatic</i> |  | RimO | Ribosomal protein S12 |

Figure 1.2. Overview of YcaO superfamily. Asterisk denotes putative functions.

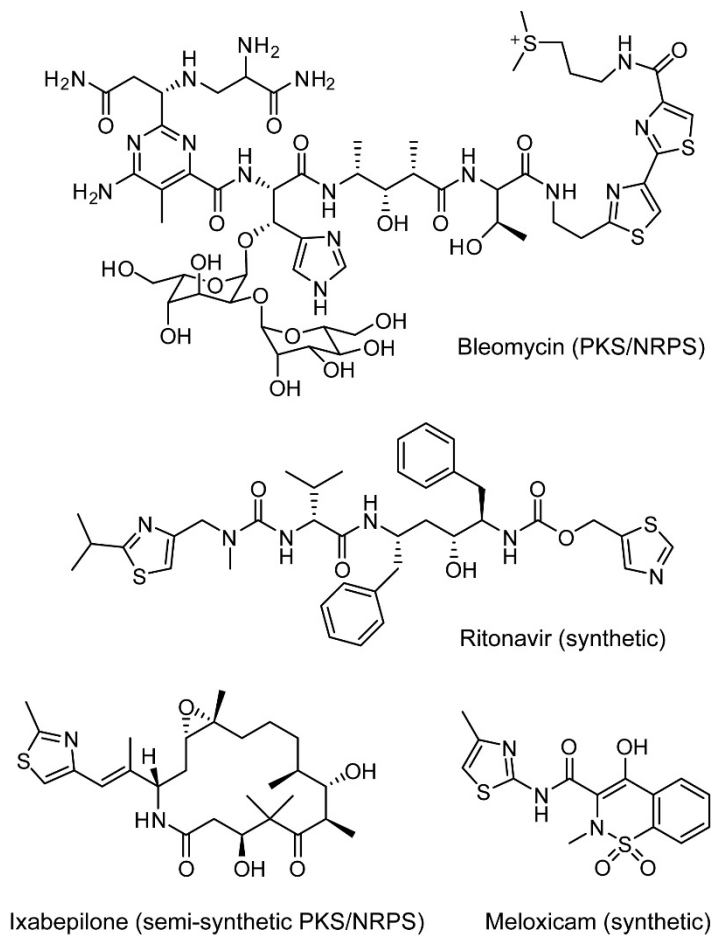


Figure 1.3 Representative thiazole-containing compounds that are of non-ribosomal or synthetic origin.

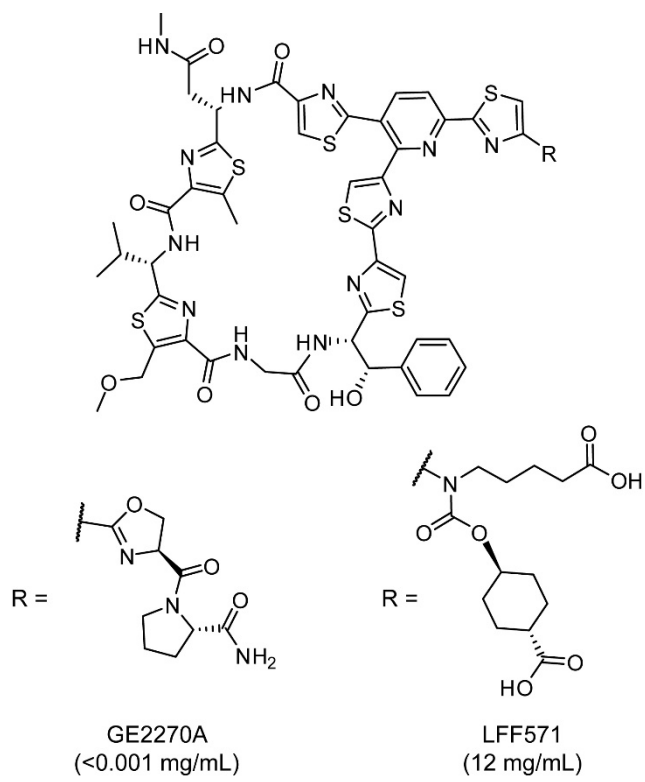


Figure 1.4 Derivatization of the thiopeptide GE2270A enhances water solubility (in parentheses).

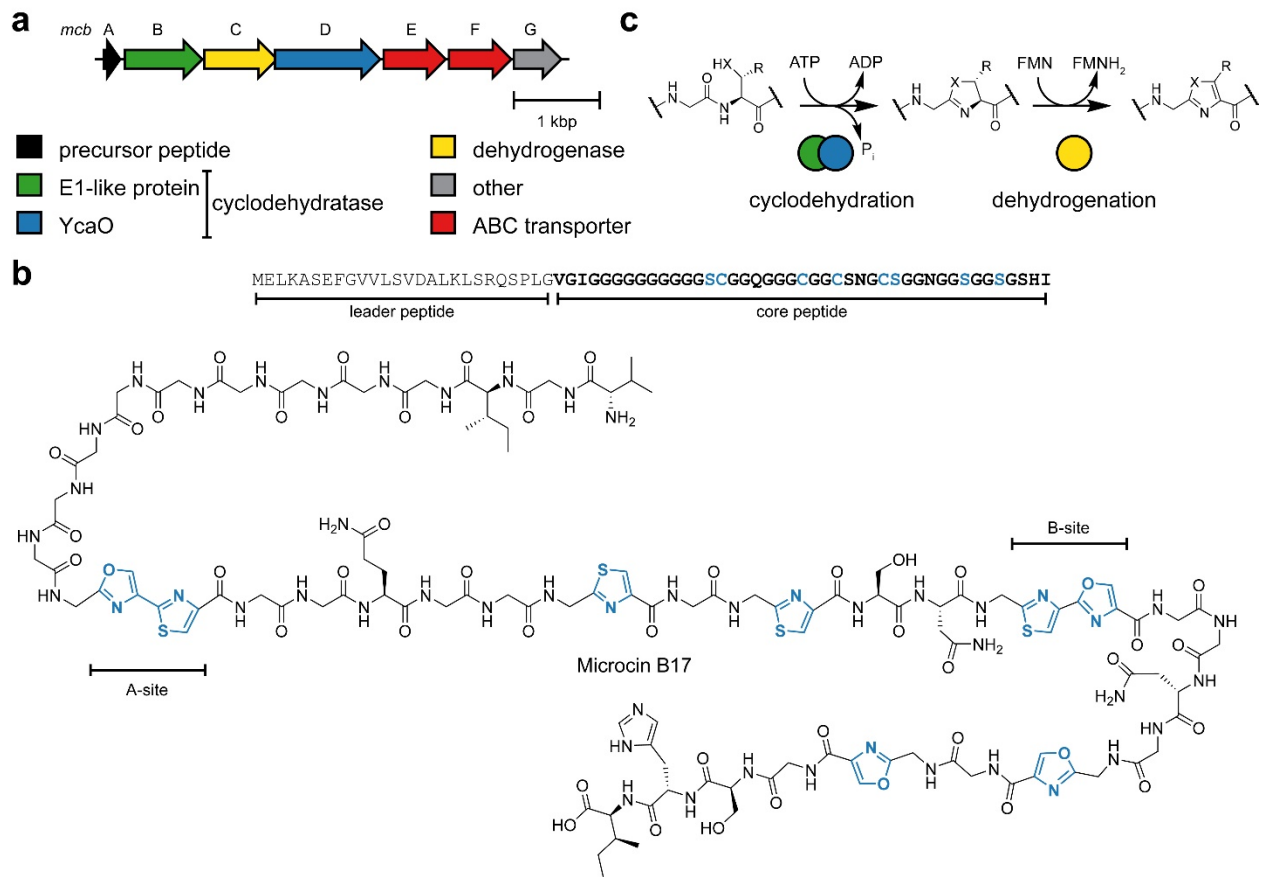


Figure 1.5 Microcin B17 biosynthesis and structure. (a) BGC for MccB17 (b) Precursor peptide sequence of MccB17 and structure of MccB17. The two bis-heterocyclic sites are known as the A- and B-sites. (c) Biosynthetic scheme for the cyclodehydratase/dehydrogenase.

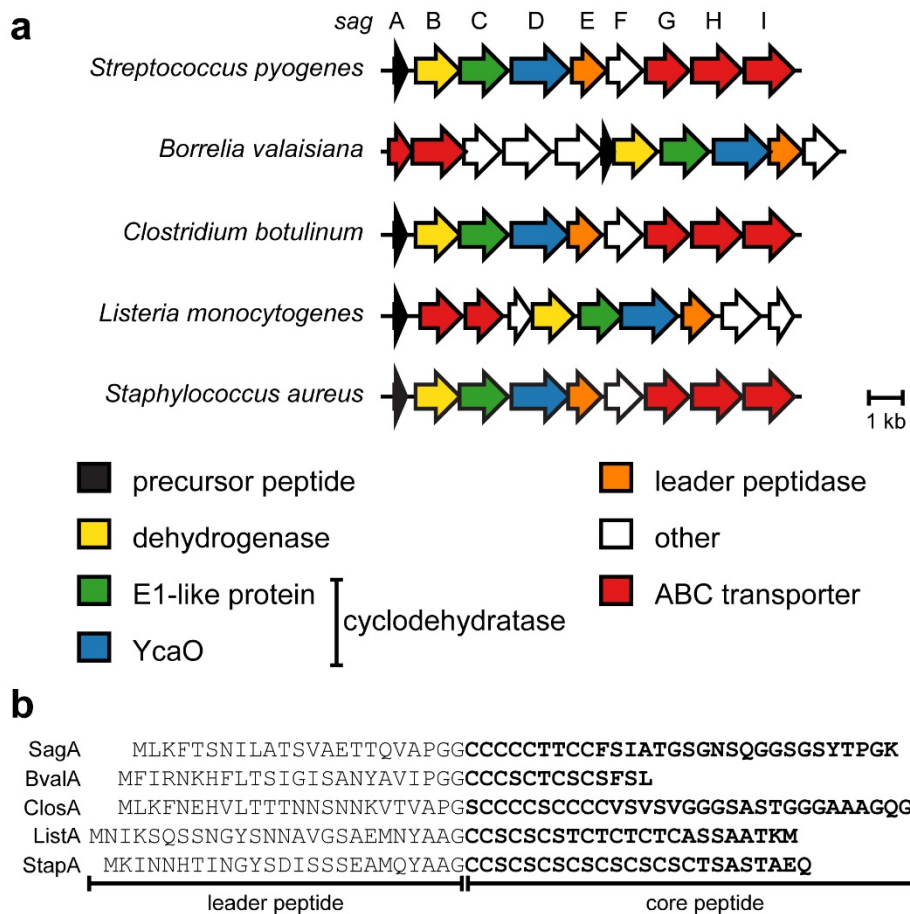


Figure 1.6 Streptolysin S biosynthetic components. (a) Representative BGCs for members of the SLS-like cytolysin family. (b) Precursor peptide sequences. The leader peptide cleavage site is unknown but suspected to be directly N-terminal to the NPH region in the core region. The final structure for any cytolysin remains unknown.

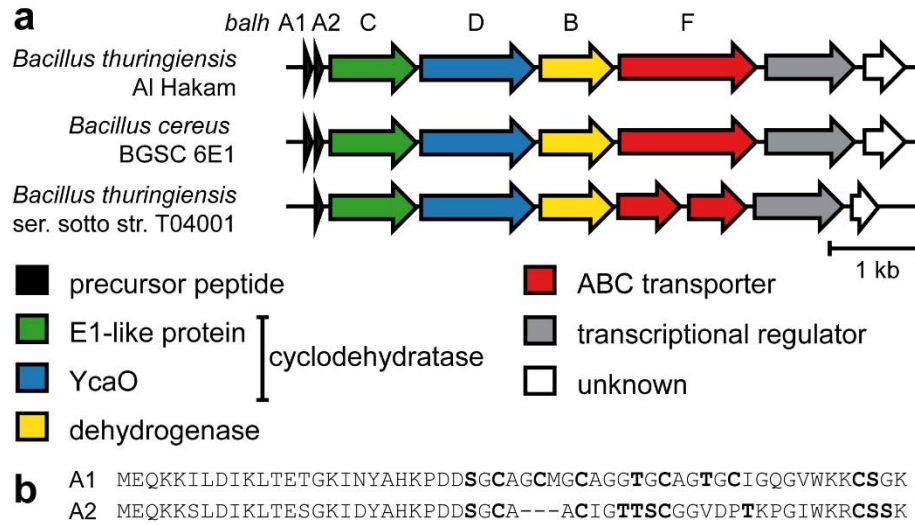


Figure 1.9 Hakacin biosynthetic components. (a) Representative hakacin BGCs (b) Precursor peptide sequences for BalhA1 and BalhA2 with emphasized heterocyclizable residues. The final structure for hakacin is also unknown.

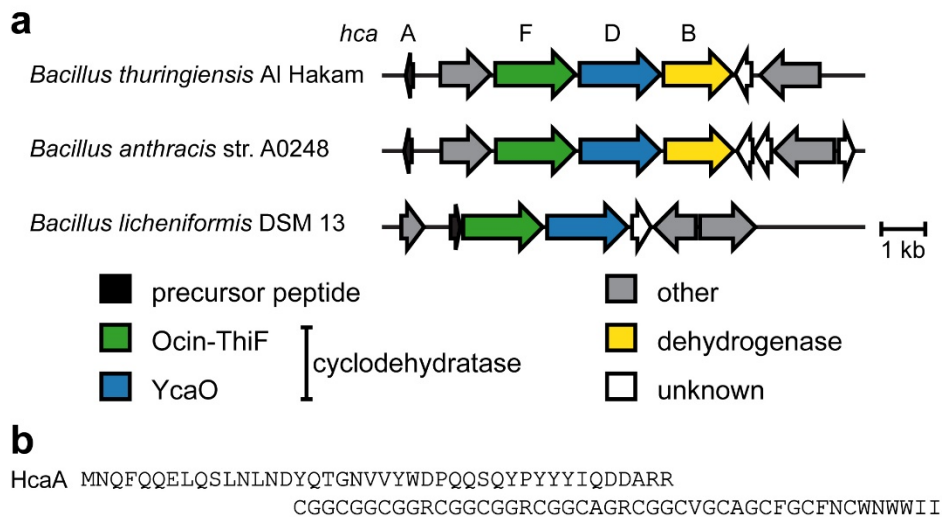


Figure 1.10 HCA biosynthetic components. (a) Heterocycloanthracin (HCA) gene clusters from related *Bacillus* organisms. (b) Sequence of HcaA from *B. sp.* Al Hakam.

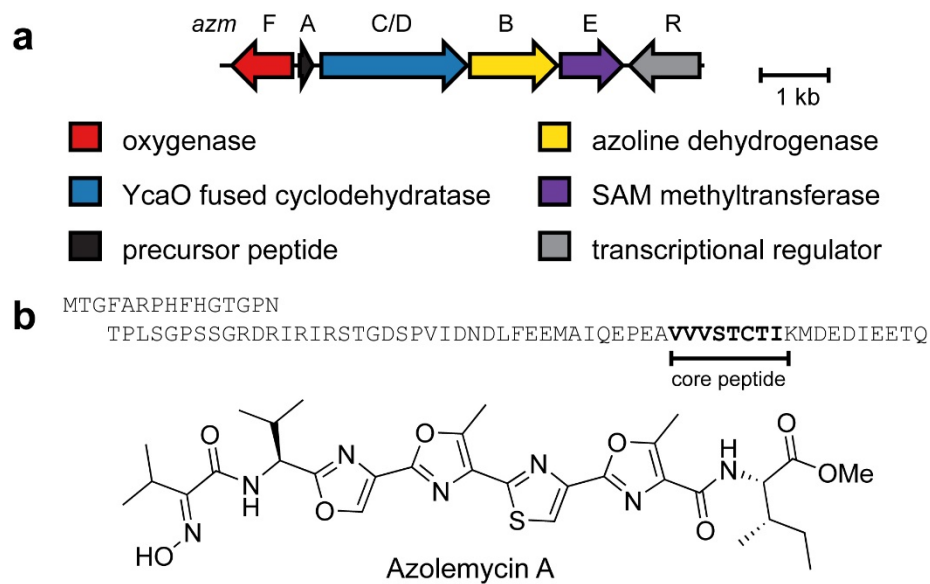


Figure 1.11 Azolemycin biosynthesis and structure. (a) Azolemycin BGC. (b) Precursor peptide and structure for azolemycin A.

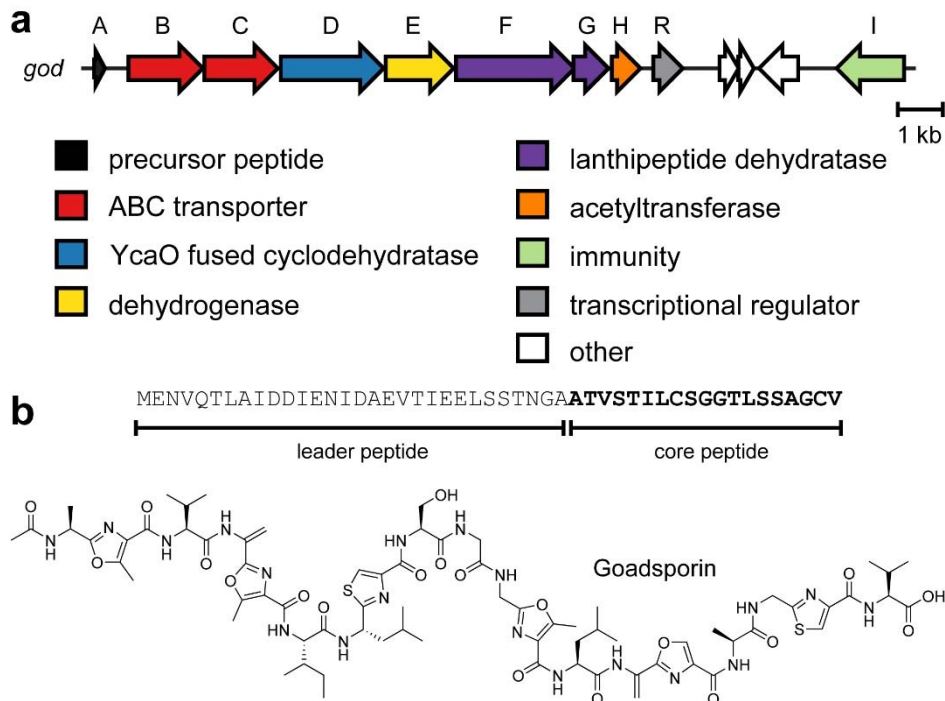


Figure 1.12 Goadsporin biosynthesis and structure. (a) The goadsporin BGC. (b) Precursor peptide and structure for goadsporin.

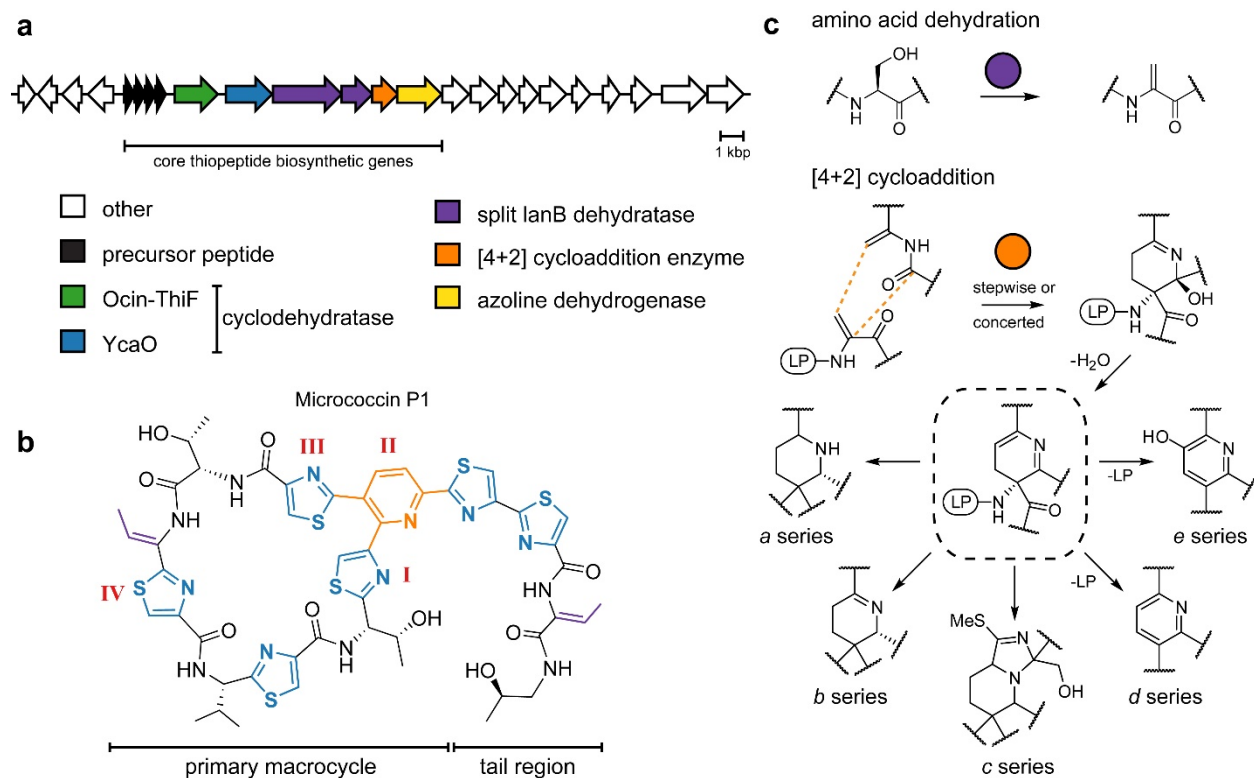


Figure 1.13 Overview of thiopeptide biosynthesis and macrocycle architecture. (a) Thiopeptide BGC highlighting the enzymes that form the minimal thiopeptide scaffold. (b) Structure of micrococcin P1 with hallmark modifications color-coded: YcaO-catalyzed cyclodehydration (blue), lanthipeptide-like dehydrations (purple) and [4+2] cycloaddition (orange). The four-ring motif important for TipA recognition is also highlighted (red). (c) Biosynthesis of the 6-membered nitrogenous heterocycle and corresponding series designations.

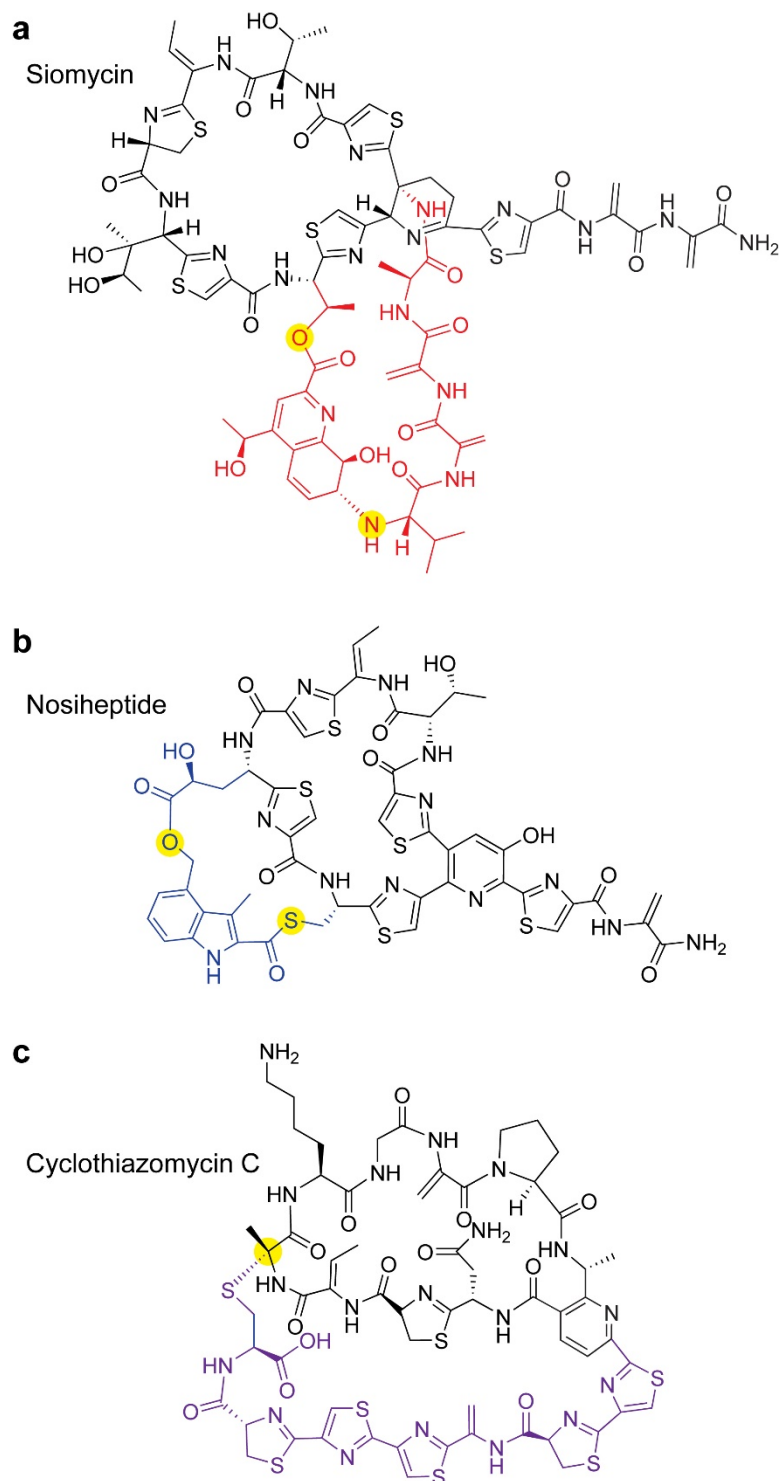


Figure 1.14 Thiopeptide secondary macrocycles. (a) Quinaldic acid derived ring in siomycin (red). (b) Indolic acid derived ring in nosiheptide (blue). (c) Thioether crosslink formed ring in cyclothiazomycin (purple). Attachment points to the main scaffold are highlighted (yellow).

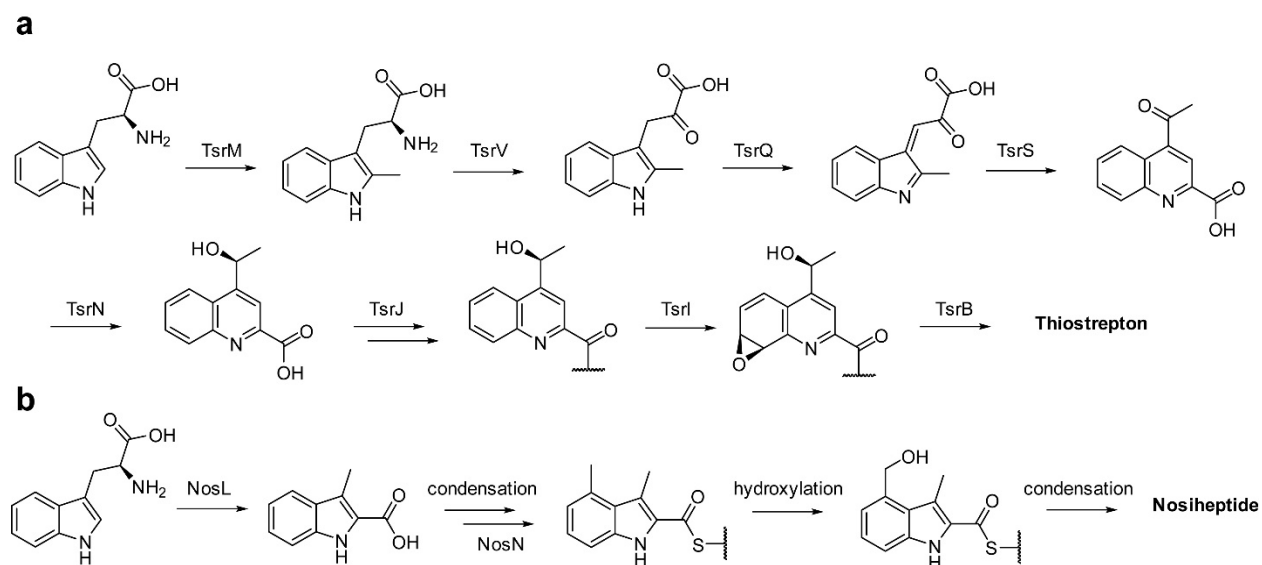


Figure 1.15 Formation of quinaldic (a) and indolic (b) acid moieties.

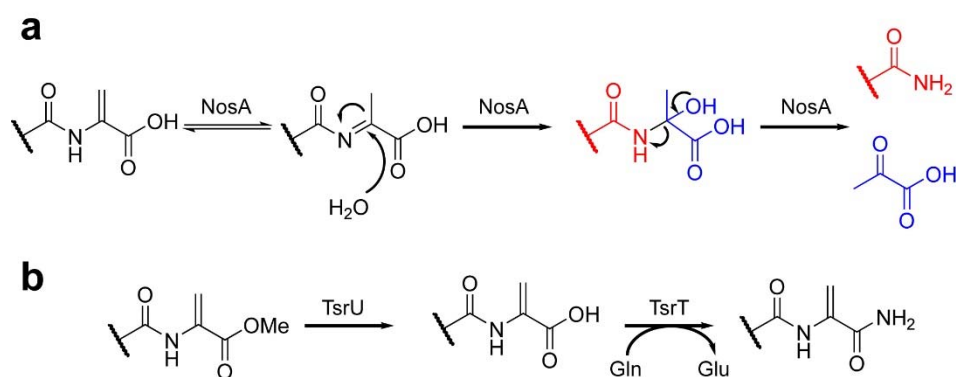


Figure 1.16 Two mechanisms of thiopeptide C-terminal amidation illustrated from nosiheptide (a) and thioestrepton (b) biosynthesis.

| Thiopeptide ^a | Core peptide ^b | Ring size ^c | Series | Secondary ring ^d |
|--|---------------------------|------------------------|--------|-----------------------------|
| Siomycin ^e | VSSASCTTCICTCSCSS----- | 26 | b | quinaldic acid |
| Thiopeptin | VASASCTTCICTCSCSS----- | 26 | a/b | quinaldic acid |
| Sch 18640 | IASASCTTCICTCSCSS----- | 26 | a | quinaldic acid |
| Thiostrepton ^e | IASASCTTCICTCSCSS----- | 26 | b | quinaldic acid |
| Sch 40832 | TSSSSCTTCICTCSSSS----- | 26 | c | quinaldic acid |
| Glycothiohexide | ----SCTTCECCSCS----- | 26 | e | indolic acid |
| Philipimycin | ----SCTTCECCSCSS----- | 26 | e | indolic acid |
| Nosiheptide ^e | ----SCTTCECCSCSS----- | 26 | e | indolic acid |
| Nocathiacin ^e | ----SCTTCECSCSCSS----- | 26 | e | indolic acid |
| Thiazomycin | ----SCTTCECSCSCSS----- | 26 | e | indolic acid |
| S-54832 | ----SCTTCECSCSCSS----- | 26 | e | indolic acid |
| Nocardithiocin ^e | ----SCTSCVVICSCCT----- | 26 | d | |
| Micrococcin ^e | ----SCTTCVCTCSCCTT----- | 26 | d | |
| YM-266183 ^e | ----SCTTCVCTCSCCTT----- | 26 | d | |
| QN3323 | ----SCTTCVCTCSCCTT----- | 26 | d | |
| Thiocillin ^e | ----SCTTCVCTCSCCTT----- | 26 | d | |
| Lactocillin ^e | ----SCTTCTCCSCCA----- | 26 | d | |
| Promothiocin | ----SCVGTACATSSSS----- | 26 | d | |
| JBIR-83 | ----SCVATACATSSS----- | 26 | d | |
| GE2270 ^e | ----SCNCVCGFCCSCSPA--- | 29 | d | |
| Amythiamicin | ----SCNCVCGVCCSCSPA--- | 29 | d | |
| Baringolin/Kocurin | ----STNCFYCPCSCSAPSSS- | 29 | d | |
| GE37468 ^e | ----STNCFYICCCSCSSN---- | 29 | d | |
| Thiomuracin ^e | ----SCNCFYICCCSCSSA--- | 29 | d | |
| Cyclothiazomycin A ^e | ----SNCTSTGTPASCCSCCC- | 29 | d | thioether |
| Cyclothiazomycin B ^e | ----SNCTSRGTPASCCSCCC- | 29 | d | thioether |
| Cyclothiazomycin C ^e | ----SNCTSKGSPASCCSCCC- | 29 | d | thioether |
| Lactazole ^e | ----SWGSCSQASSSCAQPQDM | 32 | d | |
| Berninamycin ^e | ----SCTTTSVSTSSSSSSS--- | 35 | d | |
| Geninthiocin | ----SCTTSSVSTSSSSSSS--- | 35 | d | |
| Sulfomycin | ----SCSTTGCTTSSSSSSS--- | 35 | d | |
| Thiotipin | ----SCTTIGCTTSSSSSSS--- | 35 | d | |
| Methylsulfomycin | ----SCTTIGCTTSSSTSSSSS--- | 35 | d | |
| Promoinducin | ----SCTTIGCTTSSSSSSS--- | 35 | d | |
| TP-1161 ^e | ----SCTTIGCACSSSSSST--- | 35 | d | |
| A10255 | ----SCTTSGCACSSSSSSS--- | 35 | d | |
| Thioactin | ----SCTCSGCACSSSSSSS--- | 35 | d | |
| Thioxamycin | ----SCTCSGCACSSSSSSS--- | 35 | d | |
| Radamycin | ----SCVGTACACSSSTSSSSS--- | 35 | d | |

Figure 1.17 Sequence comparison of the core regions for all known thiopeptides. Thiopeptides with the same precursor peptide vary in ancillary tailoring to give different compounds. Groups of related thiopeptides are separated based on azol(in)e position (gray shading), ring size, and secondary ring. Group representatives are in bold font. Azol(in)e, blue; dehydroamino acids, orange; piperidine/pyridine, green; secondary ring, magenta. ^aName of the most common congener or parent compound. ^bCore sequences are from BGCs or inferred from the structure of the mature product. ^cThe number of atoms in the primary macrocycle. ^dThe moiety that serves to link the secondary side ring to the primary macrocycle. ^eKnown BGC.

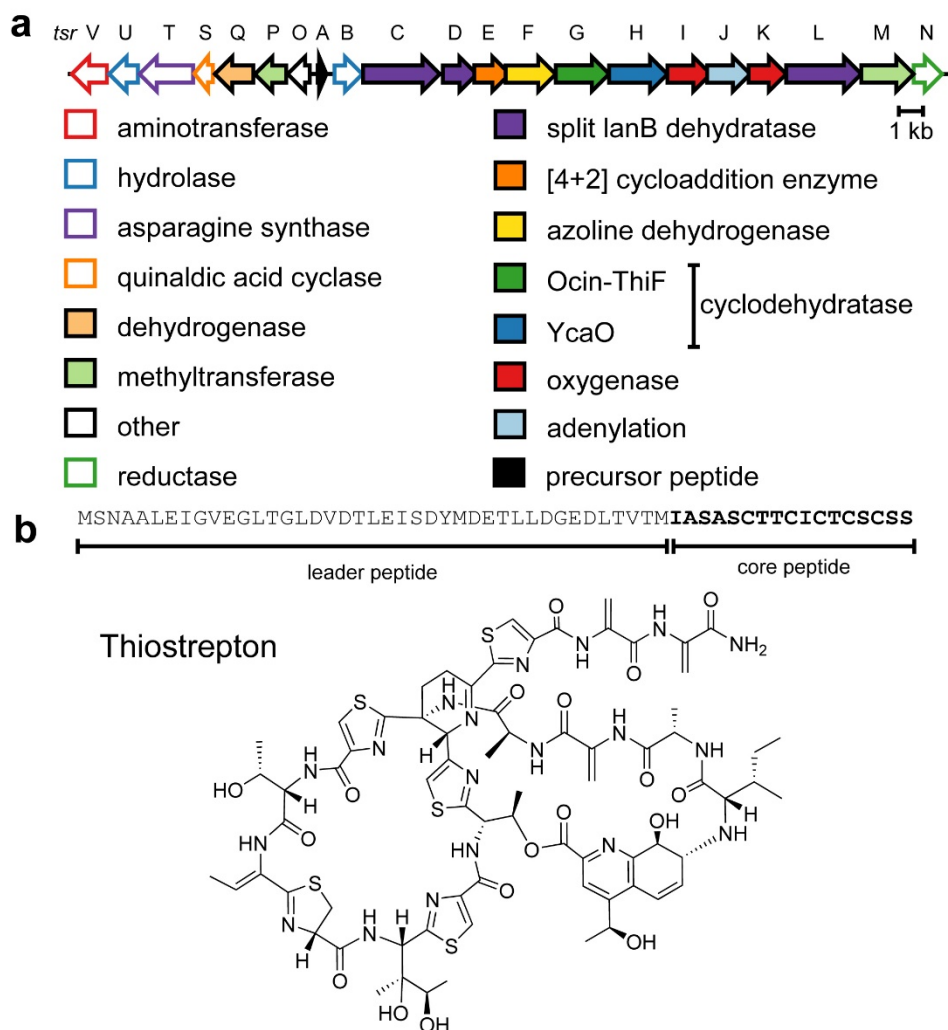


Figure 1.18 ThioStrepton biosynthesis and structure. (a) ThioStrepton BGC. (b) Precursor peptide and structure for thioStrepton A.

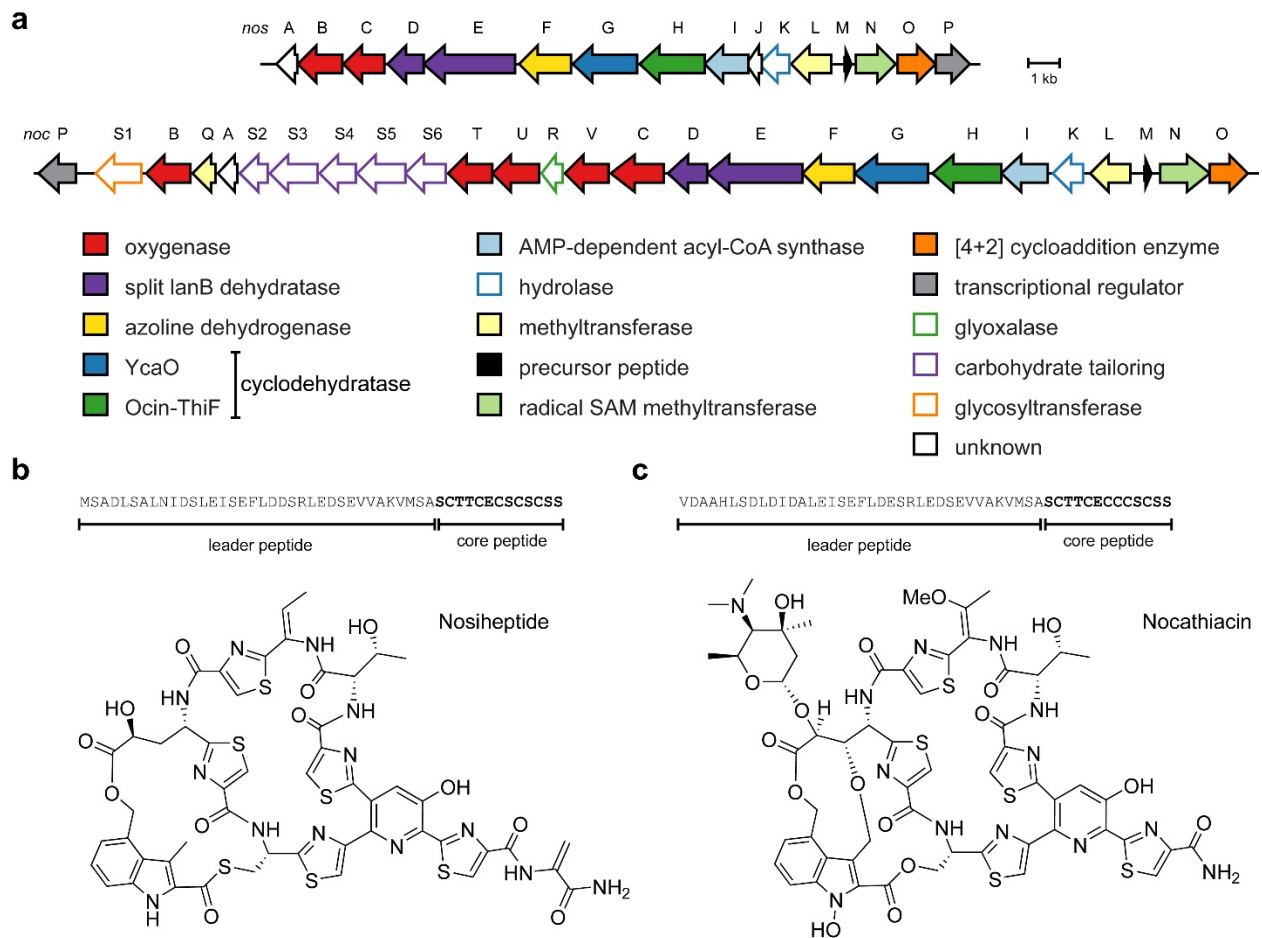


Figure 1.19 Nosiheptide and nocathiacin biosynthesis and structure. (a) BGCs for nosiheptide and nocathiacin. (b) Precursor peptide and structure of nosiheptide with characteristic, thioester-linked indole macrocycle. (c) Precursor peptide and structure of nocathiacin with characteristic glycone and additional oxidations when compared to nosiheptide.

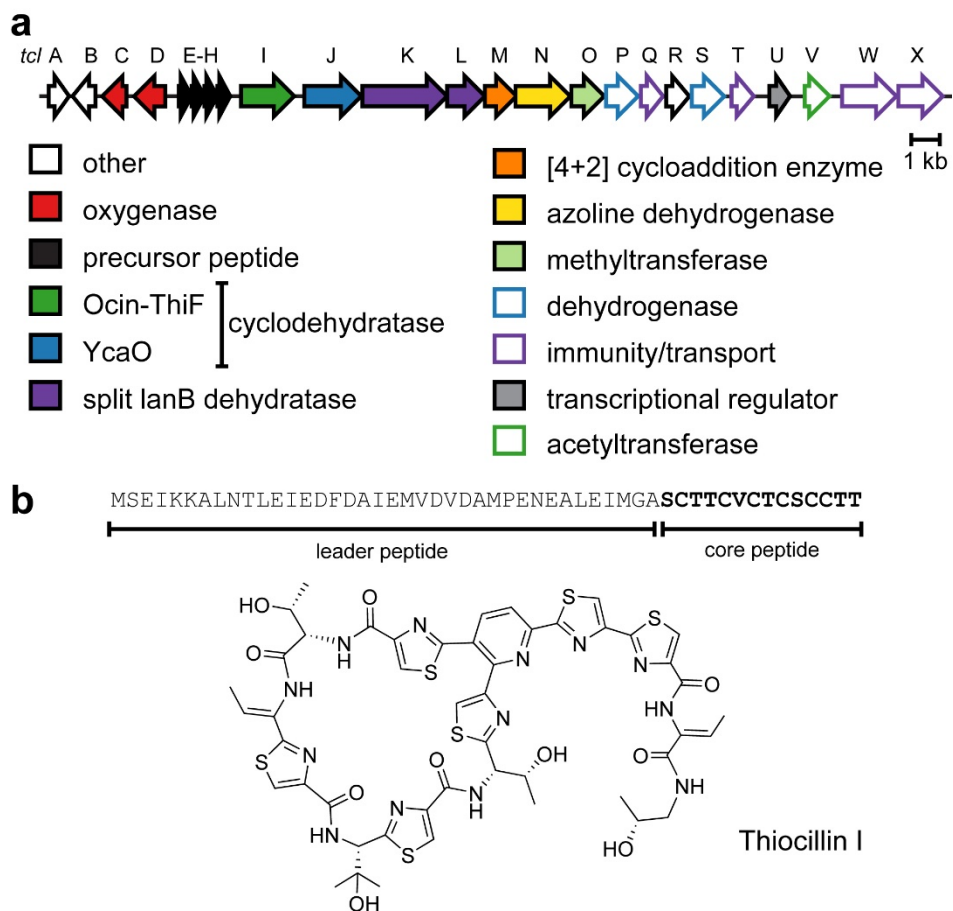


Figure 1.20 Thiocillin biosynthesis and structure. (a) BGC of thiocillin. (b) Precursor peptide and chemical structure of thiocillin I.

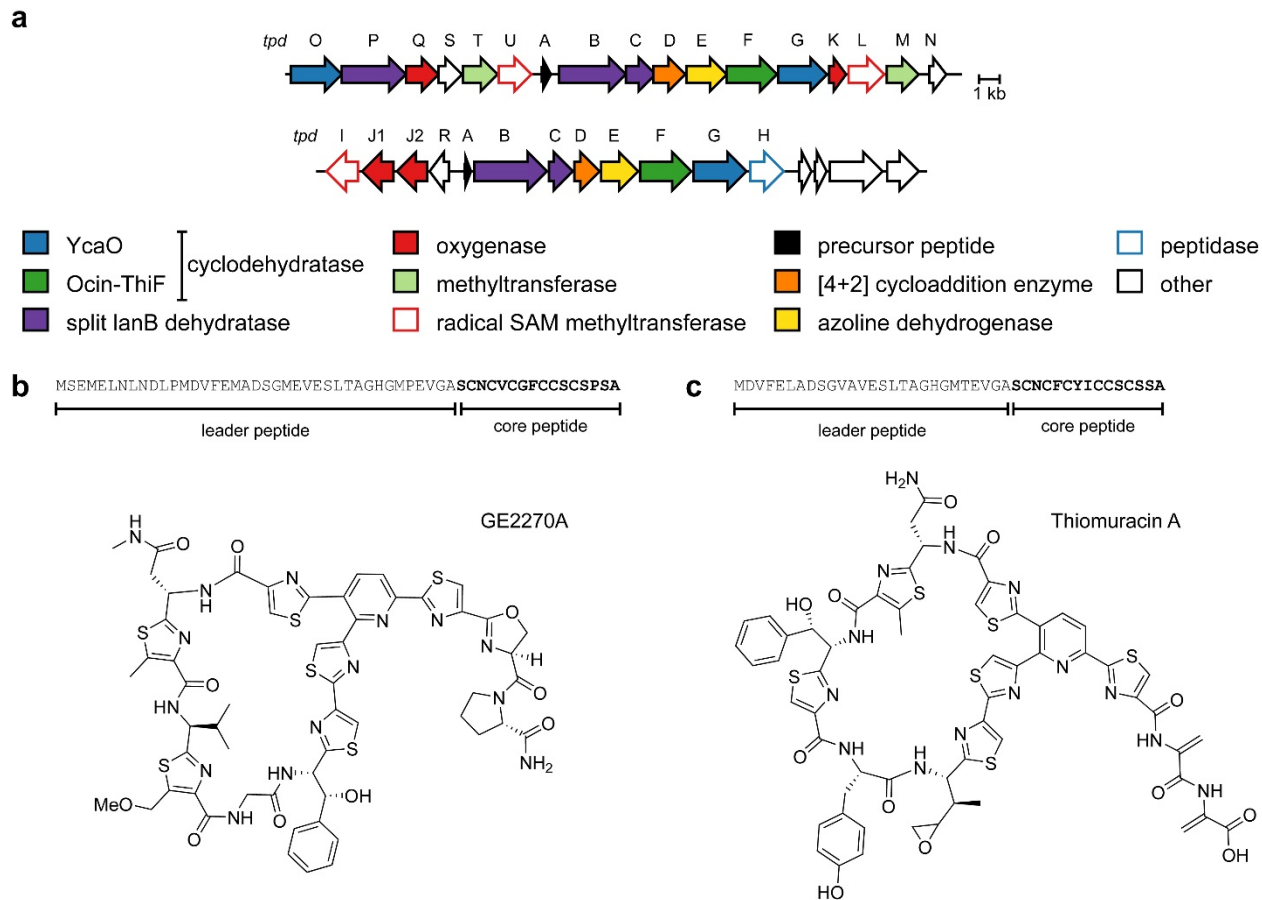


Figure 1.21 GE2270A and thiomuracin A biosynthesis and structure. (a) BGC for GE2270 and thiomuracin A showing ancillary tailoring enzymes that C-methylate thiazoles and further decorate various side chains. (b) Precursor peptide and structure of GE2270A. (c) Precursor peptide and structure of thiomuracin A.

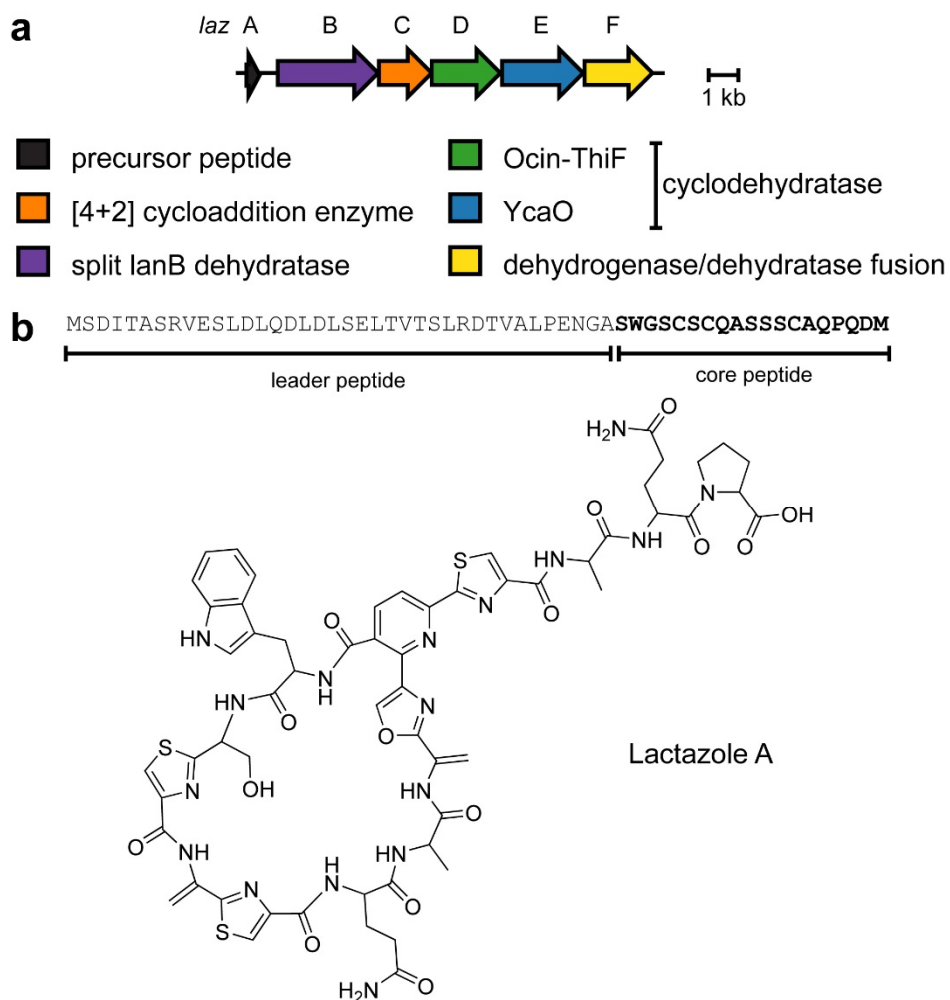


Figure 1.23 Lactazole biosynthesis and structure. (a) The lactazole BGC. (b) Precursor peptide and structure of lactazole A. Stereochemistry has not been rigorously established but each amino acid is presumed to be in natural L configuration.

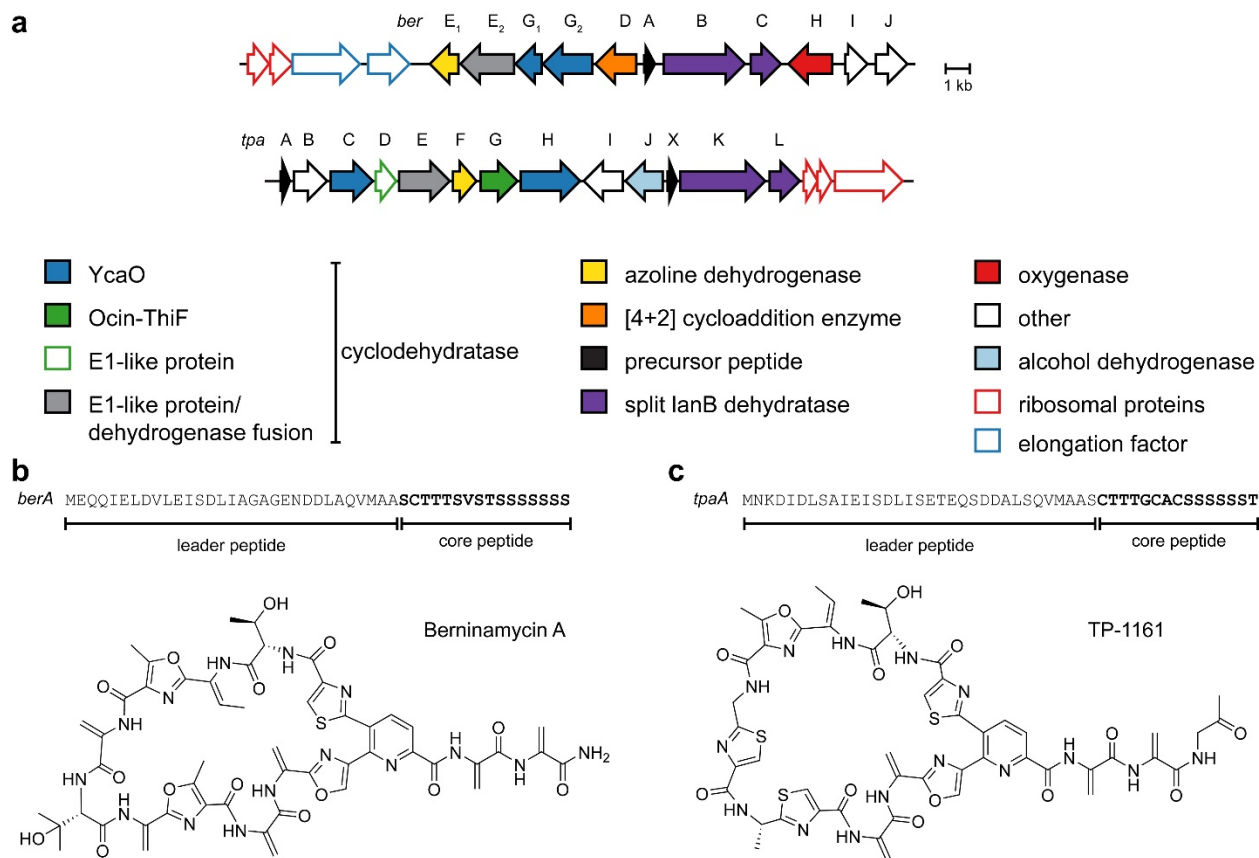


Figure 1.24 Berninamycin and TP-1161 biosynthesis and structure. (a) BGCs for berninamycin/TP-1161 (b) Precursor sequence and structure for berninamycin A. (c) Precursor sequence and structure for TP-1161.

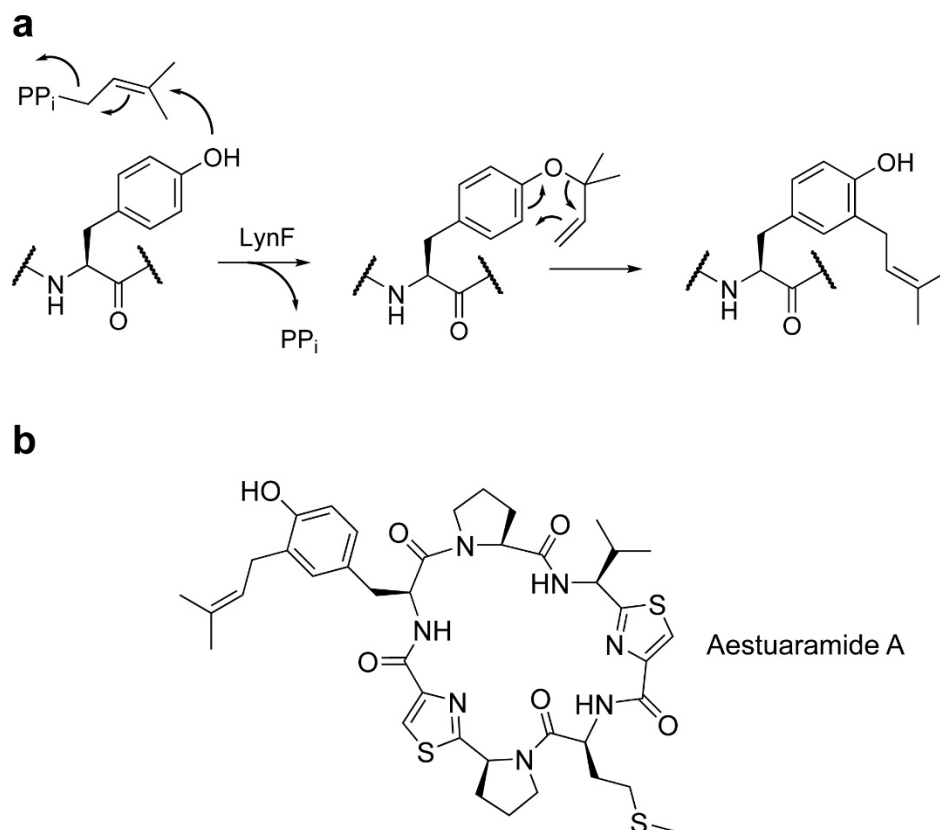


Figure 1.25 Tyrosine C-prenylation. (a) After *O*-prenylation by a LynF-homolog, a spontaneous Claisen rearrangement generates an *ortho* C-prenylated product. (b) Example of C-prenylation in the aestuaramide A.

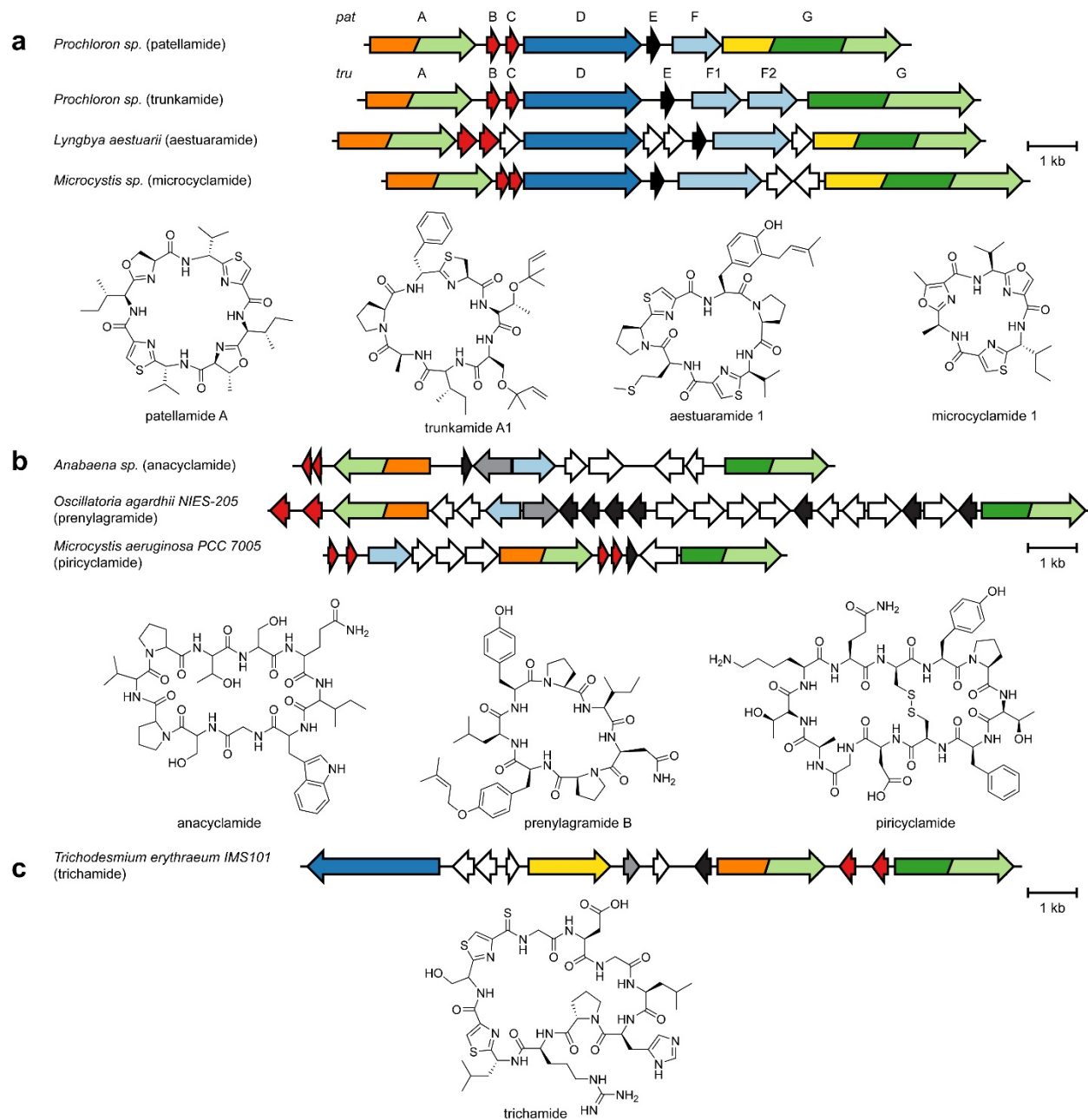


Figure 1.26 Cyanobactin genotypes I-IV and associated structures. (a) Genotype I. (b) Genotype II. The piricyclamides are likely genotype II, but this has not been strictly classified. (c) Genotype III. (Continued on next page) (d) Genotype IV.

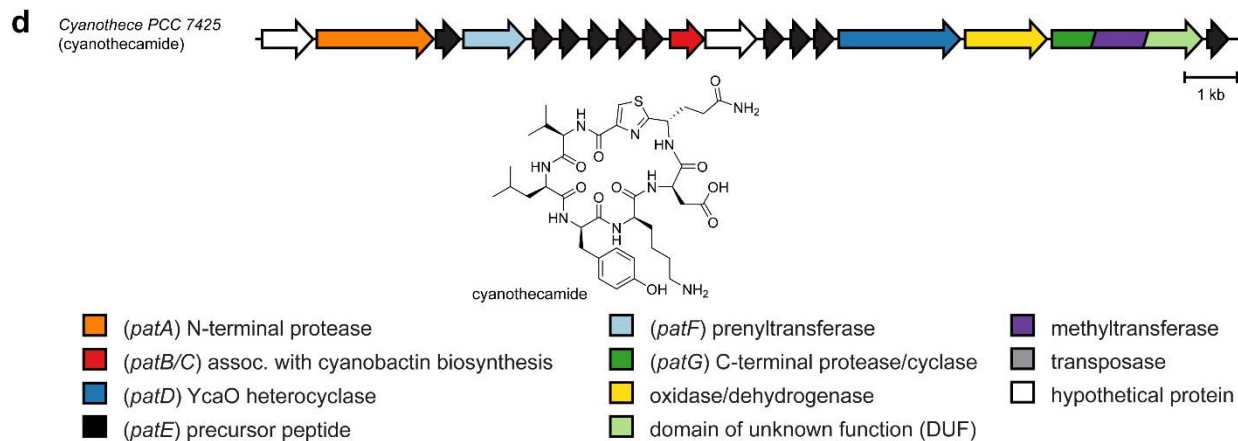


Figure 1.26 (Continued)

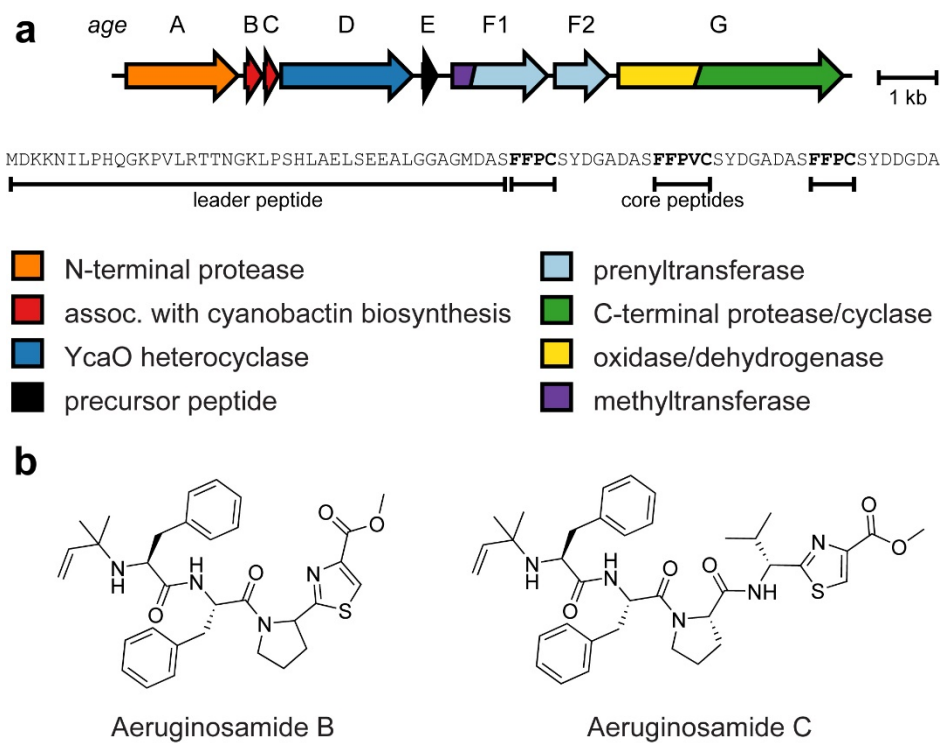


Figure 1.27 Linear cyanobactin biosynthesis. (a) BGC for the aeruginosamides. (b) Structure of aeruginosamide B and C.

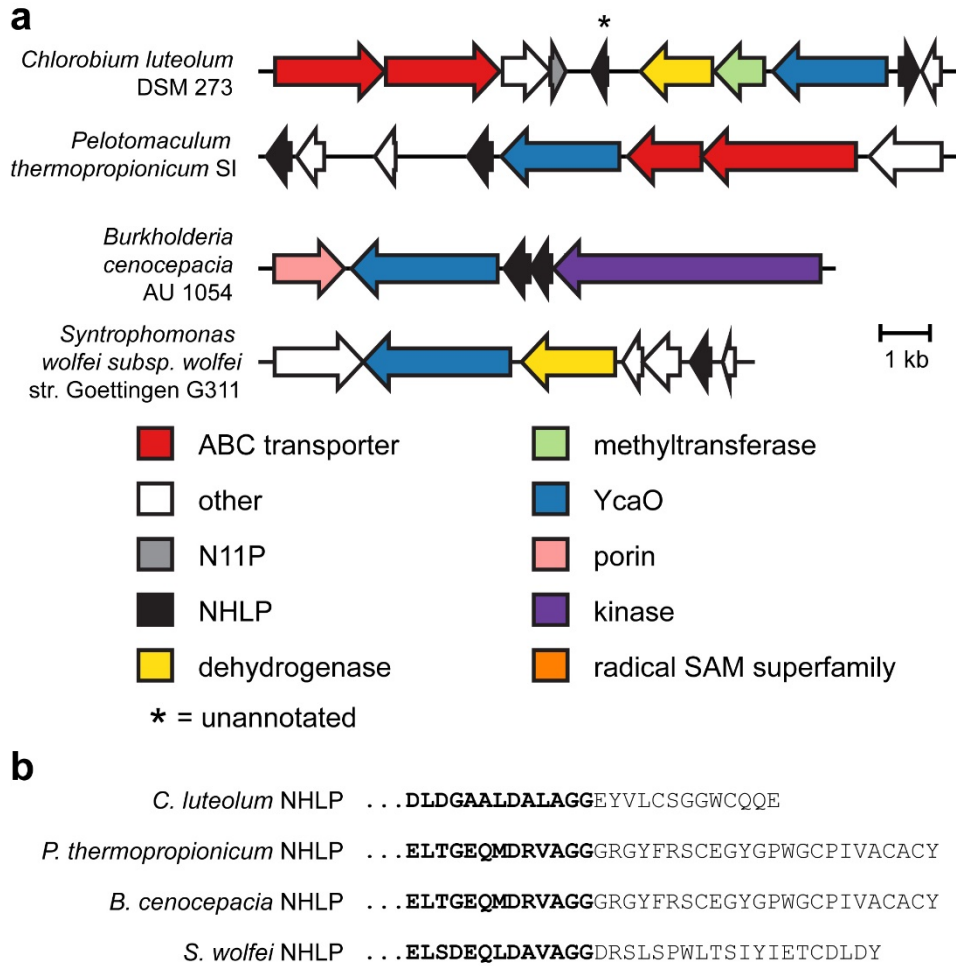


Figure 1.28 Cyclodehydratase-associated NHLP and Nif11 biosynthesis and structure. (a) Representative BGCs putatively encoding azol(in)e-containing RiPP products. (b) C-terminal portion of precursor sequences from the NHLP family, highlighting the proposed leader peptide cleavage site (after VAGG).

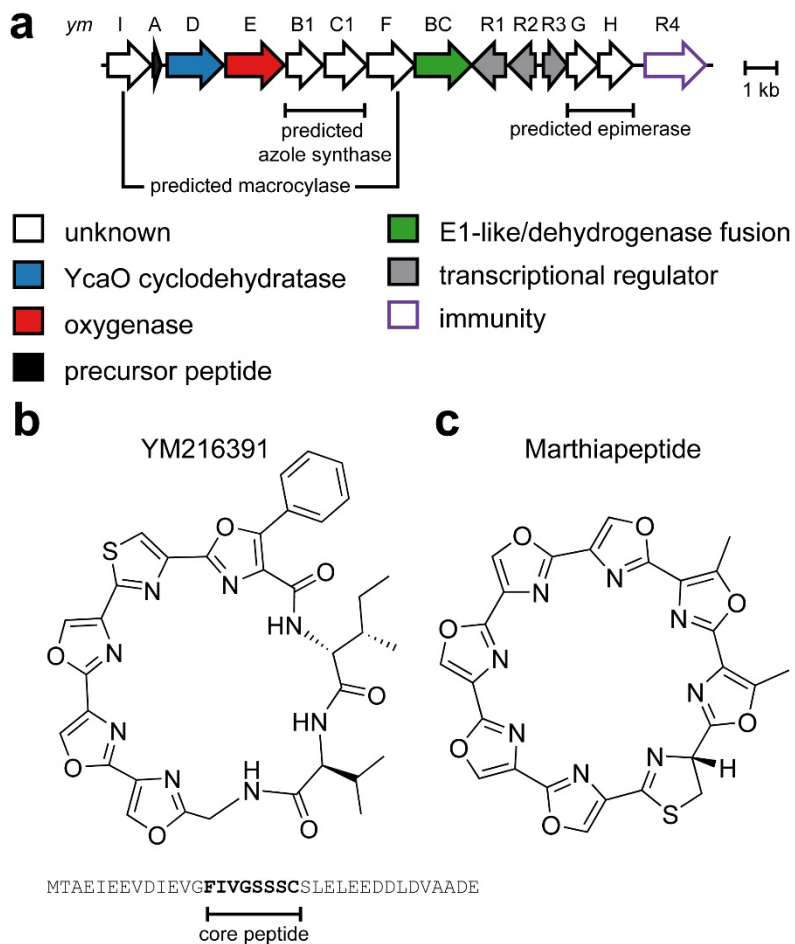






Figure 1.29 YM-216391 biosynthesis and structure. (a) The BGC for YM-216391 (b) Precursor peptide and structure for YM-216391 (c) Structure of marthiapeptide.

| Types of RiPP cyclodehydratases | | |
|---------------------------------|---|-----------------------------|
| Designation | BGC architecture | BGC distribution |
| <i>Discrete</i> |  | LAPs Thiopeptides |
| <i>Fused</i> |  | LAPs Cyanobactins |
| <i>F-dependent</i> |  | LAPs Thiopeptides |
| <i>Standalone</i> |  | Bottromycin Trifolitoxin |




 YcaO
 E1-like activating protein
 Ocni-ThiF

Figure 1.30 Overview of different types of RiPP cyclodehydratases. Until the bottromycin and trifolitoxin YcaOs are reconstituted in vitro, their status as a “standalone YcaO” remains a bioinformatic prediction.

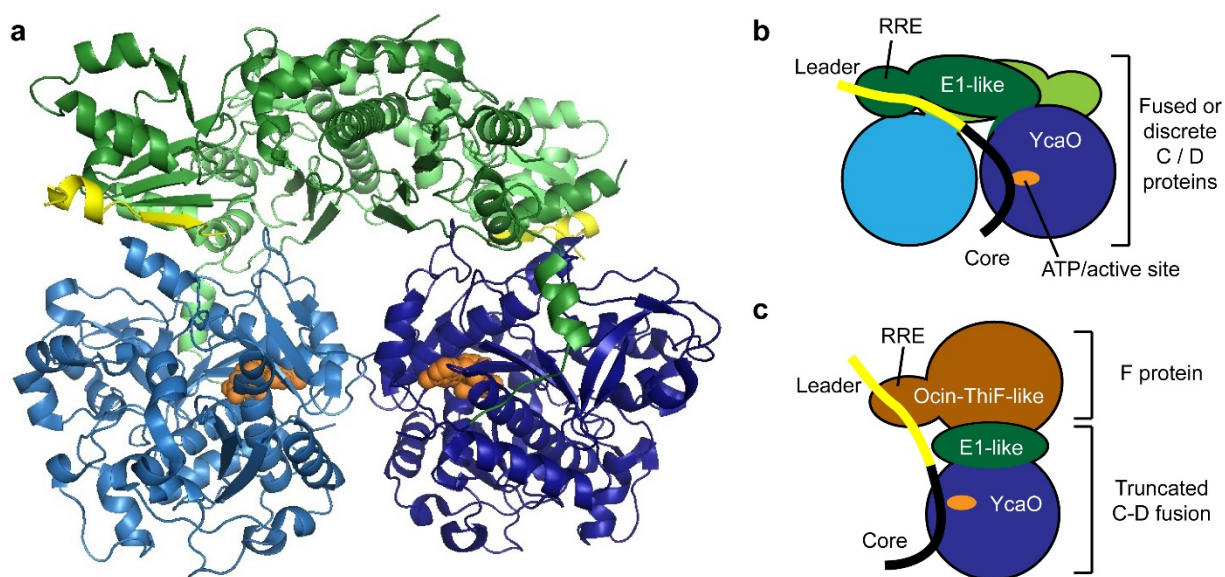


Figure 1.31 Diagram of active canonical RiPP cyclodehydratases. (a) Crystal structure of LynD (PDB ID: 4VLT). Colored regions represent the YcaO (blue), E1-like (green), Ocni-ThiF-like (orange), leader peptide (yellow), core (black) regions. Dark and light colors denote different monomers. (b) Schematic model of the dimeric cyclodehydratase shown in panel A. (c) Schematic model of the F-dependent cyclodehydratase.

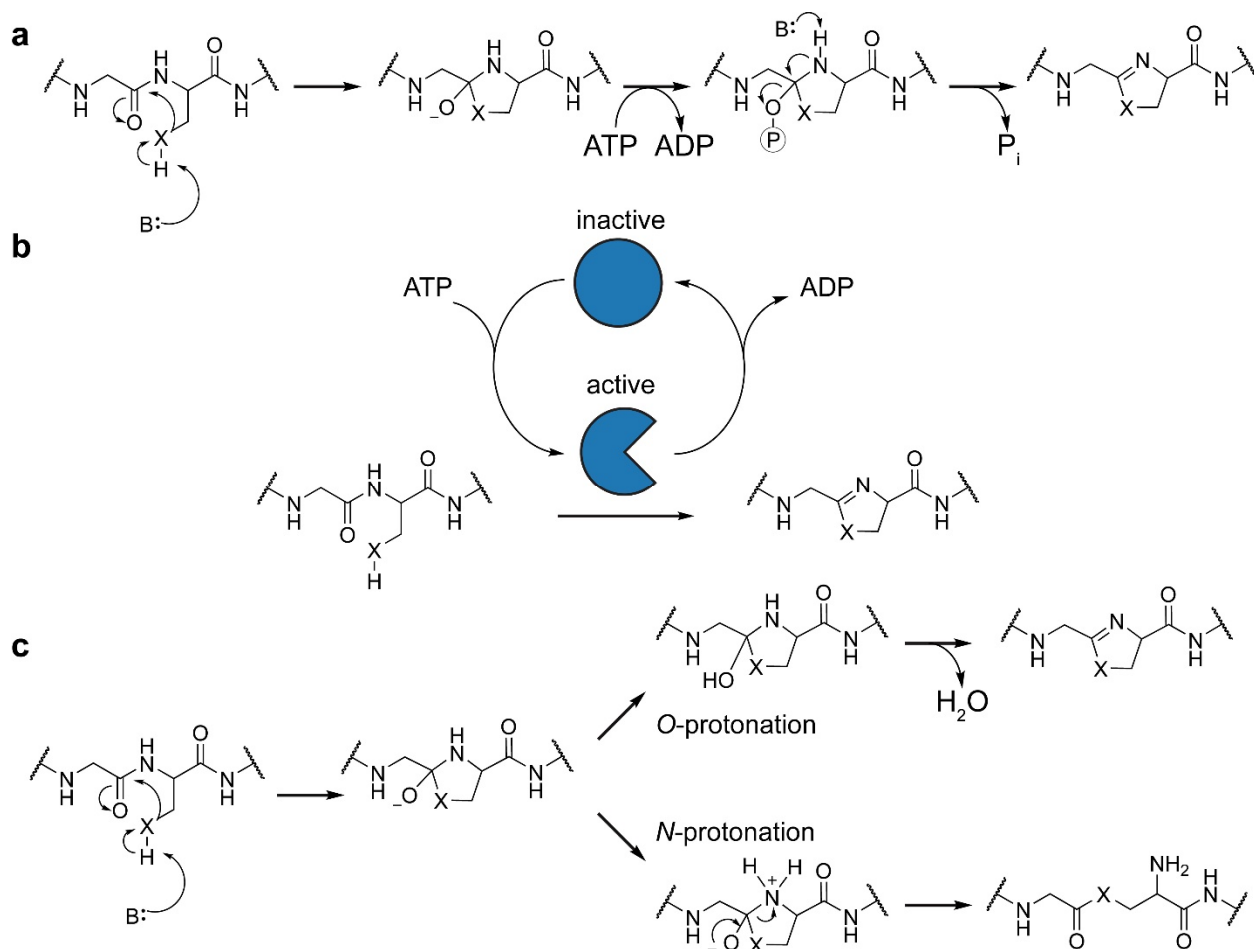


Figure 1.32 Proposed mechanisms for azoline biosynthesis by YcaO domains. (a) Direct activation suggests the incorporation of phosphate(s) directly into the heterocyclized intermediate, providing a leaving group (P_i) for elimination. (b) The molecular machine hypothesis proposes an allosteric activation role for ATP in which an active conformation of the cyclodehydratase is generated by ATP hydrolysis. (c) An intein-splicing-like mechanism, by which protonation of the hemioorthoamide intermediate allows for elimination of water to generate the azoline in a manner independent of ATP.

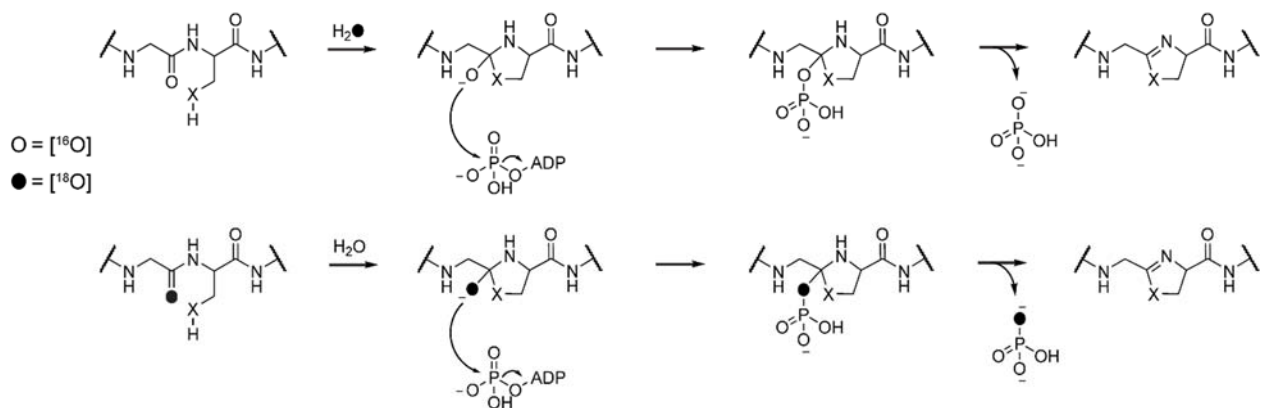


Figure 1.33 Stable isotope labeling experiments support the direct activation mechanism. ^{31}P -NMR was used to monitor the oxygen isotope status of the phosphate byproduct. Cyclodehydratase reactions in $[^{18}\text{O}]\text{-H}_2\text{O}$ (top) showed no $[^{18}\text{O}]$ enrichment in the phosphate. When $[^{18}\text{O}]\text{-BalhA}$ was used as the substrate (bottom), the ^{18}O label was incorporated into the free phosphate.

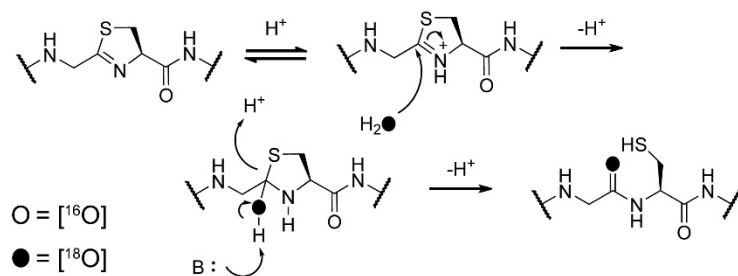


Figure 1.34 Acidic hydrolysis of thiazoline in $[^{18}\text{O}]\text{-H}_2\text{O}$ site-selectively introduces an $[^{18}\text{O}]$ label into the carbonyl of the adjacent amino acid. This is the key aspect underlying the AMPL method.

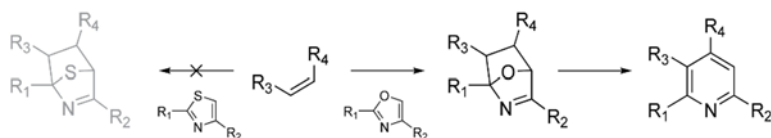


Figure 1.35 Oxazoles are used in organic synthesis in Diels-Alder [4 + 2] cycloaddition reactions. This chemistry is a feature of a conjugated di-ene, and not of an aromatic ring. Thiazoles, which exhibit greater aromatic nature than oxazoles, do not readily undergo this chemistry.

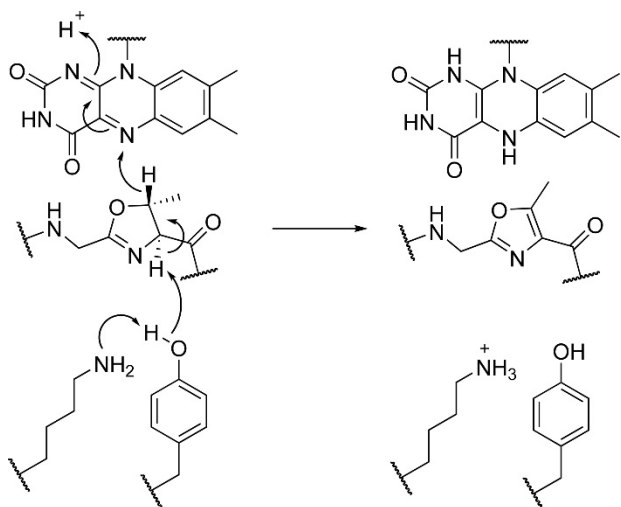


Figure 1.36 A plausible mechanism for the dehydrogenation of methyloxazoline catalyzed by B enzymes. The conserved Lys-Tyr motif is likely involved in deprotonating the C α of the azoline, which is followed by hydride transfer, in what is essentially an E2 mechanism.

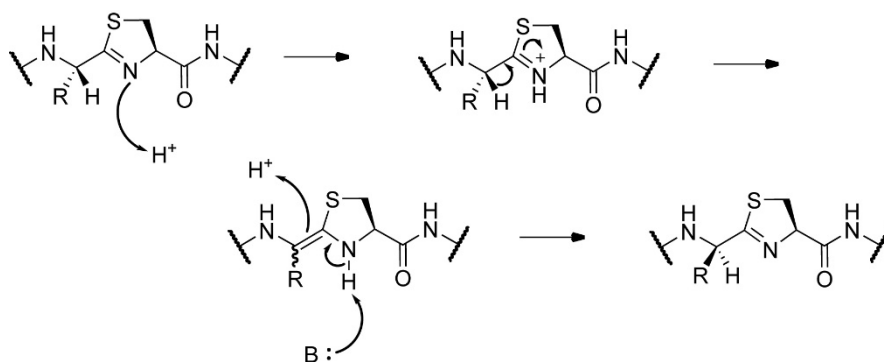


Figure 1.37 A plausible mechanism for cyanobactin epimerization.

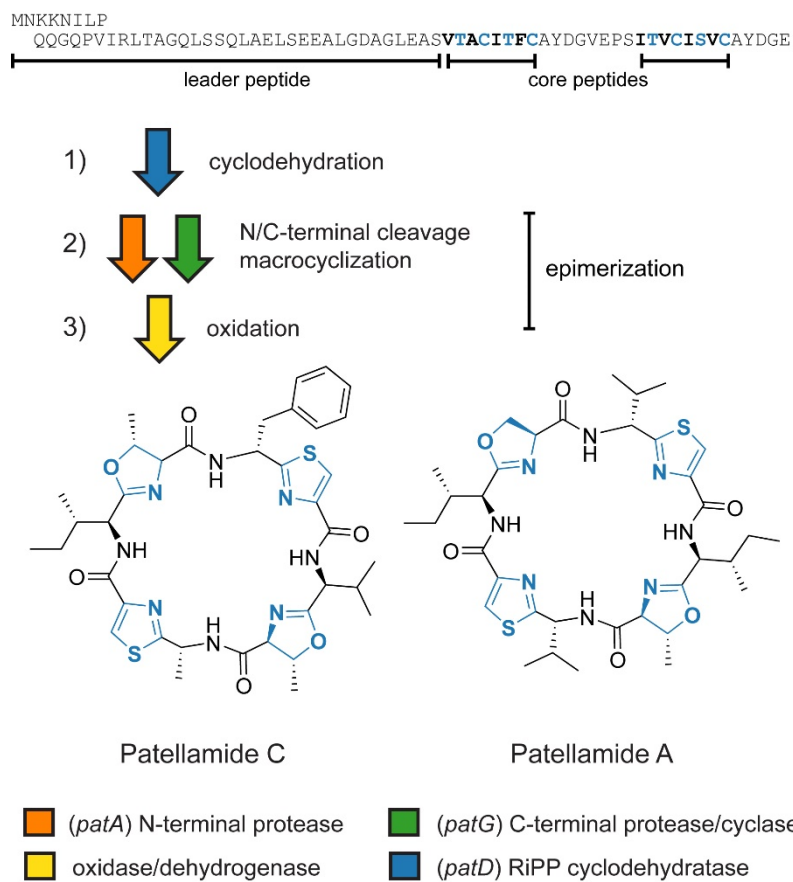


Figure 1.38 Order of biosynthetic enzymes in patellamide biosynthesis. The exact timing of epimerization is not known.

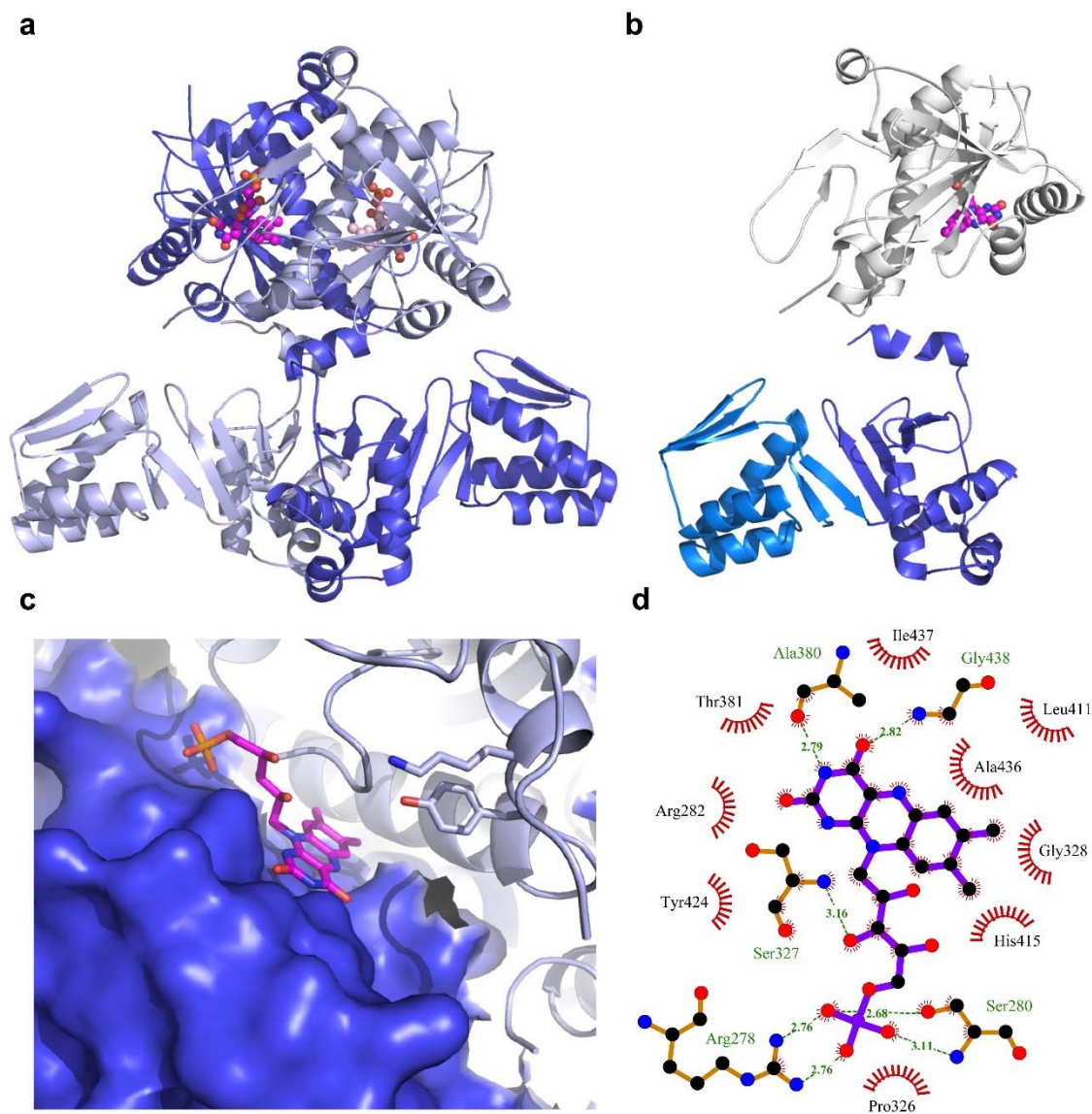


Figure 1.39 X-ray crystal structure of the dehydrogenase from cyanotheceamide biosynthesis. (a) Structure of the dimeric dehydrogenase or (b) as a single monomer. (c) The FMN molecule (magenta) is bound at the dimer interface. (d) Ligand interaction diagram of FMN binding. Putative hydrogen bonds are shown in green with distances, light red arcs indicate hydrophobic interactions.

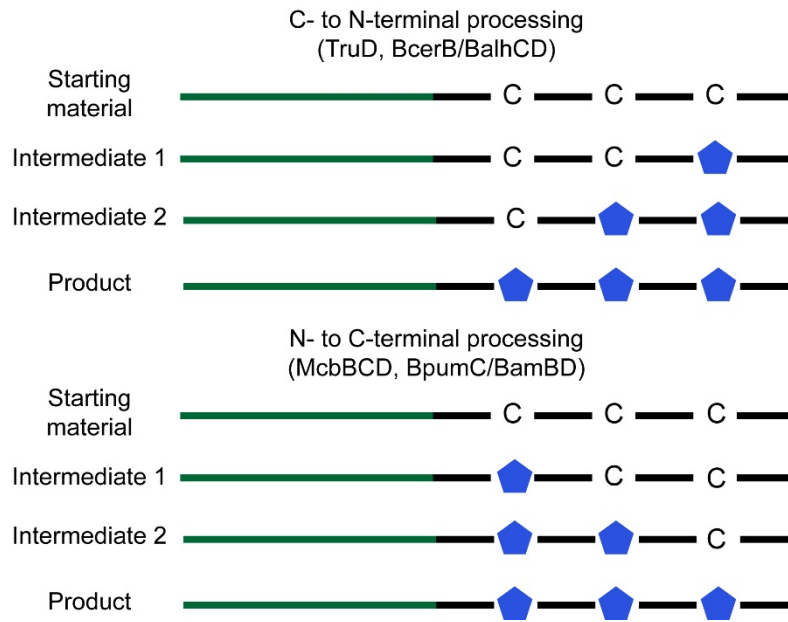


Figure 1.40 Cartoon illustrating cyclodehydratase processing of cysteine residues. The precursor peptide, composed of the leader (green) and core (peptide), can be cyclodehydrated in an ordered fashion that depends on the cyclodehydratase although the preferred order can non-linear as well.

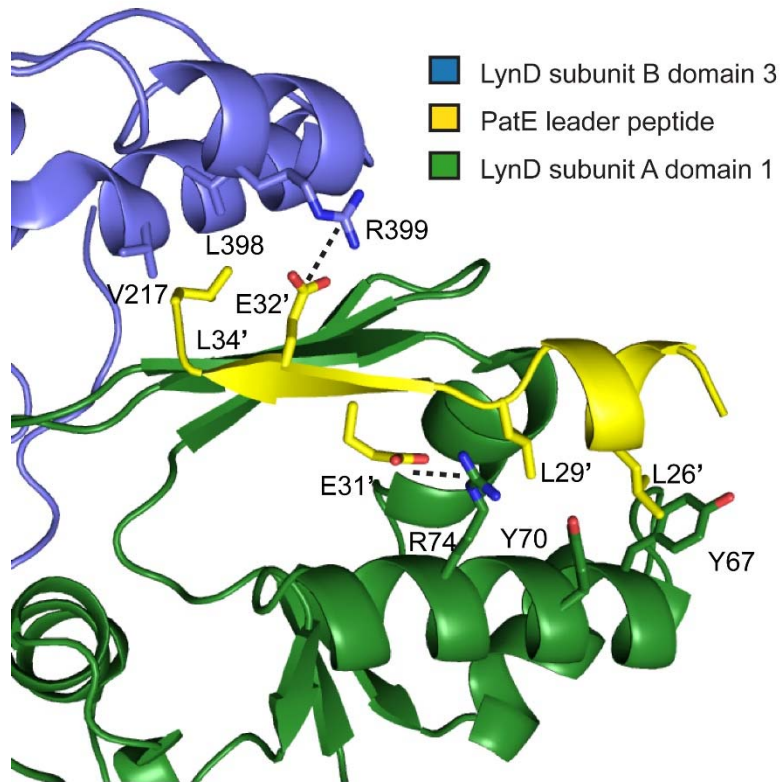


Figure 1.41 PatE leader peptide binding by LynD. Ordered residues (sticks) of the leader region (yellow) interact with the RRE (green, domain 1). PatE residues are marked with an apostrophe. Presumed salt bridges are shown with dashed lines.

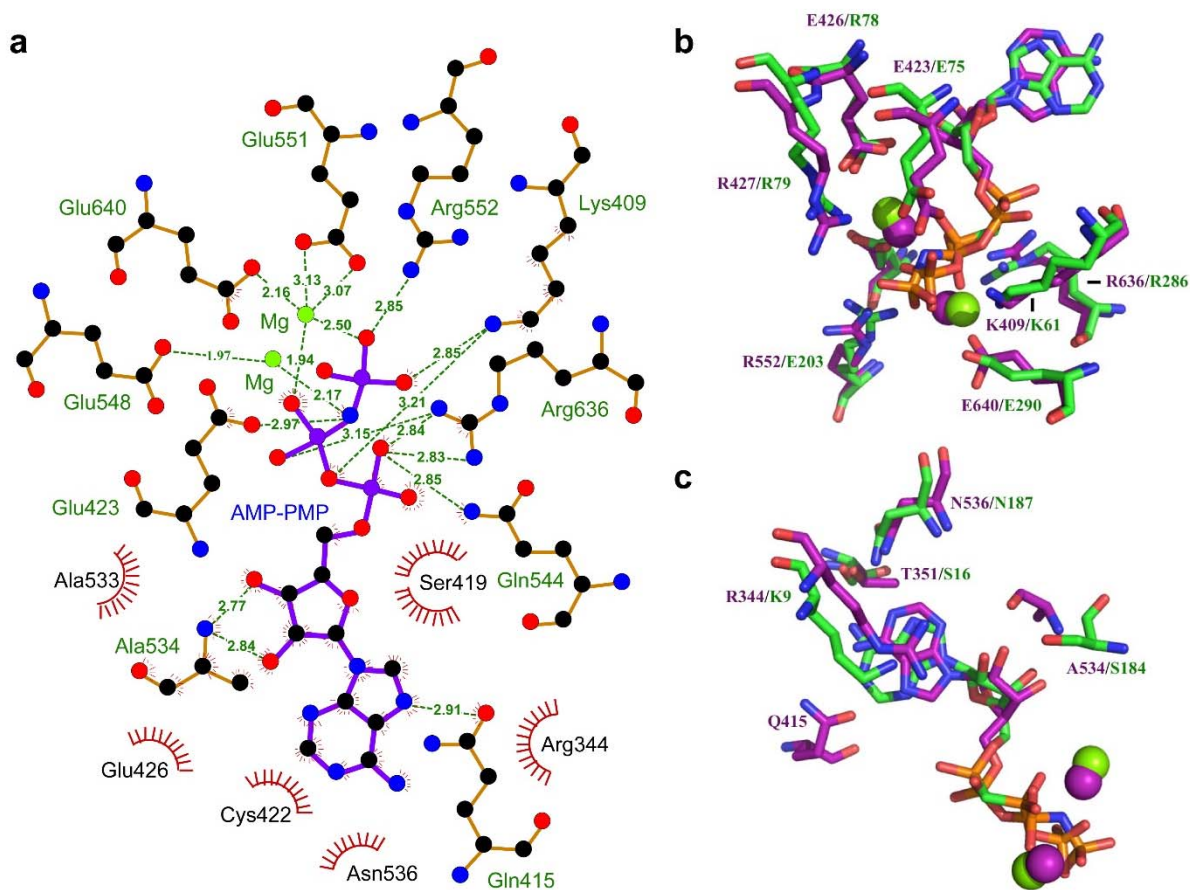


Figure 1.42 Nucleotide binding in YcaO domains. (a) A ligand interaction diagram for AMPPNP bound to LysD. (b) Residues involved in adenosine and ribose recognition in LysD-AMPPNP (purple) and Ec-YcaO-AMPCPP (green) complex structures. (c) Residues involved in Mg²⁺ and phosphate coordination in LysD-AMPPNP (purple) and Ec-YcaO-AMPCPP (green) complex structures.

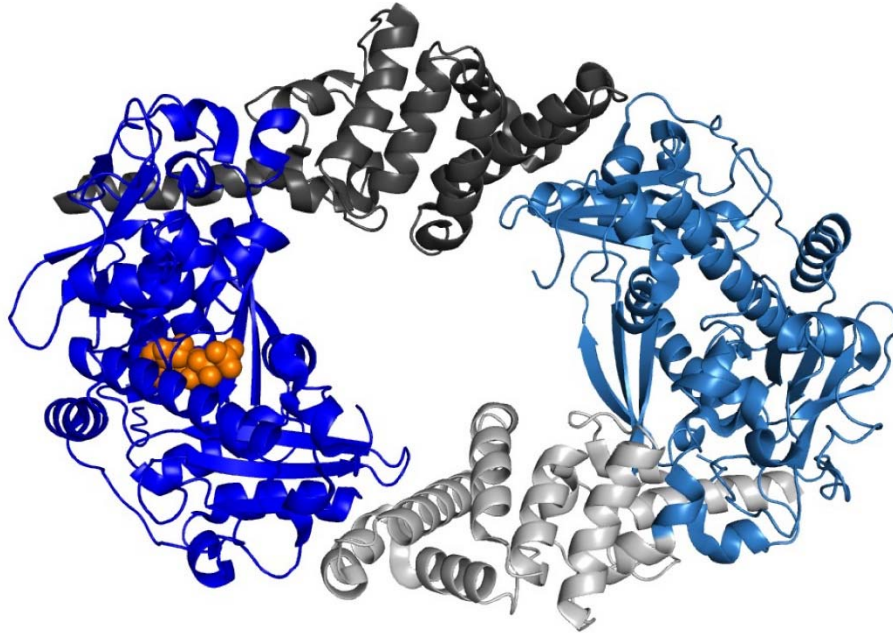


Figure 1.43 Ribbon structure of dimeric Ec-YcaO. Colored regions represent the YcaO domain (blue), tetratricopeptide-like domain (gray), ATP (orange spheres), or different monomers (dark vs. light coloring).

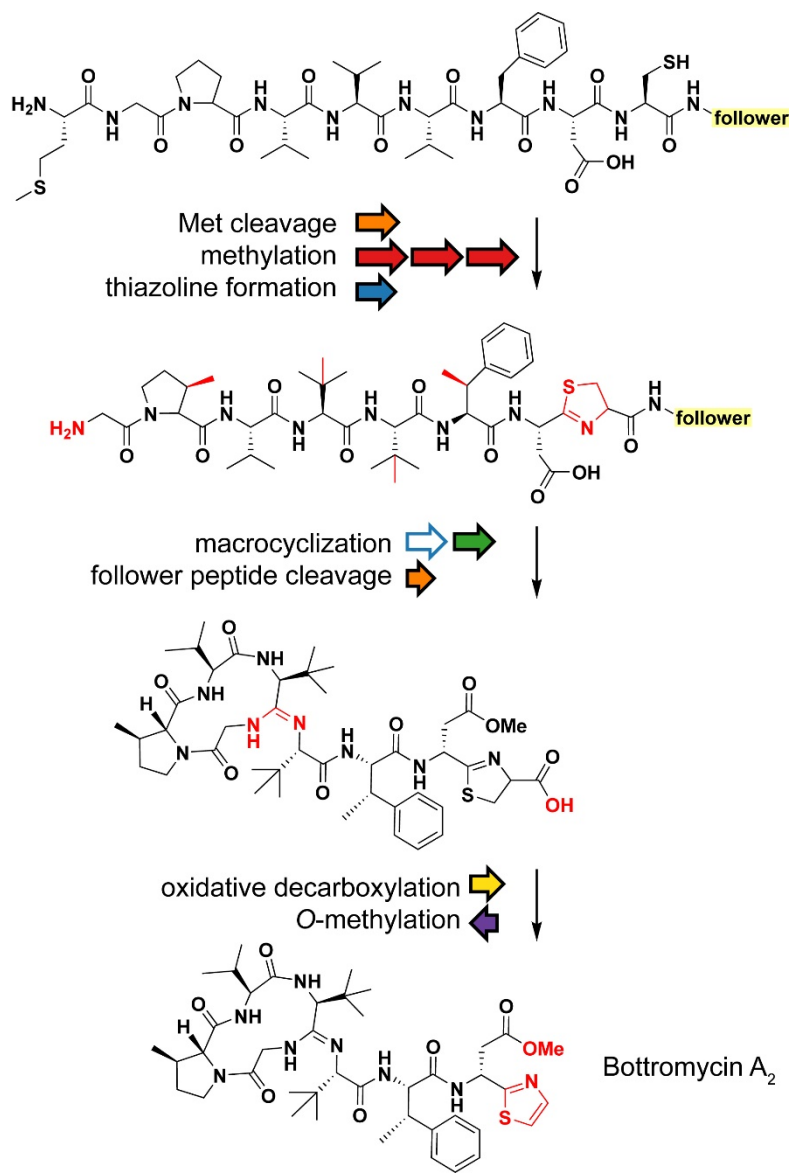


Figure 1.45 Biosynthesis of bottromycin determined through gene deletion in a heterologous host (*S. coelicolor*) and analysis of detected intermediates.

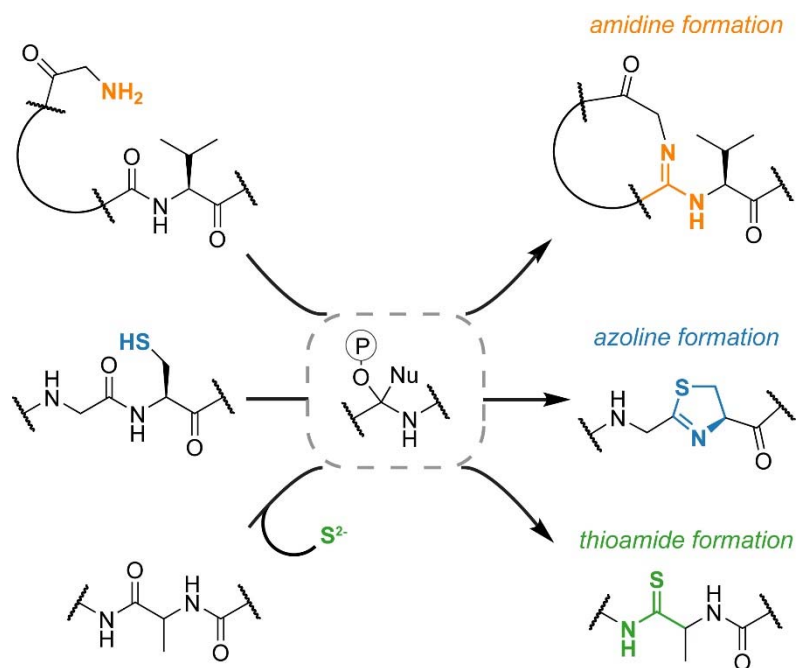


Figure 1.46 YcaOs catalyze diverse reactions through a similar backbone-activating mechanism. We hypothesize that YcaOs ubiquitously use ATP to O-phosphorylate peptide backbones, generating a common hemiorthoamide intermediate upon nucleophilic attack. Phosphate elimination yields different functional groups dependent on the identity of the nucleophile.

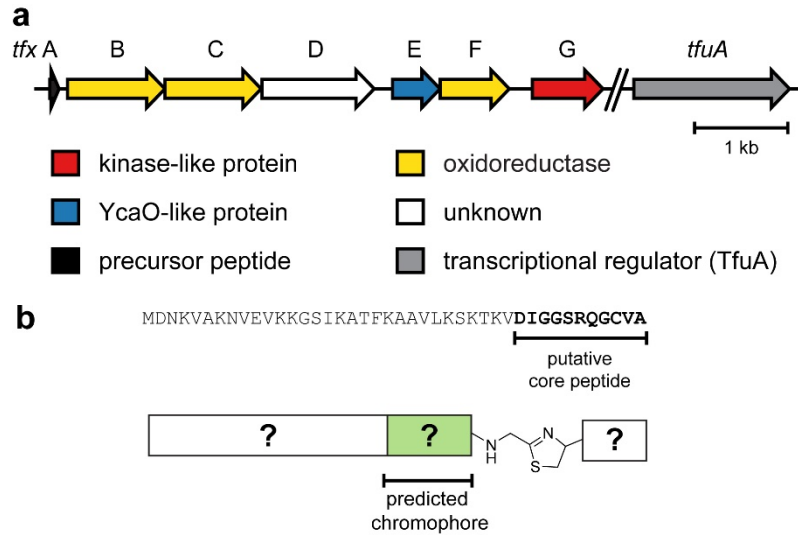


Figure 1.48 Trifolitoxin biosynthesis and partial structure. (a) BGC for TFX. (b) Precursor sequence and partial structure of TFX.

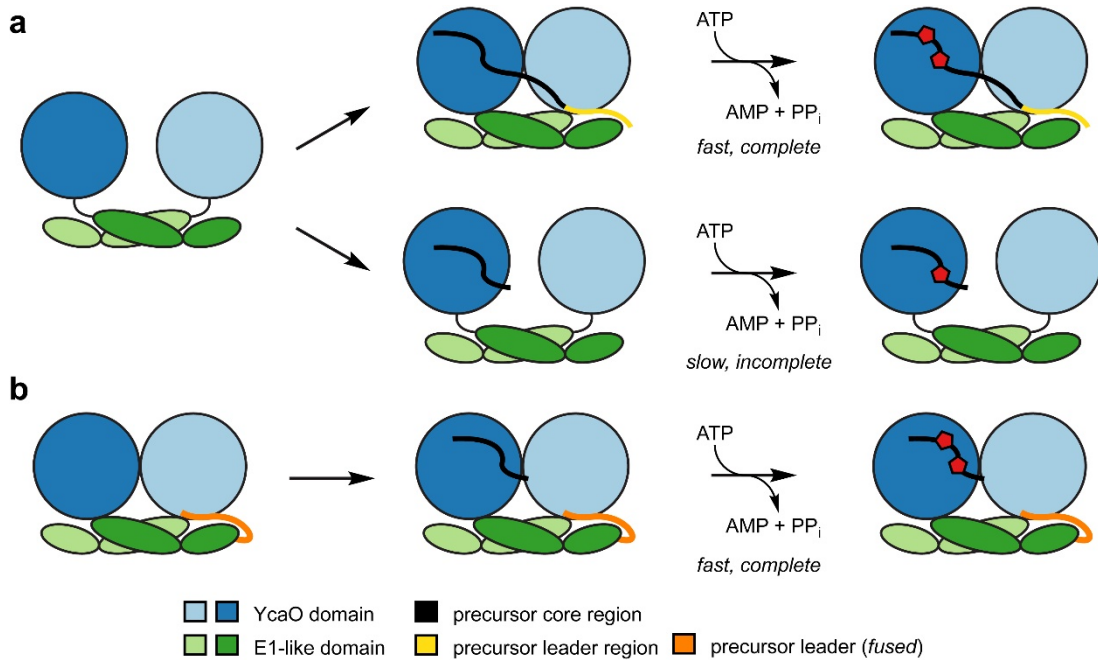


Figure 1.49 Cartoon schematic of LynD and the engineered “activated” cyclodehydratase (AcLynD). (a) The LynD dimer undergoes a conformational rearrangement (blue and green move closer together) following binding of PatE (yellow/black) generating an “active” enzyme. In the absence of leader peptide, the conformational rearrangement is not possible and heterocyclization of the core peptide is inefficient. (b) Covalent attachment of the leader peptide to the enzyme (orange) in AcLynD mimics the “activated” native LynD, allowing for efficient wild-type-like processing of the core peptide.

1.12 References

- (1) Eddy, S. R. Profile Hidden Markov Models. *Bioinformatics* **1998**, *14*.
- (2) Soding, J. Protein Homology Detection by HMM-HMM Comparison. *Bioinformatics* **2005**, *21*.
- (3) Bashton, M.; Chothia, C. The Generation of New Protein Functions by the Combination of Domains. *Structure* **2007**, *15* (1), 85-99.
- (4) Das, S.; Orengo, C. A. Protein Function Annotation Using Protein Domain Family Resources. *Methods* **2016**, *93*, 24-34.
- (5) Lee, D.; Redfern, O.; Orengo, C. Predicting Protein Function from Sequence and Structure. *Nat. Rev. Mol. Cell Biol.* **2007**, *8* (12), 995-1005.
- (6) Schnoes, A. M.; Brown, S. D.; Dodevski, I.; Babbitt, P. C. Annotation Error in Public Databases: Misannotation of Molecular Function in Enzyme Superfamilies. *PLoS Comput. Biol.* **2009**, *5* (12), e1000605.
- (7) Gerlt, J. A.; Babbitt, P. C. Divergent Evolution of Enzymatic Function: Mechanistically Diverse Superfamilies and Functionally Distinct Suprafamilies. *Annu. Rev. Biochem.* **2001**, *70*.
- (8) Gerlt, J. A.; Babbitt, P. C.; Jacobson, M. P.; Almo, S. C. Divergent Evolution in Enolase Superfamily: Strategies for Assigning Functions. *J. Biol. Chem.* **2012**, *287* (1), 29-34.
- (9) Das, S.; Dawson, N. L.; Orengo, C. A. Diversity in Protein Domain Superfamilies. *Curr. Opin. Genet. Dev.* **2015**, *35*, 40-9.
- (10) Broderick, J. B.; Duffus, B. R.; Duschene, K. S.; Shepard, E. M. Radical S-Adenosylmethionine Enzymes. *Chem. Rev.* **2014**, *114* (8), 4229-4317.
- (11) Bateman, A.; Coghill, P.; Finn, R. D. Dufs: Families in Search of Function. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2010**, *66*.
- (12) Mudgal, R.; Sandhya, S.; Chandra, N.; Srinivasan, N. De-DUFing the DUFs: Deciphering Distant Evolutionary Relationships of Domains of Unknown Function Using Sensitive Homology Detection Methods. *Biol. Direct* **2015**, *10* (1), 38.
- (13) Dunbar, K. L.; Mitchell, D. A. Revealing Nature's Synthetic Potential through the Study of Ribosomal Natural Product Biosynthesis. *ACS Chem. Biol.* **2013**, *8* (3), 473-487.
- (14) Corre, C.; Challis, G. L. New Natural Product Biosynthetic Chemistry Discovered by Genome Mining. *Nat. Prod. Rep.* **2009**, *26* (8), 977-986.
- (15) Guengerich, F. P.; Munro, A. W. Unusual Cytochrome P450 Enzymes and Reactions. *J. Biol. Chem.* **2013**, *288* (24), 17065-17073.
- (16) Zhang, Q.; van der Donk, W. A.; Liu, W. Radical-Mediated Enzymatic Methylation: A Tale of Two SAMs. *Acc. Chem. Res.* **2012**, *45* (4), 555-64.

- (17) Dunbar, K. L.; Melby, J. O.; Mitchell, D. A. YcaO Domains Use ATP to Activate Amide Backbones During Peptide Cyclodehydrations. *Nat. Chem. Biol.* **2012**, *8* (6), 569-575.
- (18) Dunbar, K. L.; Chekan, J. R.; Cox, C. L.; Burkhart, B. J.; Nair, S. K.; Mitchell, D. A. Discovery of a New ATP-Binding Motif Involved in Peptidic Azoline Biosynthesis. *Nat. Chem. Biol.* **2014**, *10* (10), 823-829.
- (19) Rudd, K. E. Linkage Map of Escherichia Coli K-12, Edition 10: The Physical Map. *Microbiol. Mol. Biol. Rev.* **1998**, *62* (3), 985-1019.
- (20) Bork, P. Powers and Pitfalls in Sequence Analysis: The 70% Hurdle. *Genome Res.* **2000**, *10* (4), 398-400.
- (21) Galperin, M. Y.; Koonin, E. V. From Complete Genome Sequence to 'Complete' Understanding? *Trends Biotechnol.* **2010**, *28* (8), 398-406.
- (22) Karp, P. D.; Keseler, I. M.; Shearer, A.; Latendresse, M.; Krummenacker, M.; Paley, S. M.; Paulsen, I.; Collado-Vides, J.; Gama-Castro, S.; Peralta-Gil, M., et al. Multidimensional Annotation of the *Escherichia Coli* K-12 Genome. *Nucleic Acids Res.* **2007**, *35* (22), 7577-90.
- (23) Strader, M. B.; Costantino, N.; Elkins, C. A.; Chen, C. Y.; Patel, I.; Makusky, A. J.; Choy, J. S.; Court, D. L.; Markey, S. P.; Kowalak, J. A. A Proteomic and Transcriptomic Approach Reveals New Insight into Beta-Methylthiolation of Escherichia Coli Ribosomal Protein S12. *Mol. Cell. Proteomics* **2011**, *10* (3), M110 005199.
- (24) Strader, M. B.; Hervey, W. J. t.; Costantino, N.; Fujigaki, S.; Chen, C. Y.; Akal-Strader, A.; Ihunnah, C. A.; Makusky, A. J.; Court, D. L.; Markey, S. P., et al. A Coordinated Proteomic Approach for Identifying Proteins that Interact with the E. Coli Ribosomal Protein S12. *J. Proteome Res.* **2013**, *12* (3), 1289-99.
- (25) Tenorio, E.; Saeki, T.; Fujita, K.; Kitakawa, M.; Baba, T.; Mori, H.; Isono, K. Systematic Characterization of Escherichia Coli Genes/ORFs Affecting Biofilm Formation. *FEMS Microbiol. Lett.* **2003**, *225* (1), 107-114.
- (26) Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J., et al. Ribosomally Synthesized and Post-Translationally Modified Peptide Natural Products: Overview and Recommendations for a Universal Nomenclature. *Nat. Prod. Rep.* **2013**, *30* (1), 108-160.
- (27) 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.
- (28) Dong, S. H.; Tang, W.; Lukk, T.; Yu, Y.; Nair, S. K.; van der Donk, W. A. The Enterococcal Cytolysin Synthetase Has an Unanticipated Lipid Kinase Fold. *eLife* **2015**, *4*.

- (29) Li, K.; Concurso, H. L.; Li, G.; Ding, Y.; Bruner, S. D. Structural Basis for Precursor Protein-Directed Ribosomal Peptide Macrocyclization. *Nat. Chem. Biol.* **2016**, *12* (11), 973-979.
- (30) McIntosh, J. A.; Donia, M. S.; Schmidt, E. W. Ribosomal Peptide Natural Products: Bridging the Ribosomal and Nonribosomal Worlds. *Nat. Prod. Rep.* **2009**, *26* (4), 537-559.
- (31) Walsh, C. T. Blurring the Lines between Ribosomal and Nonribosomal Peptide Scaffolds. *ACS Chem. Biol.* **2014**, *9* (8), 1653-61.
- (32) Bagley, M. C.; Dale, J. W.; Merritt, E. A.; Xiong, X. Thiopeptide Antibiotics. *Chem. Rev.* **2005**, *105* (2), 685-714.
- (33) Davagnino, J.; Herrero, M.; Furlong, D.; Moreno, F.; Kolter, R. The DNA Replication Inhibitor Microcin B17 Is a Forty-Three-Amino-Acid Protein Containing Sixty Percent Glycine. *Proteins* **1986**, *1* (3), 230-8.
- (34) Bayer, A.; Freund, S.; Nicholson, G.; Jung, G. Posttranslational Backbone Modifications in the Ribosomal Biosynthesis of the Glycine-Rich Antibiotic Microcin-B17. *Angew. Chem., Int. Ed.* **1993**, *32* (9), 1336-1339.
- (35) Bayer, A.; Freund, S.; Jung, G. Post-Translational Heterocyclic Backbone Modifications in the 43-Peptide Antibiotic Microcin B17. Structure Elucidation and NMR Study of a ¹³C,¹⁵N-Labeled Gyrase Inhibitor. *Eur. J. Biochem.* **1995**, *234* (2), 414-26.
- (36) Yorgey, P.; Lee, J.; Koedel, J.; Vivas, E.; Warner, P.; Jebaratnam, D.; Kolter, R. Posttranslational Modifications in Microcin B17 Define an Additional Class of DNA Gyrase Inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91* (10), 4519-23.
- (37) Li, Y. M.; Milne, J. C.; Madison, L. L.; Kolter, R.; Walsh, C. T. From Peptide Precursors to Oxazole and Thiazole-Containing Peptide Antibiotics: Microcin B17 Synthase. *Science* **1996**, *274* (5290), 1188-93.
- (38) Gomez-Escribano, J. P.; Song, L.; Bibb, M. J.; Challis, G. L. Posttranslational β -Methylation and Macrolactamidation in the Biosynthesis of the Bottromycin Complex of Ribosomal Peptide Antibiotics. *Chem. Sci.* **2012**, *3* (12), 3522-3525.
- (39) Crone, W. J. K.; Leeper, F. J.; Truman, A. W. Identification and Characterization of the Gene Cluster for the Anti-MRSA Antibiotic Bottromycin: Expanding the Biosynthetic Diversity of Ribosomal Peptides. *Chem. Sci.* **2012**, *3* (12), 3516-3521.
- (40) Huo, L.; Rachid, S.; Stadler, M.; Wenzel, Silke C.; Müller, R. Synthetic Biotechnology to Study and Engineer Ribosomal Bottromycin Biosynthesis. *Chem. Biol.* **2012**, *19* (10), 1278-1287.
- (41) Melby, J. O.; Nard, N. J.; Mitchell, D. A. Thiazole/Oxazole-Modified Microcins: Complex Natural Products from Ribosomal Templates. *Curr. Opin. Chem. Biol.* **2011**, *15* (3), 369-78.

- (42) Sudek, S.; Haygood, M. G.; Youssef, D. T.; Schmidt, E. W. Structure of Trichamide, a Cyclic Peptide from the Bloom-Forming Cyanobacterium *Trichodesmium Erythraeum*, Predicted from the Genome Sequence. *Appl. Environ. Microbiol.* **2006**, *72*, 4382 - 4387.
- (43) Lee, J.; McIntosh, J.; Hathaway, B. J.; Schmidt, E. W. Using Marine Natural Products to Discover a Protease That Catalyzes Peptide Macrocyclization of Diverse Substrates. *J. Am. Chem. Soc.* **2009**, *131* (6), 2122-4.
- (44) Bowers, A. A.; Walsh, C. T.; Acker, M. G. Genetic Interception and Structural Characterization of Thiopeptide Cyclization Precursors from *Bacillus Cereus*. *J. Am. Chem. Soc.* **2010**, *132* (35), 12182-4.
- (45) Hudson, G. A.; Zhang, Z.; Tietz, J. I.; Mitchell, D. A.; van der Donk, W. A. In Vitro Biosynthesis of the Core Scaffold of the Thiopeptide Thiomuracin. *J. Am. Chem. Soc.* **2015**, *137* (51), 16012-16015.
- (46) Yorgey, P.; Lee, J.; Kolter, R., The Structure and Maturation Pathway of Microcin B17. In *Bacteriocins, Microcins and Lantibiotics*, James, R.; Lazdunski, C.; Pattus, F., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 1992; pp 19-31.
- (47) Yorgey, P.; Davagnino, J.; Kolter, R. The Maturation Pathway of Microcin B17, a Peptide Inhibitor of DNA Gyrase. *Mol. Microbiol.* **1993**, *9* (4), 897-905.
- (48) Schmidt, E. W.; Nelson, J. T.; Rasko, D. A.; Sudek, S.; Eisen, J. A.; Haygood, M. G.; Ravel, J. Patellamide A and C Biosynthesis by a Microcin-Like Pathway in *Prochloron Didemni*, the Cyanobacterial Symbiont of *Lissoclinum Patella*. *Proc. Nat. Acad. Sci. USA* **2005**, *102* (20), 7315-20.
- (49) Donia, M. S.; Hathaway, B. J.; Sudek, S.; Haygood, M. G.; Rosovitz, M. J.; Ravel, J.; Schmidt, E. W. Natural Combinatorial Peptide Libraries in Cyanobacterial Symbionts of Marine Ascidians. *Nat. Chem. Biol.* **2006**, *2* (12), 729-35.
- (50) Donia, M. S.; Ravel, J.; Schmidt, E. W. A Global Assembly Line for Cyanobactins. *Nat. Chem. Biol.* **2008**, *4* (6), 341-343.
- (51) Lee, S. W.; Mitchell, D. A.; Markley, A. L.; Hensler, M. E.; Gonzalez, D.; Wohlrab, A.; Dorrestein, P. C.; Nizet, V.; Dixon, J. E. Discovery of a Widely Distributed Toxin Biosynthetic Gene Cluster. *Proc. Nat. Acad. Sci. USA* **2008**, *105* (15), 5879-84.
- (52) Wieland Brown, L. C.; Acker, M. G.; Clardy, J.; Walsh, C. T.; Fischbach, M. A. Thirteen Posttranslational Modifications Convert a 14-Residue Peptide into the Antibiotic Thiocillin. *Proc. Nat. Acad. Sci. USA* **2009**, *106* (8), 2549-53.

- (53) Yu, Y.; Duan, L.; Zhang, Q.; Liao, R.; Ding, Y.; Pan, H.; Wendt-Pienkowski, E.; Tang, G.; Shen, B.; Liu, W. Nosiheptide Biosynthesis Featuring a Unique Indole Side Ring Formation on the Characteristic Thiopeptide Framework. *ACS Chem. Biol.* **2009**, *4* (10), 855-64.
- (54) Liao, R.; Duan, L.; Lei, C.; Pan, H.; Ding, Y.; Zhang, Q.; Chen, D.; Shen, B.; Yu, Y.; Liu, W. Thiopeptide Biosynthesis Featuring Ribosomally Synthesized Precursor Peptides and Conserved Posttranslational Modifications. *Chem. Biol.* **2009**, *16* (2), 141-7.
- (55) Morris, R. P.; Leeds, J. A.; Naegeli, H. U.; Oberer, L.; Memmert, K.; Weber, E.; LaMarche, M. J.; Parker, C. N.; Burrer, N.; Esterow, S., et al. Ribosomally Synthesized Thiopeptide Antibiotics Targeting Elongation Factor Tu. *J. Am. Chem. Soc.* **2009**, *131* (16), 5946-55.
- (56) Kelly, W.; Pan, L.; Li, C. Thiostrepton Biosynthesis: Prototype for a New Family of Bacteriocins. *J. Am. Chem. Soc.* **2009**, *131*, 4327 - 4334.
- (57) Cox, C. L.; Doroghazi, J. R.; Mitchell, D. A. The Genomic Landscape of Ribosomal Peptides Containing Thiazole and Oxazole Heterocycles. *BMC Genomics* **2015**, *16* (1), 778.
- (58) Jin, Z. Muscarine, Imidazole, Oxazole and Thiazole Alkaloids. *Nat. Prod. Rep.* **2013**, *30* (6), 869-915.
- (59) Walsh, C. T. Insights into the Chemical Logic and Enzymatic Machinery of Nrps Assembly Lines. *Nat. Prod. Rep.* **2016**, *33* (2), 127-35.
- (60) Stubbe, J.; Kozarich, J. W.; Wu, W.; Vanderwall, D. E. Bleomycins: A Structural Model for Specificity, Binding, and Double Strand Cleavage. *Acc. Chem. Res.* **1996**, *29* (7), 322-330.
- (61) Puhalla, S.; Brufsky, A. Ixabepilone: A New Chemotherapeutic Option for Refractory Metastatic Breast Cancer. *Biologics* **2008**, *2* (3), 505-515.
- (62) Siddiqui, N.; Arya, S. K.; Ahsan, W.; Azad, B. Diverse Biological Activities of Thiazoles: A Retrospect. *Int. J. Drug Dev. Res.* **2011**, *3* (4), 55-67.
- (63) Morigi, R.; Locatelli, A.; Leoni, A.; Rambaldi, M. Recent Patents on Thiazole Derivatives Endowed with Antitumor Activity. *Recent Pat. Anti-canc.* **2015**, *10* (3), 280-297.
- (64) Chhabria, M. T.; Patel, S.; Modi, P.; Brahmikshatriya, P. S. Thiazole: A Review on Chemistry, Synthesis and Therapeutic Importance of Its Derivatives. *Curr. Top. Med. Chem.* **2016**.
- (65) Just-Baringo, X.; Albericio, F.; Álvarez, M. Thiopeptide Antibiotics: Retrospective and Recent Advances. *Mar. Drugs* **2014**, *12* (1), 317-351.
- (66) Benazet, F.; Cartier, J. R. Effect of Nosiheptide as a Feed Additive in Chicks on the Quantity, Duration, Prevalence of Excretion, and Resistance to Antibacterial Agents of Salmonella Typhimurium; on the Proportion of Escherichia Coli and Other Coliforms Resistant to Antibacterial Agents; and on Their Degree and Spectrum of Resistance. *Poultry Sci.* **1980**, *59* (7), 1405-15.

- (67) McGinnis, C. H.; Johnson, C. A.; Fox, J. E. The Effect of Nosiheptide, a New Antibiotic, on Body Weight Gain and Feed Efficiency in Broiler Chickens. *Poultry Sci.* **1978**, *57* (6), 1641-1645.
- (68) Sachdeva, M.; Leeds, J. A. Subinhibitory Concentrations of Lff571 Reduce Toxin Production by *Clostridium Difficile*. *Antimicrob. Agents Chemother.* **2015**, *59* (2), 1252-7.
- (69) Bhansali, S. G.; Mullane, K.; Ting, L. S.; Leeds, J. A.; Dabovic, K.; Praestgaard, J.; Pertel, P. Pharmacokinetics of Lff571 and Vancomycin in Patients with Moderate *Clostridium Difficile* Infections. *Antimicrob. Agents Chemother.* **2015**, *59* (3), 1441-5.
- (70) Mullane, K.; Lee, C.; Bressler, A.; Buitrago, M.; Weiss, K.; Dabovic, K.; Praestgaard, J.; Leeds, J. A.; Blais, J.; Pertel, P. Multicenter, Randomized Clinical Trial to Compare the Safety and Efficacy of LFF571 and Vancomycin for *Clostridium Difficile* Infections. *Antimicrob. Agents Chemother.* **2015**, *59* (3), 1435-40.
- (71) Skinnider, M. A.; Johnston, C. W.; Edgar, R. E.; Dejong, C. A.; Merwin, N. J.; Rees, P. N.; Magarvey, N. A. Genomic Charting of Ribosomally Synthesized Natural Product Chemical Space Facilitates Targeted Mining. *Proc. Natl. Acad. Sci. USA* **2016**, *113* (42), E6343-E6351.
- (72) Melby, J. O.; Li, X.; Mitchell, D. A. Orchestration of Enzymatic Processing by Thiazole/Oxazole-Modified Microcin Dehydrogenases. *Biochemistry* **2014**, *53* (2), 413-422.
- (73) Severinov, K.; Semenova, E.; Kazakov, A.; Kazakov, T.; Gelfand, M. S. Low-Molecular-Weight Post-Translationally Modified Microcins. *Mol. Microbiol.* **2007**, *65* (6), 1380-94.
- (74) Baquero, F.; Moreno, F. The Microcins. *FEMS Microbiol. Lett.* **1984**, *23* (2-3), 117-124.
- (75) Nissen-Meyer, J.; Nes, F. I. Ribosomally Synthesized Antimicrobial Peptides: Their Function, Structure, Biogenesis, and Mechanism of Action. *Arch. Microbiol.* **1997**, *167* (2), 67-77.
- (76) Cotter, P. D.; Hill, C.; Ross, R. P. Bacteriocins: Developing Innate Immunity for Food. *Nat Rev Micro* **2005**, *3* (10), 777-788.
- (77) Cotter, P. D.; Ross, R. P.; Hill, C. Bacteriocins - a Viable Alternative to Antibiotics? *Nat. Rev. Microbiol.* **2013**, *11*.
- (78) Asensio, C.; Pérez-Díaz, J. C.; Martínez, M. C.; Baquero, F. A New Family of Low Molecular Weight Antibiotics from Enterobacteria. *Biochem. Bioph. Res. Co.* **1976**, *69* (1), 7-14.
- (79) Vizán, J. L.; Hernández-Chico, C.; Del Castillo, I.; Moreno, F. The Peptide Antibiotic Microcin B17 Induces Double-Strand Cleavage of DNA Mediated by *E. Coli* DNA Gyrase. *EMBO J.* **1991**, *10* (2), 467-76.
- (80) Herrero, M.; Moreno, F. Microcin B17 Blocks DNA Replication and Induces the SOS System in *Escherichia Coli*. *J. Gen. Microbiol.* **1986**, *132* (2), 393-402.
- (81) Collin, F.; Karkare, S.; Maxwell, A. Exploiting Bacterial DNA Gyrase as a Drug Target: Current State and Perspectives. *Appl. Microbiol Biotechnol.* **2011**, *92* (3), 479-97.

- (82) Baquero, F.; Bouanchaud, D.; Martinez-Perez, M. C.; Fernandez, C. Microcin Plasmids: A Group of Extrachromosomal Elements Coding for Low-Molecular-Weight Antibiotics in *Escherichia Coli*. *J. Bacteriol.* **1978**, *135* (2), 342-7.
- (83) San Millan, J. L.; Kolter, R.; Moreno, F. Evidence That Colicin X Is Microcin B17. *J. Bacteriol.* **1987**, *169* (6), 2899-901.
- (84) Hernandez-Chico, C.; Herrero, M.; Rejas, M.; San Millan, J. L.; Moreno, F. Gene Ompr and Regulation of Microcin 17 and Colicin E2 Syntheses. *J. Bacteriol.* **1982**, *152* (2), 897-900.
- (85) Hernandez-Chico, C.; San Millan, J. L.; Kolter, R.; Moreno, F. Growth Phase and Ompr Regulation of Transcription of Microcin B17 Genes. *J. Bacteriol.* **1986**, *167* (3), 1058-65.
- (86) Kenney, L. J. Structure/Function Relationships in Ompr and Other Winged-Helix Transcription Factors. *Curr. Opin. Microbiol.* **2002**, *5* (2), 135-141.
- (87) San Millan, J. L.; Hernandez-Chico, C.; Pereda, P.; Moreno, F. Cloning and Mapping of the Genetic Determinants for Microcin B17 Production and Immunity. *J. Bacteriol.* **1985**, *163* (1), 275-81.
- (88) Genilloud, O.; Moreno, F.; Kolter, R. DNA Sequence, Products, and Transcriptional Pattern of the Genes Involved in Production of the DNA Replication Inhibitor Microcin B17. *J. Bacteriol.* **1989**, *171* (2), 1126-1135.
- (89) Rodriguez-Sainz, M. C.; Hernandez-Chico, C.; Moreno, F. Molecular Characterization of PmbA, an *Escherichia Coli* Chromosomal Gene Required for the Production of the Antibiotic Peptide Mccb17. *Mol. Microbiol.* **1990**, *4* (11), 1921-32.
- (90) Allali, N.; Afif, H.; Couturier, M.; Van Melderen, L. The Highly Conserved TldD and TldE Proteins of *Escherichia Coli* Are Involved in Microcin B17 Processing and in Ccda Degradation. *J. Bacteriol.* **2002**, *184* (12), 3224-3231.
- (91) Garrido, M. d. C.; Herrero, M.; Kolter, R.; Moreno, F. The Export of the DNA Replication Inhibitor Microcin B17 Provides Immunity for the Host Cell. *EMBO J.* **1988**, *7* (6), 1853-62.
- (92) Shah, S.; Heddle, J. G. Squaring up to DNA: Pentapeptide Repeat Proteins and DNA Mimicry. *Appl. Microbiol. Biotechnol.* **2014**, *98* (23), 9545-60.
- (93) Collin, F.; Thompson, R. E.; Jolliffe, K. A.; Payne, R. J.; Maxwell, A. Fragments of the Bacterial Toxin Microcin B17 as Gyrase Poisons. *PLoS One* **2013**, *8* (4), e61459.
- (94) Shkundina, I.; Serebryakova, M.; Severinov, K. The C-Terminal Part of Microcin B Is Crucial for DNA Gyrase Inhibition and Antibiotic Uptake by Sensitive Cells. *J. Bacteriol.* **2014**, *196* (9), 1759-67.
- (95) Metelev, M.; Serebryakova, M.; Ghilarov, D.; Zhao, Y.; Severinov, K. Structure of Microcin B-Like Compounds Produced by *Pseudomonas Syringae* and Species Specificity of Their Antibacterial Action. *J. Bacteriol.* **2013**, *195* (18), 4129-4137.

- (96) Lavina, M.; Pugsley, A. P.; Moreno, F. Identification, Mapping, Cloning and Characterization of a Gene (SbmA) Required for Microcin B17 Action on Escherichia Coli K12. *J. Gen. Microbiol.* **1986**, *132* (6), 1685-93.
- (97) Runti, G.; Lopez Ruiz, M. d. C.; Stoilova, T.; Hussain, R.; Jennions, M.; Choudhury, H. G.; Benincasa, M.; Gennaro, R.; Beis, K.; Scocchi, M. Functional Characterization of SbmA, a Bacterial Inner Membrane Transporter Required for Importing the Antimicrobial Peptide Bac7(1-35). *J. Bacteriol.* **2013**, *195* (23), 5343-5351.
- (98) Salomon, R. A.; Farias, R. N. The Peptide Antibiotic Microcin 25 Is Imported through the TonB Pathway and the SbmA Protein. *J. Bacteriol.* **1995**, *177* (11), 3323-5.
- (99) Mathavan, I.; Beis, K. The Role of Bacterial Membrane Proteins in the Internalization of Microcin Mccj25 and Mccb17. *Biochem. Soc. Trans.* **2012**, *40* (6), 1539-43.
- (100) Nöllmann, M.; Crisona, N. J.; Arimondo, P. B. Thirty Years of Escherichia Coli DNA Gyrase: From in Vivo Function to Single-Molecule Mechanism. *Biochimie* **2007**, *89* (4), 490-499.
- (101) Champoux, J. J. DNA Topoisomerases: Structure, Function, and Mechanism. *Annu. Rev. Biochem.* **2001**, *70* (1), 369-413.
- (102) Del Castillo, F. J.; Del Castillo, I.; Moreno, F. Construction and Characterization of Mutations at Codon 751 of the Escherichia Coli GyrB Gene That Confer Resistance to the Antimicrobial Peptide Microcin B17 and Alter the Activity of DNA Gyrase. *J. Bacteriol.* **2001**, *183* (6), 2137-2140.
- (103) Pierrat, O. A.; Maxwell, A. The Action of the Bacterial Toxin Microcin B17: Insight into the Cleavage-Religation Reaction of DNA Gyrase. *J. Biol. Chem.* **2003**, *278* (37), 35016-35023.
- (104) Pierrat, O. A.; Maxwell, A. Evidence for the Role of DNA Strand Passage in the Mechanism of Action of Microcin B17 on DNA Gyrase. *Biochemistry* **2005**, *44* (11), 4204-15.
- (105) Parks, W. M.; Bottrill, A. R.; Pierrat, O. A.; Durrant, M. C.; Maxwell, A. The Action of the Bacterial Toxin, Microcin B17, on DNA Gyrase. *Biochimie* **2007**, *89* (4), 500-507.
- (106) Heddle, J. G.; Blance, S. J.; Zamble, D. B.; Hollfelder, F.; Miller, D. A.; Wentzell, L. M.; Walsh, C. T.; Maxwell, A. The Antibiotic Microcin B17 Is a DNA Gyrase Poison: Characterization of the Mode of Inhibition. *J. Mol. Biol.* **2001**, *307* (5), 1223-1234.
- (107) Zamble, D. B.; Miller, D. A.; Heddle, J. G.; Maxwell, A.; Walsh, C. T.; Hollfelder, F. In Vitro Characterization of DNA Gyrase Inhibition by Microcin B17 Analogs with Altered Bisheterocyclic Sites. *Proc. Natl. Acad. Sci. USA* **2001**, *98* (14), 7712-7717.
- (108) Herrero, M.; Kolter, R.; Moreno, F. Effects of Microcin B17 on Microcin B17-Immune Cells. *J. Gen. Microbiol.* **1986**, *132* (2), 403-10.
- (109) Chatterji, M.; Nagaraja, V. Gyri: A Counter-Defensive Strategy against Proteinaceous Inhibitors of DNA Gyrase. *EMBO Rep.* **2002**, *3* (3), 261-7.

- (110) Nakanishi, A.; Oshida, T.; Matsushita, T.; Imajoh-Ohmi, S.; Ohnuki, T. Identification of DNA Gyrase Inhibitor (Gyri) in Escherichia Coli. *J. Biol. Chem.* **1998**, *273* (4), 1933-8.
- (111) Baquero, M. R.; Bouzon, M.; Varea, J.; Moreno, F. Sbm, a Stationary-Phase Induced *Sos* Escherichia Coli Gene, Whose Product Protects Cells from the DNA Replication Inhibitor Microcin B17. *Mol. Microbiol.* **1995**, *18* (2), 301-11.
- (112) Ghilarov, D.; Serebryakova, M.; Shkundina, I.; Severinov, K. A Major Portion of DNA Gyrase Inhibitor Microcin B17 Undergoes an N,O-Peptidyl Shift During Synthesis. *J. Biol. Chem.* **2011**, *286* (30), 26308-26318, S26308/1-S26308/5.
- (113) Roy, R. S.; Kelleher, N. L.; Milne, J. C.; Walsh, C. T. In Vivo Processing and Antibiotic Activity of Microcin B17 Analogs with Varying Ring Content and Altered Bisheterocyclic Sites. *Chem. Biol.* **1999**, *6* (5), 305-318.
- (114) Videnov, G.; Kaiser, D.; Brooks, M.; Jung, G. Synthesis of the DNA Gyrase Inhibitor Microcin B17, a 43-Peptide Antibiotic with Eight Aromatic Heterocycles in Its Backbone. *Angew. Chem., Int. Ed. Engl.* **1996**, *35* (13/14), 1506-1508.
- (115) Thompson, R. E.; Jolliffe, K. A.; Payne, R. J. Total Synthesis of Microcin B17 Via a Fragment Condensation Approach. *Org. Lett.* **2011**, *13* (4), 680-3.
- (116) Melby, J. O.; Nard, N. J.; Mitchell, D. A. Thiazole/Oxazole-Modified Microcins: Complex Natural Products from Ribosomal Templates. *Curr. Opin. Chem. Biol.* **2011**, *15* (3), 369-378.
- (117) Van Tyne, D.; Martin, M.; Gilmore, M. Structure, Function, and Biology of the Enterococcus Faecalis Cytolysin. *Toxins* **2013**, *5* (5), 895.
- (118) Vandenesch, F.; Lina, G.; Henry, T. Staphylococcus Aureus Hemolysins, Bi-Component Leukocidins, and Cytolytic Peptides: A Redundant Arsenal of Membrane-Damaging Virulence Factors? *Front. Cell. Infect. Microbiol.* **2012**, *2*, 12.
- (119) Alouf, J. E. Cholesterol-Binding Cytolytic Protein Toxins. *Int. J. Med. Microbiol.* **2000**, *290* (4-5), 351-6.
- (120) Braun, V.; Focareta, T. Pore-Forming Bacterial Protein Hemolysins (Cytolysins). *Crit. Rev. Microbiol.* **1991**, *18* (2), 115-158.
- (121) Cunningham, M. W. Pathogenesis of Group a Streptococcal Infections. *Clin. Microbiol. Rev.* **2000**, *13*.
- (122) Todd, E. W. The Differentiation of Two Distinct Serologic Varieties of Streptolysin, Streptolysin O and Streptolysin S. *J. Pathol. Bacteriol.* **1938**, *47*.
- (123) Marmorek, A. Le Streptocoque Et Le S rum Antistreptococcique. *Ann. Inst. Pasteur.* **1895**, *9*.
- (124) Brown, J. H., *The Use of Blood Agar for the Study of Streptococci*. Rockefeller Institute for Medical Research: New York, 1919; Vol. 9.

- (125) Ginsburg, I.; Harris, T. N. Oxygen-Stable Hemolysins of Beta-Hemolytic Streptococci. *Ergeb. Mikrobiol. Immunitätsforsch Exp. Ther.* **1964**, *38*, 198-222.
- (126) Alouf, J. E. Streptococcal Toxins (Streptolysin O, Streptolysin S, Erythrogenic Toxin). *Pharmacol. Ther.* **1980**, *11* (3), 661-717.
- (127) Ginsburg, I.; Harris, T. N. Oxygen-Stable Hemolysins of Group a Streptococci. II. Chromatographic and Electrophoretic Studies. *J. Exp. Med.* **1963**, *118* (6), 919-34.
- (128) Koyama, J.; Egami, F. Biochemical Studies on Streptolysin S' Formed in the Presence of Yeast Ribonucleic Acid. I. The Purification and Some Properties of the Toxin. *J. Biochem.* **1963**, *53*, 147-54.
- (129) Koyama, J. Biochemical Studies on Streptolysin S'. II. Properties of a Polypeptide Component and Its Role in the Toxin Activity. *J. Biochem.* **1963**, *54*, 146-51.
- (130) Bernheimer, A. W. Physical Behavior of Streptolysin S. *J. Bacteriol.* **1967**, *93* (6), 2024-5.
- (131) Borgia, S. M.; Betschel, S.; Low, D. E.; de Azavedo, J. C. Cloning of a Chromosomal Region Responsible for Streptolysin S Production in Streptococcus Pyogenes. *Adv. Exp. Med. Biol.* **1997**, *418*, 733-6.
- (132) Betschel, S. D.; Borgia, S. M.; Barg, N. L.; Low, D. E.; De Azavedo, J. C. Reduced Virulence of Group a Streptococcal Tn916 Mutants That Do Not Produce Streptolysin S. *Infect. Immun.* **1998**, *66* (4), 1671-9.
- (133) Nizet, V.; Beall, B.; Bast, D. J.; Datta, V.; Kilburn, L.; Low, D. E. Genetic Locus for Streptolysin S Production by Group a Streptococcus. *Infect. Immun.* **2000**, *68*.
- (134) Ferretti, J. J.; McShan, W. M.; Ajdic, D.; Savic, D. J.; Savic, G.; Lyon, K.; Primeaux, C.; Sezate, S.; Suvorov, A. N.; Kenton, S., et al. Complete Genome Sequence of an M1 Strain of Streptococcus Pyogenes. *Proc. Natl. Acad. Sci. USA* **2001**, *98* (8), 4658-63.
- (135) Datta, V.; Myskowski, S. M.; Kwinn, L. A.; Chiem, D. N.; Varki, N.; Kansal, R. G. Mutational Analysis of the Group a Streptococcal Operon Encoding Streptolysin S and Its Virulence Role in Invasive Infection. *Mol. Microbiol.* **2005**, *56*.
- (136) Mitchell, D. A.; Lee, S. W.; Pence, M. A.; Markley, A. L.; Limm, J. D.; Nizet, V. Structural and Functional Dissection of the Heterocyclic Peptide Cytotoxin Streptolysin S. *J. Biol. Chem.* **2009**, *284*.
- (137) Molloy, E. M.; Cotter, P. D.; Hill, C.; Mitchell, D. A.; Ross, R. P. Streptolysin S-Like Virulence Factors: The Continuing Saga. *Nat. Rev. Microbiol.* **2011**, *9*.
- (138) Maxson, T.; Deane, C. D.; Molloy, E. M.; Cox, C. L.; Markley, A. L.; Lee, S. W.; Mitchell, D. A. Hiv Protease Inhibitors Block Streptolysin S Production. *ACS Chem. Biol.* **2015**, *10* (5), 1217-1226.

- (139) Plummer, L. J.; Hildebrandt, E. R.; Porter, S. B.; Rogers, V. A.; McCracken, J.; Schmidt, W. K. Mutational Analysis of the Ras Converting Enzyme Reveals a Requirement for Glutamate and Histidine Residues. *J. Biol. Chem.* **2006**, *281* (8), 4596-4605.
- (140) Bergo, M. O.; Wahlstrom, A. M.; Fong, L. G.; Young, S. G., Genetic Analyses of the Role of Rce1 in Ras Membrane Association and Transformation. In *Methods Enzymol.*, Academic Press: 2008; Vol. Volume 438, pp 367-389.
- (141) Manolaridis, I.; Kulkarni, K.; Dodd, R. B.; Ogasawara, S.; Zhang, Z.; Bineva, G.; Reilly, N. O.; Hanrahan, S. J.; Thompson, A. J.; Cronin, N., et al. Mechanism of Farnesylated CAAX Protein Processing by the Integral Membrane Protease Rce1. *Nature* **2013**, *504* (7479), 10.1038/nature12754.
- (142) Pei, J.; Mitchell, D. A.; Dixon, J. E.; Grishin, N. V. Expansion of Type II CAAX Proteases Reveals Evolutionary Origin of Gamma-Secretase Subunit Aph-1. *J. Mol. Biol.* **2011**, *410*.
- (143) Jack, R. W.; Tagg, J. R.; Ray, B. Bacteriocins of Gram-Positive Bacteria. *Microbiol. Rev.* **1995**, *59* (2), 171-200.
- (144) Gonzalez, D. J.; Lee, S. W.; Hensler, M. E.; Markley, A. L.; Dahesh, S.; Mitchell, D. A.; Bandeira, N.; Nizet, V.; Dixon, J. E.; Dorrestein, P. C. Clostridiolysin S, a Post-Translationally Modified Biotxin from Clostridium Botulinum. *J. Biol. Chem.* **2010**, *285* (36), 28220-8.
- (145) Molloy, E. M.; Casjens, S. R.; Cox, C. L.; Maxson, T.; Ethridge, N. A.; Margos, G.; Fingerle, V.; Mitchell, D. A. Identification of the Minimal Cytolytic Unit for Streptolysin S and an Expansion of the Toxin Family. *BMC Microbiol.* **2015**, *15* (1), 1-12.
- (146) Keiser, H.; Weissmann, G.; Bernheimer, A. W. Studies on Lysosomes. IV. Solubilization of Enzymes During Mitochondrial Swelling and Disruption of Lysosomes by Streptolysin S and Other Hemolytic Agents. *J. Cell. Biol.* **1964**, *22*, 101-13.
- (147) Hryniewicz, W.; Pryjma, J. Effect of Streptolysin S on Human and Mouse T and B Lymphocytes. *Infect. Immun.* **1977**, *16* (3), 730-3.
- (148) Bernheimer, A. W.; Schwartz, L. L. Lysosomal Disruption by Bacterial Toxins. *J. Bacteriol.* **1964**, *87* (5), 1100-4.
- (149) Bernheimer, A. W.; Schwartz, L. L. Effect of Staphylococcal and Other Bacterial Toxins on Platelets in Vitro. *J. Pathol. Bacteriol.* **1965**, *89*, 209-23.
- (150) Bernheimer, A. W. Disruption of Wall-Less Bacteria by Streptococcal and Staphylococcal Toxins. *J. Bacteriol.* **1966**, *91* (5), 1677-80.
- (151) Carr, A.; Sledjeski, D. D.; Podbielski, A.; Boyle, M. D.; Kreikemeyer, B. Similarities between Complement-Mediated and Streptolysin S-Mediated Hemolysis. *J. Biol. Chem.* **2001**, *276*.

- (152) Duncan, J. L.; Mason, L. Characteristics of Streptolysin S Hemolysis. *Infect. Immun.* **1976**, *14* (1), 77-82.
- (153) Higashi, D. L.; Biais, N.; Donahue, D. L.; Mayfield, J. A.; Tessier, C. R.; Rodriguez, K.; Ashfeld, B. L.; Luchetti, J.; Ploplis, V. A.; Castellino, F. J., et al. Activation of Band 3 Mediates Group A Streptococcus Streptolysin S-Based Beta-Haemolysis. *Nat. Microbiol.* **2016**, *1*, 15004.
- (154) Flaherty, R. A.; Puricelli, J. M.; Higashi, D. L.; Park, C. J.; Lee, S. W. Streptolysin S Promotes Programmed Cell Death and Enhances Inflammatory Signaling in Epithelial Keratinocytes During Group a Streptococcal Infection. *Infect. Immun.* **2015**.
- (155) Nizet, V. Streptococcal Beta-Hemolysins: Genetics and Role in Disease Pathogenesis. *Trends Microbiol.* **2002**, *10* (12), 575-80.
- (156) Humar, D.; Datta, V.; Bast, D. J.; Beall, B.; De Azavedo, J. C.; Nizet, V. Streptolysin S and Necrotising Infections Produced by Group G Streptococcus. *Lancet* **2002**, *359* (9301), 124-9.
- (157) Fuller, J. D.; Camus, A. C.; Duncan, C. L.; Nizet, V.; Bast, D. J.; Thune, R. L. Identification of a Streptolysin S-Associated Gene Cluster and Its Role in the Pathogenesis of Streptococcus Iniae Disease. *Infect. Immun.* **2002**, *70*.
- (158) Flanagan, J.; Collin, N.; Timoney, J.; Mitchell, T.; Mumford, J. A.; Chanter, N. Characterization of the Haemolytic Activity of Streptococcus Equi. *Microb. Pathog.* **1998**, *24* (4), 211-21.
- (159) Tabata, A.; Nakano, K.; Ohkura, K.; Tomoyasu, T.; Kikuchi, K.; Whiley, R. A. Novel Twin Streptolysin S-Like Peptides Encoded in the Sag Operon Homologue of Beta-Hemolytic Streptococcus Anginosus. *J. Bacteriol.* **2013**, *195*.
- (160) Cotter, P. D.; Draper, L. A.; Lawton, E. M.; Daly, K. M.; Groeger, D. S.; Casey, P. G.; Ross, R. P.; Hill, C. Listeriolysin S, a Novel Peptide Haemolysin Associated with a Subset of Lineage I Listeria Monocytogenes. *PLoS Pathog.* **2008**, *4* (9), e1000144.
- (161) Clayton, E. M.; Hill, C.; Cotter, P. D.; Ross, R. P. Real-Time PCR Assay to Differentiate Listeriolysin S-Positive and -Negative Strains of Listeria Monocytogenes. *Appl. Environ. Microbiol.* **2011**, *77* (1), 163-71.
- (162) Clayton, E. M.; Daly, K. M.; Guinane, C. M.; Hill, C.; Cotter, P. D.; Ross, P. R. Atypical Listeria Innocua Strains Possess an Intact LIPI-3. *BMC Microbiol.* **2014**, *14*, 58.
- (163) Letzel, A.-C.; Pidot, S. J.; Hertweck, C. Genome Mining for Ribosomally Synthesized and Post-Translationally Modified Peptides (RiPPs) in Anaerobic Bacteria. *BMC Genomics* **2014**, *15* (1), 1-21.
- (164) Molohon, K. J.; Melby, J. O.; Lee, J.; Evans, B. S.; Dunbar, K. L.; Bumpus, S. B.; Kelleher, N. L.; Mitchell, D. A. Structure Determination and Interception of Biosynthetic Intermediates for the

- Plantazolicin Class of Highly Discriminating Antibiotics. *ACS Chem. Biol.* **2011**, *6* (12), 1307-1313.
- (165) Deane, C. D.; Mitchell, D. A. Lessons Learned from the Transformation of Natural Product Discovery to a Genome-Driven Endeavor. *J. Ind. Microbiol. Biotechnol.* **2014**, *41* (2), 315-331.
- (166) Tietz, J. I.; Mitchell, D. A. Using Genomics for Natural Product Structure Elucidation. *Curr. Top. Med. Chem.* **2016**, *16* (15), 1645-94.
- (167) Hao, Y.; Blair, P. M.; Sharma, A.; Mitchell, D. A.; Nair, S. K. Insights into Methyltransferase Specificity and Bioactivity of Derivatives of the Antibiotic Plantazolicin. *ACS Chem. Biol.* **2015**, *10* (5), 1209-1216.
- (168) Lee, J.; Hao, Y.; Blair, P. M.; Melby, J. O.; Agarwal, V.; Burkhart, B. J.; Nair, S. K.; Mitchell, D. A. Structural and Functional Insight into an Unexpectedly Selective N-Methyltransferase Involved in Plantazolicin Biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110* (32), 12954-12959.
- (169) Sharma, A.; Blair, P. M.; Mitchell, D. A. Synthesis of Plantazolicin Analogues Enables Dissection of Ligand Binding Interactions of a Highly Selective Methyltransferase. *Org. Lett.* **2013**, *15* (19), 5076-5079.
- (170) Piwowarska, N. A.; Banala, S.; Overkleeft, H. S.; Suessmuth, R. D. Arg-Thz Is a Minimal Substrate for the N^α,N^α-Arginyl Methyltransferase Involved in the Biosynthesis of Plantazolicin. *Chem. Commun.* **2013**, *49* (91), 10703-10705.
- (171) Scholz, R.; Molohon, K. J.; Nachtigall, J.; Vater, J.; Markley, A. L.; Suessmuth, R. D.; Mitchell, D. A.; Borriss, R. Plantazolicin, a Novel Microcin B17/Streptolysin S-Like Natural Product from *Bacillus Amyloliquefaciens* FZB42. *J. Bacteriol.* **2011**, *193* (1), 215-224.
- (172) Kalyon, B.; Helaly, S. E.; Scholz, R.; Nachtigall, J.; Vater, J.; Borriss, R.; Suessmuth, R. D. Plantazolicin a and B: Structure Elucidation of Ribosomally Synthesized Thiazole/Oxazole Peptides from *Bacillus Amyloliquefaciens* FZB42. *Org. Lett.* **2011**, *13* (12), 2996-2999.
- (173) Deane, C. D.; Burkhart, B. J.; Blair, P. M.; Tietz, J. I.; Lin, A.; Mitchell, D. A. In Vitro Biosynthesis and Substrate Tolerance of the Plantazolicin Family of Natural Products. *ACS Chem. Biol.* **2016**.
- (174) Molohon, K. J.; Blair, P. M.; Park, S.; Doroghazi, J. R.; Maxson, T.; Hershfield, J. R.; Flatt, K. M.; Schroeder, N. E.; Ha, T.; Mitchell, D. A. Plantazolicin Is an Ultranarrow-Spectrum Antibiotic That Targets the *Bacillus Anthracis* Membrane. *ACS Infect. Dis.* **2015**, *2* (3), Ahead of Print.
- (175) Unsay, J. D.; Cosentino, K.; Subburaj, Y.; Garcia-Saez, A. J. Cardiolipin Effects on Membrane Structure and Dynamics. *Langmuir* **2013**, *29* (51), 15878-87.
- (176) Romantsov, T.; Guan, Z.; Wood, J. M. Cardiolipin and the Osmotic Stress Responses of Bacteria. *BBA-Biomembranes* **2009**, *1788* (10), 2092-2100.

- (177) Mileykovskaya, E.; Dowhan, W. Cardiolipin Membrane Domains in Prokaryotes and Eukaryotes. *BBA-Biomembranes* **2009**, *1788* (10), 2084-2091.
- (178) Deane, C. D.; Melby, J. O.; Molohon, K. J.; Susarrey, A. R.; Mitchell, D. A. Engineering Unnatural Variants of Plantazolicin through Codon Reprogramming. *ACS Chem. Biol.* **2013**, *8* (9), 1998-2008.
- (179) Banala, S.; Ensle, P.; Suessmuth, R. D. Total Synthesis of the Ribosomally Synthesized Linear Azole-Containing Peptide Plantazolicin a from *Bacillus Amyloliquefaciens*. *Angew. Chem. Int. Ed.* **2013**, *52* (36), 9518-9523.
- (180) Wilson, Z. E.; Fenner, S.; Ley, S. V. Total Syntheses of Linear Polythiazole/Oxazole Plantazolicin A and Its Biosynthetic Precursor Plantazolicin B. *Angew. Chem. Int. Ed.* **2015**, *54* (4), 1284-1288.
- (181) Wada, H.; Williams, H. E. L.; Moody, C. J. Total Synthesis of the Posttranslationally Modified Polyazole Peptide Antibiotic Plantazolicin A. *Angew. Chem. Int. Ed.* **2015**, *54* (50), 15147-15151.
- (182) Melby, J. O.; Dunbar, K. L.; Trinh, N. Q.; Mitchell, D. A. Selectivity, Directionality, and Promiscuity in Peptide Processing from a *Bacillus* Sp. Al Hakam Cyclodehydratase. *J. Am. Chem. Soc.* **2012**, *134*.
- (183) Dunbar, K. L.; Mitchell, D. A. Insights into the Mechanism of Peptide Cyclodehydrations Achieved through the Chemoenzymatic Generation of Amide Derivatives. *J. Am. Chem. Soc.* **2013**, *135* (23), 8692-8701.
- (184) Haft, D. H. A Strain-Variable Bacteriocin in *Bacillus Anthracis* and *Bacillus Cereus* with Repeated Cys-Xaa-Xaa Motifs. *Biol. Direct* **2009**, *4*, 15.
- (185) Dunbar, K. L.; Tietz, J. I.; Cox, C. L.; Burkhart, B. J.; Mitchell, D. A. Identification of an Auxiliary Leader Peptide-Binding Protein Required for Azoline Formation in Ribosomal Natural Products. *J. Am. Chem. Soc.* **2015**, *137* (24), 7672-7.
- (186) Begley, T. P.; Downs, D. M.; Ealick, S. E.; McLafferty, F. W.; Van Loon, A. P.; Taylor, S.; Campobasso, N.; Chiu, H. J.; Kinsland, C.; Reddick, J. J., et al. Thiamin Biosynthesis in Prokaryotes. *Arch. Microbiol.* **1999**, *171* (5), 293-300.
- (187) Begley, T. P.; Ealick, S. E.; McLafferty, F. W. Thiamin Biosynthesis: Still Yielding Fascinating Biological Chemistry. *Biochem. Soc. Trans.* **2012**, *40* (3), 555-60.
- (188) Burroughs, A. M.; Iyer, L. M.; Aravind, L. Natural History of the E1-Like Superfamily: Implication for Adenylation, Sulfur Transfer and Ubiquitin Conjugation. *Proteins* **2009**, *75* (4), 895-910.
- (189) Chopra, L.; Singh, G.; Choudhary, V.; Sahoo, D. K. Sonorensin: An Antimicrobial Peptide, Belonging to the Heterocycloanthracin Subfamily of Bacteriocins, from a New Marine Isolate, *Bacillus Sonorensis* Mt93. *Appl. Environ. Microbiol.* **2014**, *80* (10), 2981-90.

- (190) Chopra, L.; Singh, G.; Jena, K. K.; Verma, H.; Sahoo, D. K. Bioprocess Development for the Production of Sonorensin by *Bacillus Sonorensis* MT93 and Its Application as a Food Preservative. *Bioresour. Technol.* **2014**, *175C*, 358-366.
- (191) Chopra, L.; Singh, G.; Kumar Jena, K.; Sahoo, D. K. Sonorensin: A New Bacteriocin with Potential of an Anti-Biofilm Agent and a Food Biopreservative. *Sci. Rep.* **2015**, *5*, 13412.
- (192) Divekar, P. V.; Read, G.; Vining, L. C. Caerulomycin, a New Antibiotic from *Streptomyces Caeruleus* Baldacci. II. Structure. *Can. J. Chemistry* **1967**, *45* (11), 1215-1223.
- (193) Bycroft, B. W.; Pinchin, R. Structure of Althiomycin, a Highly Modified Peptide Antibiotic. *J. Chem. Soc., Chem. Commun.* **1975**, (4), 121-122.
- (194) Shindo, K.; Yamagishi, Y.; Okada, Y.; Kawai, H. Collismycins a and B, Novel Non-Steroidal Inhibitors of Dexamethasone-Glucocorticoid Receptor Binding. *J. Antibiot.* **1994**, *47* (9), 1072-4.
- (195) Hashimoto, M.; Komori, T.; Kamiya, T. Nocardicin a, a New Monocyclic Beta-Lactam Antibiotic Ii. Structure Determination of Nocardicins a and B. *J. Antibiot.* **1976**, *29* (9), 890-901.
- (196) Liu, N.; Song, L.; Liu, M.; Shang, F.; Anderson, Z.; Fox, D. J.; Challis, G. L.; Huang, Y. Unique Post-Translational Oxime Formation in the Biosynthesis of the Azolemycin Complex of Novel Ribosomal Peptides from *Streptomyces* Sp. Fxj1.264. *Chem. Sci.* **2016**, *7* (1), 482-488.
- (197) Anderson, Z. J.; Fox, D. J. Total Synthesis of the Azolemycins. *Org. Biomol. Chem.* **2016**, *14* (4), 1450-1454.
- (198) Bentley, S. D.; Chater, K. F.; Cerdeno-Tarraga, A. M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. A.; Kieser, H.; Harper, D., et al. Complete Genome Sequence of the Model Actinomycete *Streptomyces Coelicolor* A3(2). *Nature* **2002**, *417* (6885), 141-147.
- (199) Zarins-Tutt, J. S.; Barberi, T. T.; Gao, H.; Mearns-Spragg, A.; Zhang, L.; Newman, D. J.; Goss, R. J. Prospecting for New Bacterial Metabolites: A Glossary of Approaches for Inducing, Activating and Upregulating the Biosynthesis of Bacterial Cryptic or Silent Natural Products. *Nat. Prod. Rep.* **2016**, *33* (1), 54-72.
- (200) Abdelmohsen, U. R.; Grkovic, T.; Balasubramanian, S.; Kamel, M. S.; Quinn, R. J.; Hentschel, U. Elicitation of Secondary Metabolism in Actinomycetes. *Biotechnol. Adv.* **2015**, *33* (6, Part 1), 798-811.
- (201) Onaka, H.; Tabata, H.; Igarashi, Y.; Sato, Y.; Furumai, T. Goadsporin, a Chemical Substance Which Promotes Secondary Metabolism and Morphogenesis in Streptomycetes. I. Purification and Characterization. *J. Antibiot.* **2001**, *54* (12), 1036-44.
- (202) Fujii, K.; Yahashi, Y.; Nakano, T.; Imanishi, S.; Baldia, S. F.; Harada, K.-i. Simultaneous Detection and Determination of the Absolute Configuration of Thiazole-Containing Amino Acids in a Peptide. *Tetrahedron* **2002**, *58* (34), 6873-6879.

- (203) Igarashi, Y.; Kan, Y.; Fujii, K.; Fujita, T.; Harada, K.-I.; Naoki, H.; Tabata, H.; Onaka, H.; Furumai, T. Goadsporin, a Chemical Substance Which Promotes Secondary Metabolism and Morphogenesis in Streptomycetes. II. Structure Determination. *J. Antibiot.* **2001**, *54* (12), 1045-1053.
- (204) Ozaki, T.; Kurokawa, Y.; Hayashi, S.; Oku, N.; Asamizu, S.; Igarashi, Y.; Onaka, H. Insights into the Biosynthesis of Dehydroalanines in Goadsporin. *ChemBioChem* **2016**, *17* (3), Ahead of Print.
- (205) Ortega, M. A.; Hao, Y.; Zhang, Q.; Walker, M. C.; van der Donk, W. A.; Nair, S. K. Structure and Mechanism of the tRNA-Dependent Lantibiotic Dehydratase Nisb. *Nature* **2015**, *517* (7535), 509-12.
- (206) Onaka, H. Biosynthesis of Heterocyclic Antibiotics in Actinomycetes and an Approach to Synthesize the Natural Compounds. *Actinomycetologica* **2006**, *20* (2), 62-71.
- (207) Onaka, H. Biosynthesis of Indolocarbazole and Goadsporin, Two Different Heterocyclic Antibiotics Produced by Actinomycetes. *Biosci. Biotechnol. Biochem.* **2009**, *73* (10), 2149-2155.
- (208) Haginaka, K.; Asamizu, S.; Ozaki, T.; Igarashi, Y.; Furumai, T.; Onaka, H. Genetic Approaches to Generate Hyper-Producing Strains of Goadsporin: The Relationships between Productivity and Gene Duplication in Secondary Metabolite Biosynthesis. *Biosci. Biotechnol. Biochem.* **2014**, *78* (3), 394-399.
- (209) Onaka, H.; Ozaki, T.; Mori, Y.; Izawa, M.; Hayashi, S.; Asamizu, S. Mycolic Acid-Containing Bacteria Activate Heterologous Secondary Metabolite Expression in Streptomyces Lividans. *J. Antibiot.* **2015**, *68* (9), 594-597.
- (210) Chatterjee, C.; Paul, M.; Xie, L.; van der Donk, W. A. Biosynthesis and Mode of Action of Lantibiotics. *Chem Rev* **2005**, *105*.
- (211) Knerr, P. J.; van der Donk, W. A. Discovery, Biosynthesis, and Engineering of Lantipeptides. *Annu. Rev. Biochem.* **2012**, *81*, 479-505.
- (212) Repka, L. M.; Chekan, J. R.; Nair, S. K.; van der Donk, W. A. Mechanistic Understanding of Lanthipeptide Biosynthetic Enzymes. *Chem. Rev.* **2017**.
- (213) Onaka, H.; Nakaho, M.; Hayashi, K.; Igarashi, Y.; Furumai, T. Cloning and Characterization of the Goadsporin Biosynthetic Gene Cluster from Streptomyces Sp. TP-A0584. *Microbiology* **2005**, *151*, 3923-3933.
- (214) Davies, J. Are Antibiotics Naturally Antibiotics? *J. Ind. Microbiol. Biotechnol.* **2006**, *33* (7), 496-499.
- (215) Romero, D.; Traxler, M. F.; López, D.; Kolter, R. Antibiotics as Signal Molecules. *Chem. Rev.* **2011**, *111* (9), 5492-5505.
- (216) Berdy, J. Recent Developments of Antibiotic Research and Classification of Antibiotics According to Chemical Structure. *Adv. Appl. Microbiol.* **1974**, *18* (0), 309-406.

- (217) Pucci, M. J.; Bronson, J. J.; Barrett, J. F.; DenBleyker, K. L.; Discotto, L. F.; Fung-Tomc, J. C.; Ueda, Y. Antimicrobial Evaluation of Nocathiacins, a Thiazole Peptide Class of Antibiotics. *Antimicrob. Agents Chemother.* **2004**, *48* (10), 3697-3701.
- (218) Haste, N. M.; Thienphrapa, W.; Tran, D. N.; Loesgen, S.; Sun, P.; Nam, S.-J.; Jensen, P. R.; Fenical, W.; Sakoulas, G.; Nizet, V., et al. Activity of the Thiopeptide Antibiotic Nosiheptide against Contemporary Strains of Methicillin-Resistant Staphylococcus Aureus. *J. Antibiot.* **2012**, *65* (12), 593-598.
- (219) Su, T. L. Micrococcin, an Antibacterial Substance Formed by a Strain of Micrococcus. *Br. J. Exp. Pathol.* **1948**, *29* (5), 473-81.
- (220) Hughes, R. A.; Moody, C. J. From Amino Acids to Heteroaromatics—Thiopeptide Antibiotics, Nature's Heterocyclic Peptides. *Angew. Chem. Int. Ed.* **2007**, *46* (42), 7930-7954.
- (221) Nicolaou, K. C. How Thiostrepton Was Made in the Laboratory. *Angew. Chem. Int. Ed. Engl.* **2012**, *51* (50), 12414-36.
- (222) Nicolaou, K. C.; Chen, J. S.; Edmonds, D. J.; Estrada, A. A. Recent Advances in the Chemistry and Biology of Naturally Occurring Antibiotics. *Angew. Chem. Int. Ed.* **2009**, *48* (4), 660-719.
- (223) Xavier, J.-B.; Fernando, A.; Mercedes, A. Chiral Thiazoline and Thiazole Building Blocks for the Synthesis of Peptide-Derived Natural Products. *Curr. Top. Med. Chem.* **2014**, *14* (10), 1244-1256.
- (224) Just-Baringo, X.; Albericio, F.; Alvarez, M. Thiopeptide Engineering: A Multidisciplinary Effort Towards Future Drugs. *Angew. Chem. Int. Ed.* **2014**, *53* (26), 6602-16.
- (225) Mocek, U.; Zeng, Z.; O'Hagan, D.; Zhou, P.; Fan, L. D. G.; Beale, J. M.; Floss, H. G. Biosynthesis of the Modified Peptide Antibiotic Thiostrepton in *Streptomyces Azureus* and *Streptomyces Laurentii*. *J. Am. Chem. Soc.* **1993**, *115* (18), 7992-8001.
- (226) Li, C.; Kelly, W. L. Recent Advances in Thiopeptide Antibiotic Biosynthesis. *Nat. Prod. Rep.* **2010**, *27* (2), 153-64.
- (227) Walsh, C. T.; Acker, M. G.; Bowers, A. A. Thiazolyl Peptide Antibiotic Biosynthesis: A Cascade of Post-Translational Modifications on Ribosomal Nascent Proteins. *J. Biol. Chem.* **2010**, *285* (36), 27525-31.
- (228) Wang, S.; Zhou, S.; Liu, W. Opportunities and Challenges from Current Investigations into the Biosynthetic Logic of Nosiheptide-Represented Thiopeptide Antibiotics. *Curr. Opin. Chem. Biol.* **2013**, *17* (4), 626-34.
- (229) Zhang, Q.; Liu, W. Biosynthesis of Thiopeptide Antibiotics and Their Pathway Engineering. *Nat. Prod. Rep.* **2013**, *30* (2), 218-26.

- (230) Luo, X.; Zambaldo, C.; Liu, T.; Zhang, Y.; Xuan, W.; Wang, C.; Reed, S. A.; Yang, P.-Y.; Wang, R. E.; Javahishvili, T., et al. Recombinant Thiopeptides Containing Noncanonical Amino Acids. *Proc. Nat. Acad. Sci. U.S.A.* **2016**.
- (231) Pearce, C. J.; Rinehart, K. L. Berninamycin Biosynthesis. 1. Origin of the Dehydroalanine Residues. *J. Am. Chem. Soc.* **1979**, *101* (17), 5069-5070.
- (232) Houck, D. R.; Chen, L. C.; Keller, P. J.; Beale, J. M.; Floss, H. G. Biosynthesis of the Modified Peptide Antibiotic Nosiheptide in *Streptomyces Actuosus*. *J. Am. Chem. Soc.* **1987**, *109* (4), 1250-1252.
- (233) Houck, D. R.; Chen, L. C.; Keller, P. J.; Beale, J. M.; Floss, H. G. Biosynthesis of the Modified Peptide Antibiotic Nosiheptide in *Streptomyces Actuosus*. *J. Am. Chem. Soc.* **1988**, *110* (17), 5800-5806.
- (234) Zhou, P.; O'Hagan, D.; Mocek, U.; Zeng, Z.; Yuen, L. D.; Frenzel, T.; Unkefer, C. J.; Beale, J. M.; Floss, H. G. Biosynthesis of the Antibiotic Thiostrepton. Methylation of Tryptophan in the Formation of the Quinaldic Acid Moiety by Transfer of the Methionine Methyl Group with Net Retention of Configuration. *J. Am. Chem. Soc.* **1989**, *111* (18), 7274-7276.
- (235) Favret, M. E.; Paschal, J. W.; Elzey, T. K.; Boeck, L. D. Biosynthesis of Thiopeptide Antibiotic A10255: Incorporation of Isotopically-Labeled Precursors. *J. Antibiot.* **1992**, *45* (9), 1499-511.
- (236) Mocek, U.; Knaggs, A. R.; Tsuchiya, R.; Nguyen, T.; Beale, J. M.; Floss, H. G. Biosynthesis of the Modified Peptide Antibiotic Nosiheptide in *Streptomyces Actuosus*. *J. Am. Chem. Soc.* **1993**, *115* (17), 7557-7568.
- (237) Lau, R. C. M.; Rinehart, K. L. Biosynthesis of Berninamycin: Incorporation of ¹³C-Labeled Amino Acids. *J. Am. Chem. Soc.* **1995**, *117* (29), 7606-7610.
- (238) Fate, G. D.; Benner, C. P.; Grode, S. H.; Gilbertson, T. J. The Biosynthesis of Sulfomycin Elucidated by Isotopic Labeling Studies. *J. Am. Chem. Soc.* **1996**, *118* (46), 11363-11368.
- (239) Priestley, N. D.; Smith, T. M.; Shipley, P. R.; Floss, H. G. Studies on the Biosynthesis of Thiostrepton: 4-(1-Hydroxyethyl)Quinoline-2-Carboxylate as a Free Intermediate on the Pathway to the Quinaldic Acid Moiety. *Bioorg. Med. Chem.* **1996**, *4* (7), 1135-1147.
- (240) De Pietro, M. T.; Marazzi, A.; Sosio, M.; Donadio, S.; Lancini, G. Biosynthesis of the Thiazolylpeptide Antibiotic GE2270. *J. Antibiot (Tokyo)* **2001**, *54* (12), 1066-71.
- (241) Singh, S. B.; Herath, K.; Yu, N. X.; Walker, A. A.; Connors, N. Biosynthetic Studies of Nocathiacin-I. *Tetrahedron Lett.* **2008**, *49* (43), 6265-6268.
- (242) Bycroft, B. W.; Gowland, M. S. The Structures of the Highly Modified Peptide Antibiotics Micrococin P1 and P2. *J. Chem. Soc., Chem. Commun.* **1978**, (6), 256-258.

- (243) Li, Y.; Dosch, D. C.; Woodman, R. H.; Floss, H. G.; Strohl, W. R. Transcriptional Organization and Regulation of the Nosiheptide Resistance Gene in *Streptomyces Actuosus*. *J. Ind. Microbiol.* **1991**, *8* (1), 1-12.
- (244) Floss, H. G.; Beale, J. M. Biosynthetic Studies on Antibiotics. *Angew. Chem. Int. Ed.* **1989**, *28* (2), 146-177.
- (245) Smith, T. M.; Priestley, N. D.; Knaggs, A. R.; Nguyen, T.; Floss, H. G. 3,4-Dimethylindole-2-Carboxylate and 4-(1-Hydroxyethyl)Quinoline-2-Carboxylate Activating Enzymes from the Nosiheptide and Thiostrepton Producers, *Streptomyces Actuosus* and *Streptomyces Laurentii*. *J. Chem. Soc., Chem. Commun.* **1993**, (21), 1612-1614.
- (246) Carnio, M. C.; Stachelhaus, T.; Francis, K. P.; Scherer, S. Pyridinyl Polythiazole Class Peptide Antibiotic Micrococcin P1, Secreted by Foodborne *Staphylococcus Equorum* Ws2733, Is Biosynthesized Nonribosomally. *Eur. J. Biochem.* **2001**, *268* (24), 6390-401.
- (247) Roy, R. S.; Gehring, A. M.; Milne, J. C.; Belshaw, P. J.; Walsh, C. T. Thiazole and Oxazole Peptides: Biosynthesis and Molecular Machinery. *Nat. Prod. Rep.* **1999**, *16* (2), 249-63.
- (248) Marahiel, M. A.; Stachelhaus, T.; Mootz, H. D. Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem. Rev.* **1997**, *97* (7), 2651-2674.
- (249) Bennallack, P. R.; Bewley, K. D.; Burlingame, M. A.; Robison, R. A.; Miller, S. M.; Griffiths, J. S. Reconstitution and Minimization of a Micrococcin Biosynthetic Pathway in *Bacillus Subtilis*. *J. Bacteriol.* **2016**.
- (250) Bennallack, P. R.; Burt, S. R.; Heder, M. J.; Robison, R. A.; Griffiths, J. S. Characterization of a Novel Plasmid-Borne Thiopeptide Gene Cluster in *Staphylococcus Epidermidis* Strain 115. *J. Bacteriol.* **2014**, *196* (24), 4344-4350, 8 pp.
- (251) Liao, R.; Liu, W. Thiostrepton Maturation Involving a Deesterification–Amidation Way to Process the C-Terminally Methylated Peptide Backbone. *J. Am. Chem. Soc.* **2011**, *133* (9), 2852-2855.
- (252) Liu, W.; Xue, Y.; Ma, M.; Wang, S.; Liu, N.; Chen, Y. Multiple Oxidative Routes Towards the Maturation of Nosiheptide. *ChemBioChem* **2013**, *14* (13), 1544-1547.
- (253) Yu, Y.; Guo, H.; Zhang, Q.; Duan, L.; Ding, Y.; Liao, R.; Lei, C.; Shen, B.; Liu, W. Nosa Catalyzing Carboxyl-Terminal Amide Formation in Nosiheptide Maturation Via an Enamine Dealkylation on the Serine-Extended Precursor Peptide. *J. Am. Chem. Soc.* **2010**, *132* (46), 16324-16326.
- (254) Flinspach, K.; Kapitzke, C.; Tocchetti, A.; Sosio, M.; Apel, A. K. Heterologous Expression of the Thiopeptide Antibiotic GE2270 from *Planobispora Rosea* ATCC 53733 in *Streptomyces Coelicolor* Requires Deletion of Ribosomal Genes from the Expression Construct. *PLoS One* **2014**, *9* (3), e90499.

- (255) Tocchetti, A.; Maffioli, S.; Iorio, M.; Alt, S.; Mazzei, E.; Brunati, C.; Sosio, M.; Donadio, S. Capturing Linear Intermediates and C-Terminal Variants During Maturation of the Thiopeptide GE2270. *Chem. Biol.* **2013**, *20* (8), 1067-1077.
- (256) Ding, Y.; Yu, Y.; Pan, H.; Guo, H.; Li, Y.; Liu, W. Moving Posttranslational Modifications Forward to Biosynthesize the Glycosylated Thiopeptide Nocathiacin I in *Nocardia* Sp. ATCC 202099. *Mol. BioSyst.* **2010**, *6* (7), 1180-1185.
- (257) Wei, M.; Deng, J.; Wang, S.; Liu, N.; Chen, Y. A Simple Reverse Genetics Approach to Elucidating the Biosynthetic Pathway of Nocathiacin. *Biotechnol. Lett.* **2011**, *33* (3), 585-591.
- (258) Wang, J.; Yu, Y.; Tang, K.; Liu, W.; He, X.; Huang, X.; Deng, Z. Identification and Analysis of the Biosynthetic Gene Cluster Encoding the Thiopeptide Antibiotic Cyclothiazomycin in *Streptomyces Hygroscopicus* 10-22. *Appl. Environ. Microbiol.* **2010**, *76* (7), 2335-2344.
- (259) Zhang, P.; Wu, H.; Chen, X. L.; Deng, Z.; Bai, L.; Pang, X. Regulation of the Biosynthesis of Thiopeptide Antibiotic Cyclothiazomycin by the Transcriptional Regulator Shjg8833 in *Streptomyces Hygroscopicus* 5008. *Microbiology* **2014**, *160*, 1379-92.
- (260) Engelhardt, K.; Degnes, K. F.; Zotchev, S. B. Isolation and Characterization of the Gene Cluster for Biosynthesis of the Thiopeptide Antibiotic TP-1161. *Appl. Environ. Microbiol.* **2010**, *76* (21), 7093-7101.
- (261) Young, T. S.; Walsh, C. T. Identification of the Thiazolyl Peptide GE37468 Gene Cluster from *Streptomyces* ATCC 55365 and Heterologous Expression in *Streptomyces Lividans*. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108* (32), 13053-8.
- (262) Malcolmson, S. J.; Young, T. S.; Ruby, J. G.; Skewes-Cox, P.; Walsh, C. T. The Posttranslational Modification Cascade to the Thiopeptide Berninamycin Generates Linear Forms and Altered Macrocyclic Scaffolds. *Proc. Nat. Acad. Sci. U.S.A.* **2013**, *110* (21), 8483-8.
- (263) Donia, Mohamed S.; Cimermanic, P.; Schulze, Christopher J.; Wieland Brown, Laura C.; Martin, J.; Mitreva, M.; Clardy, J.; Linington, Roger G.; Fischbach, Michael A. A Systematic Analysis of Biosynthetic Gene Clusters in the Human Microbiome Reveals a Common Family of Antibiotics. *Cell* **2014**, *158* (6), 1402-1414.
- (264) Sakai, K.; Komaki, H.; Gono, T. Identification and Functional Analysis of the Nocardithiocin Gene Cluster in *Nocardia Pseudobrasiliensis*. *PLoS One* **2015**, *10* (11), e0143264/1-e0143264/13.
- (265) Hayashi, S.; Ozaki, T.; Asamizu, S.; Ikeda, H.; Omura, S.; Oku, N.; Igarashi, Y.; Tomoda, H.; Onaka, H. Genome Mining Reveals a Minimum Gene Set for the Biosynthesis of 32-Membered Macrocyclic Thiopeptides Lactazoles. *Chem. Biol.* **2014**, *21* (5), 679-688.
- (266) Cox, C. L.; Tietz, J. I.; Sokolowski, K.; Melby, J. O.; Doroghazi, J. R.; Mitchell, D. A. Nucleophilic 1,4-Additions for Natural Product Discovery. *ACS Chem. Biol.* **2014**, *9* (9), 2014-22.

- (267) Bowers, A. A.; Acker, M. G.; Koglin, A.; Walsh, C. T. Manipulation of Thiocillin Variants by Prepeptide Gene Replacement: Structure, Conformation, and Activity of Heterocycle Substitution Mutants. *J. Am. Chem. Soc.* **2010**, *132* (21), 7519-27.
- (268) Xi, J.; Ge, Y.; Kinsland, C.; McLafferty, F. W.; Begley, T. P. Biosynthesis of the Thiazole Moiety of Thiamin in Escherichia Coli: Identification of an Acyldisulfide-Linked Protein-Protein Conjugate That Is Functionally Analogous to the Ubiquitin/E1 Complex. *Proc. Nat. Acad. Sci. U.S.A.* **2001**, *98* (15), 8513-8518.
- (269) Garg, N.; Salazar-Ocampo, L. M. A.; van der Donk, W. A. In Vitro Activity of the Nisin Dehydratase Nisb. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110* (18), 7258-7263, S7258/1-S7258/11.
- (270) Wever, W. J.; Bogart, J. W.; Baccile, J. A.; Chan, A. N.; Schroeder, F. C.; Bowers, A. A. Chemoenzymatic Synthesis of Thiazolyl Peptide Natural Products Featuring an Enzyme-Catalyzed Formal [4 + 2] Cycloaddition. *J. Am. Chem. Soc.* **2015**, *137* (10), 3494-7.
- (271) Guo, H.; Wang, J.; Li, Y.; Yu, Y.; Zheng, Q.; Wu, J.; Liu, W. Insight into Bicyclic Thiopeptide Biosynthesis Benefited from Development of a Uniform Approach for Molecular Engineering and Production Improvement. *Chem. Sci.* **2014**, *5* (1), 240-246.
- (272) Pierre, S.; Guillot, A.; Benjdia, A.; Sandström, C.; Langella, P.; Berteau, O. Thiostrepton Tryptophan Methyltransferase Expands the Chemistry of Radical SAM Enzymes. *Nat. Chem. Biol.* **2012**, *8* (12), 957-959.
- (273) Benjdia, A.; Pierre, S.; Gherasim, C.; Guillot, A.; Carmona, M.; Amara, P.; Banerjee, R.; Berteau, O. The Thiostrepton a Tryptophan Methyltransferase TsrM Catalyses a Cob(II)Alamin-Dependent Methyl Transfer Reaction. *Nat. Commun.* **2015**, *6*, 8377.
- (274) Frenzel, T.; Zhou, P.; Floss, H. G. Formation of 2-Methyltryptophan in the Biosynthesis of Thiostrepton: Isolation of S-Adenosylmethionine: Tryptophan 2-Methyltransferase. *Arch. Biochem. Biophys.* **1990**, *278* (1), 35-40.
- (275) Duan, L.; Wang, S.; Liao, R.; Liu, W. Insights into Quinaldic Acid Moiety Formation in Thiostrepton Biosynthesis Facilitating Fluorinated Thiopeptide Generation. *Chem. Biol.* **2012**, *19* (4), 443-448.
- (276) Li, C.; Zhang, F.; Kelly, W. L. Mutagenesis of the Thiostrepton Precursor Peptide at Thr7 Impacts Both Biosynthesis and Function. *Chem. Comm.* **2012**, *48* (4), 558-560.
- (277) Zhang, F.; Li, C.; Kelly, W. L. Thiostrepton Variants Containing a Contracted Quinaldic Acid Macrocycle Result from Mutagenesis of the Second Residue. *ACS Chem. Biol.* **2016**, *11* (2), 415-24.

- (278) Zheng, Q.; Wang, S.; Duan, P.; Liao, R.; Chen, D.; Liu, W. An α/β -Hydrolase Fold Protein in the Biosynthesis of Thiostrepton Exhibits a Dual Activity for Endopeptidyl Hydrolysis and Epoxide Ring Opening/Macrocyclization. *Proc. Natl. Acad. Sci. USA* **2016**, *113* (50), 14318-14323.
- (279) Sicoli, G.; Mouesca, J. M.; Zeppieri, L.; Amara, P.; Martin, L.; Barra, A. L.; Fontecilla-Camps, J. C.; Gambarelli, S.; Nicolet, Y. Fine-Tuning of a Radical-Based Reaction by Radical S-Adenosyl-L-Methionine Tryptophan Lyase. *Science* **2016**, *351* (6279), 1320-3.
- (280) Ji, X.; Li, Y.; Jia, Y.; Ding, W.; Zhang, Q. Mechanistic Insights into the Radical S-Adenosyl-L-Methionine Enzyme NosI from a Substrate Analogue and the Shunt Products. *Angew. Chem. Int. Ed. Engl.* **2016**, *55* (10), 3334-7.
- (281) Nicolet, Y.; Zeppieri, L.; Amara, P.; Fontecilla-Camps, J. C. Crystal Structure of Tryptophan Lyase (NosI): Evidence for Radical Formation at the Amino Group of Tryptophan. *Angew. Chem. Int. Ed.* **2014**, *53* (44), 11840-4.
- (282) Zhang, Q.; Li, Y.; Chen, D.; Yu, Y.; Duan, L.; Shen, B.; Liu, W. Radical-Mediated Enzymatic Carbon Chain Fragmentation-Recombination. *Nat. Chem. Biol.* **2011**, *7* (3), 154-60.
- (283) Ji, X.; Li, Y.; Ding, W.; Zhang, Q. Substrate-Tuned Catalysis of the Radical S-Adenosyl-L-Methionine Enzyme NosI Involved in Nosiheptide Biosynthesis. *Angew. Chem. Int. Ed.* **2015**, *54* (31), 9021-4.
- (284) Zhang, Q.; Chen, D.; Lin, J.; Liao, R.; Tong, W.; Xu, Z.; Liu, W. Characterization of NosI Involved in Thiopeptide Nocathiacin I Biosynthesis: A [4Fe-4S] Cluster and the Catalysis of a Radical S-Adenosylmethionine Enzyme. *J. Biol. Chem.* **2011**, *286* (24), 21287-94.
- (285) Wang, Y.; Liu, S.; Yao, P.; Yu, Y.; Zhang, Y.; Lan, W.; Wang, C.; Ding, J.; Liu, W.; Cao, C. Crystallographic Analysis of Nosa, Which Catalyzes Terminal Amide Formation in the Biosynthesis of Nosiheptide. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2015**, *71* (Pt 8), 1033-7.
- (286) Liu, S.; Guo, H.; Zhang, T.; Han, L.; Yao, P.; Zhang, Y.; Rong, N.; Yu, Y.; Lan, W.; Wang, C., et al. Structure-Based Mechanistic Insights into Terminal Amide Synthase in Nosiheptide-Represented Thiopeptides Biosynthesis. *Sci. Rep.* **2015**, *5*, 12744.
- (287) Kulathila, R.; Merkler, K. A.; Merkler, D. J. Enzymatic Formation of C-Terminal Amides. *Nat. Prod. Rep.* **1999**, *16* (2), 145-54.
- (288) Müller, I.; Weinig, S.; Steinmetz, H.; Kunze, B.; Veluthoor, S.; Mahmud, T.; Müller, R. A Unique Mechanism for Methyl Ester Formation Via an Amide Intermediate Found in Myxobacteria. *ChemBioChem* **2006**, *7* (8), 1197-1205.

- (289) Merkler, D. J. C-Terminal Amidated Peptides: Production by the in Vitro Enzymatic Amidation of Glycine-Extended Peptides and the Importance of the Amide to Bioactivity. *Enzyme Microb. Tech.* **1994**, *16* (6), 450-456.
- (290) Donovan, R.; Pagano, J. F.; Stout, H. A.; Weinstein, M. J. Thiostrepton, a New Antibiotic. I. In Vitro Studies. *Antibiot. Annu.* **1955**, *3*, 554-9.
- (291) Dutcher, J. D.; Vandeputte, J. Thiostrepton, a New Antibiotic. II. Isolation and Chemical Characterization. *Antibiot. Annu.* **1955**, *3*, 560-1.
- (292) Jambor, W. P.; Steinberg, B. A.; Suydam, L. O. Thiostrepton, a New Antibiotic. III. In Vivo Studies. *Antibiot. Annu.* **1955**, *3*, 562-5.
- (293) Anderson, B.; Hodgkin, D. C.; Viswamitra, M. A. The Structure of Thiostrepton. *Nature* **1970**, *225* (5229), 233-5.
- (294) Cron, M. J.; Whitehead, D. G.; Hooper, I. R.; Heinemann, B.; Lein, J. Bryamycin, a New Antibiotic. *Antibiot. Chemother.* **1956**, *6* (1), 63-7.
- (295) Jones, W. F., Jr.; Finland, M. In Vitro Observations on a-8506 (Bryamycin). *Antibiot. Chemother.* **1958**, *8*, 387-91.
- (296) Heinemann, B.; Hooper, I. R.; Cron, M. J. Thiactin (Bryamycin). GB790521, 1958.
- (297) Bodanszky, M.; Dutcher, J. D.; Williams, N. J. The Establishment of the Identity of Thiostrepton with Thiactin (Bryamycin). *J. Antibiot., Ser. A* **1963**, *Ser. A 16* (2), 76-9.
- (298) Tori, K.; Tokura, K.; Yoshimura, Y.; Okabe, K.; Otsuka, H.; Inagaki, F.; Miyazawa, T. ¹H NMR Spectral Evidence for the Structure and Conformation of Peptide Antibiotic Siomycin-A. *J. Antibiot.* **1979**, *32* (10), 1072-7.
- (299) Ebata, M.; Miyazaki, K.; Otsuka, H. Studies on Siomycin. I. Physicochemical Properties of Siomycins a, B and C. *J. Antibiot.* **1969**, *22* (8), 364-8.
- (300) Puar, M. S.; Ganguly, A. K.; Afonso, A.; Brambilla, R.; Mangiaracina, P.; Sarre, O.; MacFarlane, R. D. Sch 18640. A New Thiostrepton-Type Antibiotic. *J. Am. Chem. Soc.* **1981**, *103* (17), 5231-5233.
- (301) Hensens, O. D.; Albers-Schönberg, G. Total Structure of the Peptide Antibiotic Components of Thiopeptin by ¹H and ¹³C NMR Spectroscopy. *Tetrahedron Lett.* **1978**, *19* (39), 3649-3652.
- (302) Miyairi, N.; Miyoshi, T.; Aoki, H.; Kosaka, M.; Ikushima, H. Studies on Thiopeptin Antibiotics. I. Characteristics of Thiopeptin B. *J. Antibiot.* **1970**, *23* (3), 113-9.
- (303) Puar, M. S.; Chan, T. M.; Hegde, V.; Patel, M.; Bartner, P.; Ng, K. J.; Pramanik, B. N.; MacFarlane, R. D. Sch 40832: A Novel Thiostrepton from *Micromonospora Carbonacea*. *J. Antibiot.* **1998**, *51* (2), 221-4.

- (304) Schoof, S.; Arndt, H.-D. D-Cysteine Occurrence in Thiostrepton May Not Necessitate an Epimerase. *Chem. Comm.* **2009**, (46), 7113-7115.
- (305) Zhang, F.; Kelly, W. L. Saturation Mutagenesis of Tsra Ala4 Unveils a Highly Mutable Residue of Thiostrepton A. *ACS Chem. Biol.* **2015**, *10* (4), 998-1009.
- (306) Wang, S.; Zheng, Q.; Wang, J.; Zhao, Z.; Li, Q.; Yu, Y.; Wang, R.; Liu, W. Target-Oriented Design and Biosynthesis of Thiostrepton-Derived Thiopeptide Antibiotics with Improved Pharmaceutical Properties. *Org. Chem. Front.* **2015**, *2* (2), 106-109.
- (307) Li, C.; Zhang, F.; Kelly, W. L. Heterologous Production of Thiostrepton A and Biosynthetic Engineering of Thiostrepton Analogs. *Mol. Biosyst.* **2011**, *7* (1), 82-90.
- (308) Zheng, Q.; Wang, Q.; Wang, S.; Wu, J.; Gao, Q.; Liu, W. Thiopeptide Antibiotics Exhibit a Dual Mode of Action against Intracellular Pathogens by Affecting Both Host and Microbe. *Chem. Biol.* **2015**, *22* (8), 1002-1007.
- (309) Endo, T.; Yonehara, H. Identity of Multhiomycin with Nosiheptide. *J. Antibiot.* **1978**, *31* (6), 623-5.
- (310) Pascard, C.; Ducruix, A.; Lunel, J.; Prange, T. Highly Modified Cysteine-Containing Antibiotics. Chemical Structure and Configuration of Nosiheptide. *J. Am. Chem. Soc.* **1977**, *99* (19), 6418-23.
- (311) Prange, T.; Ducruix, A.; Pascard, C.; Lunel, J. Structure of Nosiheptide, a Polythiazole-Containing Antibiotic. *Nature* **1977**, *265* (5590), 189-90.
- (312) Tanaka, T.; Endo, T.; Shimazu, A.; Yoshida, R.; Suzuki, Y.; Otake, N.; Yonehara, H. New Antibiotic, Multhiomycin. *J. Antibiot.* **1970**, *23* (5), 231-7.
- (313) Ma, M.; Xue, Y.; Liu, W.; Zhang, H.; Kong, L.; Wang, S.; Chen, Y. Directly Utilizing an Endogenous Gene to Dissect Regulatory Elements in the Biosynthetic Gene Cluster of Nosiheptide. *Chem. Comm.* **2014**, *50* (72), 10430-10433.
- (314) Steinberg, D. A.; Bernan, V. S.; Montenegro, D. A.; Abbanat, D. R.; Pearce, C. J.; Korshalla, J. D.; Jacobus, N. V.; Petersen, P. J.; Mroczenski-Willey, M. J.; Maiese, W. M., et al. Glycothiohexides, Novel Antibiotics Produced by "Sebekia" Sp. LL-14E605. I. Taxonomy, Fermentation and Biological Evaluation. *J. Antibiot.* **1994**, *47* (8), 887-93.
- (315) Northcote, P. T.; Siegel, M.; Borders, D. B.; Lee, M. D. Glycothiohexide Alpha, a Novel Antibiotic Produced by "Sebekia" Sp., LL-14E605. III. Structural Elucidation. *J. Antibiot.* **1994**, *47* (8), 901-8.
- (316) Northcote, P. T.; Williams, D.; Manning, J. K.; Borders, D. B.; Maiese, W. M.; Lee, M. D. Glycothiohexide Alpha, a Novel Antibiotic Produced by "Sebekia" Sp., LL-14E605. II. Isolation and Physical-Chemical Characterization. *J. Antibiot.* **1994**, *47* (8), 894-900.

- (317) Zhang, C.; Occi, J.; Masurekar, P.; Barrett, J. F.; Zink, D. L.; Smith, S.; Onishi, R.; Ha, S.; Salazar, O.; Genilloud, O., et al. Isolation, Structure, and Antibacterial Activity of Philipimycin, a Thiazolyl Peptide Discovered from Actinoplanes Philippinensis MA7347. *J. Am. Chem. Soc.* **2008**, *130* (36), 12102-12110.
- (318) Leet, J. E.; Li, W.; Ax, H. A.; Matson, J. A.; Huang, S.; Huang, R.; Cantone, J. L.; Drexler, D.; Dalterio, R. A.; Lam, K. S. Nocathiacins, New Thiazolyl Peptide Antibiotics from Nocardia Sp. II. Isolation, Characterization, and Structure Determination. *J. Antibiot (Tokyo)* **2003**, *56* (3), 232-42.
- (319) Li, W.; Leet, J. E.; Ax, H. A.; Gustavson, D. R.; Brown, D. M.; Turner, L.; Brown, K.; Clark, J.; Yang, H.; Fung-Tomc, J., et al. Nocathiacins, New Thiazolyl Peptide Antibiotics from Nocardia Sp. I. Taxonomy, Fermentation and Biological Activities. *J. Antibiot.* **2003**, *56* (3), 226-31.
- (320) Sasaki, T.; Otani, T.; Matsumoto, H.; Unemi, N.; Hamada, M.; Takeuchi, T.; Hori, M. MJ347-81F4 A & B, Novel Antibiotics from Amycolatopsis Sp.: Taxonomic Characteristics, Fermentation, and Antimicrobial Activity. *J. Antibiot.* **1998**, *51* (8), 715-21.
- (321) Zhang, C.; Herath, K.; Jayasuriya, H.; Ondeyka, J. G.; Zink, D. L.; Occi, J.; Birdsall, G.; Venugopal, J.; Ushio, M.; Burgess, B., et al. Thiazomycins, Thiazolyl Peptide Antibiotics from Amycolatopsis Fastidiosa. *J. Nat. Prod.* **2009**, *72* (5), 841-7.
- (322) Keller-Juslen, C.; Kuhn, M.; King, H. D. Antibiotics, Pharmaceutical Compositions and Their Use. US4478831A, 1984.
- (323) Xu, L.; Farthing, A. K.; Dropinski, J. F.; Meinke, P. T.; McCallum, C.; Hickey, E.; Liu, K. Synthesis and Antibacterial Activity of Novel Water-Soluble Nocathiacin Analogs. *Bioorg. Med. Chem. Lett.* **2013**, *23* (1), 366-9.
- (324) Xu, L.; Farthing, A. K.; Dropinski, J. F.; Meinke, P. T.; McCallum, C.; Leavitt, P. S.; Hickey, E. J.; Colwell, L.; Barrett, J.; Liu, K. Nocathiacin Analogs: Synthesis and Antibacterial Activity of Novel Water-Soluble Amides. *Bioorg. Med. Chem. Lett.* **2009**, *19* (13), 3531-3535.
- (325) Naidu, B. N.; Sorenson, M. E.; Matiskella, J. D.; Li, W.; Sausker, J. B.; Zhang, Y.; Connolly, T. P.; Lam, K. S.; Bronson, J. J.; Pucci, M. J., et al. Synthesis and Antibacterial Activity of Nocathiacin I Analogues. *Bioorg Med Chem Lett* **2006**, *16* (13), 3545-9.
- (326) Naidu, B. N.; Sorenson, M. E.; Bronson, J. J.; Pucci, M. J.; Clark, J. M.; Ueda, Y. Synthesis, In Vitro, and In Vivo Antibacterial Activity of Nocathiacin I Thiol-Michael Adducts. *Bioorg. Med. Chem. Lett.* **2005**, *15* (8), 2069-72.
- (327) Naidu, B. N.; Sorenson, M. E.; Zhang, Y.; Kim, O. K.; Matiskella, J. D.; Wichtowski, J. A.; Connolly, T. P.; Li, W.; Lam, K. S.; Bronson, J. J., et al. Nocathiacin I Analogues: Synthesis, In Vitro and In Vivo Biological Activity of Novel Semi-Synthetic Thiazolyl Peptide Antibiotics. *Bioorg. Med. Chem. Lett.* **2004**, *14* (22), 5573-7.

- (328) Naidu, B. N.; Sorenson, M. E.; Hudyma, T.; Zheng, X.; Zhang, Y.; Bronson, J. J.; Pucci, M. J.; Clark, J. M.; Ueda, Y. Synthesis and Antibacterial Activity of O-Substituted Nocathiacin I Derivatives. *Bioorg. Med. Chem. Lett.* **2004**, *14* (14), 3743-6.
- (329) Regueiro-Ren, A.; Naidu, B. N.; Zheng, X.; Hudyma, T. W.; Connolly, T. P.; Matiskella, J. D.; Zhang, Y.; Kim, O. K.; Sorenson, M. E.; Pucci, M., et al. Novel Semi-Synthetic Nocathiacin Antibiotics: Synthesis and Antibacterial Activity of Bis- and Mono-O-Alkylated Derivatives. *Bioorg. Med. Chem. Lett.* **2004**, *14* (1), 171-5.
- (330) Li, W.; Huang, S.; Liu, X.; Leet, J. E.; Cantone, J. L.; Lam, K. S. N-Demethylation of Nocathiacin I Via Photo-Oxidation. *Bioorg. Med. Chem. Lett.* **2008**, *18* (14), 4051-3.
- (331) Liu, W.; Ma, M.; Xue, Y.; Liu, N.; Wang, S.; Chen, Y. The C-Terminal Extended Serine Residue Is Absolutely Required in Nosiheptide Maturation. *Chembiochem* **2013**, *14* (5), 573-6.
- (332) Mo, Q.; Zhang, H.; Liu, Q.; Tang, X.; Zhao, L.; Zan, X.; Song, Y. Enhancing Nosiheptide Production in *Streptomyces Actuosus* by Heterologous Expression of Haemoprotein from *Sinorhizobium Meliloti*. *Lett. Appl. Microbiol.* **2016**, *62* (6), 480-7.
- (333) Jiang, L.; Xue, Y. J.; Liu, W. Y.; Ma, M.; Wu, X. R.; Wang, S. Z.; Chen, Y. J. The Importance of Start Codon of NosM in Nosiheptide Production. *Chin. J. Nat. Med.* **2015**, *13* (11), 854-60.
- (334) Zhou, W.; Liu, X.; Zhang, P.; Zhou, P.; Shi, X. Effect Analysis of Mineral Salt Concentrations on Nosiheptide Production by *Streptomyces Actuosus* Z-10 Using Response Surface Methodology. *Molecules* **2014**, *19* (10), 15507-20.
- (335) Wang, Q.; Zhang, D.; Li, Y.; Zhang, F.; Wang, C.; Liang, X. Genome Shuffling and Ribosome Engineering of *Streptomyces Actuosus* for High-Yield Nosiheptide Production. *Appl. Biochem. Biotechnol.* **2014**, *173* (6), 1553-63.
- (336) Walker, J.; Olesker, A.; Valente, L.; Rabanal, R.; Lukacs, G. Total Structure of the Polythiazole-Containing Antibiotic Micrococcin P. A Carbon-13 Nuclear Magnetic Resonance Study. *J. Chem. Soc., Chem. Commun.* **1977**, (20), 706-8.
- (337) Shoji, J.; Kato, T.; Yoshimura, Y.; Tori, K. Studies on Antibiotics from the Genus *Bacillus*. XXIX. Structural Studies on Thiocillins I, II and III. *J. Antibiot.* **1981**, *34* (9), 1126-36.
- (338) Shoji, J.; Hino, H.; Wakisaka, Y.; Koizumi, K.; Mayama, M.; Matsuura, S.; Matsumoto, K. Studies on Antibiotics from the Genus *Bacillus*. VIII. Isolation of Three New Antibiotics, Thiocillins I, II and III, Related to Micrococcin P. *J. Antibiot.* **1976**, *29* (4), 366-74.
- (339) Nagai, K.; Kamigiri, K.; Arao, N.; Suzumura, K.; Kawano, Y.; Yamaoka, M.; Zhang, H.; Watanabe, M.; Suzuki, K. YM-266183 and YM-266184, Novel Thiopeptide Antibiotics Produced by *Bacillus Cereus* Isolated from a Marine Sponge. I. Taxonomy, Fermentation, Isolation, Physico-Chemical Properties and Biological Properties. *J. Antibiot.* **2003**, *56* (2), 123-8.

- (340) Suzumura, K.-i.; Yokoi, T.; Funatsu, M.; Nagai, K.; Tanaka, K.; Zhang, H.; Suzuki, K. Ym-266183 and YM-266184, Novel Thiopeptide Antibiotics Produced by *Bacillus Cereus* Isolated from a Marine Sponge II. Structure Elucidation. *J Antibiot* **2003**, *56* (2), 129-34.
- (341) Kamigiri, K.; Watanabe, M.; Nagai, K.; Arao, N.; Suzumura, K.; Suzuki, K.; Kurane, R.; Yamaoka, M.; Kawano, Y. Thiopeptide Compounds Suitable for Treatment of Multidrug Resistant Bacteria Infection. WO2002072617A1, 2002.
- (342) Acker, M. G.; Bowers, A. A.; Walsh, C. T. Generation of Thiocillin Variants by Prepeptide Gene Replacement and in Vivo Processing by *Bacillus Cereus*. *J. Am. Chem. Soc.* **2009**, *131* (48), 17563-5.
- (343) Mukai, A.; Fukai, T.; Hoshino, Y.; Yazawa, K.; Harada, K.-i.; Mikami, Y. Nocardithiocin, a Novel Thiopeptide Antibiotic, Produced by Pathogenic *Nocardia Pseudobrasiliensis* LFM 0757. *J. Antibiot.* **2009**, *62* (11), 613-619.
- (344) Bowers, A. A.; Acker, M. G.; Young, T. S.; Walsh, C. T. Generation of Thiocillin Ring Size Variants by Prepeptide Gene Replacement and in Vivo Processing by *Bacillus Cereus*. *J. Am. Chem. Soc.* **2012**, *134* (25), 10313-6.
- (345) Baumann, S.; Schoof, S.; Bolten, M.; Haering, C.; Takagi, M.; Shin-ya, K.; Arndt, H. D. Molecular Determinants of Microbial Resistance to Thiopeptide Antibiotics. *J Am Chem Soc* **2010**, *132* (20), 6973-81.
- (346) Habazettl, J.; Allan, M.; Jensen, P. R.; Sass, H. J.; Thompson, C. J.; Grzesiek, S. Structural Basis and Dynamics of Multidrug Recognition in a Minimal Bacterial Multidrug Resistance System. *Proc. Nat. Acad. Sci. U.S.A.* **2014**, *111* (51), E5498-507.
- (347) Harms, J. M.; Wilson, D. N.; Schluenzen, F.; Connell, S. R.; Stachelhaus, T.; Zaborowska, Z.; Spahn, C. M. T.; Fucini, P. Translational Regulation Via L11: Molecular Switches on the Ribosome Turned on and Off by Thiostrepton and Micrococcin. *Mol. Cell* **2008**, *30* (1), 26-38.
- (348) Anborgh, P. H.; Parmeggiani, A. New Antibiotic That Acts Specifically on the GTP-Bound Form of Elongation Factor Tu. *EMBO J.* **1991**, *10* (4), 779-784.
- (349) Selva, E.; Beretta, G.; Montanini, N.; Saddler, G. S.; Gastaldo, L.; Ferrari, P.; Lorenzetti, R.; Landini, P.; Ripamonti, F.; Goldstein, B. P., et al. Antibiotic GE2270 A: A Novel Inhibitor of Bacterial Protein Synthesis. I. Isolation and Characterization. *J. Antibiot.* **1991**, *44* (7), 693-701.
- (350) Tavecchia, P.; Gentili, P.; Kurz, M.; Sottani, C.; Bonfichi, R.; Lociuro, S.; Selva, E. Revised Structure of the Antibiotic Ge 2270a. *J. Antibiot.* **1994**, *47* (12), 1564-7.
- (351) Shimanaka, K.; Iinuma, H.; Hamada, M.; Ikeno, S.; Tsuchiya, K. S.; Arita, M.; Hori, M. Novel Antibiotics, Amythiamicins. IV. A Mutation in the Elongation Factor Tu Gene in a Resistant Mutant of *B. Subtilis*. *J. Antibiot.* **1995**, *48* (2), 182-4.

- (352) Shimanaka, K.; Takahashi, Y.; Inuma, H.; Naganawa, H.; Takeuchi, T. Novel Antibiotics, Amythiamicins. Iii. Structure Elucidations of Amythiamicins A, B and C. *J. Antibiot.* **1994**, *47* (10), 1153-9.
- (353) Shimanaka, K.; Kinoshita, N.; Inuma, H.; Hamada, M.; Takeuchi, T. Novel Antibiotics, Amythiamicins. I. Taxonomy, Fermentation, Isolation, Physico-Chemical Properties, and Antimicrobial Activity. *J. Antibiot.* **1994**, *47* (6), 668-74.
- (354) Shimanaka, K.; Takahashi, Y.; Inuma, H.; Naganawa, H.; Takeuchi, T. Novel Antibiotics, Amythiamicins. II. Structure Elucidation of Amythiamicin D. *J. Antibiot.* **1994**, *47* (10), 1145-52.
- (355) Stella, S.; Montanini, N.; Le Monnier, F.; Ferrari, P.; Colombo, L.; Marinelli, F.; Landini, P.; Ciciliato, I.; Goldstein, B. P.; Selva, E., et al. Antibiotic Ge37468 A: A New Inhibitor of Bacterial Protein Synthesis. I. Isolation and Characterization. *J. Antibiot.* **1995**, *48* (8), 780-6.
- (356) Ferrari, P.; Colombo, L.; Stella, S.; Selva, E.; Zerilli, L. F. Antibiotic GE37468 A: A Novel Inhibitor of Bacterial Protein Synthesis. II. Structure Elucidation. *J. Antibiot.* **1995**, *48* (11), 1304-11.
- (357) Just-Baringo, X.; Bruno, P.; Ottesen, L. K.; Cañedo, L. M.; Albericio, F.; Álvarez, M. Total Synthesis and Stereochemical Assignment of Baringolin. *Angew. Chem. Int. Ed.* **2013**, *125* (30), 7972-7975.
- (358) Martin, J.; da, S. S. T.; Crespo, G.; Palomo, S.; Gonzalez, I.; Tormo, J. R.; de la Cruz, M.; Anderson, M.; Hill, R. T.; Vicente, F., et al. Kocurin, the True Structure of Pm181104, an Anti-Methicillin-Resistant Staphylococcus Aureus (MRSA) Thiazolyl Peptide from the Marine-Derived Bacterium Kocuria Palustris. *Mar. Drugs* **2013**, *11* (2), 387-98.
- (359) Palomo, S.; Gonzalez, I.; de la Cruz, M.; Martin, J.; Tormo, J. R.; Anderson, M.; Hill, R. T.; Vicente, F.; Reyes, F.; Genilloud, O. Sponge-Derived Kocuria and Micrococcus Spp. As Sources of the New Thiazolyl Peptide Antibiotic Kocurin. *Mar. Drugs* **2013**, *11* (4), 1071-86.
- (360) Tocchetti, A.; Bordoni, R.; Gallo, G.; Petiti, L.; Corti, G.; Alt, S.; Cruz, J. C. S.; Salzano, A. M.; Scaloni, A.; Puglia, A. M., et al. A Genomic, Transcriptomic and Proteomic Look at the GE2270 Producer Planobispora Rosea, an Uncommon Actinomycete. *PLoS ONE* **2015**, *10* (7), e0133705.
- (361) Young, Travis S.; Dorrestein, Pieter C.; Walsh, Christopher T. Codon Randomization for Rapid Exploration of Chemical Space in Thiopeptide Antibiotic Variants. *Chem. Biol.* **2012**, *19* (12), 1600-1610.
- (362) LaMarche, M. J.; Leeds, J. A.; Brewer, J.; Dean, K.; Ding, J.; Dzik-Fox, J.; Gamber, G.; Jain, A.; Kerrigan, R.; Krastel, P., et al. Antibacterial and Solubility Optimization of Thiomuracin A. *J. Med. Chem.* **2016**.

- (363) LaMarche, M. J.; Leeds, J. A.; Dzink-Fox, J.; Gangl, E.; Krastel, P.; Neckermann, G.; Palestrant, D.; Patane, M. A.; Rann, E. M.; Tiamfook, S., et al. Antibiotic Optimization and Chemical Structure Stabilization of Thiomuracin A. *J. Med. Chem.* **2012**, *55* (15), 6934-41.
- (364) LaMarche, M. J.; Leeds, J. A.; Dzink-Fox, J.; Mullin, S.; Patane, M. A.; Rann, E. M.; Tiamfook, S. 4-Aminothiazolyl Analogs of GE2270 A: Design, Synthesis and Evaluation of Imidazole Analogs. *Bioorg. Med. Chem. Lett.* **2011**, *21* (11), 3210-3215.
- (365) LaMarche, M. J.; Leeds, J. A.; Amaral, K.; Brewer, J. T.; Bushell, S. M.; Dewhurst, J. M.; Dzink-Fox, J.; Gangl, E.; Goldovitz, J.; Jain, A., et al. Antibacterial Optimization of 4-Aminothiazolyl Analogues of the Natural Product GE2270 A: Identification of the Cycloalkylcarboxylic Acids. *J. Med. Chem.* **2011**, *54* (23), 8099-109.
- (366) LaMarche, M. J.; Leeds, J. A.; Dzink-Fox, J.; Gunderson, K.; Krastel, P.; Memmert, K.; Patane, M. A.; Rann, E. M.; Schmitt, E.; Tiamfook, S., et al. 4-Aminothiazolyl Analogues of GE2270 A: Antibacterial Lead Finding. *J. Med. Chem.* **2011**, *54* (7), 2517-21.
- (367) Leeds, J. A.; LaMarche, M. J.; Brewer, J. T.; Bushell, S. M.; Deng, G.; Dewhurst, J. M.; Dzink-Fox, J.; Gangl, E.; Jain, A.; Lee, L., et al. In Vitro and in Vivo Activities of Novel, Semisynthetic Thiopeptide Inhibitors of Bacterial Elongation Factor Tu. *Antimicrob. Agents Chemother.* **2011**, *55* (11), 5277-83.
- (368) Debast, S. B.; Bauer, M. P.; Sanders, I. M.; Wilcox, M. H.; Kuijper, E. J. Antimicrobial Activity of Lff571 and Three Treatment Agents against Clostridium Difficile Isolates Collected for a Pan-European Survey in 2008: Clinical and Therapeutic Implications. *J. Antimicrob. Chemother.* **2013**, *68* (6), 1305-11.
- (369) LaMarche, M. J.; Leeds, J. A.; Amaral, A.; Brewer, J. T.; Bushell, S. M.; Deng, G.; Dewhurst, J. M.; Ding, J.; Dzink-Fox, J.; Gamber, G., et al. Discovery of Lff571: An Investigational Agent for Clostridium Difficile Infection. *J. Med. Chem.* **2012**, *55* (5), 2376-87.
- (370) Leeds, J. A.; Sachdeva, M.; Mullin, S.; Dzink-Fox, J.; Lamarche, M. J. Mechanism of Action of and Mechanism of Reduced Susceptibility to the Novel Anti-Clostridium Difficile Compound LFF571. *Antimicrob. Agents Chemother.* **2012**, *56* (8), 4463-5.
- (371) Trzasko, A.; Leeds, J. A.; Praestgaard, J.; Lamarche, M. J.; McKenney, D. Efficacy of LFF571 in a Hamster Model of Clostridium Difficile Infection. *Antimicrob. Agents Chemother.* **2012**, *56* (8), 4459-62.
- (372) Ting, L. S.; Praestgaard, J.; Grunenber, N.; Yang, J. C.; Leeds, J. A.; Pertel, P. A First-in-Human, Randomized, Double-Blind, Placebo-Controlled, Single- and Multiple-Ascending Oral Dose Study to Assess the Safety and Tolerability of LFF571 in Healthy Volunteers. *Antimicrob. Agents Chemother.* **2012**, *56* (11), 5946-51.

- (373) Citron, D. M.; Tyrrell, K. L.; Merriam, C. V.; Goldstein, E. J. Comparative in Vitro Activities of LFF571 against *Clostridium Difficile* and 630 Other Intestinal Strains of Aerobic and Anaerobic Bacteria. *Antimicrob. Agents Chemother.* **2012**, *56* (5), 2493-503.
- (374) Just-Baringo, X.; Bruno, P.; Pitart, C.; Vila, J.; Albericio, F.; Alvarez, M. Dissecting the Structure of Thiopeptides: Assessment of Thiazoline and Tail Moieties of Baringolin and Antibacterial Activity Optimization. *J. Med. Chem.* **2014**, *57* (10), 4185-95.
- (375) Aoki, M.; Ohtsuka, T.; Itezono, Y.; Yokose, K.; Furihata, K.; Seto, H. Structure of Cyclothiazomycin, a Unique Polythiazole-Containing Peptide with Renin Inhibitory Activity. Part 1. Chemistry and Partial Structures of Cyclothiazomycin. *Tetrahedron Lett.* **1991**, *32* (2), 217-220.
- (376) Aoki, M.; Ohtsuka, T.; Itezono, Y.; Yokose, K.; Furihata, K.; Seto, H. Structure of Cyclothiazomycin, a Unique Polythiazole-Containing Peptide with Renin Inhibitory Activity. Part 2. Total Structure. *Tetrahedron Lett.* **1991**, *32* (2), 221-224.
- (377) Aoki, M.; Ohtsuka, T.; Yamada, M.; Ohba, Y.; Yoshizaki, H.; Yasuno, H.; Sano, T.; Watanabe, J.; Yokose, K.; Seto, H. Cyclothiazomycin, a Novel Polythiazole-Containing Peptide with Renin Inhibitory Activity. Taxonomy, Fermentation, Isolation and Physico-Chemical Characterization. *J. Antibiot.* **1991**, *44* (6), 582-8.
- (378) Hashimoto, M.; Murakami, T.; Funahashi, K.; Tokunaga, T.; Nihei, K.; Okuno, T.; Kimura, T.; Naoki, H.; Himeno, H. An RNA Polymerase Inhibitor, Cyclothiazomycin B1, and Its Isomer. *Bioorg. Med. Chem.* **2006**, *14* (24), 8259-70.
- (379) Mizuhara, N.; Kuroda, M.; Ogita, A.; Tanaka, T.; Usuki, Y.; Fujita, K. Antifungal Thiopeptide Cyclothiazomycin B1 Exhibits Growth Inhibition Accompanying Morphological Changes Via Binding to Fungal Cell Wall Chitin. *Bioorg. Med. Chem.* **2011**, *19* (18), 5300-10.
- (380) Abe, H.; Kushida, K.; Shiobara, Y.; Kodama, M. The Structures of Sulfomycin I and Berninamycin A. *Tetrahedron Lett.* **1988**, *29* (12), 1401-1404.
- (381) Reusser, F. Mode of Action of Berninamycin. An Inhibitor of Protein Biosynthesis. *Biochemistry* **1969**, *8* (8), 3303-8.
- (382) Egawa, Y.; Umino, K.; Tamura, Y.; Shimizu, M.; Kaneko, K. Sulfomycins, a Series of New Sulfur-Containing Antibiotics. I. Isolation, Purification and Properties. *J. Antibiot.* **1969**, *22* (1), 12-7.
- (383) Yun, B. S.; Hidaka, T.; Furihata, K.; Seto, H. Microbial Metabolites with TipA Promoter Inducing Activity. II. Geninthiocin, a Novel Thiopeptide Produced by *Streptomyces* Sp. DD84. *J. Antibiot.* **1994**, *47* (9), 969-75.
- (384) Yun, B. S.; Hidaka, T.; Furihata, K.; Seto, H. Promothiocins A and B Novel Thiopeptides with a TipA Promoter Inducing Activity Produced by *Streptomyces* Sp. Sf2741. *J. Antibiot.* **1994**, *47* (4), 510-4.

- (385) Takagi, M.; Motohashi, K.; Nagai, A.; Hashimoto, J.; Shin-Ya, K. JBIR-83 and JBIR-84, New Promothiocin Derivatives, Isolated from Streptomyces Sp. RI19. *J. Antibiot.* **2010**, *63* (7), 405-8.
- (386) Gonzalez Holgado, G.; Castro Rodriguez, J.; Canedo Hernandez, L. M.; Diaz, M.; Fernandez-Abalos, J. M.; Trujillano, I.; Santamaria, R. I. Radamycin, a Novel Thiopeptide Produced by Streptomyces Sp. RSP9. I. Taxonomy, Fermentation, Isolation and Biological Activities. *J. Antibiot.* **2002**, *55* (4), 383-90.
- (387) Castro Rodriguez, J.; Gonzalez Holgado, G.; Santamaria Sanchez, R. I.; Canedo, L. M. Radamycin, a Novel Thiopeptide Produced by Streptomyces Sp. RSP9. II. Physico-Chemical Properties and Structure Determination. *J Antibiot (Tokyo)* **2002**, *55* (4), 391-5.
- (388) Kumar, E. K. S. V.; Kenia, J.; Mukhopadhyay, T.; Nadkarni, S. R. Methylsulfomycin I, a New Cyclic Peptide Antibiotic from Streptomyces Sp. HIL Y-9420704. *J. Nat. Prod.* **1999**, *62* (11), 1562-1564.
- (389) Yun, B. S.; Seto, H. Promoinducin, a Novel Thiopeptide Produced by Streptomyces Sp. SF2741. *Biosci. Biotechnol. Biochem.* **1995**, *59* (5), 876-80.
- (390) Yun, B.-S.; Hidaka, T.; Furihata, K.; Seto, H. Thiotipin, a Novel Thiopeptide with a TipA Promoter Inducing Activity Produced by Streptomyces Sp. DT31. *Tetrahedron* **1994**, *50* (40), 11659-11664.
- (391) Debono, M.; Molloy, R. M.; Ocolowitz, J. L.; Paschal, J. W.; Hunt, A. H.; Michel, K. H.; Martin, J. W. The Structures of A10255 B, -G and -J: New Thiopeptide Antibiotics Produced by Streptomyces Gardneri. *J. Org. Chem.* **1992**, *57* (19), 5200-8.
- (392) Boeck, L. D.; Hoehn, M. M.; Kirst, H. A.; Michel, K. H.; Seno, E. T.; Godfrey, O. W., Jr. Antibiotic A10255 Complex and Factors and Their Manufacture with Streptomyces Garderi. EP274873A2, 1988.
- (393) Matsumoto, M.; Kawamura, Y.; Yasuda, Y.; Tanimoto, T.; Matsumoto, K.; Yoshida, T.; Shoji, J. Isolation and Characterization of Thioxamycin. *J. Antibiot.* **1989**, *42* (10), 1465-9.
- (394) Yun, B. S.; Hidaka, T.; Furihata, K.; Seto, H. Microbial Metabolites with Tipa Promoter Inducing Activity. III. Thioxamycin and Its Novel Derivative, Thioactin, Two Thiopeptides Produced by Streptomyces Sp. Dp94. *J. Antibiot.* **1994**, *47* (12), 1541-5.
- (395) Steitz, T. A. A Structural Understanding of the Dynamic Ribosome Machine. *Nat Rev Mol Cell Biol* **2008**, *9* (3), 242-253.
- (396) Spahn, C. M. T.; Prescott, C. D. Throwing a Spanner in the Works: Antibiotics and the Translation Apparatus. *J. Mol. Med.* **1996**, *74* (8), 423-439.
- (397) Ling, C.; Ermolenko, D. N. Structural Insights into Ribosome Translocation. *Wiley Interdiscip. Rev. RNA* **2016**, *7* (5), 620-36.

- (398) Selva, E.; Montanini, N.; Stella, S.; Soffientini, A.; Gastaldo, L.; Denaro, M. Targeted Screening for Elongation Factor Tu Binding Antibiotics. *J. Antibiot.* **1997**, *50* (1), 22-6.
- (399) Landini, P.; Bandera, M.; Goldstein, B. P.; Ripamonti, F.; Soffientini, A.; Islam, K.; Denaro, M. Inhibition of Bacterial Protein Synthesis by Elongation-Factor-Tu-Binding Antibiotics MDL 62,879 and Efrotomycin. *Biochem. J.* **1992**, *283* (Pt 3), 649-52.
- (400) Gross, S.; Nguyen, F.; Bierschenk, M.; Sohmen, D.; Menzel, T.; Antes, I.; Wilson, D. N.; Bach, T. Amythiamicin D and Related Thiopeptides as Inhibitors of the Bacterial Elongation Factor Ef-Tu: Modification of the Amino Acid at Carbon Atom C2 of Ring C Dramatically Influences Activity. *ChemMedChem* **2013**, *8* (12), 1954-62.
- (401) Anborgh, P. H.; Parmeggiani, A. Probing the Reactivity of the GTP- and Gdp-Bound Conformations of Elongation Factor Tu in Complex with the Antibiotic GE2270 A. *J. Biol. Chem.* **1993**, *268* (33), 24622-8.
- (402) Heffron, S. E.; Jurnak, F. Structure of an Ef-Tu Complex with a Thiazolyl Peptide Antibiotic Determined at 2.35 Å Resolution: Atomic Basis for GE2270A Inhibition of Ef-Tu. *Biochemistry* **2000**, *39* (1), 37-45.
- (403) Parmeggiani, A.; Krab, I. M.; Okamura, S.; Nielsen, R. C.; Nyborg, J.; Nissen, P. Structural Basis of the Action of Pulvomycin and GE2270 a on Elongation Factor Tu. *Biochemistry* **2006**, *45* (22), 6846-6857.
- (404) Sosio, M.; Amati, G.; Cappellano, C.; Sarubbi, E.; Monti, F.; Donadio, S. An Elongation Factor Tu (Ef-Tu) Resistant to the Ef-Tu Inhibitor GE2270 in the Producing Organism Planobispora Rosea. *Mol. Microbiol.* **1996**, *22* (1), 43-51.
- (405) Mohrle, V. G.; Tieleman, L. N.; Kraal, B. Elongation Factor Tu1 of the Antibiotic GE2270A Producer Planobispora Rosea Has an Unexpected Resistance Profile against Ef-Tu Targeted Antibiotics. *Biochem. Bioph. Res. Co.* **1997**, *230* (2), 320-6.
- (406) Hogg, T.; Mesters, J. R.; Hilgenfeld, R. Inhibitory Mechanisms of Antibiotics Targeting Elongation Factor Tu. *Curr. Protein Pept. Sci.* **2002**, *3* (1), 121-131.
- (407) Parmeggiani, A.; Nissen, P. Elongation Factor Tu-Targeted Antibiotics: Four Different Structures, Two Mechanisms of Action. *FEBS Lett.* **2006**, *580* (19), 4576-4581.
- (408) Thompson, J.; Cundliffe, E.; Stark, M. Binding of Thiostrepton to a Complex of 23-S rRNA with Ribosomal Protein L11. *Eur. J. Biochem.* **1979**, *98* (1), 261-5.
- (409) Pestka, S.; Weiss, D.; Vince, R. Partition of Ribosomes in Two-Polymer Aqueous Phase Systems. *Anal. Biochem.* **1976**, *71* (1), 137-142.

- (410) Cundliffe, E., Involvement of Specific Portions of Ribosomal RNA in Defined Ribosomal Functions: A Study Utilizing Antibiotics. In *Structure, Function, and Genetics of Ribosomes*, Hardesty, B.; Kramer, G., Eds. Springer New York: New York, NY, 1986; pp 586-604.
- (411) Thompson, J.; Cundliffe, E. The Binding of Thiostrepton to 23s Ribosomal RNA. *Biochimie* **1991**, *73* (7), 1131-1135.
- (412) Thompson, J.; Musters, W.; Cundliffe, E.; Dahlberg, A. E. Replacement of the L11 Binding Region within E.Coli 23s Ribosomal RNA with Its Homologue from Yeast: In Vivo and in Vitro Analysis of Hybrid Ribosomes Altered in the GTPase Centre. *EMBO J.* **1993**, *12* (4), 1499-504.
- (413) Xing, Y.; Draper, D. E. Cooperative Interactions of RNA and Thiostrepton Antibiotic with Two Domains of Ribosomal Protein L11. *Biochemistry* **1996**, *35* (5), 1581-1588.
- (414) Blyn, L. B.; Risen, L. M.; Griffey, R. H.; Draper, D. E. The RNA-Binding Domain of Ribosomal Protein L11 Recognizes an rRNA Tertiary Structure Stabilized by Both Thiostrepton and Magnesium Ion. *Nucleic Acids Res.* **2000**, *28* (8), 1778-84.
- (415) Rosendahl, G.; Douthwaite, S. The Antibiotics Micrococcin and Thiostrepton Interact Directly with 23S rRNA Nucleotides 1067A and 1095A. *Nucleic Acids Res.* **1994**, *22* (3), 357-63.
- (416) Hummel, H.; Bock, A. Thiostrepton Resistance Mutations in the Gene for 23S Ribosomal RNA of Halobacteria. *Biochimie* **1987**, *69* (8), 857-61.
- (417) Mankin, A. S.; Leviev, I.; Garrett, R. A. Cross-Hypersensitivity Effects of Mutations in 23 S rRNA Yield Insight into Aminoacyl-tRNA Binding. *J. Mol. Biol.* **1994**, *244* (2), 151-157.
- (418) Ryan, P. C.; Lu, M.; Draper, D. E. Recognition of the Highly Conserved GTPase Center of 23 S Ribosomal RNA by Ribosomal Protein L11 and the Antibiotic Thiostrepton. *J. Mol. Biol.* **1991**, *221* (4), 1257-1268.
- (419) Spedding, G.; Cundliffe, E. Identification of the Altered Ribosomal Component Responsible for Resistance to Micrococcin in Mutants of Bacillus Megaterium. *Eur. J. Biochem.* **1984**, *140* (3), 453-459.
- (420) Cameron, D. M.; Thompson, J.; Gregory, S. T.; March, P. E.; Dahlberg, A. E. Thiostrepton-Resistant Mutants of Thermus Thermophilus. *Nucleic Acids Res.* **2004**, *32* (10), 3220-3227.
- (421) Cundliffe, E.; Dixon, P.; Stark, M.; Stoffler, G.; Ehrlich, R.; Stoffler-Meilicke, M.; Cannon, M. Ribosomes in Thiostrepton-Resistant Mutants of Bacillus Megaterium Lacking a Single 50 S Subunit Protein. *J. Mol. Biol.* **1979**, *132* (2), 235-52.
- (422) Thompson, J.; Schmidt, F.; Cundliffe, E. Site of Action of a Ribosomal RNA Methylase Conferring Resistance to Thiostrepton. *J. Biol. Chem.* **1982**, *257* (14), 7915-7.
- (423) Cundliffe, E.; Thompson, J. Concerning the Mode of Action of Micrococcin Upon Bacterial Protein Synthesis. *Eur. J. Biochem.* **1981**, *118* (1), 47-52.

- (424) Cundliffe, E.; Thompson, J. Ribose Methylation and Resistance to Thiostrepton. *Nature* **1979**, *278* (5707), 859-861.
- (425) Skinner, R.; Cundliffe, E.; Schmidt, F. J. Site of Action of a Ribosomal RNA Methylase Responsible for Resistance to Erythromycin and Other Antibiotics. *J. Biol. Chem.* **1983**, *258* (20), 12702-6.
- (426) Dunstan, M. S.; Hang, P. C.; Zelinskaya, N. V.; Honek, J. F.; Conn, G. L. Structure of the Thiostrepton Resistance Methyltransferase-S-Adenosyl-L-Methionine Complex and Its Interaction with Ribosomal RNA. *J. Biol. Chem.* **2009**, *284* (25), 17013-20.
- (427) Bechthold, A.; Floss, H. G. Overexpression of the Thiostrepton-Resistance Gene from *Streptomyces Azureus* in *Escherichia Coli* and Characterization of Recognition Sites of the 23S rRNA A1067 2'-Methyltransferase in the Guanosine Triphosphatase Center of 23S Ribosomal RNA. *Eur. J. Biochem.* **1994**, *224* (2), 431-7.
- (428) Kuiper, E. G.; Conn, G. L. Binding Induced RNA Conformational Changes Control Substrate Recognition and Catalysis by the Thiostrepton Resistance Methyltransferase (Tsr). *J. Biol. Chem.* **2014**, *289* (38), 26189-26200.
- (429) Yin, S.; Jiang, H.; Chen, D.; Murchie, A. I. Substrate Recognition and Modification by the Nosiheptide Resistance Methyltransferase. *PLoS One* **2015**, *10* (4), e0122972.
- (430) Myers, C. L.; Kuiper, E. G.; Grant, P. C.; Hernandez, J.; Conn, G. L.; Honek, J. F. Functional Roles in S-Adenosyl-L-Methionine Binding and Catalysis for Active Site Residues of the Thiostrepton Resistance Methyltransferase. *FEBS Lett.* **2015**, *589* (21), 3263-70.
- (431) Cundliffe, E. Mechanism of Resistance to Thiostrepton in the Producing-Organism *Streptomyces Azureus*. *Nature* **1978**, *272* (5656), 792-795.
- (432) Thompson, C. J.; Skinner, R. H.; Thompson, J.; Ward, J. M.; Hopwood, D. A.; Cundliffe, E. Biochemical Characterization of Resistance Determinants Cloned from Antibiotic-Producing *Streptomyces*. *J. Bacteriol.* **1982**, *151* (2), 678-685.
- (433) Li, Y.; Dosch, D. C.; Strohl, W. R.; Floss, H. G. Nucleotide Sequence and Transcriptional Analysis of the Nosiheptide-Resistance Gene from *Streptomyces Actuosus*. *Gene* **1990**, *91* (1), 9-17.
- (434) Egebjerg, J.; Douthwaite, S. R.; Liljas, A.; Garrett, R. A. Characterization of the Binding Sites of Protein L11 and the L10.(L12)₄ Pentameric Complex in the GTPase Domain of 23 S Ribosomal RNA from *Escherichia Coli*. *J. Mol. Biol.* **1990**, *213* (2), 275-288.
- (435) Rosendahl, G.; Douthwaite, S. Ribosomal Proteins L11 and L10.(L12)₄ and the Antibiotic Thiostrepton Interact with Overlapping Regions of the 23 S rRNA Backbone in the Ribosomal GTPase Centre. *J. Mol. Biol.* **1993**, *234* (4), 1013-1020.

- (436) Wolf, A.; Schoof, S.; Baumann, S.; Arndt, H.-D.; Kirschner, K. N. Structure-Activity Relationships of Thiostrepton Derivatives: Implications for Rational Drug Design. *J. Comput. Aided Mol. Des.* **2014**, *28* (12), 1205-1215.
- (437) Wimberly, B. T.; Guymon, R.; McCutcheon, J. P.; White, S. W.; Ramakrishnan, V. A Detailed View of a Ribosomal Active Site: The Structure of the L11-Rna Complex. *Cell* **1999**, *97* (4), 491-502.
- (438) Lee, D.; Walsh, J. D.; Yu, P.; Markus, M. A.; Choli-Papadopoulou, T.; Schwieters, C. D.; Krueger, S.; Draper, D. E.; Wang, Y.-X. The Structure of Free L11 and Functional Dynamics of L11 in Free, L11-rRNA(58 Nt) Binary and L11-r (58 Nt)-Thiostrepton Ternary Complexes. *J. Mol. Biol.* **2007**, *367* (4), 1007-1022.
- (439) Jonker, H. R. A.; Baumann, S.; Wolf, A.; Schoof, S.; Hiller, F.; Schulte, K. W.; Kirschner, K. N.; Schwalbe, H.; Arndt, H.-D. NMR Structures of Thiostrepton Derivatives for Characterization of the Ribosomal Binding Site. *Angew. Chem., Int. Ed.* **2011**, *50* (14), 3308-3312, S3308/1-S3308/23.
- (440) Lentzen, G.; Klinck, R.; Matassova, N.; Aboul-ela, F.; Murchie, A. I. H. Structural Basis for Contrasting Activities of Ribosome Binding Thiazole Antibiotics. *Chem. Biol.* **2003**, *10* (8), 769-778.
- (441) Bowen, W. S.; Van Dyke, N.; Murgola, E. J.; Lodmell, J. S.; Hill, W. E. Interaction of Thiostrepton and Elongation Factor-G with the Ribosomal Protein L11-Binding Domain. *J. Biol. Chem.* **2005**, *280* (4), 2934-2943.
- (442) Jonker, H. R. A.; Ilin, S.; Grimm, S. K.; Wöhnert, J.; Schwalbe, H. L11 Domain Rearrangement Upon Binding to RNA and Thiostrepton Studied by NMR Spectroscopy. *Nucleic Acids Res.* **2007**, *35* (2), 441-454.
- (443) Baumann, S.; Schoof, S.; Harkal, S. D.; Arndt, H.-D. Mapping the Binding Site of Thiopeptide Antibiotics by Proximity-Induced Covalent Capture. *J. Am. Chem. Soc.* **2008**, *130* (17), 5664-5666.
- (444) Porse, B. T.; Garrett, R. A. Ribosomal Mechanics, Antibiotics, and GTP Hydrolysis. *Cell* **1999**, *97* (4), 423-426.
- (445) Mohr, D.; Wintermeyer, W.; Rodnina, M. V. GTPase Activation of Elongation Factors Tu and G on the Ribosome. *Biochemistry* **2002**, *41* (41), 12520-12528.
- (446) Xu, W.; Pagel, F. T.; Murgola, E. J. Mutations in the GTPase Center of Escherichia Coli 23S rRNA Indicate Release Factor 2-Interactive Sites. *Journal of Bacteriology* **2002**, *184* (4), 1200-1203.
- (447) Villa, E.; Sengupta, J.; Trabuco, L. G.; LeBarron, J.; Baxter, W. T.; Shaikh, T. R.; Grassucci, R. A.; Nissen, P.; Ehrenberg, M.; Schulten, K., et al. Ribosome-Induced Changes in Elongation Factor Tu Conformation Control GTP Hydrolysis. *Proc. Natl. Acad. Sci. USA* **2009**, *106* (4), 1063-1068.

- (448) Li, W.; Sengupta, J.; Rath, B. K.; Frank, J. Functional Conformations of the L11–Ribosomal RNA Complex Revealed by Correlative Analysis of Cryo-Em and Molecular Dynamics Simulations. *RNA* **2006**, *12* (7), 1240-1253.
- (449) Allen, G. S.; Zavialov, A.; Gursky, R.; Ehrenberg, M.; Frank, J. The Cryo-Em Structure of a Translation Initiation Complex from Escherichia Coli. *Cell* **2005**, *121* (5), 703-712.
- (450) Schuwirth, B. S.; Borovinskaya, M. A.; Hau, C. W.; Zhang, W.; Vila-Sanjurjo, A.; Holton, J. M.; Cate, J. H. D. Structures of the Bacterial Ribosome at 3.5 Å Resolution. *Science* **2005**, *310* (5749), 827-834.
- (451) Mazumder, R. Effect of Thiostrepton on Recycling of Escherichia Coli Initiation Factor 2. *Proc. Nat. Acad. Sci. U.S.A.* **1973**, *70* (7), 1939-42.
- (452) Lockwood, A. H.; Sarkar, P.; Maitra, U.; Brot, N.; Weissbach, H. Effect of Thiostrepton on Polypeptide Chain Initiation in Escherichia Coli. *J. Biol. Chem.* **1974**, *249* (18), 5831-5834.
- (453) Sarkar, P.; Stringer, E. A.; Maitra, U. Thiostrepton Inhibition of Initiation Factor 1 Activity in Polypeptide Chain Initiation in Escherichia Coli. *Proc. Nat. Acad. Sci. U.S.A.* **1974**, *71* (12), 4986-90.
- (454) Grunberg-Manago, M.; Dondon, J.; Graffe, M. Inhibition by Thiostrepton of the If-2-Dependent Ribosomal GTPase. *FEBS Lett.* **1972**, *22* (2), 217-221.
- (455) Naaktgeboren, N.; Roobol, K.; Gubbens, J.; Voorma, H. O. The Mode of Action of Thiostrepton in the Initiation of Protein Synthesis. *Eur. J. Biochem.* **1976**, *70* (1), 39-47.
- (456) Brandi, L.; Marzi, S.; Fabbretti, A.; Fleischer, C.; Hill, W. E.; Gualerzi, C. O.; Stephen Lodmell, J. The Translation Initiation Functions of If2: Targets for Thiostrepton Inhibition. *J Mol Biol* **2004**, *335* (4), 881-94.
- (457) Brot, N.; Tate, W. P.; Caskey, C. T.; Weissbach, H. The Requirement for Ribosomal Proteins L7 and L12 in Peptide-Chain Termination. *Proc. Nat. Acad. Sci. U.S.A.* **1974**, *71* (1), 89-92.
- (458) Cameron, D. M.; Thompson, J.; March, P. E.; Dahlberg, A. E. Initiation Factor If2, Thiostrepton and Micrococin Prevent the Binding of Elongation Factor G to the Escherichia Coli Ribosome. *Journal of Molecular Biology* **2002**, *319* (1), 27-35.
- (459) Kinoshita, T.; Liou, Y.; Tanaka, N. Inhibition by Thiopeptin of Ribosomal Functions Associated with T and G Factors. *Biochem. Biophys. Res. Co.* **1971**, *44* (4), 859-63.
- (460) Pestka, S.; Brot, N. Studies on the Formation of Transfer Ribonucleic Acid-Ribosome Complexes. Iv. Effect of Antibiotics on Steps of Bacterial Protein Synthesis: Some New Ribosomal Inhibitors of Translocation. *J. Biol. Chem.* **1971**, *246* (24), 7715-22.
- (461) Cundliffe, E.; Thompson, J. The Mode of Action of Nosiheptide (Multhiomycin) and the Mechanism of Resistance in the Producing Organism. *J. Gen. Microbiol.* **1981**, *126* (1), 185-92.

- (462) Hornig, H.; Woolley, P.; Lührmann, R. Decoding at the Ribosomal a Site: Antibiotics, Misreading and Energy of Aminoacyl-tRNA Binding. *Biochimie* **1987**, *69* (8), 803-813.
- (463) Weisblum, B.; Demohn, V. Inhibition by Thiostrepton of the Formation of a Ribosome-Bound Guanine Nucleotide Complex. *FEBS Lett.* **1970**, *11* (3), 149-152.
- (464) Hausner, T. P.; Geigenmüller, U.; Nierhaus, K. H. The Allosteric Three-Site Model for the Ribosomal Elongation Cycle. New Insights into the Inhibition Mechanisms of Aminoglycosides, Thiostrepton, and Viomycin. *J. Biol. Chem.* **1988**, *263* (26), 13103-13111.
- (465) Highland, J. H.; Lin, L.; Bodley, J. W. Protection of Ribosomes from Thiostrepton Inactivation by the Binding of G Factor and Guanosine Diphosphate. *Biochemistry* **1971**, *10* (24), 4404-9.
- (466) Bodley, J. W.; Lin, L.; Highland, J. H. Studies on Translocation. Vi. Thiostrepton Prevents the Formation of a Ribosome-G Factor-Guanine Nucleotide Complex. *Biochem. Biophys. Res. Co.* **1970**, *41* (6), 1406-11.
- (467) Li, J. M.; Cabrer, B.; Parmeggiani, A.; Azquez, D. V. Inhibition by Siomycin and Thiostrepton of Both Aminoacyl-Trna and Factor G Binding to Ribosomes. *Proc. Nat. Acad. Sci. U.S.A.* **1971**, *68* (8), 1796-1800.
- (468) Weisblum, B.; Demohn, V. Thiostrepton, an Inhibitor of 50S Ribosome Subunit Function. *J. Bacteriol.* **1970**, *101* (3), 1073-1075.
- (469) Tanaka, K.; Watanabe, S.; Teraoka, H.; Tamaki, M. Effect of Siomycin on Protein Synthesizing Activity of Escherichia Coli Ribosomes. *Biochem. Biophys. Res. Co.* **1970**, *39* (6), 1189-1193.
- (470) Pestka, S. Thiostrepton: A Ribosomal Inhibitor of Translocation. *Biochem. Biophys. Res. Co.* **1970**, *40* (3), 667-74.
- (471) Cundliffe, E. The Mode of Action of Thiostreption In Vivo. *Biochem. Biophys. Res. Co.* **1971**, *44* (4), 912-7.
- (472) Otaka, T.; Kaji, A. Micrococcin: Acceptor-Site-Specific Inhibitor of Protein Synthesis. *Eur. J. Biochem.* **1974**, *50* (1), 101-106.
- (473) Gonzalez, R. L.; Chu, S.; Puglisi, J. D. Thiostrepton Inhibition of Trna Delivery to the Ribosome. *RNA* **2007**, *13* (12), 2091-2097.
- (474) Cundliffe, E.; Dixon, P. D. Inhibition of Ribosomal a Site Functions by Sporangiomycin and Micrococcin. *Antimicrob. Agents Chemother.* **1975**, *8* (1), 1-4.
- (475) Egebjerg, J.; Douthwaite, S.; Garrett, R. A. Antibiotic Interactions at the GTPase-Associated Centre within Escherichia Coli 23S rRNA. *EMBO J.* **1989**, *8* (2), 607-11.
- (476) Skold, S. E. Chemical Crosslinking of Elongation Factor G to the 23S RNA in 70S Ribosomes from Escherichia Coli. *Nucleic Acids Res.* **1983**, *11* (14), 4923-32.

- (477) Walter, J. D.; Hunter, M.; Cobb, M.; Traeger, G.; Spiegel, P. C. Thiostrepton Inhibits Stable 70S Ribosome Binding and Ribosome-Dependent GTPase Activation of Elongation Factor G and Elongation Factor 4. *Nucleic Acids Res.* **2012**, *40* (1), 360-370.
- (478) Porse, B. T.; Leviev, I.; Mankin, A. S.; Garrett, R. A. The Antibiotic Thiostrepton Inhibits a Functional Transition within Protein L11 at the Ribosomal GTPase Centre1. *J. Mol. Biol.* **1998**, *276* (2), 391-404.
- (479) Rodnina, M. V.; Savelsbergh, A.; Matassova, N. B.; Katunin, V. I.; Semenov, Y. P.; Wintermeyer, W. Thiostrepton Inhibits the Turnover but Not the GTPase of Elongation Factor G on the Ribosome. *Proc. Natl. Acad. Sci. USA* **1999**, *96* (17), 9586-9590.
- (480) Pan, D.; Kirillov, S. V.; Cooperman, B. S. Kinetically Competent Intermediates in the Translocation Step of Protein Synthesis. *Mol Cell* **2007**, *25* (4), 519-29.
- (481) Gao, Y.-G.; Selmer, M.; Dunham, C. M.; Weixlbaumer, A.; Kelley, A. C.; Ramakrishnan, V. The Structure of the Ribosome with Elongation Factor G Trapped in the Posttranslocational State. *Science* **2009**, *326* (5953), 694-699.
- (482) Stark, H.; Rodnina, M. V.; Wieden, H.-J.; van Heel, M.; Wintermeyer, W. Large-Scale Movement of Elongation Factor G and Extensive Conformational Change of the Ribosome During Translocation. *Cell* **2000**, *100* (3), 301-309.
- (483) Lin, J.; Gagnon, M. G.; Bulkeley, D.; Steitz, T. A. Conformational Changes of Elongation Factor G on the Ribosome During tRNA Translocation. *Cell* **2015**, *160* (0), 219-227.
- (484) Seo, H.-S.; Kiel, M.; Pan, D.; Raj, V. S.; Kaji, A.; Cooperman, B. S. Kinetics and Thermodynamics of Rrf, Ef-G, and Thiostrepton Interaction on the Escherichia Coli Ribosome. *Biochemistry* **2004**, *43* (40), 12728-12740.
- (485) Seo, H.-S.; Abedin, S.; Kamp, D.; Wilson, D. N.; Nierhaus, K. H.; Cooperman, B. S. EF-G-Dependent GTPase on the Ribosome. Conformational Change and Fusidic Acid Inhibition. *Biochemistry* **2006**, *45* (8), 2504-2514.
- (486) Porse, B. T.; Cundliffe, E.; Garrett, R. A. The Antibiotic Micrococcin Acts on Protein L11 at the Ribosomal GTPase Centre1. *J. Mol. Biol.* **1999**, *287* (1), 33-45.
- (487) Mikolajka, A.; Liu, H.; Chen, Y.; Starosta, A. L.; Márquez, V.; Ivanova, M.; Cooperman, B. S.; Wilson, D. N. Differential Effects of Thiopeptide and Orthosomycin Antibiotics on Translational GTPases. *Chem. Biol.* **2011**, *18* (5), 589-600.
- (488) Thompson, J.; Cundliffe, E.; Stark, M. J. The Mode of Action of Berninamycin and Mechanism of Resistance in the Producing Organism, Streptomyces Bernensis. *J. Gen. Microbiol.* **1982**, *128* (4), 875-84.

- (489) Bleich, R.; Watrous, J. D.; Dorrestein, P. C.; Bowers, A. A.; Shank, E. A. Thiopeptide Antibiotics Stimulate Biofilm Formation in *Bacillus Subtilis*. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112* (10), 3086-91.
- (490) Murakami, T.; Holt, T. G.; Thompson, C. J. Thiostrepton-Induced Gene Expression in *Streptomyces Lividans*. *J. Bacteriol.* **1989**, *171* (3), 1459-66.
- (491) Holmes, D. J.; Caso, J. L.; Thompson, C. J. Autogenous Transcriptional Activation of a Thiostrepton-Induced Gene in *Streptomyces Lividans*. *EMBO J.* **1993**, *12* (8), 3183-91.
- (492) Chiu, M. L.; Viollier, P. H.; Katoh, T.; Ramsden, J. J.; Thompson, C. J. Ligand-Induced Changes in the *Streptomyces Lividans* Tipal Protein Imply an Alternative Mechanism of Transcriptional Activation for Merr-Like Proteins. *Biochemistry* **2001**, *40* (43), 12950-8.
- (493) Chiu, M. L.; Folcher, M.; Katoh, T.; Puglia, A. M.; Vohradsky, J.; Yun, B. S.; Seto, H.; Thompson, C. J. Broad Spectrum Thiopeptide Recognition Specificity of the *Streptomyces Lividans* Tipal Protein and Its Role in Regulating Gene Expression. *J. Biol. Chem.* **1999**, *274* (29), 20578-86.
- (494) Yun, B. S.; Hidaka, T.; Kuzuyama, T.; Seto, H. Thiopeptide Non-Producing *Streptomyces* Species Carry the Tipa Gene: A Clue to Its Function. *J Antibiot* **2001**, *54* (4), 375-8.
- (495) Chiu, M. L.; Folcher, M.; Griffin, P.; Holt, T.; Klatt, T.; Thompson, C. J. Characterization of the Covalent Binding of Thiostrepton to a Thiostrepton-Induced Protein from *Streptomyces Lividans*. *Biochemistry* **1996**, *35* (7), 2332-41.
- (496) Kahmann, J. D.; Sass, H. J.; Allan, M. G.; Seto, H.; Thompson, C. J.; Grzesiek, S. Structural Basis for Antibiotic Recognition by the Tipa Class of Multidrug-Resistance Transcriptional Regulators. *EMBO J* **2003**, *22* (8), 1824-34.
- (497) Myers, C. L.; Harris, J.; Yeung, J. C.; Honek, J. F. Molecular Interactions between Thiostrepton and the Tipas Protein from *Streptomyces Lividans*. *Chembiochem* **2014**, *15* (5), 681-7.
- (498) Ralph, S. A.; van Dooren, G. G.; Waller, R. F.; Crawford, M. J.; Fraunholz, M. J.; Foth, B. J.; Tonkin, C. J.; Roos, D. S.; McFadden, G. I. Tropical Infectious Diseases: Metabolic Maps and Functions of the *Plasmodium Falciparum* Apicoplast. *Nat Rev Micro* **2004**, *2* (3), 203-216.
- (499) McConkey, G. A.; Rogers, M. J.; McCutchan, T. F. Inhibition of *Plasmodium Falciparum* Protein Synthesis. Targeting the Plastid-Like Organelle with Thiostrepton. *J. Biol. Chem.* **1997**, *272* (4), 2046-9.
- (500) Clough, B.; Strath, M.; Preiser, P.; Denny, P.; Wilson, I. R. Thiostrepton Binds to Malarial Plastid rRNA. *FEBS Lett.* **1997**, *406* (1-2), 123-5.
- (501) Rogers, M. J.; Bukhman, Y. V.; McCutchan, T. F.; Draper, D. E. Interaction of Thiostrepton with an RNA Fragment Derived from the Plastid-Encoded Ribosomal RNA of the Malaria Parasite. *RNA* **1997**, *3* (8), 815-20.

- (502) Clough, B.; Rangachari, K.; Strath, M.; Preiser, P. R.; Wilson, R. J. Antibiotic Inhibitors of Organellar Protein Synthesis in Plasmodium Falciparum. *Protist* **1999**, *150* (2), 189-95.
- (503) Chaubey, S.; Kumar, A.; Singh, D.; Habib, S. The Apicoplast of Plasmodium Falciparum Is Translationally Active. *Mol. Microbiol.* **2005**, *56* (1), 81-89.
- (504) Goodman, C. D.; Su, V.; McFadden, G. I. The Effects of Anti-Bacterials on the Malaria Parasite Plasmodium Falciparum. *Mol. Biochem. Parasit.* **2007**, *152* (2), 181-91.
- (505) Schlitzer, M. Malaria Chemotherapeutics Part I: History of Antimalarial Drug Development, Currently Used Therapeutics, and Drugs in Clinical Development. *ChemMedChem* **2007**, *2* (7), 944-986.
- (506) Sullivan, M.; Li, J.; Kumar, S.; Rogers, M. J.; McCutchan, T. F. Effects of Interruption of Apicoplast Function on Malaria Infection, Development, and Transmission. *Mol. Biochem. Parasit.* **2000**, *109* (1), 17-23.
- (507) Rogers, M. J.; Cundliffe, E.; McCutchan, T. F. The Antibiotic Micrococcin Is a Potent Inhibitor of Growth and Protein Synthesis in the Malaria Parasite. *Antimicrob. Agents Chemother.* **1998**, *42* (3), 715-716.
- (508) Sharma, I.; Sullivan, M.; McCutchan, T. F. In Vitro Antimalarial Activity of Novel Semisynthetic Nocathiacin I Antibiotics. *Antimicrob. Agents Chemother.* **2015**, *59* (6), 3174-9.
- (509) Dahl, E. L.; Rosenthal, P. J. Apicoplast Translation, Transcription and Genome Replication: Targets for Antimalarial Antibiotics. *Trends in Parasitol.* **2008**, *24* (6), 279-284.
- (510) Wilson, D. W.; Goodman, C. D.; Sleeb, B. E.; Weiss, G. E.; de Jong, N. W. M.; Angrisano, F.; Langer, C.; Baum, J.; Crabb, B. S.; Gilson, P. R., et al. Macrolides Rapidly Inhibit Red Blood Cell Invasion by the Human Malaria Parasite, Plasmodium Falciparum. *BMC Biology* **2015**, *13*, 52.
- (511) Botté, C. Y.; Dubar, F.; McFadden, G. I.; Maréchal, E.; Biot, C. Plasmodium Falciparum Apicoplast Drugs: Targets or Off-Targets? *Chem. Rev.* **2012**, *112* (3), 1269-1283.
- (512) Kreidenweiss, A.; Kremsner, P. G.; Mordmüller, B. Comprehensive Study of Proteasome Inhibitors against Plasmodium Falciparum Laboratory Strains and Field Isolates from Gabon. *Malaria J.* **2008**, *7* (1), 1-8.
- (513) Gantt, S. M.; Myung, J. M.; Briones, M. R.; Li, W. D.; Corey, E. J.; Omura, S.; Nussenzweig, V.; Sinnis, P. Proteasome Inhibitors Block Development of Plasmodium Spp. *Antimicrob. Agents Chemother.* **1998**, *42*.
- (514) Lindenthal, C.; Weich, N.; Chia, Y. S.; Heussler, V.; Klinkert, M. Q. The Proteasome Inhibitor MLN-273 Blocks Exoerythrocytic and Erythrocytic Development of Plasmodium Parasites. *Parasitology* **2005**, *131*.

- (515) Schoof, S.; Pradel, G.; Aminake, M. N.; Ellinger, B.; Baumann, S.; Potowski, M.; Najajreh, Y.; Kirschner, M.; Arndt, H.-D. Antiplasmodial Thiostrepton Derivatives: Proteasome Inhibitors with a Dual Mode of Action. *Angew. Chem., Int. Ed.* **2010**, *49* (19), 3317-3321.
- (516) Gartel, A. L. Thiostrepton, Proteasome Inhibitors and Foxm1. *Cell Cycle* **2011**, *10* (24), 4341-4342.
- (517) Pandit, B.; Bhat, U. G.; Gartel, A. L. Proteasome Inhibitory Activity of Thiazole Antibiotics. *Cancer Biol. Ther.* **2011**, *11* (1), 43-7.
- (518) Jonghee, K. Cancer Therapeutic Agent Comprising Thiopeptide with Multiple Thiazole Rings. WO2002066046A1, 2002.
- (519) Nicolaou, K. C.; Zak, M.; Rahimipour, S.; Estrada, A. A.; Lee, S. H.; O'Brate, A.; Giannakakou, P.; Ghadiri, M. R. Discovery of a Biologically Active Thiostrepton Fragment. *J. Am. Chem. Soc.* **2005**, *127* (43), 15042-15044.
- (520) Bhat, U. G.; Zipfel, P. A.; Tyler, D. S.; Gartel, A. L. Novel Anticancer Compounds Induce Apoptosis in Melanoma Cells. *Cell Cycle* **2008**, *7* (12), 1851-5.
- (521) Radhakrishnan, S. K.; Bhat, U. G.; Hughes, D. E.; Wang, I.-C.; Costa, R. H.; Gartel, A. L. Identification of a Chemical Inhibitor of the Oncogenic Transcription Factor Forkhead Box M1. *Cancer Res.* **2006**, *66* (19), 9731-9735.
- (522) Kwok, J. M.-M.; Myatt, S. S.; Marson, C. M.; Coombes, R. C.; Constantinidou, D.; Lam, E. W.-F. Thiostrepton Selectively Targets Breast Cancer Cells through Inhibition of Forkhead Box M1 Expression. *Molecular Cancer Therapeutics* **2008**, *7* (7), 2022-2032.
- (523) Bhat, U. G.; Halasi, M.; Gartel, A. L. Thiazole Antibiotics Target FoxM1 and Induce Apoptosis in Human Cancer Cells. *PLoS ONE* **2009**, *4* (5), e5592.
- (524) Halasi, M.; Gartel, A. L. Fox(M1) News – It Is Cancer. *Mol. Cancer Ther.* **2013**, *12* (3), 245-254.
- (525) Halasi, M.; Gartel, A. L. Targeting FoxM1 in Cancer. *Biochem. Pharmacol.* **2013**, *85* (5), 644-52.
- (526) Bhat, U. G.; Halasi, M.; Gartel, A. L. Foxm1 Is a General Target for Proteasome Inhibitors. *PLoS ONE* **2009**, *4* (8), e6593.
- (527) Sayanjali, B.; Christensen, G. J. M.; Al-Zeer, M. A.; Mollenkopf, H.-J.; Meyer, T. F.; Brüggemann, H. Propionibacterium Acnes Inhibits FoxM1 and Induces Cell Cycle Alterations in Human Primary Prostate Cells. *Int. J. Med. Microbiol.* **2016**.
- (528) Tomko, R. J.; Hochstrasser, M. Molecular Architecture and Assembly of the Eukaryotic Proteasome. *Annu. Rev. Biochem.* **2013**, *82*, 415-45.
- (529) Sandu, C.; Chandramouli, N.; Glickman, J. F.; Molina, H.; Kuo, C.-L.; Kukushkin, N.; Goldberg, A. L.; Steller, H. Thiostrepton Interacts Covalently with Rpt Subunits of the 19s Proteasome and Proteasome Substrates. *J. Cell. Mol. Med.* **2015**, *19* (9), 2181-2192.

- (530) Halasi, M.; Varaljai, R.; Benevolenskaya, E.; Gartel, A. L. A Novel Function of Molecular Chaperone Hsp70: Suppression of Oncogenic FoxM1 after Proteotoxic Stress. *J. Biol. Chem.* **2016**, *291* (1), 142-8.
- (531) Hegde, N. S.; Sanders, D. A.; Rodriguez, R.; Balasubramanian, S. The Transcription Factor FoxM1 Is a Cellular Target of the Natural Product Thiostrepton. *Nat. Chem.* **2011**, *3* (9), 725-731.
- (532) Zhang, L.; Ging, N. C.; Komoda, T.; Hanada, T.; Suzuki, T.; Watanabe, K. Antibiotic Susceptibility of Mammalian Mitochondrial Translation. *FEBS Letters* **2005**, *579* (28), 6423-6427.
- (533) Bowling, B. D.; Doudican, N.; Manga, P.; Orlow, S. J. Inhibition of Mitochondrial Protein Translation Sensitizes Melanoma Cells to Arsenic Trioxide Cytotoxicity Via a Reactive Oxygen Species Dependent Mechanism. *Cancer Chemoth. Pharm.* **2008**, *63* (1), 37-43.
- (534) Ueno, M.; Furukawa, S.; Abe, F.; Ushioda, M.; Fujine, K.; Johki, S.; Hatori, H.; Ueda, H. Suppressive Effect of Antibiotic Siomycin on Antibody Production. *J Antibiot* **2004**, *57* (9), 590-6.
- (535) Miyajima, A.; Kosaka, T.; Kikuchi, E.; Oya, M. Renin-Angiotensin System Blockade: Its Contribution and Controversy. *Int. J. Urol.* **2015**, *22* (8), 721-30.
- (536) Kehoe, P. G.; Wilcock, G. K. Is Inhibition of the Renin-Angiotensin System a New Treatment Option for Alzheimer's Disease? *Lancet. Neurology* **2007**, *6* (4), 373-8.
- (537) Silva, L. S.; Silva-Filho, J. L.; Caruso-Neves, C.; Pinheiro, A. A. S. New Concepts in Malaria Pathogenesis: The Role of the Renin-Angiotensin System. *Front. Cell. Infect. Microbiol.* **2015**, *5*, 103.
- (538) Sparks, M. A.; Crowley, S. D.; Gurley, S. B.; Mirotsoy, M.; Coffman, T. M. Classical Renin-Angiotensin System in Kidney Physiology. *Compr. Physiol.* **2014**, *4* (3), 1201-1228.
- (539) Sajid, I.; Shaaban, K. A.; Frauendorf, H.; Hasnain, S.; Laatsch, H. Val-Geninthiocin: Structure Elucidation and MSN Fragmentation of Thiopeptide Antibiotics Produced by *Streptomyces* Sp. Rsf18. *Z. Naturforsch., B: Chem. Sci.* **2008**, *63* (10), 1223-1230.
- (540) Longstaff, C.; Kolev, K. Basic Mechanisms and Regulation of Fibrinolysis. *J. Thromb. Haemost.* **2015**, *13* (S1), S98-105.
- (541) Miles, L. A.; Parmer, R. J. Plasminogen Receptors: The First Quarter Century. *Semin. Thromb. Hemost.* **2013**, *39* (4), 329-337.
- (542) Zhang, F.; Kelly, W. L. In Vivo Production of Thiopeptide Variants. *Methods Enzymol.* **2012**, *516*, 3-24.
- (543) Sivonen, K.; Leikoski, N.; Fewer, D. P.; Jokela, J. Cyanobactins-Ribosomal Cyclic Peptides Produced by Cyanobacteria. *Appl. Microbiol. Biotechnol.* **2010**, *86* (5), 1213-25.

- (544) Leikoski, N.; Liu, L.; Jokela, J.; Wahlsten, M.; Gugger, M.; Calteau, A.; Permi, P.; Kerfeld, C. A.; Sivonen, K.; Fewer, D. P. Genome Mining Expands the Chemical Diversity of the Cyanobactin Family to Include Highly Modified Linear Peptides. *Chem. Biol.* **2013**, *20* (8), 1033-1043.
- (545) Leikoski, N.; Fewer, D. P.; Jokela, J.; Wahlsten, M.; Rouhiainen, L.; Sivonen, K. Highly Diverse Cyanobactins in Strains of the Genus *Anabaena*. *Appl. Environ. Microbiol.* **2010**, *76* (3), 701-709.
- (546) Martins, J.; Vasconcelos, V. Cyanobactins from Cyanobacteria: Current Genetic and Chemical State of Knowledge. *Mar. Drugs* **2015**, *13* (11), 6910-6946.
- (547) Houssen, W. E.; Jaspars, M. Azole-Based Cyclic Peptides from the Sea Squirt *Lissoclinum Patella*: Old Scaffolds, New Avenues. *ChemBioChem* **2010**, *11* (13), 1803-15.
- (548) McIntosh, J. A.; Donia, M. S.; Schmidt, E. W. Insights into Heterocyclization from Two Highly Similar Enzymes. *J. Am. Chem. Soc.* **2010**, *132* (12), 4089-4091.
- (549) Ziemert, N.; Ishida, K.; Quillardet, P.; Bouchier, C.; Hertweck, C.; de Marsac, N. T.; Dittmann, E. Microcyclamide Biosynthesis in Two Strains of *Microcystis Aeruginosa*: From Structure to Genes and Vice Versa. *Appl. Environ. Microbiol.* **2008**, *74* (6), 1791-1797.
- (550) Zheng, L.-H.; Wang, Y.-J.; Sheng, J.; Wang, F.; Zheng, Y.; Lin, X.-K.; Sun, M. Antitumor Peptides from Marine Organisms. *Mar. Drugs* **2011**, *9* (10), 1840-1859.
- (551) Zabriskie, M.; Foster, M. P.; Stout, T. J.; Clardy, J.; Ireland, C. M. Studies on the Solution- and Solid-State Structure of Patellin 2. *J. Am. Chem. Soc.* **1990**, *112* (22).
- (552) Parajuli, A.; Kwak, D. H.; Dalponte, L.; Leikoski, N.; Galica, T.; Umeobika, U.; Trembleau, L.; Bent, A.; Sivonen, K.; Wahlsten, M., et al. A Unique Tryptophan C-Prenyltransferase from the Kawaguchi-peptin Biosynthetic Pathway. *Angew. Chem. Int. Ed.* **2016**, *55* (11), 3596-3599.
- (553) McIntosh, J. A.; Lin, Z.; Tianero, M. D. B.; Schmidt, E. W. Aestuaramides, a Natural Library of Cyanobactin Cyclic Peptides Resulting from Isoprene-Derived Claisen Rearrangements. *ACS Chem. Biol.* **2013**, *8* (5), 877-883.
- (554) Schmitz, F. J.; Ksebati, M. B.; Chang, J. S.; Wang, J. L.; Hossain, M. B.; Van der Helm, D.; Engel, M. H.; Serban, A.; Silfer, J. A. Cyclic Peptides from the Ascidian *Lissoclinum Patella*: Conformational Analysis of Patellamide D by X-Ray Analysis and Molecular Modeling. *J. Org. Chem.* **1989**, *54* (14), 3463-3472.
- (555) Wipf, P.; Uto, Y. Total Synthesis and Revision of Stereochemistry of the Marine Metabolite Trunkamide A. *J. Org. Chem.* **2000**, *65* (4), 1037-1049.
- (556) Caba, J. M.; Rodriguez, I. M.; Manzanares, I.; Giralt, E.; Albericio, F. Solid-Phase Total Synthesis of Trunkamide A(1). *J. Org. Chem.* **2001**, *66* (23), 7568-7574.
- (557) Ireland, C.; Scheuer, P. J. Ulicyclamide and Ulithiacyclamide, Two New Small Peptides from a Marine Tunicate. *J. Am. Chem. Soc.* **1980**, *102* (17), 5688-5691.

- (558) Williams, D. E.; Moore, R. E.; Paul, V. J. The Structure of Ulithiacyclamide B. Antitumor Evaluation of Cyclic Peptides and Macrolides from *Lissoclinum Patella*. *J. Nat. Prod.* **1989**, *52* (4), 732-739.
- (559) Fu, X.; Do, T.; Schmitz, F. J.; Andrushevich, V.; Engel, M. H. New Cyclic Peptides from the Ascidian *Lissoclinum Patella*. *J. Nat. Prod.* **1998**, *61* (12), 1547-1551.
- (560) Williams, R. S. J. A Marine Natural Product, Patellamide D, Reverses Multidrug Resistance in a Human Leukemic Cell Line. *Cancer Lett.* **1993**, *71* (1-3), 97-102.
- (561) Garcia, G. D.; Bowden, B. F. A Cyclic Hepta-Peptide Derivative from Colonial Ascidiaceans, *Lissoclinum* Sp. CA2252099 A1, 1997/10/23/, 1997.
- (562) Portmann, C.; Blom, J. F.; Kaiser, M.; Brun, R.; Jüttner, F.; Gademann, K. Isolation of Aerucyclamides C and D and Structure Revision of Microcyclamide 7806A: Heterocyclic Ribosomal Peptides from *Microcystis Aeruginosa* PCC 7806 and Their Antiparasite Evaluation. *J. Nat. Prod.* **2008**, *71* (11), 1891-1896.
- (563) Ishida, K.; Matsuda, H.; Murakami, M.; Yamaguchi, K. Kawaguchi-peptin B, an Antibacterial Cyclic Undecapeptide from the Cyanobacterium *Microcystis Aeruginosa*. *J. Nat. Prod.* **1997**, *60* (7), 724-726.
- (564) García-Reynaga, P.; VanNieuwenhze, M. S. A New Total Synthesis of Patellamide A. *Organic Letters* **2008**, *10* (20), 4621-4623.
- (565) Sugiyama, Y.; Eguchi, F.; Miyazaki, A.; Hayashi, K.; Takahashi, H.; Hamamoto, H.; Shioiri, T.; Matsugi, M. A Synthesis of All Stereoisomers of Tenucyclamide a Employing a Fluorous-Fmoc Strategy. *J. Org. Chem.* **2013**, *78* (20), 10264-10272.
- (566) You, S.-L.; Deechongkit, S.; Kelly, J. W. Solid-Phase Synthesis and Stereochemical Assignments of Tenucyclamides a-D Employing Heterocyclic Amino Acids Derived from Commercially Available Fmoc Alpha-Amino Acids. *Organic Letters* **2004**, *6* (15), 2627-2630.
- (567) Schmidt, U.; Griesser, H. Total Synthesis and Structure Determination of Patellamide B. *Tetrahedron Lett.* **1986**, *27* (2), 163-166.
- (568) Wipf, P.; Fritch, P. C. Total Synthesis and Assignment of Configuration of *Lissoclinamide 7*. *J. Am. Chem. Soc.* **1996**, *118* (49), 12358-12367.
- (569) Boden, C. D. J.; Pattenden, G. Cyclopeptides from Ascidiaceans. Total Synthesis of *Lissoclinamide 4*, and a General Strategy for the Synthesis of Chiral Thiazoline-Containing Macrocyclic Peptides. *Tetrahedron Lett.* **1995**, *36* (34), 6153-6156.
- (570) Boden, C.; Pattenden, G. Total Synthesis of *Lissoclinamide 5*, a Cytotoxic Cyclic Peptide from the Tunicate *Lissoclinum Patella*. *Tetrahedron Lett.* **1994**, *35* (44), 8271-8274.
- (571) Leeson, P. Drug Discovery: Chemical Beauty Contest. *Nature* **2012**, *481* (7382), 455-456.

- (572) Hewitt, W. M.; Leung, S. S. F.; Pye, C. R.; Ponkey, A. R.; Bednarek, M.; Jacobson, M. P.; Lokey, R. S. Cell-Permeable Cyclic Peptides from Synthetic Libraries Inspired by Natural Products. *J. Am. Chem. Soc.* **2015**, *137* (2), 715-721.
- (573) Ahlback, C. L.; Lexa, K. W.; Bockus, A. T.; Chen, V.; Crews, P.; Jacobson, M. P.; Lokey, R. S. Beyond Cyclosporine A: Conformation-Dependent Passive Membrane Permeabilities of Cyclic Peptide Natural Products. *Future Med. Chem.* **2015**, *7* (16), 2121-2130.
- (574) Doak, B. C.; Zheng, J.; Dobritzsch, D.; Kihlberg, J. How Beyond Rule of 5 Drugs and Clinical Candidates Bind to Their Targets. *J. Med. Chem.* **2016**, *59* (6), 2312-2327.
- (575) Ko, S. Y.; Dalvit, C. Conformation of Cyclosporin a in Polar Solvents. *Int. J. Pept. Prot. Res.* **1992**, *40* (5), 380-382.
- (576) Augustijns, P. F.; Brown, S. C.; Willard, D. H.; Consler, T. G.; Annaert, P. P.; Hendren, R. W.; Bradshaw, T. P. Hydration Changes Implicated in the Remarkable Temperature-Dependent Membrane Permeation of Cyclosporin A. *Biochemistry* **2000**, *39* (25), 7621-7630.
- (577) Nielsen, D. S.; Hoang, H. N.; Lohman, R.-J.; Hill, T. A.; Lucke, A. J.; Craik, D. J.; Edmonds, D. J.; Griffith, D. A.; Rotter, C. J.; Ruggeri, R. B., et al. Improving on Nature: Making a Cyclic Heptapeptide Orally Bioavailable. *Angew. Chem. Int. Ed.* **2014**, *53* (45), 12059-12063.
- (578) Bockus, A. T.; McEwen, C. M.; Lokey, R. S. Form and Function in Cyclic Peptide Natural Products: A Pharmacokinetic Perspective. *Curr. Top. Med. Chem.* **2013**, *13* (7), 821-836.
- (579) Heinis, C. Drug Discovery: Tools and Rules for Macrocycles. *Nat. Chem. Biol.* **2014**, *10* (9), 696-698.
- (580) Namjoshi, S.; Benson, H. A. E. Cyclic Peptides as Potential Therapeutic Agents for Skin Disorders. *Biopolymers* **2010**, *94* (5), 673-680.
- (581) Bindman, N. A.; Der Donk, W. A. V., Ripps: Ribosomally Synthesized and Posttranslationally Modified Peptides. In *Natural Products*, Osbourn, A.; Goss, R. J.; Carter, G. T., Eds. John Wiley & Sons, Inc.: 2014; pp 195-217.
- (582) de Ruyter, P. G.; Kuipers, O. P.; Beerthuyzen, M. M.; van Alen-Boerrigter, I.; de Vos, W. M. Functional Analysis of Promoters in the Nisin Gene Cluster of *Lactococcus Lactis*. *J. Bacteriol.* **1996**, *178* (12), 3434-3439.
- (583) Severinov, K.; Nair, S. K. Microcin C: Biosynthesis and Mechanisms of Bacterial Resistance. *Future Microbiol.* **2012**, *7* (2), 281-289.
- (584) Ruffner, D. E.; Schmidt, E. W.; Heemstra, J. R. Assessing the Combinatorial Potential of the RiPP Cyanobactin Tru Pathway. *ACS Synth. Biol.* **2015**, *4* (4), 482-492.
- (585) Koehnke, J.; Bent, A. F.; Houssen, W. E.; Mann, G.; Jaspars, M.; Naismith, J. H. The Structural Biology of Patellamide Biosynthesis. *Curr. Opin. Struct. Biol.* **2014**, *29*, 112-121.

- (586) Houssen, W. E.; Koehnke, J.; Zollman, D.; Vendome, J.; Raab, A.; Smith, M. C. M.; Naismith, J. H.; Jaspars, M. The Discovery of New Cyanobactins from Cyanotheca PCC 7425 Defines a New Signature for Processing of Patellamides. *ChemBiochem* **2012**, *13* (18), 2683-2689.
- (587) Agarwal, V.; Pierce, E.; McIntosh, J.; Schmidt, E. W.; Nair, S. K. Structures of Cyanobactin Maturation Enzymes Define a Family of Transamidating Proteases. *Chem. Biol.* **2012**, *19* (11), 1411-1422.
- (588) Mann, G.; Koehnke, J.; Bent, A. F.; Graham, R.; Houssen, W.; Jaspars, M.; Schwarz-Linek, U.; Naismith, J. H. The Structure of the Cyanobactin Domain of Unknown Function from PatG in the Patellamide Gene Cluster. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2014**, *70* (12), 1597-1603.
- (589) Koehnke, J.; Bent, A.; Houssen, W. E.; Zollman, D.; Morawitz, F.; Shirran, S.; Vendome, J.; Nneoyiegbe, A. F.; Trembleau, L.; Botting, C. H., et al. The Mechanism of Patellamide Macrocyclization Revealed by the Characterization of the Patg Macrocyclase Domain. *Nat. Struct. Mol. Biol.* **2012**, *19* (8), 767-772.
- (590) Leikoski, N.; Fewer, D. P.; Jokela, J.; Alakoski, P.; Wahlsten, M.; Sivonen, K. Analysis of an Inactive Cyanobactin Biosynthetic Gene Cluster Leads to Discovery of New Natural Products from Strains of the Genus *Microcystis*. *PLoS One* **2012**, *7* (8), e43002.
- (591) Regni, C. A.; Roush, R. F.; Miller, D. J.; Nourse, A.; Walsh, C. T.; Schulman, B. A. How the Mccb Bacterial Ancestor of Ubiquitin E1 Initiates Biosynthesis of the Microcin C7 Antibiotic. *EMBO J.* **2009**, *28*, 1953 - 1964.
- (592) Koehnke, J.; Bent, A. F.; Zollman, D.; Smith, K.; Houssen, W. E.; Zhu, X.; Mann, G.; Lebl, T.; Scharff, R.; Shirran, S., et al. The Cyanobactin Heterocyclase Enzyme: A Processive Adenylase That Operates with a Defined Order of Reaction. *Angew. Chem. Int. Ed.* **2013**, *52* (52), 13991-13996.
- (593) Koehnke, J.; Mann, G.; Bent, A. F.; Ludewig, H.; Shirran, S.; Botting, C.; Lebl, T.; Houssen, W. E.; Jaspars, M.; Naismith, J. H. Structural Analysis of Leader Peptide Binding Enables Leader-Free Cyanobactin Processing. *Nat. Chem. Biol.* **2015**, *11* (8), 558-563.
- (594) McIntosh, J. A.; Schmidt, E. W. Marine Molecular Machines: Heterocyclization in Cyanobactin Biosynthesis. *ChemBioChem* **2010**, *11* (10), 1413-1421.
- (595) Sardar, D.; Pierce, E.; McIntosh, J. A.; Schmidt, E. W. Recognition Sequences and Substrate Evolution in Cyanobactin Biosynthesis. *ACS Synth. Biol.* **2015**, *4* (2), 167-176.
- (596) Houssen, W. E.; Bent, A. F.; McEwan, A. R.; Pieiller, N.; Tabudravu, J.; Koehnke, J.; Mann, G.; Adaba, R. I.; Thomas, L.; Hawas, U. W., et al. An Efficient Method for the in Vitro Production of Azol(in)e-Based Cyclic Peptides. *Angew. Chem. Int. Ed.* **2014**, *53* (51), 14171-14174.

- (597) McIntosh, J. A.; Donia, M. S.; Nair, S. K.; Schmidt, E. W. Enzymatic Basis of Ribosomal Peptide Prenylation in Cyanobacteria. *J. Am. Chem. Soc.* **2011**, *133* (34), 13698-13705.
- (598) Bent, A. F.; Koehnke, J.; Houssen, W. E.; Smith, M. C. M.; Jaspars, M.; Naismith, J. H. Structure of Patf from Prochloron Didemni. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2013**, *69* (6), 618-623.
- (599) Tianero, M. D. B.; Donia, M. S.; Young, T. S.; Schultz, P. G.; Schmidt, E. W. Ribosomal Route to Small-Molecule Diversity. *J. Am. Chem. Soc.* **2012**, *134* (1), 418-425.
- (600) Hao, Y.; Pierce, E.; Roe, D.; Morita, M.; McIntosh, J. A.; Agarwal, V.; Cheatham, T. E.; Schmidt, E. W.; Nair, S. K. Molecular Basis for the Broad Substrate Selectivity of a Peptide Prenyltransferase. *Proc. Natl. Acad. Sci. USA* **2016**, *113* (49), 14037-14042.
- (601) Donia, M. S.; Schmidt, E. W. Linking Chemistry and Genetics in the Growing Cyanobactin Natural Products Family. *Chem. Biol.* **2011**, *18* (4), 508-519.
- (602) McIntosh, J. A.; Robertson, C. R.; Agarwal, V.; Nair, S. K.; Bulaj, G. W.; Schmidt, E. W. Circular Logic: Nonribosomal Peptide-Like Macrocyclization with a Ribosomal Peptide Catalyst. *J. Am. Chem. Soc.* **2010**, *132* (44), 15499-15501.
- (603) Oman, T. J.; Donk, W. A. Follow the Leader: The Use of Leader Peptides to Guide Natural Product Biosynthesis. *Nat. Chem. Biol.* **2010**, *6*, 9 - 18.
- (604) Haft, D. H.; Basu, M. K.; Mitchell, D. A. Expansion of Ribosomally Produced Natural Products: A Nitrile Hydratase- and Nif11-Related Precursor Family. *BMC Biol.* **2010**, *8*.
- (605) Kobayashi, M.; Shimizu, S. Metalloenzyme Nitrile Hydratase: Structure, Regulation and Application to Biotechnology. *Nat. Biotechnol.* **1998**, *16*, 733 - 736.
- (606) Harrop, T. C.; Mascharak, P. K. Fe(III) and Co(III) Centers with Carboxamido Nitrogen and Modified Sulfur Coordination: Lessons Learned from Nitrile Hydratase. *Acc Chem Res* **2004**, *37* (4), 253-60.
- (607) Jacobson, M. R.; Brigle, K. E.; Bennett, L. T.; Setterquist, R. A.; Wilson, M. S.; Cash, V. L.; Beynon, J.; Newton, W. E.; Dean, D. R. Physical and Genetic Map of the Major Nif Gene Cluster from *Azotobacter Vinelandii*. *J. Bacteriol.* **1989**, *171* (2), 1017-27.
- (608) Kaneko, T.; Nakamura, Y.; Wolk, C. P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, M.; Ishikawa, A.; Kawashima, K.; Kimura, T., et al. Complete Genomic Sequence of the Filamentous Nitrogen-Fixing Cyanobacterium *Anabaena* Sp. Strain PCC 7120. *DNA Res.* **2001**, *8* (5), 205-13; 227-53.
- (609) Li, B.; Sher, D.; Kelly, L.; Shi, Y.; Huang, K.; Knerr, P. J.; Joewono, I.; Rusch, D.; Chisholm, S. W.; van der Donk, W. A. Catalytic Promiscuity in the Biosynthesis of Cyclic Peptide Secondary

- Metabolites in Planktonic Marine Cyanobacteria. *Proc. Nat. Acad. Sci. USA* **2010**, *107* (23), 10430-5.
- (610) Freeman, M. F.; Gurgui, C.; Helf, M. J.; Morinaka, B. I.; Uria, A. R.; Oldham, N. J.; Sahl, H. G.; Matsunaga, S.; Piel, J. Metagenome Mining Reveals Polytheonamides as Posttranslationally Modified Ribosomal Peptides. *Science* **2012**, *338* (6105), 387-90.
- (611) Tang, W.; van der Donk, W. A. Structural Characterization of Four Prochlorosins: A Novel Class of Lantipeptides Produced by Planktonic Marine Cyanobacteria. *Biochemistry* **2012**, *51* (21), 4271-4279.
- (612) Zhang, Q.; Yang, X.; Wang, H.; van der Donk, W. A. High Divergence of the Precursor Peptides in Combinatorial Lanthipeptide Biosynthesis. *ACS Chem. Biol.* **2014**, *9* (11), 2686-2694.
- (613) Sohda, K. Y.; Nagai, K.; Yamori, T.; Suzuki, K.; Tanaka, A. Ym-216391, a Novel Cytotoxic Cyclic Peptide from *Streptomyces Nobilis*. I. Fermentation, Isolation and Biological Activities. *J Antibiot* **2005**, *58* (1), 27-31.
- (614) Sohda, K. Y.; Hiramoto, M.; Suzumura, K.; Takebayashi, Y.; Suzuki, K.; Tanaka, A. Ym-216391, a Novel Cytotoxic Cyclic Peptide from *Streptomyces Nobilis*. II. Physico-Chemical Properties and Structure Elucidation. *J Antibiot* **2005**, *58* (1), 32-6.
- (615) Deeley, J.; Pattenden, G. Synthesis and Establishment of Stereochemistry of the Unusual Polyoxazole-Thiazole Based Cyclopeptide Ym-216391 Isolated from *Streptomyces Nobilis*. *Chem. Commun.* **2005**, (6), 797-9.
- (616) Deeley, J.; Bertram, A.; Pattenden, G. Novel Polyoxazole-Based Cyclopeptides from *Streptomyces* Sp. Total Synthesis of the Cyclopeptide YM-216391 and Synthetic Studies Towards Telomestatin. *Org Biomol Chem* **2008**, *6* (11), 1994-2010.
- (617) Shin-ya, K.; Wierzba, K.; Matsuo, K.-i.; Ohtani, T.; Yamada, Y.; Furihata, K.; Hayakawa, Y.; Seto, H. Telomestatin, a Novel Telomerase Inhibitor from *Streptomyces Anulatus*. *J. Am. Chem. Soc.* **2001**, *123* (6), 1262-1263.
- (618) Jian, X. H.; Pan, H. X.; Ning, T. T.; Shi, Y. Y.; Chen, Y. S.; Li, Y.; Zeng, X. W.; Xu, J.; Tang, G. L. Analysis of YM-216391 Biosynthetic Gene Cluster and Improvement of the Cyclopeptide Production in a Heterologous Host. *ACS Chem. Biol.* **2012**, *7* (4), 646-51.
- (619) Zhou, X.; Huang, H.; Chen, Y.; Tan, J.; Song, Y.; Zou, J.; Tian, X.; Hua, Y.; Ju, J. Marthiapeptide a, an Anti-Infective and Cytotoxic Polythiazole Cyclopeptide from a 60 L Scale Fermentation of the Deep Sea-Derived *Marinactinospora Thermotolerans* Scsio 00652. *J. Nat. Prod.* **2012**, *75* (12), 2251-2255.
- (620) Milne, J. C.; Roy, R. S.; Eliot, A. C.; Kelleher, N. L.; Wokhlu, A.; Nickels, B.; Walsh, C. T. Cofactor Requirements and Reconstitution of Microcin B17 Synthetase: A Multienzyme Complex

- That Catalyzes the Formation of Oxazoles and Thiazoles in the Antibiotic Microcin B17. *Biochemistry* **1999**, *38*, 4768 - 4781.
- (621) Milne, J. C.; Eliot, A. C.; Kelleher, N. L.; Walsh, C. T. ATP/GTP Hydrolysis Is Required for Oxazole and Thiazole Biosynthesis in the Peptide Antibiotic Microcin B17. *Biochemistry* **1998**, *37* (38), 13250-13261.
- (622) Zamble, D. B.; McClure, C. P.; Penner-Hahn, J. E.; Walsh, C. T. The Mcbb Component of Microcin B17 Synthetase Is a Zinc Metalloprotein. *Biochemistry* **2000**, *39* (51), 16190-16199.
- (623) Kelleher, N. L.; Hendrickson, C. L.; Walsh, C. T. Posttranslational Heterocyclization of Cysteine and Serine Residues in the Antibiotic Microcin B17: Distributivity and Directionality. *Biochemistry* **1999**, *38*, 15623 - 15630.
- (624) Huang, Y.; Gan, H.; Li, S.; Xu, J.; Wu, X.; Yao, H. Oxidation of 4-Carboxylate Thiazolines to 4-Carboxylate Thiazoles by Molecular Oxygen. *Tetrahedron Lett.* **2010**, *51* (13), 1751-1753.
- (625) Balaban, A. T.; Oniciu, D. C.; Katritzky, A. R. Aromaticity as a Cornerstone of Heterocyclic Chemistry. *Chemical Reviews* **2004**, *104* (5), 2777-2812.
- (626) Boger, D. L. Diels-Alder Reactions of Azadienes. *Tetrahedron* **1983**, *39* (18), 2869-2939.
- (627) Boger, D. L. Diels-Alder Reactions of Heterocyclic Aza Dienes. Scope and Applications. *Chem. Rev.* **1986**, *86* (5), 781-793.
- (628) Banker, R.; Carmeli, S. Tenucyclamides a-D, Cyclic Hexapeptides from the Cyanobacterium *Nostoc Spongiaeforme* Var. *Tenue*. *J. Nat. Prod.* **1998**, *61* (10), 1248-1251.
- (629) Milne, B. F.; Long, P. F.; Starcevic, A.; Hranueli, D.; Jaspars, M. Spontaneity in the Patellamide Biosynthetic Pathway. *Org. Biomol. Chem.* **2006**, *4* (4), 631-8.
- (630) Shimamura, H.; Gouda, H.; Nagai, K.; Hirose, T.; Ichioka, M.; Furuya, Y.; Kobayashi, Y.; Hirono, S.; Sunazuka, T.; Omura, S. Structure Determination and Total Synthesis of Bottromycin A2:A Potent Antibiotic against Mrsa and Vre. *Angew. Chem., Int. Ed.* **2009**, *48* (5), 914-917.
- (631) Crone, W. J.; Vior, N. M.; Santos-Aberturas, J.; Schmitz, L. G.; Leeper, F. J.; Truman, A. W. Dissecting Bottromycin Biosynthesis Using Comparative Untargeted Metabolomics. *Angew. Chem. Int. Ed.* **2016**.
- (632) Zhang, Z.; Hudson, G. A.; Mahanta, N.; Tietz, J. I.; van der Donk, W. A.; Mitchell, D. A. Biosynthetic Timing and Substrate Specificity for the Thiopeptide Thiomuracin. *J. Am. Chem. Soc.* **2016**, *138* (48), 15511-15514.
- (633) Belshaw, P. J.; Roy, R. S.; Kelleher, N. L.; Walsh, C. T. Kinetics and Regioselectivity of Peptide-to-Heterocycle Conversions by Microcin B17 Synthetase. *Chem. Biol.* **1998**, *5* (7), 373-84.

- (634) Koehnke, J.; Morawitz, F.; Bent, A. F.; Houssen, W. E.; Shirran, S. L.; Fuszard, M. A.; Smellie, I. A.; Botting, C. H.; Smith, M. C. M.; Jaspars, M., et al. An Enzymatic Route to Selenazolines. *ChemBioChem* **2013**, *14* (5), 564-567.
- (635) Roy, R. S.; Allen, O.; Walsh, C. T. Expressed Protein Ligation to Probe Regiospecificity of Heterocyclization in the Peptide Antibiotic Microcin B17. *Chem. Biol.* **1999**, *6* (11), 789-99.
- (636) Goto, Y.; Ito, Y.; Kato, Y.; Tsunoda, S.; Suga, H. One-Pot Synthesis of Azoline-Containing Peptides in a Cell-Free Translation System Integrated with a Posttranslational Cyclodehydratase. *Chem. Biol.* **2014**, *21* (6), 766-774.
- (637) Kohda, K.; Ohta, Y.; Yokoyama, Y.; Kawazoe, Y.; Kato, T.; Suzumura, Y.; Hamada, Y.; Shioiri, T. Mechanistic Aspects of the Cytocidal Action of Ulithiacyclamide on Mouse Leukemia L1210 Cells in Vitro. *Biochem. Pharmacol.* **1989**, *38* (24), 4497-4500.
- (638) Roy, R. S.; Belshaw, P. J.; Walsh, C. T. Mutational Analysis of Posttranslational Heterocycle Biosynthesis in the Gyrase Inhibitor Microcin B17: Distance Dependence from Propeptide and Tolerance for Substitution in a Gscg Cyclizable Sequence. *Biochemistry* **1998**, *37* (12), 4125-4136.
- (639) Kelleher, N. L.; Belshaw, P. J.; Walsh, C. T. Regioselectivity and Chemoselectivity Analysis of Oxazole and Thiazole Ring Formation by the Peptide-Heterocyclizing Microcin B17 Synthetase Using High-Resolution MS/MS. *J. Am. Chem. Soc.* **1998**, *120* (37), 9716-9717.
- (640) Nov, Y. When Second Best Is Good Enough: Another Probabilistic Look at Saturation Mutagenesis. *Appl. Environ. Microbiol.* **2012**, *78* (1), 258-262.
- (641) Oueis, E.; Adamson, C.; Mann, G.; Ludewig, H.; Westwood, N. J.; Naismith, J. H.; Redpath, P.; Migaud, M.; Naismith, J. H. Derivatisable Cyanobactin Analogues: A Semisynthetic Approach. *Chembiochem* **2015**, *16* (18), 2646-50.
- (642) Schmidt, E. W. The Hidden Diversity of Ribosomal Peptide Natural Products. *BMC Biol.* **2010**, *8*, 83.
- (643) Madison, L. L.; Vivas, E. I.; Li, Y. M.; Walsh, C. T.; Kolter, R. The Leader Peptide Is Essential for the Post-Translational Modification of the DNA-Gyrase Inhibitor Microcin B17. *Mol. Microbiol.* **1997**, *23* (1), 161-8.
- (644) Sardar, D.; Lin, Z.; Schmidt, E. W. Modularity of Ripp Enzymes Enables Designed Synthesis of Decorated Peptides. *Chem. Biol.* **2015**, *22* (7), 907-16.
- (645) Oman, T. J.; Knerr, P. J.; Bindman, N. A.; Velasquez, J. E.; van der Donk, W. A. An Engineered Lantibiotic Synthetase That Does Not Require a Leader Peptide on Its Substrate. *J. Am. Chem. Soc.* **2012**, *134* (16), 6952-5.
- (646) Dunbar, K. L.; Melby, J. O.; Mitchell, D. A. Ycao Domains Use ATP to Activate Amide Backbones During Peptide Cyclodehydrations. *Nat Chem Biol* **2012**, *8* (6), 569-75.

- (647) Cox, C. L.; Doroghazi, J. R.; Mitchell, D. A. The Genomic Landscape of Ribosomal Peptides Containing Thiazole and Oxazole Heterocycles. *BMC Genomics* **2015**, *16* (1), 1-16.
- (648) Dunbar, K. L.; Tietz, J. I.; Cox, C. L.; Burkhart, B. J.; Mitchell, D. A. Identification of an Auxiliary Leader Peptide-Binding Protein Required for Azoline Formation in Ribosomal Natural Products. *J. Am. Chem. Soc.* **2015**, *137* (24), 7672-7.
- (649) Strader, M. B.; VerBerkmoes, N. C.; Tabb, D. L.; Connelly, H. M.; Barton, J. W.; Bruce, B. D.; Pelletier, D. A.; Davison, B. H.; Hettich, R. L.; Larimer, F. W., et al. Characterization of the 70S Ribosome from *Rhodospseudomonas Palustris* Using an Integrated “Top-Down” and “Bottom-up” Mass Spectrometric Approach. *J. Proteome Res.* **2004**, *3* (5), 965-978.
- (650) Suh, M.-J.; Hamburg, D.-M.; Gregory, S. T.; Dahlberg, A. E.; Limbach, P. A. Extending Ribosomal Protein Identifications to Unsequenced Bacterial Strains Using Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Proteomics* **2005**, *5* (18), 4818-4831.
- (651) Noeske, J.; Wasserman, M. R.; Terry, D. S.; Altman, R. B.; Blanchard, S. C.; Cate, J. H. D. High-Resolution Structure of the *Escherichia Coli* Ribosome. *Nat. Struct. Mol. Biol.* **2015**, *22* (4), 336-341.
- (652) Louie, D. F.; Resing, K. A.; Lewis, T. S.; Ahn, N. G. Mass Spectrometric Analysis of 40 S Ribosomal Proteins from Rat-1 Fibroblasts. *J. Biol. Chem.* **1996**, *271* (45), 28189-28198.
- (653) Lee, K. H.; Saleh, L.; Anton, B. P.; Madinger, C. L.; Benner, J. S.; Iwig, D. F.; Roberts, R. J.; Krebs, C.; Booker, S. J. Characterization of RimO, a New Member of the Methylthiotransferase Subclass of the Radical SAM Superfamily. *Biochemistry* **2009**, *48* (42), 10162-74.
- (654) Anton, B. P.; Saleh, L.; Benner, J. S.; Raleigh, E. A.; Kasif, S.; Roberts, R. J. Rimo, a MiaB-Like Enzyme, Methylthiolates the Universally Conserved Asp88 Residue of Ribosomal Protein S12 in *Escherichia Coli*. *Proc. Natl. Acad. Sci. USA* **2008**, *105* (6), 1826-1831.
- (655) Strader, M. B.; Costantino, N.; Elkins, C. A.; Chen, C. Y.; Patel, I.; Makusky, A. J.; Choy, J. S.; Court, D. L.; Markey, S. P.; Kowalak, J. A. A Proteomic and Transcriptomic Approach Reveals New Insight into Beta-Methylthiolation of *Escherichia Coli* Ribosomal Protein S12. *Mol. Cell. Proteomics* **2011**, *10* (3), M110 005199.
- (656) Koehnke, J.; Bent, A. F.; Zollman, D.; Smith, K.; Houssen, W. E.; Zhu, X.; Mann, G.; Lebl, T.; Scharff, R.; Shirran, S., et al. The Cyanobactin Heterocyclase Enzyme: A Processive Adenylase That Operates with a Defined Order of Reaction. *Angew. Chem. Int. Ed. Engl.* **2013**, *52* (52), 13991-6.
- (657) Waisvisz, J. M.; van der Hoeven, M. G.; van Peppen, J.; Zwennis, W. C. M. Bottromycin. I. A New Sulfur-Containing Antibiotic. *J. Am. Chem. Soc.* **1957**, *79* (16), 4520-4521.

- (658) Nakamura, S.; Umezawa, H. The Structure of Bottromycin A2, a New Component of Bottromycins. *Chem. Pharm. Bull.* **1966**, *14* (9), 981-6.
- (659) Nakamura, S.; Omura, S.; Nishimura, T.; Tanaka, N.; Umezawa, H. Derivatives of Bottromycin A2 and Their Biological Activity. *J. Antibiot.* **1967**, *20* (3), 162-6.
- (660) Miller, W. J.; Chalet, L.; Rasmussen, G.; Christensen, B.; Hannah, J.; Miller, A. K.; Wolf, F. J. Bottromycin. Separation of Biologically Active Compounds and Preparation and Testing of Amide Derivatives. *J. Med. Chem.* **1968**, *11* (4), 746-9.
- (661) Takahashi, Y.; Naganawa, H.; Takita, T.; Umezawa, H.; Nakamura, S. The Revised Structure of Bottromycin A2. *J. Antibiot.* **1976**, *29* (10), 1120-3.
- (662) Schipper, D. The Revised Structure of Bottromycin A2. *J. Antibiot.* **1983**, *36* (8), 1076-7.
- (663) Kaneda, M. Studies on Bottromycins. II. Structure Elucidation of Bottromycins B2 and C2. *J. Antibiot.* **2002**, *55* (10), 924-8.
- (664) Shimamura, H.; Gouda, H.; Nagai, K.; Hirose, T.; Ichioka, M.; Furuya, Y.; Kobayashi, Y.; Hirono, S.; Sunazuka, T.; Omura, S. Structure Determination and Total Synthesis of Bottromycin A2: A Potent Antibiotic against MRSA and VRE. *Angew. Chem. Int. Ed. Engl.* **2009**, *48* (5), 914-7.
- (665) Tanaka, N.; Nishimura, T.; Nakamura, S.; Umezawa, H. Activity of Bottromycin against Mycoplasma Gallisepticum. *J. Antibiot.* **1968**, *21* (1), 75-6.
- (666) Otaka, T.; Kaji, A. Mode of Action of Bottromycin A2. Release of Aminoacyl- or Peptidyl-tRNA from Ribosomes. *J Biol Chem* **1976**, *251* (8), 2299-306.
- (667) Kinoshita, T.; Tanaka, N. On the Site of Action of Bottromycin A2. *J. Antibiot.* **1970**, *23* (6), 311-2.
- (668) Otaka, T.; Kaji, A. Mode of Action of Bottromycin A2: Effect of Bottromycin A2 on Polysomes. *FEBS Lett.* **1983**, *153* (1), 53-9.
- (669) Otaka, T.; Kaji, A. Mode of Action of Bottromycin A2: Effect on Peptide Bond Formation. *FEBS letters* **1981**, *123* (2), 173-6.
- (670) Kobayashi, Y.; Ichioka, M.; Hirose, T.; Nagai, K.; Matsumoto, A.; Matsui, H.; Hanaki, H.; Masuma, R.; Takahashi, Y.; Omura, S., et al. Bottromycin Derivatives: Efficient Chemical Modifications of the Ester Moiety and Evaluation of Anti-MRSA and Anti-VRE Activities. *Bioorg. Med. Chem. Lett.* **2010**, *20* (20), 6116-20.
- (671) Zhang, H.; Zhou, W.; Zhuang, Y.; Liang, X.; Liu, T. Draft Genome Sequence of Streptomyces Bottropensis ATCC 25435, a Bottromycin-Producing Actinomycete. *Genome Announc.* **2013**, *1* (2), e0001913.

- (672) Hou, Y.; Tianero, M. D.; Kwan, J. C.; Wyche, T. P.; Michel, C. R.; Ellis, G. A.; Vazquez-Rivera, E.; Braun, D. R.; Rose, W. E.; Schmidt, E. W., et al. Structure and Biosynthesis of the Antibiotic Bottromycin D. *Org. Lett.* **2012**, *14* (19), 5050-3.
- (673) Crone, W. J. K.; Leeper, F. J.; Truman, A. W. Identification and Characterisation of the Gene Cluster for the Anti-MRSA Antibiotic Bottromycin: Expanding the Biosynthetic Diversity of Ribosomal Peptides. *Chem. Sci.* **2012**, *3* (12), 3516-3521.
- (674) Huo, L.; Rachid, S.; Stadler, M.; Wenzel, S. C.; Muller, R. Synthetic Biotechnology to Study and Engineer Ribosomal Bottromycin Biosynthesis. *Chem. Biol.* **2012**, *19* (10), 1278-87.
- (675) Hayakawa, Y.; Sasaki, K.; Adachi, H.; Furihata, K.; Nagai, K.; Shin-ya, K. Thioviridamide, a Novel Apoptosis Inducer in Transformed Cells from *Streptomyces Olivoviridis*. *J. Antibiot.* **2006**, *59* (1), 1-5.
- (676) Hayakawa, Y.; Sasaki, K.; Nagai, K.; Shin-ya, K.; Furihata, K. Structure of Thioviridamide, a Novel Apoptosis Inducer from *Streptomyces Olivoviridis*. *J. Antibiot.* **2006**, *59* (1), 6-10.
- (677) Sit, C. S.; Yoganathan, S.; Vederas, J. C. Biosynthesis of Aminovinyl-Cysteine-Containing Peptides and Its Application in the Production of Potential Drug Candidates. *Acc. Chem. Res.* **2011**, *44* (4), 261-268.
- (678) Spaink, H. P. Root Nodulation and Infection Factors Produced by Rhizobial Bacteria. *Annu. REv. Microbiol.* **2000**, *54* (1), 257-288.
- (679) Freiberg, C.; Fellay, R.; Bairoch, A.; Broughton, W. J.; Rosenthal, A.; Perret, X. Molecular Basis of Symbiosis between *Rhizobium* and Legumes. *Nature* **1997**, *387* (6631), 394-401.
- (680) Schwinghamer, E. A.; Belkengren, R. P. Inhibition of Rhizobia by a Strain of *Rhizobium Trifolii*: Some Properties of the Antibiotic and of the Strain. *Arch. Mikrobiol.* **1968**, *64* (2), 130-145.
- (681) Hodgson, A. L. M.; Roberts, W. P.; Waid, J. S. Regulated Modulation of *Trifolium Subterraneum* Inoculated with Bacteriocinproducing Strains of *Rhizobium Trifolii*. *Soil Biol. Biochem.* **1985**, *17* (4), 475-478.
- (682) Triplett, E. W.; Barta, T. M. Trifolitoxin Production and Nodulation Are Necessary for the Expression of Superior Nodulation Competitiveness by *Rhizobium Leguminosarum* Bv. *Trifolii* Strain T24 on Clover. *Plant Physiol.* **1987**, *85* (2), 335-42.
- (683) Triplett, E. W. Isolation of Genes Involved in Nodulation Competitiveness from *Rhizobium Leguminosarum* Bv. *Trifolii* T24. *Proc. Nat. Acad. Sci. U.S.A.* **1988**, *85* (11), 3810-4.
- (684) Triplett, E. W.; Vogelzang, R. D. A Rapid Bioassay for the Activity of the Antirhizobial Peptide, Trifolitoxin. *J. Microbiol. Methods* **1989**, *10* (3), 177-82.

- (685) Triplett, E. W. Construction of a Symbiotically Effective Strain of *Rhizobium Leguminosarum* Bv. *Trifolii* with Increased Nodulation Competitiveness. *Appl. Environ. Microbiol.* **1990**, *56* (1), 98-103.
- (686) Robleto, E. A.; Kmiecik, K.; Oplinger, E. S.; Nienhuis, J.; Triplett, E. W. Trifolitoxin Production Increases Nodulation Competitiveness of *Rhizobium Etli* Ce3 under Agricultural Conditions. *Appl. Environ. Microbiol.* **1998**, *64* (7), 2630-2633.
- (687) Triplett, E. W.; Schink, M. J.; Noeldner, K. L., Mapping and Subcloning of Genes Involved in Nodulation Competitiveness from *Rhizobium Leguminosarum* Bv. *Trifolii* T24. In *The Rhizosphere and Plant Growth: Papers Presented at a Symposium Held May 8–11, 1989, at the Beltsville Agricultural Research Center (Barc), Beltsville, Maryland*, Keister, D. L.; Cregan, P. B., Eds. Springer Netherlands: Dordrecht, 1991; pp 191-191.
- (688) Breil, B. T.; Ludden, P. W.; Triplett, E. W. DNA Sequence and Mutational Analysis of Genes Involved in the Production and Resistance of the Antibiotic Peptide Trifolitoxin. *J. Bacteriol.* **1993**, *175* (12), 3693-702.
- (689) Triplett, E. W.; Breil, B. T.; Splitter, G. A. Expression of Tfx and Sensitivity to the Rhizobial Peptide Antibiotic Trifolitoxin in a Taxonomically Distinct Group of α -Proteobacteria Including the Animal Pathogen *Brucella Abortus*. *Appl. Environ. Microbiol.* **1994**, *60* (11), 4163-6.
- (690) Scupham, A. J.; Dong, Y.; Triplett, E. W. Role of Tfxe, but Not Tfxg, in Trifolitoxin Resistance. *Appl. Environ. Microbiol.* **2002**, *68* (9), 4334-4340.
- (691) Tang, H.; Wang, E.; Sui, X.; Man, C.; Jia, R.; Lin, D.; Qu, Z.; Chen, W. The Novel Alkali Tolerance Function of Tfxg in *Sinorhizobium Meliloti*. *Res. Microbiol.* **2007**, *158* (6), 501-505.
- (692) Breil, B.; Borneman, J.; Triplett, E. W. A Newly Discovered Gene, TfuA, Involved in the Production of the Ribosomally Synthesized Peptide Antibiotic Trifolitoxin. *J. Bacteriol.* **1996**, *178* (14), 4150-6.
- (693) Triplett, E. W.; Herlache, T. C., Biological Control of Crown Gall Disease. Google Patents: 2006.
- (694) Scupham, A. J.; Triplett, E. W. Determination of the Amino Acid Residues Required for the Activity of the Anti-Rhizobial Peptide Antibiotic Trifolitoxin. *J. Appl. Microbiol.* **2006**, *100* (3), 500-507.
- (695) Tianero, M. D.; Pierce, E.; Raghuraman, S.; Sardar, D.; McIntosh, J. A.; Heemstra, J. R.; Schonrock, Z.; Covington, B. C.; Maschek, J. A.; Cox, J. E., et al. Metabolic Model for Diversity-Generating Biosynthesis. *Proc. Nat. Acad. Sci. U.S.A.* **2016**, *113* (7), 1772-1777.
- (696) Sardar, D.; Schmidt, E. W. Combinatorial Biosynthesis of Ripps: Docking with Marine Life. *Curr Opin Chem Biol* **2016**, *31*, 15-21.

- (697) Young, T. S.; Schultz, P. G. Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon. *J. Biol. Chem.* **2010**, *285* (15), 11039-11044.
- (698) Oueis, E.; Jaspars, M.; Westwood, N. J.; Naismith, J. H. Enzymatic Macrocyclization of 1,2,3-Triazole Peptide Mimetics. *Angew. Chem. Int. Ed.* **2016**, *55* (19), 5842-5845.
- (699) Doroghazi, J. R.; Albright, J. C.; Goering, A. W.; Ju, K. S.; Haines, R. R.; Tchalukov, K. A.; Labeda, D. P.; Kelleher, N. L.; Metcalf, W. W. A Roadmap for Natural Product Discovery Based on Large-Scale Genomics and Metabolomics. *Nat. Chem. Biol.* **2014**, *10* (11), 963-8.
- (700) Cimermancic, P.; Medema, M. H.; Claesen, J.; Kurita, K.; Wieland Brown, L. C.; Mavrommatis, K.; Pati, A.; Godfrey, P. A.; Koehrsen, M.; Clardy, J., et al. Insights into Secondary Metabolism from a Global Analysis of Prokaryotic Biosynthetic Gene Clusters. *Cell* **2014**, *158* (2), 412-21.

Chapter 2: A Prevalent Peptide-Binding Domain Guides Ribosomal Natural Product Biosynthesis²

2.1 Introduction

Natural products, in particular polyketides and non-ribosomal peptides, have provided a wealth of pharmaceutically important molecules¹ and chemical probes for biology.² More recently, ribosomally synthesized and posttranslationally modified peptides (RiPPs) have been recognized as another major sector of natural products³ with a similar capacity for probing biological systems⁴ and for serving as new drug scaffolds.⁵ RiPPs can be exceedingly complex in structure and are further categorized based on the set of modifying enzymes encoded by their biosynthetic gene clusters and the corresponding structural features present within the final products. Consequently, RiPP gene clusters are not identified or defined by a shared biosynthetic enzyme; rather, they are unified by a common biosynthetic logic of posttranslationally modifying an unrestrained, ribosomally synthesized peptide (Figure 2.1a).³ RiPP precursor peptides are usually bipartite, being composed of an *N*-terminal leader and *C*-terminal core regions. RiPP biosynthetic enzymes recognize conserved sequence motifs in the leader region and install residue-specific modifications to the core, yielding structurally complex compounds.^{3,6} Because of the bipartite nature of RiPP precursor peptides, the modifying enzymes can be specific for a particular substrate while also displaying a high tolerance towards unnatural core sequences. This fact, combined with their gene-encoded nature, makes RiPPs attractive targets for biosynthetic engineering.^{3, 7-9} Previous efforts relied on maintenance of native leader peptide sequences because omission of the conserved binding motifs significantly limits processing.^{3,10} Although some residues of the leader peptide that are recognized by the biosynthetic enzymes have been identified⁶, there are only a few RiPP classes where it is known which protein(s) physically associate with the leader and which residues of these proteins govern the interaction.

Our group and others have made progress towards the general mechanistic understanding of substrate binding and catalysis for the cyclodehydratase involved in thiazole/oxazole-modified microcin (TOMM) biosynthesis.¹¹ TOMM cyclodehydratases are found in >1,000 biosynthetic gene clusters,¹² and most are found within three RiPP classes: linear azole-containing peptides (LAPs), azole-containing cyanobactins and thiopeptides.^{3, 11} Occurring as fused or discrete polypeptides,¹³⁻¹⁵ the TOMM cyclodehydratase consists of an adenosine 5'-triphosphate (ATP)-utilizing YcaO domain (D protein)^{12, 16, 17} and a member of the E1 ubiquitin activating-like (E1-like) superfamily (C protein).¹⁸ In collaboration, the C/D protein(s) form azoline heterocycles from Cys, Ser, and Thr residues of the core peptide (Figure

² This chapter is reproduced with permission from 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*, 564-570. Copyright 2015 Nature Publishing Group. Experiments were designed by D.A.M. and B.J.B. and performed by B.J.B. K.L.D. generated FITC-HcaA-FP, and G.A.H. produced FITC-TbtA-LP. The manuscript was written and edited by D.A.M. and B.J.B.

2.1b).¹³⁻¹⁵ Often a third component (B protein) is present, which is a flavin mononucleotide (FMN)-dependent dehydrogenase that oxidizes select azolines to azoles.

Work on a TOMM cyclodehydratase from *Bacillus* sp. Al Hakam,¹⁹ which comprises discrete C and D proteins, has revealed that BalhD catalyzes azoline formation (Balh) has revealed that the D protein (BalhD) catalyzes azoline formation¹⁶ whereas the C protein (BalhC) mediates peptide substrate recognition.¹² Although the leader peptide-binding site remains unknown, the N-terminal winged helix-turn-helix (wHTH) domain has been implicated in this function on the basis of its homology to the 'peptide clamp' of MccB,²⁰ another E1-like RiPP-modifying enzyme involved in microcin C7 biosynthesis. Even though MccB and other related E1-like enzymes (such as PaaA and Rv3196) are not part of TOMM biosynthetic pathways, these E1-like enzymes possess N-terminal wHTH domains and produce ribosomal peptide-derived compounds,¹⁸ prompting the hypothesis that the wHTH domain guides the peptide to the active site. In support of this hypothesis, recently solved structures of LynD (PDB 4V1T, a fused cyclodehydratase involved in cyanobactin biosynthesis^{14, 21}) and NisB (PDB 4WD9, a lanthipeptide dehydratase involved in nisin biosynthesis²²) show that the leader peptides of their respective substrates are bound by structurally similar wHTH motifs. This suggests that additional, seemingly unrelated RiPP biosynthetic enzymes might employ the same strategy for substrate binding. However, the connection between these 'recognition' domains and the RiPP enzymes that directly engage the leader peptides remains unexplored.

Here, we bioinformatically identified a conserved domain linking the homologous wHTH motifs in NisB, LynD, and MccB. Through reevaluation of reported structures and experimental validation of additional RiPP biosynthetic enzymes, we demonstrated that this previously unrecognized conserved domain is present in the majority of prokaryotic RiPP classes. Given its broad distribution and identified role, we named this shared domain the RiPP precursor peptide recognition element (RRE). Importantly, these findings revealed previously unidentified similarities between disparate RiPP gene clusters and will guide future investigations of RiPP biosynthetic pathways.

2.2 Results

2.2.1 RREs are related to a PQQ biosynthetic protein

We began our investigation of RiPP precursor peptide recognition by searching for a link between the peptide-binding wHTH domains of NisB and LynD. Amino acid sequence homology was not detected using routine methods (*e.g.* BLAST²³), likely due to high divergence and lack of similarity to any annotated domains in the Conserved Domain Database (CDD).²⁴ However, HHpred,²⁵ a highly sensitive homology detection tool based on profile hidden Markov model (HMM) comparisons,²⁶ revealed that the wHTH domains of LynD (residues 2–81) and NisB (residues 153–223) were related to the protein PqqD

(probability > 90%). HHpred probabilities give the most relevant representation of significance with > 90% usually being considered a true positive.^{25, 27} PqqD itself is a small protein (~90–100 residues) involved in the biosynthesis of another RiPP, pyrroloquinoline quinone (PQQ).²⁸ PQQ is posttranslationally derived from Glu and Tyr residues of the PqqA precursor peptide. Although PQQ biosynthesis is not fully understood, PqqA maturation likely relies on PqqB (putative oxygenase), PqqC (oxygenase), PqqD (“peptide chaperone”), and PqqE (radical SAM).²⁸ PqqD is known to interact with PqqE, but until a recent report showing that PqqD binds to PqqA,²⁹ its function was unknown; however, this finding is consistent with its homology to the peptide-binding domains of NisB, LynD and MccB.

The structure of PqqD has been solved,³⁰ and we queried the Dali server³¹ to assess structural similarities between MccB, LynD, NisB and PqqD as an additional verification of homology. Searches with MccB, LynD and NisB did not return PqqD as a structural homolog. When we submitted only the PqqD-like region (i.e., the wHTH domain) of these proteins, however, LynD matched PqqD with a relatively weak similarity score (Z-score 5.3, with $Z < 2.0$ being viewed as spurious). The reverse search using bona fide, full-length PqqD as the query returned LynD and MccB with similar Z-scores (5.4 and 3.0, respectively). Visual comparison of the PqqD structure (PDB entry: 3G2B) with these proteins corroborates the HHpred/Dali alignment results (Figure 2.2). Thus, structural homology to PqqD links the structurally similar domains of NisB, LynD, and MccB.

2.2.2 RREs are present in > 50% of prokaryotic RiPP classes

Building on the finding that at least three enzymatically unique RiPP biosynthetic enzymes harbor PqqD-like domains with distant homology, we hypothesized that the same could be true for additional RiPP biosynthetic proteins. Indeed, PqqD-like domains have been previously identified in AlbA,³² which is involved in subtilosin A (sactipeptide) biosynthesis, and in a protein with an unknown function in lariatin (lasso peptide) biosynthesis;^{3, 33} however, the role of these domains have remained largely enigmatic. We speculated that PqqD-like domains function as an RRE. To determine the prevalence of such domains, we constructed a sequence similarity network of the ~4,000 members of the PqqD InterPro family (IPR008792) at an *e*-value threshold of 10⁻¹⁶ (Figure 2.3). Analysis of the network revealed proteins from sactipeptide and lasso peptide biosynthesis, but homologs of lanthipeptide dehydratases and TOMM cyclodehydratases were absent (i.e., they are not members of IPR008792). As an alternative approach, we queried the Conserved Domain Architecture Retrieval Tool (CDART)³⁴ for proteins containing an annotated PqqD-like domain and found over 40 unique architectures (Table 2.1). While some architectures were related to known RiPP biosynthetic proteins (such as AlbA, LynD and NisB), many did not appear to be RiPP related. Moreover, the small set of returned sequences that were RiPP related indicated that CDART³⁴ was not

sensitive enough to identify RRE-containing proteins, likely as a consequence of their general lack of annotation.

To overcome these shortcomings, we again returned to HHPred to search RiPP biosynthetic gene clusters for RREs (on the basis of homology to PqqD). A representative gene cluster for previously defined RiPP classes³ was chosen, and each protein was individually queried using HHPred (Table 2.2). Because eukaryotic RiPP biosynthesis is bioinformatically less tractable, the current analysis included only prokaryotes. Our results indicated that over half of all known prokaryotic RiPP classes harbor a PqqD-like domain, underscoring a broad distribution of RREs (Figure 2.4 and detailed results in Table 2.3).

As described above, the RRE for canonical discrete (e.g. BalhC-D) and fused (e.g. LynD) cyclodehydratases is the N-terminal wHTH domain of the C protein portion, which itself is not annotated in the CDD.²⁴ We recently identified a third cyclodehydratase architecture which is confined largely to the thiopeptides and heterocycloanthracins (Hca; a subclass of LAPs).^{35,36} These fused cyclodehydratases are N-terminally truncated by ~150 residues such that the RRE is ablated. Almost invariably, TOMM clusters that contain 'truncated' cyclodehydratases also encode a partner protein (annotated in the CDD as "ocin_ThiF_like"; hereafter referred to as the TOMM 'F protein') that contains the RRE (Figure 2.4). As predicted, we have confirmed that the responsibility for binding the leader peptide resides with the RRE-containing TOMM F protein (such as HcaF in the Hca cluster) rather than with the fused cyclodehydratase.³⁶ Consequently, we refer to this class of (truncated) enzymes as F-dependent cyclodehydratases.

Because of its shared homology to the E1 ubiquitin-activating superfamily, the TOMM C protein is related to the microcin C7 adenylase MccB. The RRE present within MccB has previously been defined as a 'peptide clamp' based on its interaction with the MccA peptide substrate.²⁰ The gene encoding MccA is one of the shortest known and produces a 7-residue leaderless peptide substrate.³ Perhaps this explains why the MccA binding mode is somewhat different from that displayed by NisB and LynD (Figure 2.2). Regardless, the MccB RRE interacts with the N-terminal residues of MccA and is essential for substrate binding. We predict that the same holds true for PaaA (pantocin biosynthesis³) owing to its similarity to MccB (microcin C7 biosynthesis; Figure 2.4).^{18,37}

In class I lanthipeptides, the LanB and LanC enzymes (NisB and NisC, respectively, for nisin biosynthesis) form the lanthionines. LanB enzymes are serine-threonine dehydratases composed of a glutamylation and elimination domain, which occur as separate ('split') polypeptides in thiopeptide gene clusters (e.g., TbtB and TbtC).²² Subsequent to LanB dehydrations, LanC cyclases catalyze the Michael-like addition between the dehydrated serine and threonine residues and particular cysteine residues. In class II lanthipeptides, dehydration and cyclization are performed by a bifunctional synthetase composed of an N-terminal dehydratase domain (with no homology to NisB) and a C-terminal LanC-like cyclase domain. Class III and IV lanthipeptide synthetases are trifunctional, with N-terminal lyase, central kinase and

variable C-terminal cyclase domains.³ HHpred was unable to identify RREs in class II–IV lanthipeptide synthetases; however, RREs appear to be ubiquitous in class I LanB dehydratases and in their 'split' counterparts in thiopeptide biosynthesis (Figures 2.2c and 2.4).

While the majority of RREs are not explicitly annotated in public databases, the CDD identifies “PqqD” in a subset of lasso peptide gene clusters.³³ Canonical lasso peptide clusters harbor a precursor peptide, a transglutaminase-like protease, an asparagine synthetase-like protein and an ABC transporter. In many recently discovered lasso peptide gene clusters, the protease appears to be encoded by two discrete, adjacent genes³⁸ (Figure 2.5). In these cases the C-terminal portions are identified by their homology to transglutaminases (e.g., LarD; annotated as “Trans_glutcore_3” in Figure 2.4) and the N-terminal domains are annotated PqqD (e.g., LarC). Consistent with the hypothesis that the larger lasso proteases are fusions of the PqqD-like and transglutaminase-like proteins, the N-terminus of the fused protease involved in the biosynthesis of the lasso peptide astexin-1 (AtxB)³ is similar to PqqD (HHpred probability >90%). Surprisingly, some lasso peptide proteases (such as McjB from microcin J25 biosynthesis) do not display detectable homology to PqqD (HHpred probability <20%); however, because AtxB and McjB fulfill the same role in their respective gene clusters, we predict that McjB also possesses an RRE, even though HHpred cannot identify it.

Another example where RREs are annotated as PqqD-like are the sactipeptide rSAM enzymes.³² Sactipeptide rSAMs (such as Alba from subtilosin biosynthesis) install C α -S thioether crosslinks and contain auxiliary [4Fe-4S] clusters within a C-terminal 'SPASM' domain (named for its presence in subtilosin, PQQ, anaerobic sulfatase and mycofactocin clusters).^{39, 40} To investigate whether RREs are prevalent among other rSAM- and SPASM-containing gene clusters with short peptide substrates, we scanned the mycofactocin gene cluster (mft) from *Thermomicrobium roseum* with HHpred. Notable homology to PqqD (probability >90%) was found in MftB, with a genetic arrangement similar to that of PqqD-E (Figure 2.4). We also detected RREs in two additional small mycofactocin-like⁴⁰ gene clusters: the SCIFF (six cysteines in forty-five residues) and KxxxW rSAM families.⁴¹ Consistent with the proposed role of the RRE in RiPP biosynthesis, anaerobic sulfatase-maturing enzymes (anSMEs) lack bioinformatically or structurally (PDB entry: 4K39)⁴² detectable RREs and are not associated with RiPP biosynthesis. These assignments are in agreement with a recent independent report that identified PqqD-like domains in rSAMs.²⁹ Additional examples that further suggest RREs co-occurrence with RiPP rSAM enzymes include two enzymes involved in proteusin biosynthesis. RREs were located in the C-termini of the rSAM epimerase PoyD and the B₁₂-dependent methyltransferase PoyB (Figure 2.4).³ PoyD has been characterized, and its function is known to be dependent on the presence of a leader peptide,⁴³ as would be expected for an RRE-bearing protein. Although the PoyB B₁₂-dependent methyltransferase has not been functionally dissected, we hypothesize that it will similarly be a leader peptide-dependent enzyme.

Lastly, we identified RREs in proteins involved in bottromycin³ and trifolitoxin biosynthesis.^{44, 45} These two gene clusters are characterized by the presence of a discrete TOMM D protein,¹² which apparently acts in the absence of a TOMM C protein. Intriguingly, these 'stand-alone' D proteins do not bear discernable RREs; however, HHpred found RREs (probability >90%) in the N-termini of two trifolitoxin oxidoreductases (TfxB and TfxC) and the C-termini of three bottromycin rSAMs (BmbB, BmbF and BmbJ), which are architecturally similar to the proteusin C-terminal methyltransferases (Figure 2.4). Unusually, the bottromycin precursor peptide harbors a 'follower' peptide rather than a leader peptide.³ Despite this unconventional arrangement, we predict that the follower peptide of bottromycin is engaged by the rSAM RRE.

2.2.3 Experimental validation of bioinformatically predicted RREs

Our findings have revealed a broad distribution of bioinformatically identified RREs among various RiPP classes and enzymes, with strong support for the RRE's role gleaned from the substrate-bound structures of the LynD cyclodehydratase (4V1T), NisB dehydratase,²² and MccB adenylase.²⁰ To provide further evidence that the predicted RREs can recruit precursor peptides during RiPP biosynthesis, we corroborated the functions of additional RREs through site-directed mutagenesis and fluorescence polarization (FP) binding assays. We surmised that specific residues within the RRE would engage the peptide and thus used site-directed mutagenesis to explore the nature of the interaction in five proteins that were derived from three additional RiPP classes (LAPs, thiopeptides and lasso peptides). On the basis of the known structures, sequence alignments and homology models, we targeted residues in the β 1– β 3 and α 1– α 3 regions for mutagenesis (Figure 2.2 and Figure 2.6). The purity and integrity of the mutant proteins were visually evaluated by Coomassie-stained SDS-PAGE gels (Figure 2.7).

2.2.4 LAP discrete cyclodehydratase

With some exceptions, the discrete cyclodehydratases (separate C and D proteins) tend to be found within LAP gene clusters.¹³ Having used an FP assay to previously narrow down that the leader peptide binds to the C protein,¹² we introduced alanine substitutions into the BalhC RRE and similarly compared binding constants of the mutants using a fluorescently labeled BalhA1 leader peptide (FITC-BalhA1-LP). The resulting data showed that most alanine substitutions within the RRE substantially (>10-fold) reduced the interaction between BalhA1 and BalhC (Table 2.4 and Figure 2.6). Three BalhC mutations (K20A, D23A and E66A) had a moderate effect, while two (D32A and H38A) had a minor effect. For comparison, residues outside of the RRE were also substituted. These had essentially no effect on leader peptide binding except for those mutations that appeared to weaken the structural stability of BalhC by rendering it more susceptible to proteolytic degradation during expression (Table 2.5 and Figure 2.7).

Because an extended region of the BalhC RRE was sensitive to substitution, we examined a second LAP cyclodehydratase, this one from *Corynebacterium urealyticum* DSM 7109 (Cur) and a member of the plantazolicin subclass (Figure 2.8).⁴⁶ When CurC was tested for binding to the FITC-labeled CurA leader (FITC-CurA-LP), a dissociation constant (K_d) of $7 \pm 1 \mu\text{M}$ was obtained. CurD did not have an appreciable binding response, nor did it enhance CurA binding by CurC (Figure 2.9). This result was consistent with previous studies of BalhC¹² and C proteins involved in LAP cytolysin biosynthesis.¹⁰ To assess the importance of the putative RRE in CurA binding, various residues were substituted with alanine. The binding data indicated that the CurC RRE was not as sensitive as BalhC to mutation as only CurC^{Y31A} in $\beta 3$ and CurC^{R68A} and CurC^{I75A} in $\alpha 3$ resulted in substantial (≥ 10 fold) reduction in affinity (Table 2.4; Figure 2.6). The dissociation constants for peptide binding for the remaining substitutions in $\beta 1$, $\beta 2$ and $\alpha 1$ were within two-fold of that for CurC^{WT}. Collectively, these data indicated that the RREs of discrete cyclodehydratases, especially the residues lining $\beta 3$ and $\alpha 3$, are essential for peptide binding and suggested a binding mode similar to that of LynD and NisB.

2.2.5 F protein-dependent cyclodehydratase

Recent work from our group³⁶ has revealed the dependence of truncated cyclodehydratases found in thiopeptide³ and Hca gene clusters³⁵ are functionally dependent on a newly recognized biosynthetic protein (TOMM F protein). For these 'F-dependent' cyclodehydratases, it has been demonstrated that the F protein (HcaF) binds the precursor peptide (HcaA), but the involvement of the RRE remained undetermined. Therefore, we probed the HcaF protein from *Bacillus* sp. Al Hakam by replacing numerous residues of the RRE with alanine. The results indicated that residues lining the groove between $\beta 3$ and $\alpha 3$ (Phe35, Arg73 and Ile80) are important for binding (Table 2.6 and Figure 2.6). Mutation of nearby residues (HcaF^{D38A}, HcaF^{N72A} and HcaF^{E79A}) had minimal effects on FITC-labeled HcaA leader peptide (FITC-HcaA-LP) affinity, probably because they are oriented away from the likely binding groove owing to their relative positioning. Asp38 is on the opposite side of $\beta 3$ compared to Phe35 while Asn72 and Glu79 are orthogonally oriented relative to Arg73 and Ile80, respectively. Given that HcaF^{D21A} (in $\beta 1$) also had nearly WT-like binding, HcaF most probably engages HcaA in a manner analogous to peptide binding by LynD and NisB.

Because HcaF is a LAP-related protein, we chose to examine a non-LAP F-dependent cyclodehydratase. *Thermobispora bispora* (Tbt) is bioinformatically predicted to produce a GE2270A-like (thiopeptide) compound and, like Hca, harbors an F-dependent cyclodehydratase (TbtG).⁴⁷ Initial experiments with TbtF (F protein) demonstrated that it could bind the FITC-labeled TbtA leader peptide (FITC-TbtA-LP), whereas TbtG had no discernable interaction with the peptide (Figure 2.10). Mutation of the TbtF RRE gave results similar to those for HcaF, with residues in $\beta 3$ and the N-terminal portion of $\alpha 3$

being important for peptide binding (Val30, Phe68 and Arg71). TbtF^{D19A} and TbtF^{L24A} in β 2 had no differences in peptide binding compared to TbtF^{WT}, suggesting that the peptide binds between β 3 and α 3. We generated a homology model for TbtF based on the structure of LynD. From this structure, we did not expect the side chains of Gln74 and Arg79 in TbtF to make significant contact with the peptide and, consistent with this model, substitution of these residues had no measurable effect on the affinity toward FITC-TbtA-LP (Table 2.6 and Figure 2.6).

2.2.6 Lasso peptide split protease

With the role of the RRE established for TOMM cyclodehydratases, we turned next to the lasso peptides to test a RiPP class where little is known about precursor peptide recognition. Recently our group reported on an unusual lasso peptide, streptomonicin (STM), from *Streptomonospora alba*.⁴⁸ Genome sequencing revealed that the STM gene cluster contains a 'split' protease (StmE and StmB, Figure 2.5). Initial binding studies using size-exclusion chromatography indicated that the RRE-containing StmE bound the StmA precursor peptide, whereas the separate protease domain (StmB) had no interaction with either StmA or StmE under the tested conditions and appeared substantially degraded after expression (Figures 2.7 and 2.11). The interaction between StmA and StmE was further corroborated by an FP binding experiment with FITC-labeled StmA leader peptide (FITC-StmA-LP, Figure 2.12). Encouraged by the correct functional prediction for StmE, we prepared eight alanine-substituted proteins and evaluated their binding to FITC-StmA-LP. In accord with earlier examples, substitution of residues along the side of α 3 nearest to β 3 (D69A and L73A) resulted in dramatically (>500-fold) reduced peptide binding while alanine replacements of the adjacent Asp68 and Asp75, with their relative positions on the opposite side of α 3, had no effect (Table 2.7). Unlike in previous examples, β 1 (StmC^{V12A}) and β 2 (StmC^{N22A}) mutants displayed reduced binding, whereas StmE^{Y28A} and StmE^{Q76A}, in which the mutations reside near the groove between β 3 and α 3, displayed nearly WT affinity. While these homology model-based mutant protein-binding results imply that lasso peptide proteases employ a slightly different subset of RRE residues to bind the leader peptide, the fact remains that irrespective of RiPP class or the RRE-containing protein's function, the RRE appears to govern substrate recognition.

2.3 Discussion

Here we implicate a conserved PqqD-like domain in precursor peptide recognition in over half of all currently known RiPP classes (reflecting multiple thousands of gene clusters; Tables 2.2 and 2.3). Highly sensitive bioinformatics, known crystal structures (Figure 2.2) and FP binding assays provide substantial experimental support for the role of the RRE in leader peptide binding. Further, albeit indirect, evidence for the importance of RREs derive from two additional observations: (i) RREs are routinely identified in

RiPP biosynthetic proteins but are lacking in homologs not involved in processing RiPPs (e.g., RiPP rSAMs and E1-like enzymes; see above) and (ii) RiPP enzymes which act subsequent to leader peptide cleavage do not harbor RREs (e.g., methyltransferases involved in plantazolicin⁴⁹ and linaridin³ biosynthesis). Based on these findings, we predict that PqqD-like domains associated with peptide-modifying enzymes will function as RREs whenever present (as has been recently reported for PqqD).²⁹

In addition to demonstrating the role of RREs, our work also provided insight into how RREs bind their respective peptides. Binding experiments with BalhC (LAP) indicated that nearly the entire RRE was affected by mutation. Accordingly, the precise orientation of BalhA1 could not be inferred; however, CurC (52% similar), which displays considerably improved solution-phase behavior, had reduced peptide-binding affinity only upon substitution of residues in $\beta 3$ and $\alpha 3$. Given that $\beta 3$ and $\alpha 3$ were the primary affected regions in the TOMM F proteins (HcaF and TbtF) and that LynD interacted with its peptide substrate using the same strategy (Figure 2.6), it appears that TOMM cyclodehydratases bind their substrates in a similar orientation irrespective of domain organization. However, there were other RREs that interacted with peptide substrates in alternate orientations (e.g., MccB). Perhaps this should not be surprising as RREs are as divergent as the proteins and pathways they are found within. Other sequence-divergent, but structurally similar, peptide-binding domains also have varied binding modes.⁵⁰ Logically RREs must bind their substrates in a manner conducive to post-translational modification; therefore, proper presentation of the peptide to the modifying enzymes is essential. Perhaps this explains why residues in $\beta 1$ and $\beta 2$ of the StmE RRE are important for binding (owing to the mechanics necessary for forming the lasso topology). It is important to note, however, that bioinformatics and binding assays alone do not reveal atomic-level details of the substrate-binding interactions.

The real power of our findings resides in the ability to accurately predict which protein(s), even those of unknown function and in pathways that require numerous enzymes, will recruit the precursor peptide to the modifying enzymes during RiPP biosynthesis. As a case in point, the identification of an RRE within a small protein (StmE) implicated in lasso peptide biosynthesis allowed us to assign a function where none had previously been proposed.^{3, 38} Similarly, the biosynthetic logic and inclusion of an “ocin_ThiF_like” (TOMM F) protein in certain TOMM biosynthetic gene clusters could be understood through identification of the RRE-containing protein. However, such insights are limited to systems where RREs are detectable and, accordingly, leader recognition remains enigmatic in many RiPP classes (Table 2.2). It is possible that additional RREs will be found as homology detection algorithms become more sensitive and more sequences become available for building HMMs. Another possibility is that RiPP pathways lacking identifiable RREs have evolved other peptide recognition modules which are unrelated to the RRE in both sequence and structure. Nonetheless, RREs appear to be the most prevalent solution to

precursor peptide recognition among prokaryotic RiPPs, and they functionally link a diverse array of biosynthetic platforms and enzymes (Figure 2.4).

These findings suggest several future applications, including the engineering of new precursor peptide specificities via RRE swapping and the use of RRE as a marker to discover new RiPP classes. Indeed, it is difficult to predict the peptide substrates and modifying enzymes of a novel RiPP class *a priori*,³ which is why most first-in-class RiPPs have been found through phenotypic screens. However, if an RRE is bioinformatically identified in a novel genomic context, the flanking regions could be scanned for potential biosynthetic proteins and precursor peptides. As a starting point, we noted that many PqqD homologs we found belong to uncharacterized gene clusters and appear ripe for the above approach (Figure 2.3 and Table 2.1).

Overall, this work advances our understanding of RiPP biosynthesis and should serve to guide future studies. More work will be required to unequivocally establish the function of RREs in additional RiPP biosynthetic pathways but, importantly, our study lays a foundation indicating which proteins should be prioritized.

2.4 Methods

2.4.1 General methods

All materials were purchased from Fisher Scientific or Sigma-Aldrich unless indicated otherwise. DNA sequencing was performed by the Roy J. Carver Biotechnology Center (University of Illinois at Urbana-Champaign) or ACGT Inc. DNA oligonucleotides were purchased from Integrated DNA Technologies (IDT). All fluorescently-labeled leader peptides were purchased from GenScript (>90% purity) with an *N*-terminal FITC-Ahx (fluorescein isothiocyanate with an amino hexyl linker) conjugated to a single glycine spacer before the leader peptide, except for the FITC-labeled leader peptides for HcaA and TbtA, which were prepared as previously described³⁶. Full sequences are given in Table 2.8.

2.4.2 Cloning

The genes encoding CurC-D and StmA-B-E were amplified from the genomic DNA of their respective host organisms. PCR was performed with Platinum Taq DNA Polymerase High Fidelity (Invitrogen) or Pfu DNA polymerase using the primers listed in Table 2.9. BamHI-HF, NotI-HF, Sall, or HindIII (New England Biolabs, Inc., NEB) restriction sites were designed into these primers to avoid internal restriction enzyme cut sites. The amplified genes were PCR purified using DNA-binding spin columns (Qiagen, following manufacture's instruction) and digested with appropriate restriction enzymes. After an additional PCR purification step, the digested inserts were ligated with T4 DNA ligase (NEB) into similarly digested and purified pET28 vector in frame with an *N*-terminal, maltose-binding protein (MBP) tag. Codon optimized

TbtF and TbtG were synthesized by Genscript with 5' BamHI and 3' XhoI cut sites, and after digestion with BamHI and XhoI, the excised genes were purified from a 1% agarose gel using a gel extraction kit (Qiagen). His6-tagged HcaF was previously generated³⁶.

2.4.3 Site-directed mutagenesis

Site-directed mutagenesis was carried out using the Quikchange (Agilent) method or a modified method where primers were designed to overlap ~11 bp upstream and ~27 bp downstream of the targeted codon for mutation (to minimize primer-primer annealing). All mutagenesis primers are given in Table 2.9.

2.4.4 Protein expression and purification

All proteins were expressed and purified as tobacco etch virus (TEV) protease-cleavable fusions with MBP as previously described¹⁶ except for HcaF and its mutants which were purified by a N-terminal His6 tag³⁶. The purity of all expressed proteins was examined by SDS-PAGE (Figure 2.7)

2.4.5 Fluorescence polarization (FP) binding assay

The interaction between fluorescently labeled leader peptides and various RiPP biosynthetic proteins was quantified using an FP assay. To maximize the polarization signal, all proteins remained MBP tagged. In general, protein was serially diluted into binding buffer (50 mM HEPES, pH 7.5, 300 mM NaCl, 2.5% (vol/vol) glycerol, 0.5 mM TCEP) and mixed with 25 nM of the appropriate fluorescently labeled leader peptide. Binding assays were carried out in nonbinding-surface, 384-black-well polystyrene microplates (Corning) and measured using a FilterMax F5 multimode microplate reader (Molecular Devices) with $\lambda_{\text{ex}} = 485$ nm and $\lambda_{\text{em}} = 538$ nm. Prior to measurement, the dilutions were equilibrated with shaking for 15 to 30 min at 23 °C. Dissociation constant (K_d) values were calculated from the 50% saturation point using a dose-response curve fit in Origin Pro 9.1 (OriginLab) with three independent titrations. Background fluorescence from the proteins alone was subtracted from the fluorescence polarization signal obtained with the fluorophore.

2.4.6 Sequence alignments

Initial sequence alignments of proteins within the same RiPP class were made using ClustalW2 using default parameters⁵¹. Due to divergence between different RiPP classes, homology model sequence alignments were generated using HHpred as described below. (<http://toolkit.tuebingen.mpg.de/hhpred>)²⁵.

26

2.4.7 Detection of protein homologs

BLAST²³ was used to identify closely related homologs of proteins. To identify more distantly related, yet homologous, regions of RiPP biosynthetic proteins, we employed HHpred²⁵, a sensitive homology detection tool based on profile hidden Markov model (HMM) comparisons²⁶. HHpred queries were performed using the pdb70_30Apr15 database. The multiple sequence alignment (MSA) for the input sequence was generated with HHblits (iterative HMM-HMM comparison, maximum of 3 iterations), and secondary structure was scored with local alignment mode. All other options were left on their default settings except the “max number of hits in hit list” was increased to 500 so all hits above the 20% probability threshold would be returned. The hit list was then searched for PqqD or matches to the *N*-terminus of MccB, LynD, or NisB; however, we discovered that PqqD was a superior marker of RREs as the others frequently had lower probability scores if they appeared at all in the results. Only hits $\geq 90\%$ probability were considered true hits. Because the MSA is dependent on the input sequence and can affect the probability score, all discovered PqqD-like domains were verified by re-submitting smaller sections of the input sequence, centered on the identified PqqD-like region. In practice, this was only necessary for the RiPP methyltransferases because the HHblits MSA generation tended to include non-RiPP rSAM enzymes (due to the similarity of their iron-sulfur binding domains), diluting the PqqD signature in the HMM. A representative gene cluster for each RiPP class (where the DNA sequence was available) was chosen from a comprehensive review of known RiPPs³, and the biosynthetic proteins were individually assessed for homology with PqqD (Table 2.2).

2.4.8 Homology model generation

Models were created using the HHpred interface, using LynD (PDB entry: 4V1T) as the manually selected template. In this method, the HHpred based alignment is automatically used as input for MODELLER^{52, 53} to generate a 3D model.

2.4.9 Sequence similarity network

A sequence similarity network of InterPro family IPR008792 was generated with the Enzyme Function Initiative Enzyme (EFI) Similarity Tool (<http://www.enzymefunction.org/>)⁵⁴. The resulting network was visualized using the organic layout of Cytoscape⁵⁵, with an e-value threshold of 10^{-16} , and each node represents sequences with 80% or greater sequence identity. The function of different clusters with at least 3 nodes was investigated using the EFI Genome Neighborhood Tool⁵⁶. A neighborhood size of 7 was chosen with 20% co-occurrence limit. Most nodes appeared to be part of uncharacterized gene clusters, but some lasso peptide gene clusters were identified by co-occurrence with “Transglut_core3” and “Asn_synthase”.

2.4.10 Size-exclusion chromatography

Size-exclusion chromatography (SEC) was performed on a Flexar HPLC (Perkin Elmer) with analytical Yarra SEC-3000 (300 x 4.6 mm, Phenomenex). The column was pre-equilibrated with 5 column volumes (CVs) of binding buffer (50 mM HEPES, pH 7.5, 300 mM NaCl, 2.5% (v/v) glycerol) at 4 °C. Protein samples (50 μM) were made up on ice in the same buffer, injected (5 μL), and monitored at 280 nm. Traces were exported to Microsoft Excel for analysis and normalized to the highest absorbance value. Oligomeric state and apparent mass were determined based on a standard curve generated from the elution times of molecular weight standards 12-200 kDa (Sigma).

2.5 Acknowledgements

I am grateful to A. Maniak, and R. Downen for cloning several proteins and to C. Cox for bioinformatics consultation. I also thank C. Deane for critical review of this manuscript. This work was supported by the US National Institutes of Health (NIH) (1R01 GM097142 to D.A.M. and 2T32 GM070421 to B.J.B. and K.L.D) and the UIUC Department of Chemistry (Robert C. and Carolyn J. Springborn Endowment to B.J.B. and Harold R. Snyder Fellowship to K.L.D. and G.A.H.). B.J.B was also funded by a National Science Foundation Graduate Research Fellowship (DGE-1144245).

2.6 Figures

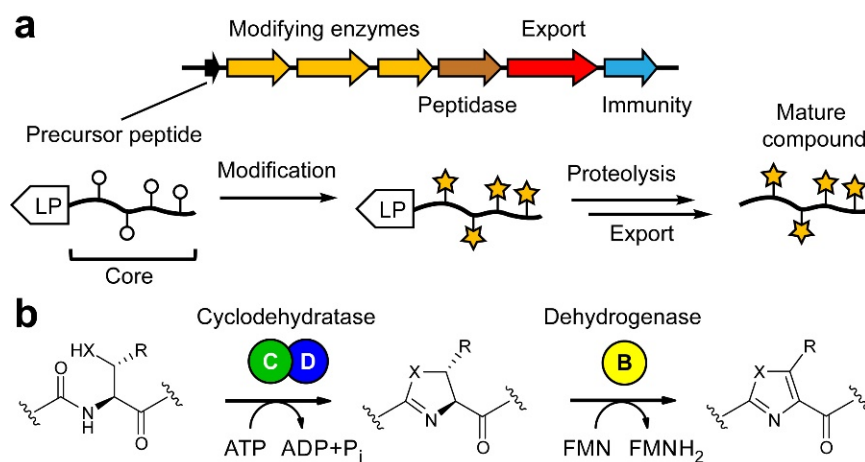


Figure 2.1 Overview of RiPP biosynthesis and the TOMM subclass. A generic RiPP biosynthetic gene cluster is displayed. The precursor peptide is composed of a leader peptide (LP) and core. While the LP contains binding motifs for the modifying enzymes, the core contains residues that undergo enzymatic processing to diverse functional groups. After removal of the LP and any additional tailoring processes, the mature RiPP is exported. (b) One RiPP biosynthetic class, the TOMMs, installs azoline-azole heterocycles. TOMMs arise from the action of an ATP-dependent cyclodehydratase (C and D proteins) and a flavin mononucleotide (FMN)-dependent dehydrogenase (B protein). X = S or O; R = H or CH₃.

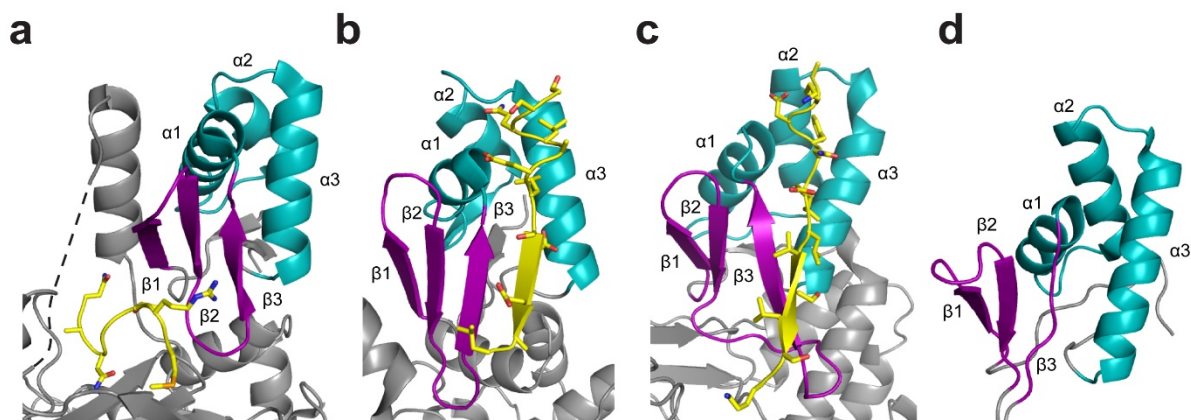


Figure 2.2 Structural comparison of four RiPP-modifying enzymes. (a–d) Shown are structurally equivalent sections of RiPP biosynthetic proteins involved in the biosynthesis of the Trojan horse antibiotic microcin C7 (MccB, an adenyating enzyme; PDB 3H9J; a), antitumor cyanobactins (LynD, a cyclodehydratase; PDB 4V1T; b), the lantibiotic nisin (NisB, a dehydratase; PDB 4WD9; c) and the bacterial dehydrogenase cofactor PQQ (PqqD, an rSAM-associated protein, 3G2B; d). The purple β -sheets and cyan α -helices constitute a conserved RRE. In a–c, the precursor peptide is shown in yellow stick formats. Dashed line indicates missing electron density.

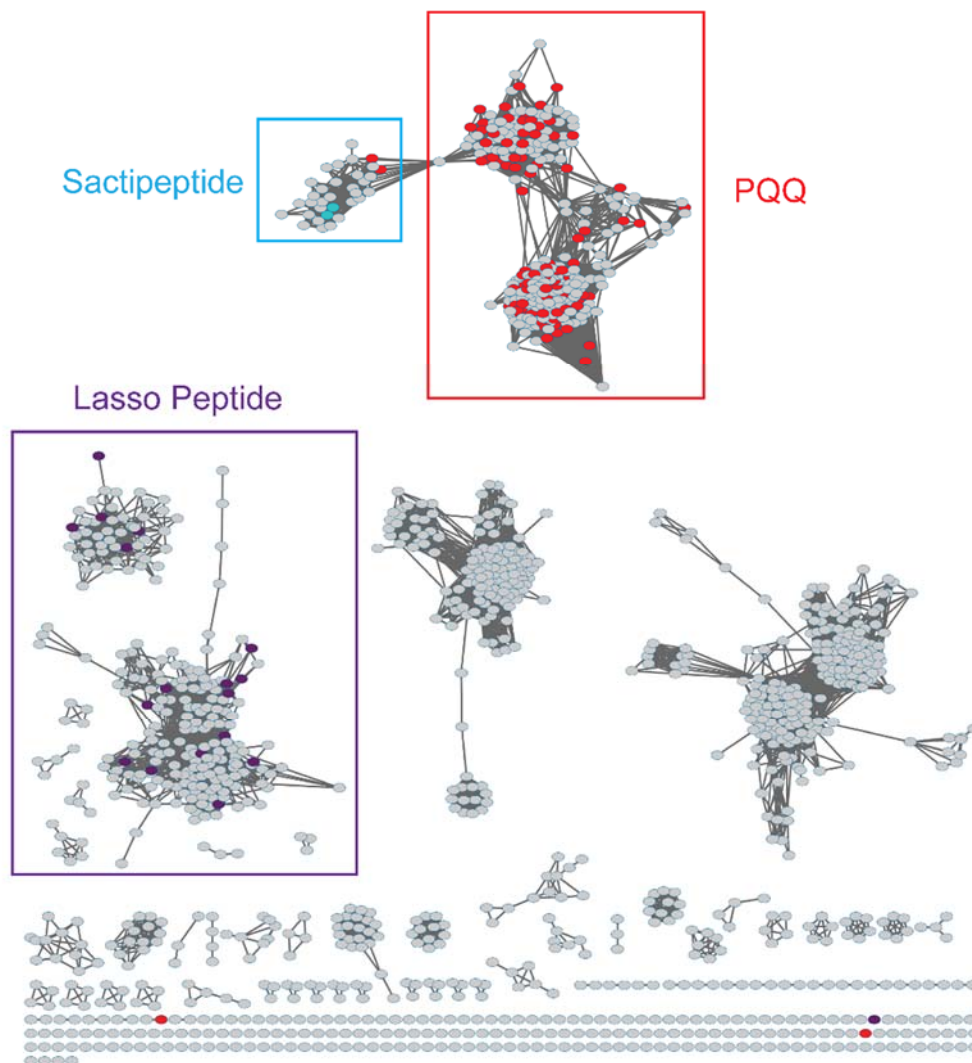


Figure 2.3 PqqD sequence similarity network. A sequence similarity network⁵⁴ for the PqqD InterPro family (IPR008792) is shown. Each circle (node) represents a single protein sequence or group sequences with $\geq 80\%$ identity while the lines (edges) denote homology at or below an E-value threshold of 10^{-16} . Proteins previously annotated as sactipeptide⁵⁷, lasso peptide^{38, 58}, and PQQ⁵⁹ biosynthetic clusters were colored in cyan, purple, and red, respectively. Groups of nodes form clusters which indicate probable isofunctionality, as indicated by the co-clustering of colored nodes. The two PQQ associated proteins (Uniprot ID: A0A017T8W2_9DELTA and S4YFB4_SORCE) in the sactipeptide cluster are PqqD-PqqE fusions, and their co-clustering with the sactipeptide group is a result of increased similarity score due to their rSAM domains (direct BLAST comparison of their PqqD-like domains with their sactipeptide neighbors do not meet the 10^{-16} threshold). The functions of other clusters were investigated by examining the local genome neighborhoods (± 7 genes)⁵⁶, but most appeared to be part of conserved, uncharacterized biosynthetic gene clusters. Regardless, a few additional lasso peptide clusters were identified by the presence of adjacent transglutaminase and asparagine synthetase genes.

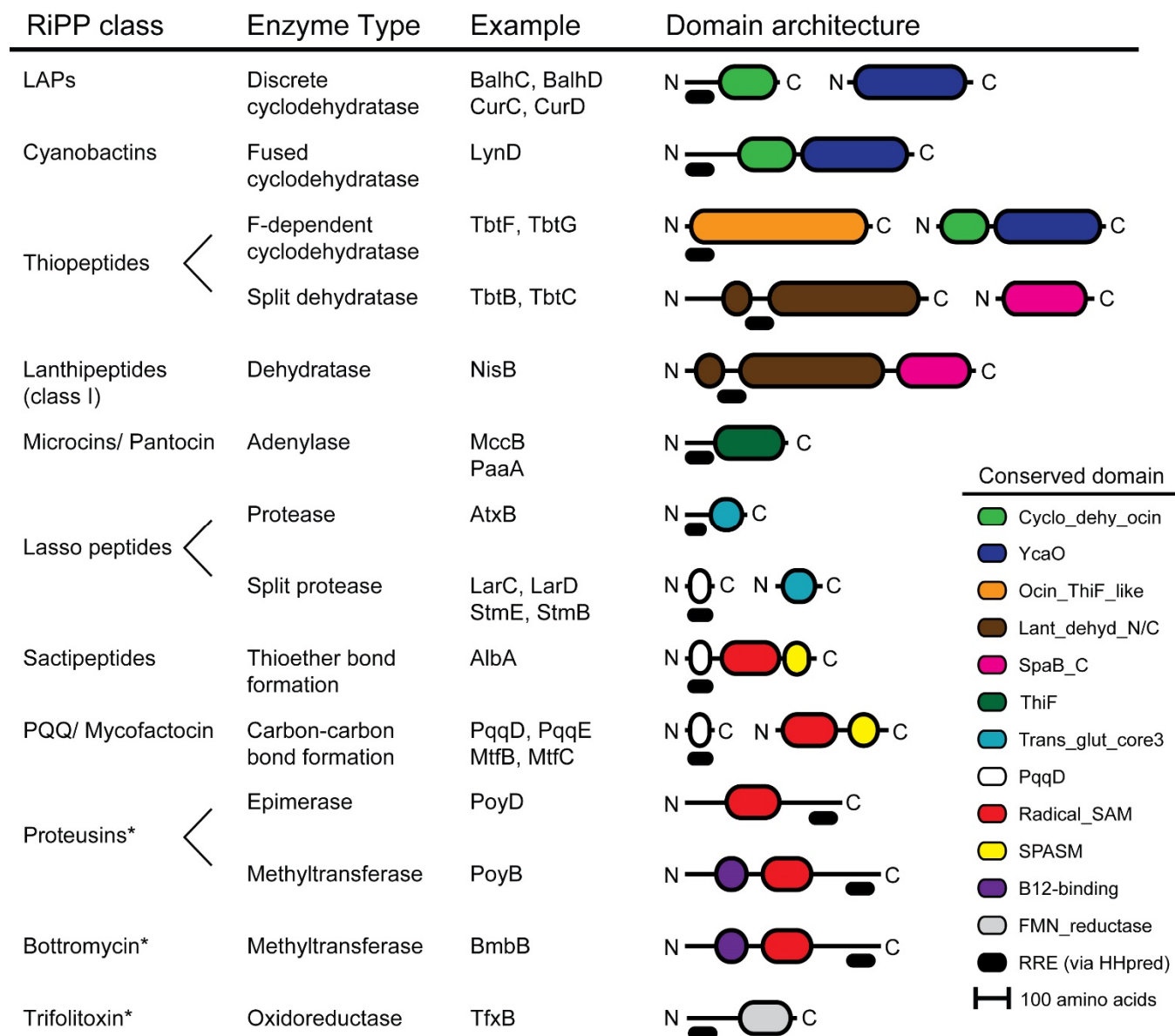


Figure 2.4 RREs are present in diverse RiPP biosynthetic proteins. RREs were found in a myriad of RiPP biosynthetic proteins using HHpred to search for PqqD homology (thick black bars). Solid lines represent protein sequences with N/C-termini labeled as “N” and “C.” Colored sections indicate the annotations for the conserved domains identified by the Conserved Domain Database²⁴. Asterisks denote RRE assignments based solely on HHpred findings. Abbreviations: LAP, linear azole-containing peptide; PQQ, pyrroloquinoline quinone; rSAM, radical SAM. More details can be found in Table 2.3.

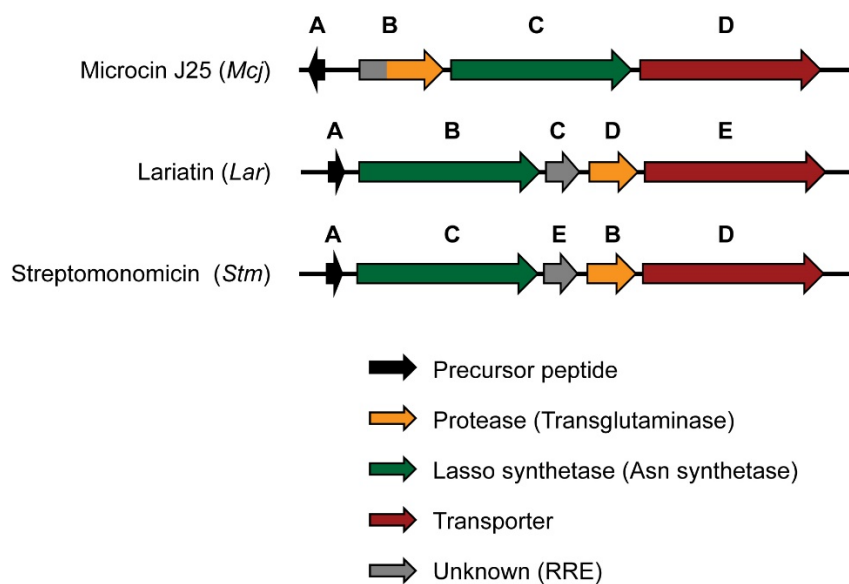


Figure 2.5 Genetic arrangement and comparison of lasso peptide biosynthetic enzymes. The microcin J25 gene cluster represents the canonical lasso peptide gene cluster and consists of four genes: a precursor peptide (A), protease (B), Asn synthetase homolog (C), and a transporter (D). A growing number of biosynthetic clusters contain five essential genes^{38, 60}, apparently due to the splitting of the protease (B protein). Note that the lariatins gene cluster does not follow the more recently established nomenclature rules³.

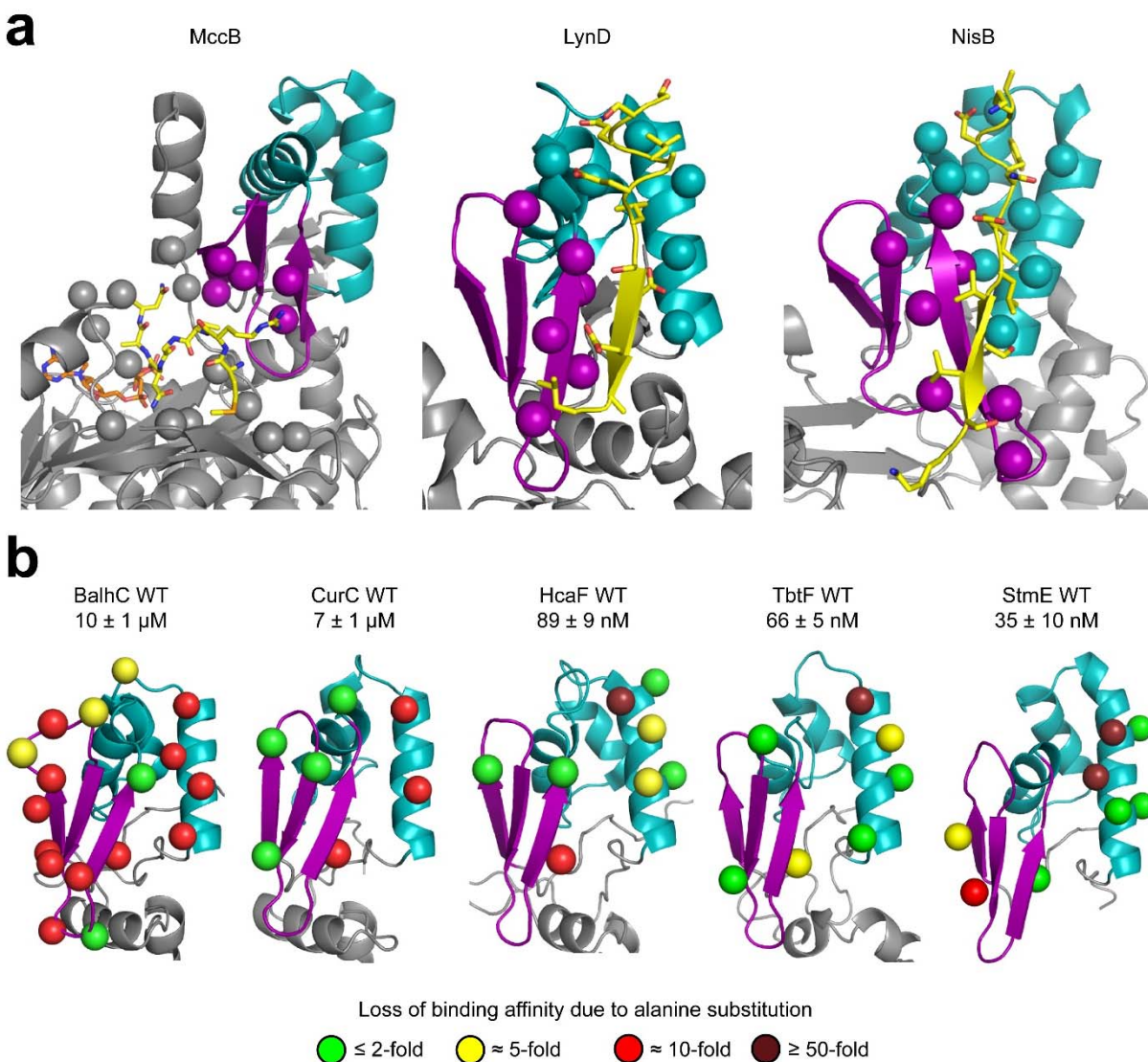
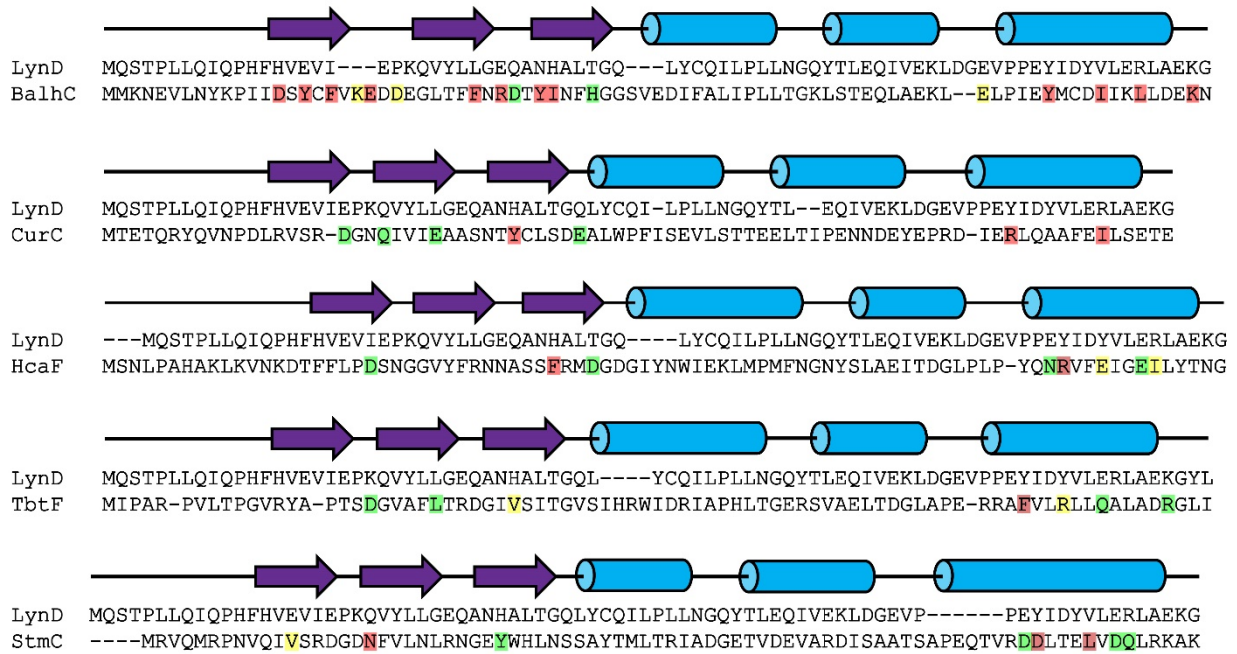


Figure 2.6 Homology models of RiPP biosynthetic protein RREs. To probe the role of the RRE in LAPs, thiopeptides, and lasso peptides, homology models of the RRE from BalhC, CurC, HcaF, TbtF and StmE were built with HHpred²⁵ using LynD (PDB entry 4V1T) as the template. Residues across various locations of the RRE were substituted with alanine, and the effect on leader peptide binding was measured. **(a)** The structures of MccB, LynD, and NisB are shown with spheres for the beta carbon of side chains within 4 Å of the peptide. The non-hydrolyzable ATP analogue (α,β -methyleneadenosine 5'-triphosphate, orange sticks) indicates the active site of MccB. Although relatively few residues near MccA are from the RRE, ablation of the RRE eliminates enzyme activity, which can be partially restored by addition of the RRE *in trans*²⁰. For clarity, 1–2 residues from outside of the RRE are not shown for NisB and LynD since they block the view many other important residues. The coloring of the spheres only indicates the secondary structure to which the residue belongs. **(b)** Homology models are shown with spheres representing the location and magnitude of the effect of alanine substitutions (see Tables 2.4, 2.6, and 2.7 for precise values). **(c)** The underlying sequence alignments on which the models were created are displayed and similarly colored. Figure continued on next page.

C**Figure 2.6 (Continued)**

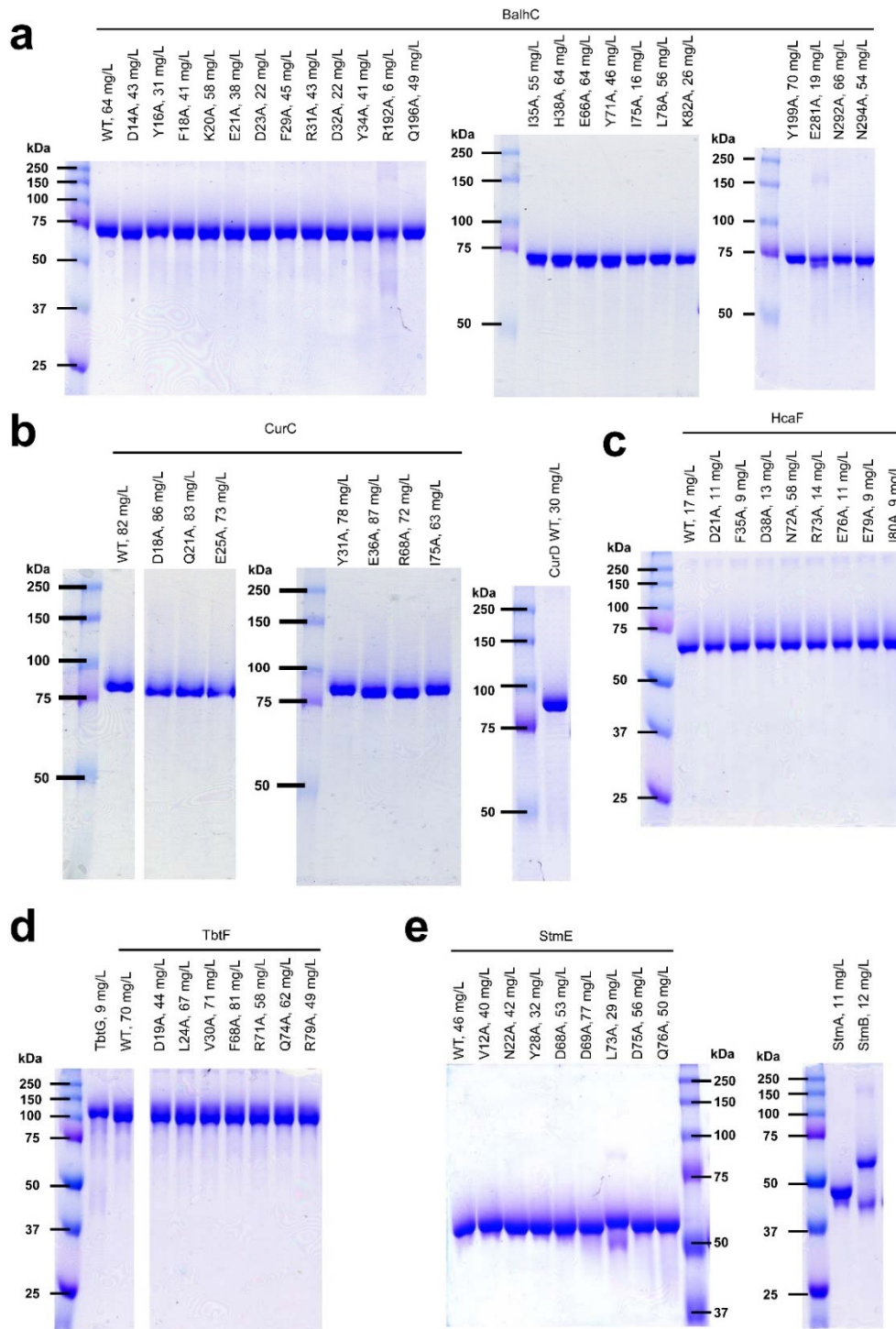


Figure 2.7 SDS-PAGE analysis of purified proteins. Heterologously expressed, MBP-tagged or His₆-tagged proteins were purified via affinity chromatography. Protein purity was assessed using Coomassie-stained 11% SDS-PAGE gels. Proteins from Balh (a), Cur (b), Hca (c), Tbt (d), and Stm (e) are labeled along with the isolated yield per liter of culture (mg/L). BalhC^{R192A} and BalhC^{E281A} were deemed structurally destabilized due to poor yields relative to WT (6 and 19 mg/L, respectively) and band distortion. To a greater extent, MBP-StmB also was proteolytically degraded as judged by the visible MBP (~45 kDa) band.

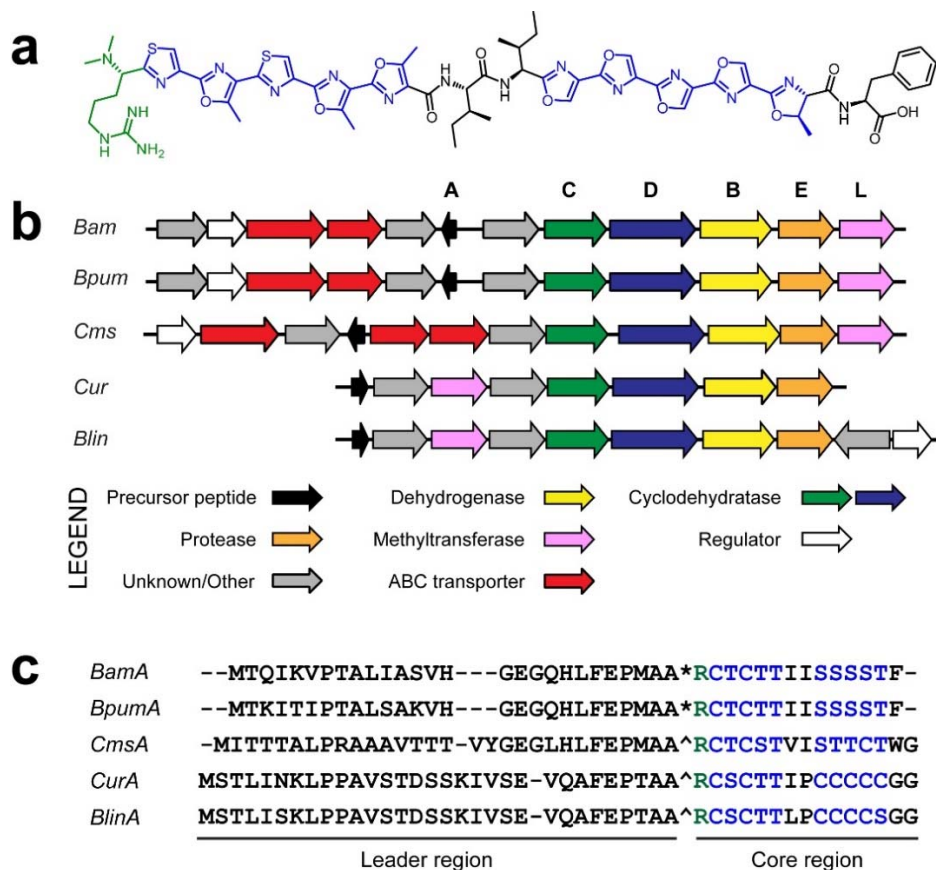


Figure 2.8 Description and genetic organization of the plantazolicins. The plantazolicins⁴⁶ are a subclass of LAPs and derive from a group of related gene clusters that are predicted to produce closely related derivatives. (a) Chemical structure of plantazolicin (PZN). Blue, heterocycles installed by the TOMM cyclodehydratase and dehydrogenase. Green, N^{α},N^{α} -dimethylarginine formed by BamL after leader peptide cleavage. (b) Genetic organization of the genes involved in PZN biosynthesis. Bam, *Bacillus amyloliquefaciens* FZB42; Bpum, *Bacillus pumilus* ATCC 7061; Cms, *Clavibacter michiganensis* subsp. sepedonicus; Cur, *Corynebacterium urealyticum* DSM 7109; Blin, *Brevibacterium linens* BL2. (c) PZN precursor peptide alignment. Color-coding is identical to panel (a) and are predicted for Cms, Cur, and Blin. * known peptidase cleavage site; ^ inferred cleavage site.

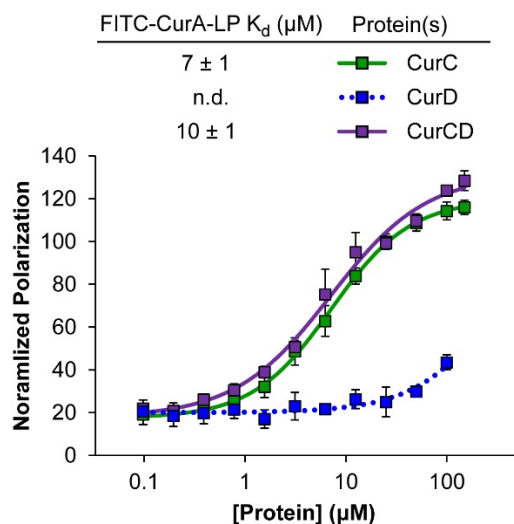


Figure 2.9 CurC is necessary and sufficient for leader peptide recognition. FP binding experiments were performed using CurC and/or CurD with FITC-labeled CurA leader peptide (FITC-CurA-LP). The rise in polarization for CurD alone is likely due to non-specific binding at high protein concentrations and dissociation constant was not able to determined (n.d.). Error bars represent s.d. of three independent data points. Error on K_d values are the s.e.m. given by regression analysis.

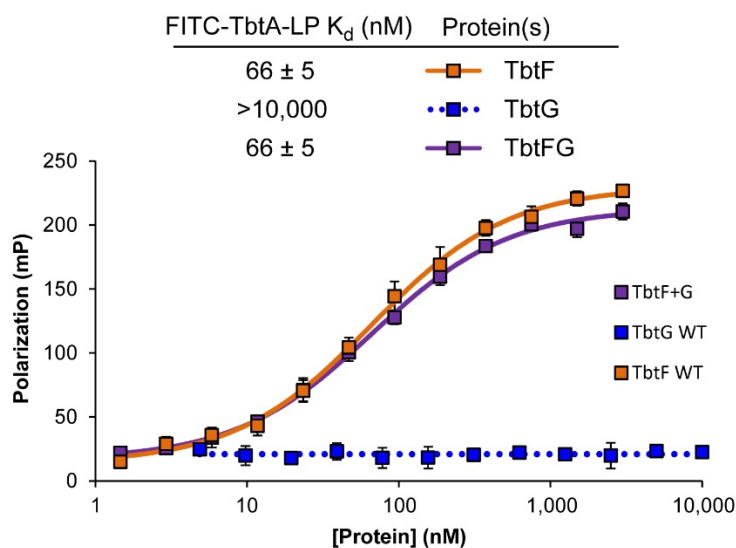
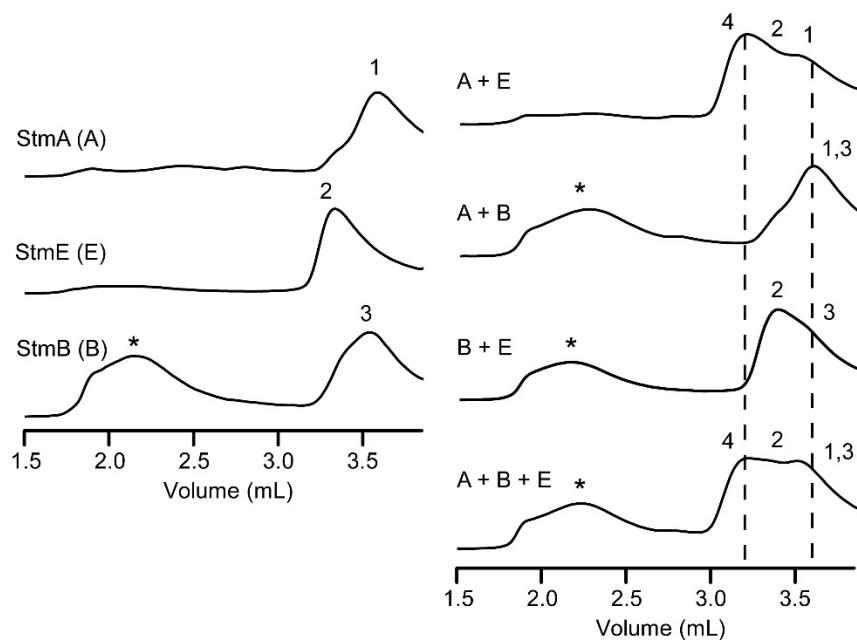


Figure 2.10 TbtF is necessary and sufficient for leader peptide recognition. FP binding experiments were performed using TbtF and/or TbtG with FITC-labeled TbtA leader peptide (FITC-TbtA-LP). The dashed line indicates no binding. Error bars represent s.d. of at three independent data points. Error on K_d values are the s.e.m. given by regression analysis.



| Peak ID | Protein(s) | V_e / V_o | Obs MW (kDa) | Actual MW (kDa) | Oligomeric State |
|---------|------------|-------------|--------------|-----------------|------------------|
| 1 | A | 1.91 | 22 | 51 | Monomer |
| 2 | E | 1.77 | 44 | 56 | Monomer |
| 3 | B | 1.88 | 25 | 61 | Monomer |
| 4 | A + E | 1.70 | 66 | 107 | 1:1 (A:C) |

Figure 2.11 Size-exclusion chromatography (SEC) of Stm proteins. The interaction between maltose-binding protein (MBP)-tagged StmA, StmB, and StmE were assessed via SEC and monitored by absorbance at 280 nm. MBP-fusions were used to maximize column separation efficiency. No visible interaction was detected between MBP-StmB and MBP-StmA; however, the clearly shifted elution volume of MBP-StmA with MBP-StmE indicated a binding interaction. Asterisk indicates aggregated StmB, which by SDS-PAGE (Figure 2.7) has some structural stability issues upon heterologous expression. Traces are representative results from two different HPLC injections.

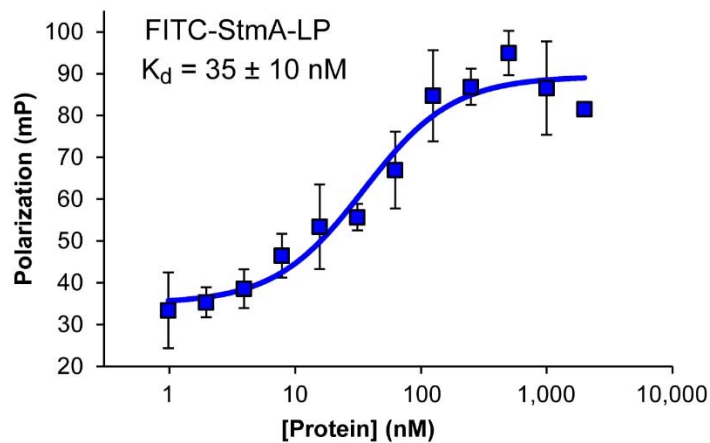


Figure 2.12 StmE is necessary and sufficient for leader peptide binding. StmE was tested for binding to the FITC-labeled StmA leader peptide (FITC-StmA-LP). StmB was not included due to significant degradation during heterologous expression and the aggregated state of the purified protein (Figures 2.7 and 2.11). Error bars represent s.d. of three independent data points. Error on K_d values is the s.e.m. given by regression analysis.

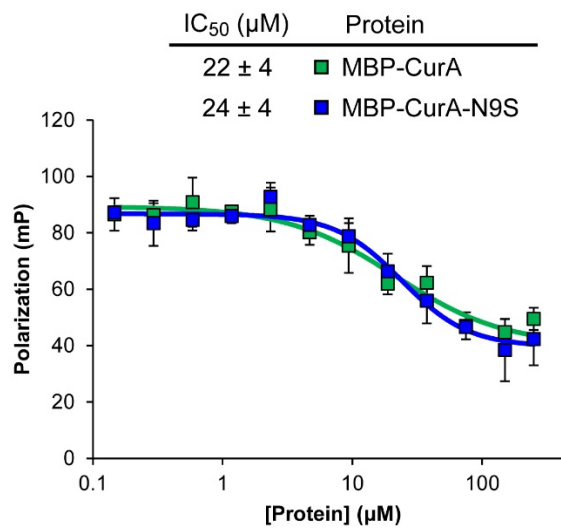


Figure 2.13 $CurA^{NS9}$ mutation does not affect binding by CurC. Using a competition FP binding assay, $CurA^{WT}$ and $CurA^{NS9}$ were shown to interact identically (within error) with CurC. MBP-tagged peptides were serially diluted and mixed with 30 nM of the FITC-labeled CurA leader peptide (FITC-CurA-LP) and 25 μ M CurC. IC_{50} values and their uncertainty were calculated using a dose response fit with Origin Pro 9.1. Error bars represent s.d. of three independent data points.

2.7 Tables

Table 2.1 PqqD domain architectures identified by CDART. CDART (conserved domain architecture retrieval tool)³⁴ was used to identify sequences and domain architectures containing an annotated PqqD domain. Many domain architectures match proteins from RiPP biosynthetic gene clusters, but unique architectures could represent proteins in unidentified RiPP pathways or false positive hits. The low number of sequences for some domain architectures is likely due to high divergence of sequences so that they no longer meet the requirements to be annotated as “PqqD.” For example, only 9 sequences match “PqqD~hypothetical_protein~YcaO” (which represents a fused cyclodehydratase), but there are hundreds of fused cyclodehydratase sequences in genbank so these values represent minimum estimates.

| Domain Architectures | Sequences | Accessions |
|--------------------------------------|-----------|---------------------------------|
| PqqD | 4242 | cl05126 |
| PqqD~Radical_SAM~SPASM | 203 | cl05126~cl18962~cl14869 |
| AdoMet_MTases~PqqD | 61 | cl17173~cl05126 |
| PqqD~ABC_ATPase | 42 | cl05126~cl21455 |
| PqqD~E1_enzyme_family | 42 | cl05126~cl22428 |
| PqqD~S2P-M50 | 41 | cl05126~cl10020 |
| PqqD~S2P-M50~Biotinyl_lipoyl_domains | 40 | cl05126~cl10020~cl11404 |
| ABC_ATPase~PqqD | 31 | cl21455~cl05126 |
| DUF59~PqqD | 25 | cl00941~cl05126 |
| ATP-grasp_4~PqqD | 20 | cl17255~cl05126 |
| HemeO~PqqD | 19 | cl15243~cl05126 |
| PqqD~Radical_SAM | 19 | cl05126~cl18962 |
| B12-binding_like~Radical_SAM~PqqD | 17 | cl00293~cl18962~cl05126 |
| PqqD~Transglut_core | 15 | cl05126~cl17362 |
| PqqD~SPASM | 11 | cl05126~cl14869 |
| PqqD~AdoMet_MTases | 11 | cl05126~cl17173 |
| Thioredoxin_like~GST_C_family~PqqD | 10 | cl00388~cl02776~cl05126 |
| SCM_chp_ScmC~PqqD | 9 | TIGR04249~cl05126 |
| PqqD~hypothetical_protein~YcaO | 9 | cl05126~TIGR03882~cl19253 |
| HTH_XRE~PqqD | 9 | cl21459~cl05126 |
| PqqD~Biotinyl_lipoyl_domains | 7 | cl05126~cl11404 |
| Lactamase_B~PqqD | 5 | cl00446~cl05126 |
| PqqD~hypothetical_protein | 5 | cl05126~TIGR03882 |
| PqqD~Extradiol_Dioxygenase_3B_like | 5 | cl05126~cl00599 |
| CAP_ED~PqqD~S2P-M50 | 4 | cl00047~cl05126~cl10020 |
| Glycosyltransferase_GTB_type~PqqD | 4 | cl10013~cl05126 |
| Radical_SAM~PqqD | 4 | cl18962~cl05126 |
| PqqD~UBN2 | 4 | cl05126~cl15874 |
| PRTases_typeI~PqqD | 3 | cl00309~cl05126 |
| FHA~PqqD~S2P-M50 | 3 | cl00062~cl05126~cl10020 |
| Lant_dehyd_N~PqqD~Lant_dehyd_C | 3 | pfam04737~cl05126~pfam04738 |
| RRM_SF~PqqD | 2 | cl17169~cl05126 |
| PqqD~SCM_chp_ScmC | 2 | cl05126~TIGR04249 |
| Radical_SAM~SPASM~PqqD | 2 | cl18962~cl14869~cl05126 |
| HSDR_N~ABC_ATPase~PqqD | 2 | cl14958~cl21455~cl05126 |
| ISOPREN_C2_like~PqqD | 2 | cl08267~cl05126 |
| PDDEXK_2~PqqD | 2 | pfam12784~cl05126 |
| terB_like~PqqD | 2 | cl11965~cl05126 |
| PqqD~Radical_SAM~RRM_SF~SPASM | 2 | cl05126~cl18962~cl17169~cl14869 |
| PqqD~YcaO | 2 | cl05126~cl19253 |
| PqqD~PHA02726 | 2 | cl05126~PHA02726 |
| PqqD~B12-binding_like~Radical_SAM | 2 | cl05126~cl00293~cl18962 |
| Glyco_tranf_GTA_type~PqqD | 1 | cl11394~cl05126 |
| GluZincin~PqqD~PDZ | 1 | cl14813~cl05126~cl00117 |

Table 2.2 Representative RiPP biosynthetic gene clusters examined for PqqD-like domains. To determine if RREs were widely distributed among additional RiPP classes, a well-studied, representative gene cluster (taken from a comprehensive RiPP review article³) was searched for proteins with homology to PqqD using HHpred²⁵. While most RiPPs are part of a specific class based on this review article, we have listed thioviridamide⁶¹ as unclassified because it has unique modifications but was discovered after publication of the review. Additionally, trifolitoxin and mycofactocin are not (fully) structurally characterized so they are left unclassified. Listed below are all the gene clusters searched with the accession code (GenBank) for the given product precursor peptide. Clusters harboring at least one protein with an identified RRE are further described in Table 2.3. Asterisk denotes compounds without fully elucidated structures.

| Class | Product | Accession | RRE present |
|---|------------------|-------------|-------------|
| Bottromycin | Bottromycin | CCM09443.1 | <u>Yes</u> |
| Cyanobactins (azole/azoline-containing) | Patellamide | AAV21154.1 | <u>Yes</u> |
| Lantipeptides (class I) | Nisin | CAA48380.1 | <u>Yes</u> |
| Lasso peptides | Astexin-1 | ADU13884.1 | <u>Yes</u> |
| Lasso peptides | Lariatins | BAL72546.1 | <u>Yes</u> |
| Linear azole-containing peptides (LAPs) | Plantazolicin | CBJ61635.1 | <u>Yes</u> |
| Microcins | Microcin C7 | CAA40808.1 | <u>Yes</u> |
| Pantocins | Pantocin A | AAD00512.2 | <u>Yes</u> |
| PQQ biosynthesis | PQQ | ACI09106.1 | <u>Yes</u> |
| Proteusins | Polytheonamide | AFS60639.1 | <u>Yes</u> |
| Sactipeptides | Subtilosin | NP_391616.1 | <u>Yes</u> |
| Thiopeptides | GE2270 | AGY49587.1 | <u>Yes</u> |
| Unclassified | Mycofactocin* | ACM06841.1 | <u>Yes</u> |
| Unclassified | Trifolitoxin* | AAA26363.1 | <u>Yes</u> |
| Autoinducing peptides | AIP-1 | AAG03056.1 | No |
| ComX and derivatives | ComX | CAB15158.1 | No |
| Glycocins | Sublancin 168 | CAB14066.1 | No |
| Head-to-tail cyclized peptides | Enterocin AS-48 | CAA72917.1 | No |
| Lantipeptides (class II) | Prochlorosin | CAE20422.1 | No |
| Lantipeptides (class III) | Labyrinthopeptin | CAX48972.1 | No |
| Lantipeptides (class IV) | Venezuelin | CCA53908.1 | No |
| Linaridins | Cypemycin | ADR72962.1 | No |
| Methanobactins | Methanobactin | E3YBA4.1 | No |
| Microviridins | Micoviridin B | CAQ16116.1 | No |
| Unclassified | Thioviridamide | BAN83916.1 | No |

Table 2.3 Detailed HHpred results. HHpred²⁵ was used to search RiPP biosynthetic proteins from well-known gene clusters (Supplementary Table 2) for homology to PqqD (additional proteins discussed in the main text are also included). Listed are the regions of query proteins that align to PqqD along with the significance of the “hit” (% Probability, E-value, and P-value). Additionally, for each protein, the number of homologous sequences is estimated to provide a measure of the prevalence for each RiPP subclass. For this estimation, a protein sequence was searched in CDART³⁴ and all retrieved sequences with the same domain architecture were counted; dashed lines indicate overlapping sequence counts due to gene cluster homology. However, CDART could not always properly distinguish RiPP homologs from non-RiPP homologs, in which case, the bioinformatics performed by others was utilized: ^aCDART³⁴, ^bMicrocins³⁷, ^cEpimerases⁴³, ^dIPR023984, ^ePQQ⁵⁹; n.d., not determined. Asterisk denotes structurally uncharacterized compounds.

| RiPP Class | Product | Protein | Accession | PqqD-like region | Probability | E-value | P-value | Sequences |
|--------------------------|-----------------------|---------|----------------|------------------|-------------|---------|---------|--------------------|
| LAPs | Plantazolicin (PZN) | BamC | CBJ61637.1 | 1–92 | 98 | 2.1E–04 | 6.2E–09 | |
| LAPs | PZN derivative* | CurC | CAQ05603.1 | 1–83 | 99 | 4.0E–08 | 1.1E–12 | |
| LAPs | Hakacin* | BalhC | ABK83869.1 | 2–86 | 98 | 1.2E–05 | 3.6E–10 | |
| LAPs | Heterocycloanthracin* | HcaF | WP_000067633.1 | 1–88 | 98 | 1.4E–05 | 3.9E–10 | |
| Cyanobactins | Patellamide | TruD | ACA04490.1 | 3–81 | 97 | 1.1E–03 | 3.2E–08 | >1500 ^a |
| Cyanobactins | Aestuaramide | LynD | WP_009787121.1 | 1–82 | 97 | 5.2E–04 | 1.5E–08 | |
| Thiopeptides | GE2270 | PbtF | AGY49592.1 | 2–83 | 98 | 1.7E–04 | 4.9E–09 | |
| Thiopeptides | GE2270 | PbtB | AGY49588.1 | 234–317 | 94 | 3.2E–01 | 9.3E–06 | |
| Thiopeptides | GE2270 derivative* | TbtF | ADG87281.1 | 3–83 | 98 | 1.9E–04 | 5.3E–09 | |
| Thiopeptides | GE2270 derivative* | TbtB | ADG87277.1 | 262–350 | 96 | 8.2E–03 | 2.4E–07 | |
| Lanthipeptides (class I) | Nisin | NisB | CAA48381.1 | 138–223 | 94 | 7.8E–02 | 2.2E–06 | ~700 ^a |
| Microcins | Microcin C | MccB | CAA40809.1 | 1–79 | 97 | 8.9E–04 | 2.6E–08 | ≥ 41 ^b |
| Pantocins | Pantocin A | PaaA | AAD00512.2 | 1–89 | 98 | 7.1E–05 | 2.0E–09 | |
| Lasso peptides | Astexin-1 | AtxB | ADU13885.1 | 2–61 | 96 | 3.7E–03 | 1.1E–07 | ~900 ^a |
| Lasso peptides | Lariatins | LarC | BAL72548.1 | 1–84 | 100 | 4.0E–25 | 1.2E–29 | |
| Lasso peptides | Streptomonicin | StmE | KIH99640.1 | 1–81 | 100 | 2E–24 | 5.8E–29 | |
| Sactipeptides | Subtilosin | AlbA | WP_003242599.1 | 8–98 | 98 | 6.3E–06 | 1.8E–10 | ~170 ^d |
| PQQ | PQQ | PqqD | ACI10478.1 | 1–88 | 100 | 2.5E–30 | 7.1E–35 | ≥ 125 ^e |
| Mycofactocins | Mycofactocin* | MftB | ACM06913.1 | 9–95 | 92 | 1.5E+00 | 4.4E–05 | ~300 ^a |
| Proteusins | Polytheonamide | PoyD | AFS60640.1 | 427–519 | 94 | 3.3E–01 | 9.6E–06 | |
| Proteusins | Polytheonamide | PoyB | AFS60637.1 | 528–609 | 97 | 7.4E–03 | 2.1E–07 | ~100 ^d |
| Botromycins | Botromycin | BmbB | CCM09442.1 | 515–587 | 96 | 1.8E–02 | 5.3E–07 | |
| Trifolitoxins | Trifolitoxin* | TfxB | AAA26364.1 | 1–85 | 94 | 9.8E–02 | 2.8E–06 | n.d. |

Table 2.4. Discrete LAP cyclodehydratase RREs bind the leader peptide. BalhC and CurC were tested for binding to FITC- BalhA1-LP and FITC-CurA-LP, respectively. The peptide binding affinity for many BalhC mutants (D14A, Y16A, F18A, E21A, F29A, R31A, Y34A, I35A, Y71A, I75A, L78A, and K82A) appeared beyond the testable range of our assay ($>150 \mu\text{M}$). Error on K_d values represents the s.e.m. from curve fitting analysis of triplicate data. Abbreviation: WT, wild-type.

| Protein | K_d (μM) | Position of mutation |
|------------|-------------------------|----------------------|
| BalhC WT | 10 ± 1 | |
| BalhC K20A | 52 ± 4 | $\beta 1$ |
| BalhC D23A | 50 ± 4 | $\beta 2$ |
| BalhC D32A | 21 ± 3 | $\beta 3$ |
| BalhC H38A | 21 ± 3 | $\beta 3$ |
| BalhC E66A | 60 ± 7 | $\alpha 2$ |
| CurC WT | 7 ± 1 | |
| CurC D18A | 3.3 ± 0.2 | $\beta 1$ |
| CurC Q21A | 8 ± 1 | $\beta 2$ |
| CurC E25A | 4.3 ± 0.3 | $\beta 2$ |
| CurC Y31A | >150 | $\beta 3$ |
| CurC E36A | 3.1 ± 0.2 | $\alpha 1$ |
| CurC R68A | 79 ± 6 | $\alpha 3$ |
| CurC I75A | 66 ± 6 | $\alpha 3$ |

Table 2.5 BalhC alanine substitutions outside of the RRE have little effect on leader peptide binding. Because most alanine substitutions tested within the BalhC RRE resulted in greatly reduced leader peptide affinity, positions outside of the RRE were also mutated for comparison. FP binding assays with these mutants indicated that residues outside the RRE contributed little to binding while some appeared to have a structural role based on low protein yield and appearance by Coomassie-stained SDS-PAGE (Figure 2.6). Error bars represent s.e.m. from regression analysis of triplicate data.

| Protein | K_d (μM) |
|-------------|-------------------------|
| BalhC WT | 10 ± 1 |
| BalhC R192A | structural |
| BalhC Q196A | 5 ± 1 |
| BalhC Y199A | 10 ± 3 |
| BalhC E281A | structural |
| BalhC N292A | 8 ± 2 |
| BalhC N294A | 7 ± 1 |

Table 2.6. F-dependent cyclodehydratase RREs bind the leader peptide. Protein components of the Hca and Tbt F-dependent cyclodehydratases were tested for binding to FITC- HcaA-LP and FITC-TbtA-LP, respectively. Error on K_d values represents the s.e.m. from curve fitting analysis of triplicate data. Abbreviation: WT, wild-type.

| Protein | K_d (nM) | Position of mutation |
|----------------|------------------------------|-----------------------------|
| HcaF WT | 89 ± 9 | |
| HcaF D21A | 130 ± 15 | β1 |
| HcaF F35A | 1,700 ± 300 | β3 |
| HcaF D38A | 77 ± 7 | β3 |
| HcaF N72A | 79 ± 7 | α3 |
| HcaF R73A | >10,000 | α3 |
| HcaF E76A | 250 ± 50 | α3 |
| HcaF E79A | 75 ± 6 | α3 |
| HcaF I80A | 500 ± 60 | α3 |
| TbtF WT | 66 ± 5 | |
| TbtG WT | >10,000 | |
| TbtF/G | 66 ± 5 | |
| TbtF D19A | 57 ± 2 | β2 |
| TbtF L24A | 65 ± 3 | β2 |
| TbtF V30A | 210 ± 15 | β3 |
| TbtF F68A | 7,700 ± 400 | α3 |
| TbtF R71A | 190 ± 22 | α3 |
| TbtF Q74A | 49 ± 3 | α3 |
| TbtF R79A | 79 ± 5 | α3 |

Table 2.7. Split lasso peptide protease RRE binds the leader peptide. StmE binding to FITC-labeled StmA leader peptide was investigated by FP. Error on K_d values represents the s.e.m. from curve fitting analysis of triplicate data. Abbreviation: WT, wild-type.

| Protein | K_d (nM) | Position of mutation |
|-----------|------------|----------------------|
| StmE WT | 35 ± 10 | |
| StmE V12A | 160 ± 50 | β1 |
| StmE N22A | 410 ± 120 | β2 |
| StmE Y28A | 28 ± 5 | β3 |
| StmE D68A | 38 ± 7 | α3 |
| StmE D69A | 2600 ± 400 | α3 |
| StmE L73A | >10,000 | α3 |
| StmE D75A | 35 ± 9 | α3 |
| StmE Q76A | 75 ± 25 | α3 |

Table 2.8 Sequence of fluorescent peptides used in this study. Fluorescein isothiocyanate (FITC)-conjugated peptides were purchased from Genscript with an amino hexyl linker (Ahx) spacer, with the exception of the FITC-labeled leader peptides (LP) for HcaA and TbtA, which were prepared by labeling the *N*-terminus of a heterologously expressed peptide (Dunbar *et al.*, submitted). FITC-labeled CurA-LP represents the CurA leader with an N6S substitution (the wild-type sequence for BlinA⁴⁶, see Figure 2.8). This peptide binds to CurC with the same affinity as the wild-type CurA peptide (Figure 2.13).

| Peptide | Sequence |
|----------------|---|
| FITC-BalhA1-LP | FITC-Ahx-GMEQKKILDIKLTETGKINYAHKPDD |
| FITC-CurA-LP | FITC-Ahx-GMSTLISKLPVAVSTDSSKIVSEVQAFEPT |
| FITC-HcaA-LP | FITC-SGSMNQFQQELQSLNLDYQTNVYVWDPQQSQYPPYYIQDDAR |
| FITC-TbtA-LP | FITC-SRRGSMNLNLSLDPMDVFEMADSGMEVESLTAGHGMPEVGA |
| FITC-StmA-LP | FITC-Ahx-MSAYEIPTLTRIGKFKDVTK |

Table 2.9 Oligonucleotides used for cloning and site-directed mutagenesis. The cloning primers are named for the amplified gene and whether the primer was for the forward (F) or reverse (R) direction. The restriction enzyme sites are also given. For the mutagenic primers, the resultant amino acid substitution is indicated in parentheses.

| Cloning Primer | Oligonucleotide Sequence | Restriction site |
|------------------|---|------------------|
| CurA F | AAAGGATCCATGTCCACACTAATCAACAAGCTGCCTC | BamHI |
| CurA R | AAAGCGCCGCTTAACCGCCGAGCAGCA | NotI |
| CurC F | AAAAGGATCCATGACTGAAACTCAGAGGTACCAAG | BamHI |
| CurC R | AAAAGCGCCGCTCACTCCTCATGAAACCAAATTTGTG | NotI |
| CurD F | AAAGTCGACATGGCCGCATCAGCAGAATTTTCG | Sall |
| CurD R | AAAGCGCCGCTTATGGGAATGGGTGGGGAAGAGG | NotI |
| StmA F | AAAAAGCTTGATGAGCGCTTACGAGATCCCCACTC | HindIII |
| StmA R | AAAGCGCCGCTCACGGGTAGATGACGATCAGAGCC | NotI |
| StmE F | AAAAAGCTTGATGAGAGTCCAGATGCGGCCCAACG | HindIII |
| StmE R | AAAGCGCCGCTCATGAGTGCCGCACCTCGACCAGC | NotI |
| StmB F | AAAAAGCTTGATGAGCTCCCCGATCTGCCCGAGG | HindIII |
| StmB R | AAAGCGCCGCTCACGCACTGGGGCGCGCCGAAC | NotI |
| Mutagenic Primer | Oligonucleotide Sequence | |
| BalhC (D14A) F | TGAGGTTCTTAATTATAAACCTATAATAGCTAGCTATTGTTTTGTAAAAGAAGATGATG | |
| BalhC (D14A) R | CATCATCTTCTTTTACAAAACAATAGCTAGCTATATAGGTTTATAAATTAAGAACCTCA | |
| BalhC (Y16A) F | CTATAATAGATAGCGCTTGTGTTTTGTAAAAGAAGATGATGAAGG | |
| BalhC (Y16A) R | CTTTTACAAAACAAGCGCTATCTATATAGGTTTATAAATTAAGAACC | |
| BalhC (F18A) F | GATAGCTATTGTGCTGTAAAAGAAGATGATGAAGGAC | |
| BalhC (F18A) R | CTTCTTTTACAGCACAATAGCTATCTATTATAGGTTTATAAATTAAGAACC | |
| BalhC (K20A) F | CTATGTTTTGTAGCAGAAGATGATGAAGGACTAACATTTTTTAATAG | |
| BalhC (K20A) R | CCTTCATCATCTTCTGCTACAAAACAATAGCTATCTATTATAGG | |
| BalhC (E21A) F | GTTTTGTAAAAGCAGATGATGAAGGACTAACATTTTTTAATAG | |
| BalhC (E21A) R | CCTTCATCATCTGCTTTTACAAAACAATAGCTATCTATTATAG | |
| BalhC (D23A) F | GTAAAAGAAGATGCTGAAGGACTAACATTTTTTAATAGGGATA | |
| BalhC (D23A) R | GTTAGTCCTTCAGCATCTTCTTTTACAAAACAATAGCTATC | |
| BalhC (F29A) F | GGACTAACATTTGCTAATAGGGATACATATATAAATTTTCATGGAG | |
| BalhC (F29A) R | TATCCCTATTAGCAAATGTTAGTCCTTCATCATCTCTTTTAC | |
| BalhC (R31A) F | AACATTTTTAATGCGGATACATATATAAATTTTCATGGAGGTTTC | |
| BalhC (R31A) R | AATTTATATATGTATCCGCATTAATAAATGTTAGTCCTTCATCATCTT | |
| BalhC (D32A) F | ATGAAGGACTAACATTTTTTAATAGGGCTACATATATAAATTTTCATGGAGGTTTC | |
| BalhC (D32A) R | GAACCTCCATGAAAATTTATATATGTAGCCCTATTAATAAATGTTAGTCCTTCAT | |
| BalhC (Y34A) F | ATAGGGATACAGCTATAAATTTTCATGGAGGTTTCAGTTGAG | |
| BalhC (Y34A) R | CATGAAAATTTATAGCTGTATCCCTATTAATAAATGTTAGTCCTTCATC | |
| BalhC (I35A) F | AGGGATACATATGCAAATTTTCATGGAGGTTTCAGTTGAGGATATC | |
| BalhC (I35A) R | CCATGAAAATTTGCATATGTATCCCTATTAATAAATGTTAGTCC | |

Table 2.9 (Continued)

| Mutagenic Primer | Oligonucleotide Sequence |
|-------------------------|--|
| BalhC (H38A) F | CATATATAAAATTTTGCTGGAGGTTTCAGTTGAGGATATCTTTGCT |
| BalhC (H38A) R | ACTGAACCTCCAGCAAATTTATATATGTATCCCTATTAATAAATGT |
| BalhC (E66A) F | CAGAAAAATTAGCATTACCTATAGAATATATGTGCGACATAAT |
| BalhC (E66A) R | TCTATAGGTAATGCTAATTTTTCTGCTAATTGTTCTGTACTTAAT |
| BalhC (Y71A) F | TACCTATAGAAGCTATGTGCGACATAATTAATTTGTTAGACG |
| BalhC (Y71A) R | ATGTCGCACATAGCTTCTATAGGTAATTTCTAATTTTTCTGCT |
| BalhC (I75A) F | GAATATATGTGCGACGCAATTAATTTAGACGAGAAAAATATTA |
| BalhC (I75A) R | AACAATTTAATTTGCGTCGCACATATATTTCTATAGGTAATCT |
| BalhC (L78A) F | ACATAATTAAGCGTTAGACGAGAAAAATATTATAAAAAATTACG |
| BalhC (L78A) R | TTCTCGTCTAACGCTTTAATTATGTGCGACATATATTTCTATAG |
| BalhC (K82A) F | TGTTAGACGAGGCAAATATTATAAAAAATTACGATTTACAAGAAAAATAT |
| BalhC (K82A) R | TTATAATATTTGCTCGCTAACAATTTAATTATGTGCGCAC |
| BalhC (R192A) F | GTACCGTTTTTAGCAGGAGTAGTACAAGAGCAATTTTTAGTA |
| BalhC (R192A) R | TGTACTACTCCTGCTAAAAACGGTACTTTGTAACATTTTCG |
| BalhC (Q196A) F | GAGGAGTAGTAGCAGAGCAATTTTTAGTATAGGGCCAATA |
| BalhC (Q196A) R | TAAAATATGCTCTGCTACTACTCCTCTTAAAAACGGTACTTTG |
| BalhC (Y199A) F | TACAAGAGCAAGCTTTTAGTATAGGGCCAATATTC |
| BalhC (Y199A) R | CTATACTAAAAGCTTGCTCTTGTACTACTCCTCTTAAAAACG |
| BalhC (E281A) F | TGTATGCCTTGCAGAAATAATTGGAAAAGCATTCACTTATAATGTATTAAATTAAGTT |
| BalhC (E281A) R | AACTTAAATTAATACATTATAAGTGAATGCTTTTCCAATTATTTTCGCAAGGCATACA |
| BalhC (N292A) F | AATTTAAGTTCAGCTTTAAATCCTGTATTAAAAGTTCCTGGG |
| BalhC (N292A) R | TACAGGATTTAAAGCTGAACTTAAATTAATACATTATAAGTGAATTTCTTTTC |
| BalhC (N294A) F | AATTTAAGTTCAAATTTAGCTCCTGTATTAAAAGTTCCTGG |
| BalhC (N294A) R | CTTTTAATACAGGAGCTAAATTTGAACCTTAAATTAATACATTATAAGTGAATTC |
| CurA (N6S) F | CCACACTAATCTCCAAGCTGCCTCCCGCAGTTTCAACTGAT |
| CurA (N6S) R | GGAGGCAGCTTGGAGATAGTGTGGACATGGATCCGGATTG |
| CurC (D18A) F | GGGTATCGCGCGCCGGCAACCAGATCGTTATAGAAGCCGCG |
| CurC (D18A) R | ATCTGGTTGCCGGCGCGGATACCCGCAGGTCTGGGTTAAC |
| CurC (Q21A) F | GCGACGGCAACGCGATCGTTATAGAAGCCGCGTCAAACACA |
| CurC (Q21A) R | TCTATAACGATCGCGTTGCCGTCGCGGATACCCGCAGGTC |
| CurC (E25A) F | AGATCGTTATAGCAGCCGCTCAAACACATATTGTCTCTCCG |
| CurC (E25A) R | TTTGACGCGGCTGCTATAACGATCTGGTTGCCGTCGCGCGAT |
| CurC (Y31A) F | GTCAAACACAGCTTGTCTCTCCGATGAGGCTCTATGGC |
| CurC (Y31A) R | CGGAGAGACAAGCTGTGTTTGACGCGGCTTCTATAACG |
| CurC (E36A) F | TCTCTCCGATGCGGCTCTATGGCCTTTTCATCTCTGAAGT |
| CurC (E36A) R | GCCATAGAGCCGATCGGAGAGACAATATGTGTTTGACG |
| CurC (R68A) F | GAGACATTGAAGCACTTCAAGCAGCATTGAAATTTCTTTTCAGA |
| CurC (R68A) R | GCTGCTTGAAGTCTTCAATGTCTCTAGGCTCATACTCGTC |

Table 2.9 (Continued)

| Mutagenic Primer | Oligonucleotide Sequence |
|-------------------------|--|
| CurC (I75A) F | CAGCATTGGAAGCTCTTTCAGAACTGAAGTTCTTATTCCATCT |
| CurC (I75A) R | GTTCCTGAAAGAGCTTCAAATGCTGCTTGAAGTCGTTCAATG |
| HcaF (D21A) F | CCTCCCCGCCTCAAACGGGGGTGTCTATTTTC |
| HcaF (D21A) R | CCGTTTGAGGCGGGGAGGAAGAAAGTATCTTTATTTAC |
| HcaF (F35A) F | CGCCAGTTCAGCTCGTATGGACGGTGACGGAATTTAT |
| HcaF (F35A) R | CGTCCATACGAGCTGAACTGGCGTTATTTTCGAAAATAGACAC |
| HcaF (D38A) F | CATTCGTATGGCCGGTGACGGAATTTATAACTGGATTGA |
| HcaF (D38A) R | TCCGTACCCGGCCATACGAAATGAACTGGCGTTATTTTC |
| HcaF (N72A) F | CTCCCCATCAAGCTCGTGTATTTGAAATTGGAGAGATTTTATATACA |
| HcaF (N72A) R | TCAAATACACGAGCTTGATAGGGGAGAGGGAGACCATC |
| HcaF (R73A) F | TCCCTATCAAATGCTGTATTTGAAATTGGAGAGATTTTATATACAAATGGGT |
| HcaF (R73A) R | AATTTCAAATACAGCATTTTGATAGGGGAGAGGGAGACCATC |
| HcaF (E76A) F | ATCGTGTATTTGCAATTGGAGAGATTTTATATACAAATGGGTTTGT |
| HcaF (E76A) R | ATCTCTCAAATTGCAAATACACGATTTTGATAGGGGAGAG |
| HcaF (E79A) F | TGAAATTGGAGCGATTTTATATACAAATGGGTTTGTCCGTG |
| HcaF (E79A) R | GTATATAAAATCGCTCCAATTTCAAATACACGATTTTGATAG |
| HcaF (I80A) F | TCGTGTATTTGCAATTGGAGAGATTTTATATACAAATGGGTTTGT |
| HcaF (I80A) R | CATTTGTATATAAAGCCTCTCCAATTTCAAATACACGATTTTGATAGGGGAG |
| TbtF (D19A) F | TCCGACCAGTGTGGCGTTGCTTTCCTGACCCGT |
| TbtF (D19A) R | GAAAGCAACGCCAGCACCTGGTCGGAGCATAACGAACG |
| TbtF (L24A) F | GCGTTGCTTTCGCGACCCGTGATGGTATTGTGAGTATTACCGGC |
| TbtF (L24A) R | CATCACGGGTCGCGAAAGCAACGCCATCACTGGTCCG |
| TbtF (V30A) F | CGTGATGGTATTGCGAGTATTACCGGCTTTCATCCACCG |
| TbtF (V30A) R | CGGTAATACTCGCAATACCATCACGGGTCAGGAAAGCAACG |
| TbtF (F68A) F | ACGTGCGCAGCCGTGCTGCGTCTGCTGCAGGCAC |
| TbtF (F68A) R | ACGCAGCACGGCTGCGCGACGTTCGGGTGCCAG |
| TbtF (R71A) F | CATTCGTGCTGGCTCTGCTGCAGGCACTGGCAGATC |
| TbtF (R71A) R | CCTGCAGCAGAGCCAGCACGAATGCGCGACGTTCC |
| TbtF (Q74A) F | GCGTCTGCTGGCGGCACTGGCAGATCGCGGTCTGATTATCG |
| TbtF (Q74A) R | CTGCCAGTGCCGCCAGCAGCAGCAGCAGCAATGCG |
| TbtF (R79A) F | CACTGGCAGATGCCGGTCTGATTATCGACGCAGCCGG |
| TbtF (R79A) R | GATAATCAGACCCGCATCTGCCAGTGCCTGCAGCAGACG |
| StmE (V12A) F | CGTGCAGATCGCTTCCCGGACGGCGACAACCTTC |
| StmE (V12A) R | GTCCCGGGAAGCGATCTGCACGTTGGGCCGCATC |
| StmE (N22A) F | CTTCGTCTCTGCCCTGCGCAACGGCGAATACTGGC |
| StmE (N22A) R | GTGCGCAGGGCGAGGACGAAGTTGTGCGCCGTCCCG |
| StmE (Y28A) F | CAACGGCGAAGCCTGGCACCTCAACTCGTCGGCG |
| StmE (Y28A) R | TGAGGTGCCAGGCTTCGCCGTTGCGCAGGTTGAG |

Table 2.9 (Continued)

| Mutagenic Primer | Oligonucleotide Sequence |
|------------------|--|
| StmE (D68A) F | AGACTGTCCGGGCGACCTCACCGAACTGGTGGACCAGCTC |
| StmE (D68A) R | TCGGTGAGGTCGGCCCGACAGTCTGCTCGGGGGCGGA |
| StmE (D69A) F | CTGTCCGGGACGCCCTCACCGAACTGGTGGACCAGCTCCGC |
| StmE (D69A) R | AGTTCGGTGAGGGCGTCCCGGACAGTCTGCTCGGGGGCG |
| StmE (L73A) F | ACCTCACCGAAGCGGTGGACCAGCTCCGCAAGGCCAAGCTGG |
| StmE (L73A) R | AGCTGGTCCACCGCTTCGGTGAGGTCGTCGCGGACAGTCTGC |
| StmE (D75A) F | GAACTGGTGGCCAGCTCCGCAAGGCCAAGCTG |
| StmE (D75A) R | GCGGAGCTGGGCCACCAAGTTCGGTGAGGTCGTC |
| StmE (Q76A) F | ACTGGTGGACGCGCTCCGCAAGGCCAAGCTGGTC |
| StmE (Q76A) R | CCTTGGGAGCGCGTCCACCAAGTTCGGTGAGGTCG |

2.8 References

- (1) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs over the 30 Years from 1981 to 2010. *J. Nat. Prod.* **2012**, *75* (3), 311-35.
- (2) Carlson, E. E. Natural Products as Chemical Probes. *ACS Chem. Biol.* **2010**, *5* (7), 639-53.
- (3) Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J., et al. Ribosomally Synthesized and Post-Translationally Modified Peptide Natural Products: Overview and Recommendations for a Universal Nomenclature. *Nat. Prod. Rep.* **2013**, *30* (1), 108-60.
- (4) Bindman, N. A.; van der Donk, W. A. A General Method for Fluorescent Labeling of the N-Termini of Lanthipeptides and Its Application to Visualize Their Cellular Localization. *J. Am. Chem. Soc.* **2013**, *135* (28), 10362-71.
- (5) Cotter, P. D.; Ross, R. P.; Hill, C. Bacteriocins – a Viable Alternative to Antibiotics? *Nat. Rev. Micro.* **2013**, *11* (2), 95-105.
- (6) Oman, T. J.; van der Donk, W. A. Follow the Leader: The Use of Leader Peptides to Guide Natural Product Biosynthesis. *Nat. Chem. Biol.* **2010**, *6* (1), 9-18.
- (7) Ruffner, D. E.; Schmidt, E. W.; Heemstra, J. R. Assessing the Combinatorial Potential of the Ripp Cyanobactin Tru Pathway. *ACS Synth. Biol.* **2015**, *4* (4), 482-92.
- (8) Goto, Y.; Ito, Y.; Kato, Y.; Tsunoda, S.; Suga, H. One-Pot Synthesis of Azoline-Containing Peptides in a Cell-Free Translation System Integrated with a Posttranslational Cyclodehydratase. *Chem. Biol.* **2014**, *21* (6), 766-74.
- (9) Deane, C. D.; Melby, J. O.; Molohon, K. J.; Susarrey, A. R.; Mitchell, D. A. Engineering Unnatural Variants of Plantazolicin through Codon Reprogramming. *ACS Chem. Biol.* **2013**, *8* (9), 1998-2008.

- (10) Mitchell, D. A.; Lee, S. W.; Pence, M. A.; Markley, A. L.; Limm, J. D.; Nizet, V.; Dixon, J. E. Structural and Functional Dissection of the Heterocyclic Peptide Cytotoxin. *J. Biol. Chem.* **2009**, *284* (19), 13004-12.
- (11) Melby, J. O.; Nard, N. J.; Mitchell, D. A. Thiazole/Oxazole-Modified Microcins: Complex Natural Products from Ribosomal Templates. *Curr. Opin. Chem. Biol.* **2011**, *15* (3), 369-78.
- (12) Dunbar, K. L.; Chekan, J. R.; Cox, C. L.; Burkhart, B. J.; Nair, S. K.; Mitchell, D. A. Discovery of a New Atp-Binding Motif Involved in Peptidic Azoline Biosynthesis. *Nat. Chem. Biol.* **2014**, *10* (10), 823-9.
- (13) Lee, S. W.; Mitchell, D. A.; Markley, A. L.; Hensler, M. E.; Gonzalez, D.; Wohlrab, A.; Dorrestein, P. C.; Nizet, V.; Dixon, J. E. Discovery of a Widely Distributed Toxin Biosynthetic Gene Cluster. *Proc. Natl. Acad. Sci. USA* **2008**, *105* (15), 5879-84.
- (14) Schmidt, E. W.; Nelson, J. T.; Rasko, D. A.; Sudek, S.; Eisen, J. A.; Haygood, M. G.; Ravel, J. Patellamide a and C Biosynthesis by a Microcin-Like Pathway in Prochloron Didemni, the Cyanobacterial Symbiont of Lissoclinum Patella. *Proc. Natl. Acad. Sci. USA* **2005**, *102* (20), 7315-7320.
- (15) Li, Y. M.; Milne, J. C.; Madison, L. L.; Kolter, R.; Walsh, C. T. From Peptide Precursors to Oxazole and Thiazole-Containing Peptide Antibiotics: Microcin B17 Synthase. *Science* **1996**, *274* (5290), 1188-1193.
- (16) Dunbar, K. L.; Melby, J. O.; Mitchell, D. A. Ycao Domains Use Atp to Activate Amide Backbones During Peptide Cyclodehydrations. *Nat. Chem. Biol.* **2012**, *8* (6), 569-75.
- (17) McIntosh, J. A.; Schmidt, E. W. Marine Molecular Machines: Heterocyclization in Cyanobactin Biosynthesis. *Chembiochem : a European journal of chemical biology* **2010**, *11* (10), 1413-21.
- (18) Burroughs, A. M.; Iyer, L. M.; Aravind, L. Natural History of the E1-Like Superfamily: Implication for Adenylation, Sulfur Transfer, and Ubiquitin Conjugation. *Proteins* **2009**, *75* (4), 895-910.
- (19) Melby, J. O.; Dunbar, K. L.; Trinh, N. Q.; Mitchell, D. A. Selectivity, Directionality, and Promiscuity in Peptide Processing from a Bacillus Sp. Al Hakam Cyclodehydratase. *J. Am. Chem. Soc.* **2012**, *134* (11), 5309-16.
- (20) Regni, C. A.; Roush, R. F.; Miller, D. J.; Nourse, A.; Walsh, C. T.; Schulman, B. A. How the Mccb Bacterial Ancestor of Ubiquitin E1 Initiates Biosynthesis of the Microcin C7 Antibiotic. *EMBO J.* **2009**, *28* (13), 1953-64.
- (21) McIntosh, J. A.; Lin, Z.; Tianero, M. D.; Schmidt, E. W. Aestuarinamides, a Natural Library of Cyanobactin Cyclic Peptides Resulting from Isoprene-Derived Claisen Rearrangements. *ACS Chem. Biol.* **2013**, *8* (5), 877-83.

- (22) Ortega, M. A.; Hao, Y.; Zhang, Q.; Walker, M. C.; van der Donk, W. A.; Nair, S. K. Structure and Mechanism of the Trna-Dependent Lantibiotic Dehydratase Nisb. *Nature* **2015**, *517* (7535), 509-12.
- (23) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215* (3), 403-410.
- (24) Marchler-Bauer, A.; Zheng, C.; Chitsaz, F.; Derbyshire, M. K.; Geer, L. Y.; Geer, R. C.; Gonzales, N. R.; Gwadz, M.; Hurwitz, D. I.; Lanczycki, C. J., et al. Cdd: Conserved Domains and Protein Three-Dimensional Structure. *Nucleic Acids Res.* **2013**, *41* (Database issue), D348-52.
- (25) Soding, J.; Biegert, A.; Lupas, A. N. The Hhpred Interactive Server for Protein Homology Detection and Structure Prediction. *Nucleic Acids Res.* **2005**, *33* (Web Server issue), W244-8.
- (26) Soding, J. Protein Homology Detection by Hmm-Hmm Comparison. *Bioinformatics* **2005**, *21* (7), 951-60.
- (27) Lopes, A.; Amarir-Bouhram, J.; Faure, G.; Petit, M. A.; Guerois, R. Detection of Novel Recombinases in Bacteriophage Genomes Unveils Rad52, Rad51 and Gp2.5 Remote Homologs. *Nucleic Acids Res.* **2010**, *38* (12), 3952-62.
- (28) Klinman, J. P.; Bonnot, F. Intrigues and Intricacies of the Biosynthetic Pathways for the Enzymatic Quinocofactors: Pqq, Ttq, Ctq, Tpq, and Ltq. *Chem. Rev.* **2014**, *114* (8), 4343-65.
- (29) Latham, J. A.; Iavarone, A. T.; Barr, I.; Juthani, P. V.; Klinman, J. P. Pqqd Is a Novel Peptide Chaperone That Forms a Ternary Complex with the Radical S-Adenosylmethionine Protein Pqqe in the Pyrroloquinoline Quinone Biosynthetic Pathway. *J. Biol. Chem.* **2015**, *290* (20), 12908-18.
- (30) Tsai, T. Y.; Yang, C. Y.; Shih, H. L.; Wang, A. H.; Chou, S. H. Xanthomonas Campestris Pqqd in the Pyrroloquinoline Quinone Biosynthesis Operon Adopts a Novel Saddle-Like Fold That Possibly Serves as a Pqq Carrier. *Proteins* **2009**, *76* (4), 1042-8.
- (31) Holm, L.; Rosenstrom, P. Dali Server: Conservation Mapping in 3d. *Nucleic Acids Res.* **2010**, *38* (Web Server issue), W545-9.
- (32) Wecksler, S. R.; Stoll, S.; Iavarone, A. T.; Imsand, E. M.; Tran, H.; Britt, R. D.; Klinman, J. P. Interaction of Pqqe and Pqqd in the Pyrroloquinoline Quinone (Pqq) Biosynthetic Pathway Links Pqqd to the Radical Sam Superfamily. *Chem. Commun. (Camb)* **2010**, *46* (37), 7031-3.
- (33) Li, Y.; Zirah, S.; Rebuffat, S., Biosynthesis, Regulation and Export of Lasso Peptides. In *Lasso Peptides*, Springer New York: 2015; pp 81-95.
- (34) Geer, L. Y.; Domrachev, M.; Lipman, D. J.; Bryant, S. H. Cdart: Protein Homology by Domain Architecture. *Genome Res.* **2002**, *12* (10), 1619-23.
- (35) Haft, D. H. A Strain-Variable Bacteriocin in Bacillus Anthracis and Bacillus Cereus with Repeated Cys-Xaa-Xaa Motifs. *Biol. Direct.* **2009**, *4* (1), 15.

- (36) Dunbar, K. L.; Tietz, J. I.; Cox, C. L.; Burkhart, B. J.; Mitchell, D. A. Identification of an Auxiliary Leader Peptide-Binding Protein Required for Azoline Formation in Ribosomal Natural Products. *J. Am. Chem. Soc.* **2015**, *Submitted*.
- (37) Bantysh, O.; Serebryakova, M.; Makarova, K. S.; Dubiley, S.; Datsenko, K. A.; Severinov, K. Enzymatic Synthesis of Bioinformatically Predicted Microcin C-Like Compounds Encoded by Diverse Bacteria. *MBio* **2014**, *5* (3), e01059-14.
- (38) Hegemann, J. D.; Zimmermann, M.; Zhu, S.; Klug, D.; Marahiel, M. A. Lasso Peptides from Proteobacteria: Genome Mining Employing Heterologous Expression and Mass Spectrometry. *Biopolymers* **2013**, *100* (5), 527-42.
- (39) Haft, D. H. Bioinformatic Evidence for a Widely Distributed, Ribosomally Produced Electron Carrier Precursor, Its Maturation Proteins, and Its Nicotinoprotein Redox Partners. *BMC Genomics* **2011**, *12* (1), 21.
- (40) Haft, D. H.; Basu, M. K. Biological Systems Discovery in Silico: Radical S-Adenosylmethionine Protein Families and Their Target Peptides for Posttranslational Modification. *J. Bacteriol.* **2011**, *193* (11), 2745-55.
- (41) Schramma, K. R.; Bushin, L. B.; Seyedsayamdost, M. R. Structure and Biosynthesis of a Macrocyclic Peptide Containing an Unprecedented Lysine-to-Tryptophan Crosslink. *Nature chemistry* **2015**, *7* (5), 431-7.
- (42) Goldman, P. J.; Grove, T. L.; Sites, L. A.; McLaughlin, M. I.; Booker, S. J.; Drennan, C. L. X-Ray Structure of an Adomet Radical Activase Reveals an Anaerobic Solution for Formylglycine Posttranslational Modification. *Proc. Natl. Acad. Sci. USA* **2013**, *110* (21), 8519-24.
- (43) Morinaka, B. I.; Vagstad, A. L.; Helf, M. J.; Gugger, M.; Kegler, C.; Freeman, M. F.; Bode, H. B.; Piel, J. Radical S-Adenosyl Methionine Epimerases: Regioselective Introduction of Diverse D-Amino Acid Patterns into Peptide Natural Products. *Angew. Chem. Int. Ed. Engl.* **2014**, *53* (32), 8503-7.
- (44) Breil, B. T.; Ludden, P. W.; Triplett, E. W. DNA Sequence and Mutational Analysis of Genes Involved in the Production and Resistance of the Antibiotic Peptide Trifolitoxin. *J. Bacteriol.* **1993**, *175* (12), 3693-702.
- (45) Breil, B.; Borneman, J.; Triplett, E. W. A Newly Discovered Gene, TfuA, Involved in the Production of the Ribosomally Synthesized Peptide Antibiotic Trifolitoxin. *J. Bacteriol.* **1996**, *178* (14), 4150-6.
- (46) Molohon, K. J.; Melby, J. O.; Lee, J.; Evans, B. S.; Dunbar, K. L.; Bumpus, S. B.; Kelleher, N. L.; Mitchell, D. A. Structure Determination and Interception of Biosynthetic Intermediates for the Plantazolicin Class of Highly Discriminating Antibiotics. *ACS Chem. Biol.* **2011**, *6* (12), 1307-13.

- (47) Morris, R. P.; Leeds, J. A.; Naegeli, H. U.; Oberer, L.; Memmert, K.; Weber, E.; LaMarche, M. J.; Parker, C. N.; Burrer, N.; Esterow, S., et al. Ribosomally Synthesized Thiopeptide Antibiotics Targeting Elongation Factor Tu. *J. Am. Chem. Soc.* **2009**, *131* (16), 5946-55.
- (48) Metelev, M.; Tietz, J. I.; Melby, J. O.; Blair, P. M.; Zhu, L.; Livnat, I.; Severinov, K.; Mitchell, D. A. Structure, Bioactivity, and Resistance Mechanism of Streptomonicin, an Unusual Lasso Peptide from an Understudied Halophilic Actinomycete. *Chem. Biol.* **2015**, *22* (2), 241-50.
- (49) Lee, J.; Hao, Y.; Blair, P. M.; Melby, J. O.; Agarwal, V.; Burkhart, B. J.; Nair, S. K.; Mitchell, D. A. Structural and Functional Insight into an Unexpectedly Selective N-Methyltransferase Involved in Plantazolicin Biosynthesis. *Proc. Natl. Acad. Sci. USA* **2013**, *110* (32), 12954-9.
- (50) Fedorov, A. A.; Fedorov, E.; Gertler, F.; Almo, S. C. Structure of Evh1, a Novel Proline-Rich Ligand-Binding Module Involved in Cytoskeletal Dynamics and Neural Function. *Nat. Struct. Mol. Biol.* **1999**, *6* (7), 661-665.
- (51) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Soding, J., et al. Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7* (1), 539.
- (52) Sali, A.; Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**, *234* (3), 779-815.
- (53) Sali, A.; Potterton, L.; Yuan, F.; van Vlijmen, H.; Karplus, M. Evaluation of Comparative Protein Modeling by Modeller. *Proteins* **1995**, *23* (3), 318-26.
- (54) Atkinson, H. J.; Morris, J. H.; Ferrin, T. E.; Babbitt, P. C. Using Sequence Similarity Networks for Visualization of Relationships across Diverse Protein Superfamilies. *PloS one* **2009**, *4* (2), e4345.
- (55) 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), 2498-504.
- (56) Zhao, S.; Sakai, A.; Zhang, X.; Vetting, M. W.; Kumar, R.; Hillerich, B.; San Francisco, B.; Solbiati, J.; Steves, A.; Brown, S., et al. Prediction and Characterization of Enzymatic Activities Guided by Sequence Similarity and Genome Neighborhood Networks. *Elife* **2014**, *3*, e03275.
- (57) Murphy, K.; O'Sullivan, O.; Rea, M. C.; Cotter, P. D.; Ross, R. P.; Hill, C. Genome Mining for Radical Sam Protein Determinants Reveals Multiple Sactibiotic-Like Gene Clusters. *PloS one* **2011**, *6* (7), e20852.
- (58) Maksimov, M. O.; Pelczer, I.; Link, A. J. Precursor-Centric Genome-Mining Approach for Lasso Peptide Discovery. *Proc. Natl. Acad. Sci. USA* **2012**, *109* (38), 15223-8.

- (59) Shen, Y. Q.; Bonnot, F.; Imsand, E. M.; RoseFigura, J. M.; Sjolander, K.; Klinman, J. P. Distribution and Properties of the Genes Encoding the Biosynthesis of the Bacterial Cofactor, Pyrroloquinoline Quinone. *Biochemistry* **2012**, *51* (11), 2265-75.
- (60) Inokoshi, J.; Matsuhama, M.; Miyake, M.; Ikeda, H.; Tomoda, H. Molecular Cloning of the Gene Cluster for Lariatins Biosynthesis of *Rhodococcus Jostii* K01-B0171. *Appl. Microbiol. Biotechnol.* **2012**, *95* (2), 451-60.
- (61) Izawa, M.; Kawasaki, T.; Hayakawa, Y. Cloning and Heterologous Expression of the Thioviridamide Biosynthesis Gene Cluster from *Streptomyces Olivoviridis*. *Appl. Microbiol. Biotechnol.* **2013**, *79* (22), 7110-7113.

Chapter 3: Chimeric Leader Peptides for the Generation of Non-Natural Hybrid RiPP Products³

3.1 Introduction

With the considerable success of natural products as drugs,¹⁻³ great efforts have been invested in using synthetic biology to engineer analogs of natural products with desired chemical and biological properties.⁴⁻⁶ One popular approach has been to combine enzymes from different biosynthetic pathways to generate new products in a process known as combinatorial biosynthesis.^{7, 8} Most commonly, this procedure was applied to polyketide synthase (PKS) and non-ribosomal peptide synthetase (NRPS) pathways where different domains or modules were swapped or deleted. This approach enabled alteration of the loaded amino acid in non-ribosomal peptides or the loading and/or extender unit and its subsequent oxidation state in polyketides.^{9, 10} Despite these feats, the approach has been largely limited to producing new variants of a given pathway rather than building entirely new products or pathways because of the incompatibility of interdomain linker regions, module interfaces, as well as other requirements.^{9, 10}

Ribosomally synthesized and posttranslationally modified peptides (RiPPs)¹¹ represent another group of natural products that provide an attractive starting point for engineering new molecules. The potential for engineering RiPPs derives from their leader peptide-guided biosynthetic logic and the reduced size of the biosynthetic gene clusters.¹² Despite their genetic simplicity, RiPPs are structurally and functionally diverse^{11, 13, 14} and their biosynthesis entails a well-orchestrated modification of a ribosomally produced precursor peptide (Figure 3.1a). RiPP biosynthetic proteins bind their respective precursor peptide(s) through specific recognition sequences (RS) typically located in the N-terminal leader region.¹² The majority of prokaryotic RiPP pathways rely on a ~90-residue PqqD-like domain, known as the RiPP recognition element (RRE),¹⁵ to engage the RS, but some notable examples of leader peptide-dependent enzymes lack this domain or do not use it.¹⁶⁻¹⁸ Once the leader peptide is bound, the C-terminal core region of the precursor peptide undergoes posttranslationally modification, sometimes extensively, to form azol(in)e heterocycles, lanthionine or sactonine crosslinks, D-amino acids, various types of macrocycles, and other modifications.¹¹ The modified core region ultimately becomes the natural product as the unmodified leader region is enzymatically removed during biosynthetic maturation. Beyond the primary, class-defining modifications installed by leader-dependent enzymes, RiPPs can also be endowed with multiple secondary modifications by “tailoring” enzymes. In most cases, these tailoring

³ This chapter is reproduced with permission from Burkhart, B.J.; Kakkar, N.; Hudson, G.A.; van der Donk, W.A.; Mitchell, D.A. Chimeric leader peptides for the generation of non-natural hybrid RiPP products. *ACS Cent. Sci.* **2017**, *6*, 629-638. Copyright 2017 American Chemical Society. Experiments were designed by B.J.B., N.K., W.A.v.d.D, and D.A.M. and performed by B.J.B., N.K., and G.A.H. B.J.B. wrote this work with contribution from N.K. for section 3.2.3 and 3.2.5. All coauthors reviewed and edited this work.

enzymes act on the core region independently of the leader peptide to increase the chemical diversity of the core peptide.¹¹

The physical separation of leader peptide binding and sites of modification allow RiPP enzymes to solve the paradoxical problem of being both specific for a substrate yet promiscuously act upon it.¹⁹⁻²⁵ Indeed, many natural RiPP pathways have highly variable core regions but retain nearly identical biosynthetic enzymes and leader peptides.²⁶⁻²⁹ RiPPs also increase structural diversity by acquiring new enzymes from different pathways to create natural “hybrid” RiPPs,³⁰ sometimes leading to entirely new classes as exemplified by the elaboration of linear azol(in)e-containing peptides into the cyanobactin and thiopeptide classes (Figure 3.2).^{11,31} A few reports have demonstrated artificial combinations of RiPP enzymes,³²⁻³⁶ but these examples have not led to combinations of modifications from different RiPP classes because they primarily swap enzymes between pathways that produce structurally related molecules.³²⁻³⁶ Ideally, more diverse combinations of structural modifications can be harnessed using RiPP enzymes, but no general approach has been reported for the production of peptides with posttranslational modifications from different RiPP classes.

In order to address this unmet challenge, we envisioned a “chimeric leader peptide” strategy that would enable the rational combination of RiPP modifications. In this method, RSs from two leader peptides are concatenated to form a new bifunctional leader peptide within a single precursor peptide (Figure 3.1b). With their native RSs present, we hypothesized that two primary modifying enzymes would bind to their respective regions on the chimeric leader peptide and sequentially install posttranslational modifications, resulting in a hybrid RiPP product. We demonstrate that, after optimizing the core sequence and properly designing the co-expression plasmids, this approach successfully combines a thiazoline-forming cyclodehydratase (HcaD/F) with primary modifying enzymes from several different RiPP classes including sactipeptides (AlbA) and class I and II lanthipeptides (NisB/C and ProcM, respectively; Figure 3.1c-e). We also demonstrate how leader-independent tailoring enzymes such as MibD and NpnJ_A (Figure 3.1f) can be included. Thus, this work lays a broad foundation for the generation of designer RiPP hybrids.

3.2 Results

3.2.1 Thiazoline-lanthipeptide (class I) hybrid

To probe the feasibility of a chimeric leader peptide strategy to mix-and-match disparate RiPP biosynthetic pathways (Figure 3.1b), we first sought to combine an azoline-forming cyclodehydratase (HcaD/F) with a lanthipeptide synthetase (NisB/C). HcaD/F is a thiazoline-forming cyclodehydratase³¹ (Figure 3.1d) and was chosen in part because the RS of the precursor peptide (HcaA) was previously defined.³⁷ The other enzyme, NisB/C, originates from the nisin biosynthetic gene cluster and was chosen

because of the large body of knowledge pertaining to how the peptide substrate (NisA) is recognized and processed (Figure 3.3).³⁸ As a class-defining enzyme, NisB/C forms the characteristic lanthionine (Lan) or methyllanthionine (MeLan) residues found in lanthipeptides through a two-step process.³⁸ First, NisB catalyzes dehydration of Ser/Thr residues to form the corresponding dehydroalanine (Dha) or dehydrobutyrine (Dhb). Next, NisC catalyzes a 1,4-nucleophilic addition of a Cys thiol to the C β of the dehydro amino acid to form a thioether ring (Figure 3.1e). Being a class I lanthipeptide dehydratase, NisB forms Dha/Dhb moieties in a tRNA-dependent manner while other classes of lanthipeptide dehydratases employ an ATP-dependent mechanism.³⁸⁻⁴⁰

A chimeric precursor peptide was constructed from the HcaA and NisA precursor peptides to serve as a potential substrate for both HcaD/F and NisB/C (Figure 3.4a). The sequence of this hybrid peptide (Hyp1.1) was designed to resemble key segments of each parent peptide. HcaD/F and NisB/C both have an RRE domain for leader peptide recognition and the RS residues of their substrates are known (Figure 3.4). To incorporate both the RS from HcaA and NisA, we replaced the non-essential C-terminal region of the HcaA leader peptide³⁷ with the essential region of the NisA leader peptide.^{41, 42} We next engineered the core region of the chimeric precursor peptide. Leader-dependent enzymes are known to require a certain minimum spacer region between the RS and the core peptide,^{36, 40, 41} whereas extended core sequences are generally accepted.⁴³⁻⁴⁶ Therefore, we first introduced a “GGRCG” motif, which derives from HcaA, at a position that would be appropriately distanced from the RS of HcaA to facilitate HcaD/F processing but too close to the NisA RS to allow for NisB/C processing. This motif was then followed by a segment of native NisA core sequence so that it would approximate the natural spacing from the NisA RS on the chimeric leader peptide (Figure 3.4a). Only twelve residues of the NisA core peptide were used to simplify analysis by mass spectrometry.

As diagrammed in Figure 3.4, the hybrid was tested by expressing HcaD/F, NisB/C, and His₆-Hyp1.1 in *Escherichia coli* (see Table 3.1 for plasmids). The modified peptide was affinity purified, digested with endoproteinase AspN, and analyzed via matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF-MS). A mixture of three products was evident. The most extensively processed peptide was dubbed Hyp1.1a and contained one thiazoline, two Dhb, one MeLan, and one Lan (Figure 3.4). To confirm the structure, we leveraged the fact that thiazolines are readily hydrolyzed to Cys under mild acid treatment with a concomitant gain of 18 Da. Treatment with iodoacetamide under reducing conditions, which would alkylate free Cys, yielded very little product suggesting that the thioether rings were formed in high yield.

While previous nisin biosynthetic manipulation introduced new tailoring enzymes from related pathways^{32, 33} and hybrids have been synthetically prepared,^{47, 48} the enzymatic incorporation of a non-native modification (thiazoline) has not been reported. Our data show the feasibility of combining

disparate RiPP pathways to generate new-to-nature hybrids by rationally designing a chimeric leader peptide. These results also demonstrate that two RiPP enzymes that compete for modification of the same residue (Cys) can still be rendered compatible. In this first example, discrete regions in Hyp1.1 were used to mimic the native peptides and a similar approach could be used to append different modified structures to the N-terminus of other lanthipeptides.

3.2.2 Thiazoline-sactipeptide hybrid

Building on these initial results, we next sought to combine HcaD/F with an enzyme from another RiPP class that also modifies Cys. For this, the radical S-adenosylmethionine (rSAM) enzyme, AlbA, from subtilosin biosynthesis was chosen (Figure 3.5). AlbA is oxygen-sensitive and forms thioether crosslinks between Cys thiols and the $C\alpha$ of an acceptor residue. These linkages are called sactionines and gave rise to the RiPP class of sactipeptides (Figure 3.1c, note the difference to $C\beta$ -linked Lan).^{49,50} AlbA and HcaD/F both harbor RREs,¹⁵ and hence this would be a second test of whether structurally similar binding domains on two enzymes that both act on Cys could be used to site-selectively modify a peptide substrate. Given the short natural leader peptide of SboA, the native substrate for AlbA, a simpler hybrid peptide (Hyp2.1) was designed that merely appended the minimal HcaA leader sequence to the N-terminus of SboA (Figure 3.6a). A C-terminal Arg was included in the core peptide to enhance MS detection. Because the core peptide of SboA has three Cys residues, we expected to see a mixture of modified products, which would allow us to assess competition between the two enzymes in the absence of any core peptide engineering.

The potential thiazoline-sactipeptide hybrid system was first tested by expressing AlbA, HcaD/F, and Hyp2.1 in *E. coli* (Table 3.1). Under high culture aeration (shaking at 250 rpm in baffled flasks), only cyclodehydration was observed. Upon reducing the culture aeration (shaking at 100 rpm in non-baffled flasks),⁵⁰ we observed either sactionine-containing (strong signal) or thiazoline-containing (weak signal) products. No observable products contained both modifications (Figure 3.7). Although both enzymes were active under lower aeration conditions, AlbA appeared to outcompete HcaD/F given the low intensity signals for cyclodehydrated products (Figure 3.7). Prior work indicates that both HcaD/F and AlbA are highly processive with their native precursor peptides (i.e. substrate processing is faster than intermediate release).^{37,51} We reasoned that intermediates with Hyp2.1 may also not be released during processing, thus preventing formation of a hybrid product. Additionally, if AlbA was acting faster because its native core sequence was used in Hyp2.1, thioether formation could preclude subsequent azoline formation.

In order to overcome this challenge, we enhanced the rate of azoline-formation by editing the core peptide to include a more native-like “RCGGC” motif resulting in a second hybrid peptide with four

Cys (Hyp2.2; Figure 3.6a). Presumably, the “RCGGC” motif would be more efficiently processed by HcaD/F while the remaining two Cys native to SboA would remain available for modification by AlbA. Upon co-expression of Hyp2.2 with the enzymes (Table 3.1), a mass was detected that was consistent with two sactionines and two dehydrations (Figure 3.6c). Mild acid treatment supported the presence of two thiazolines, as indicated by a gain of 36 Da. The presence of two sactionines was supported by lack of reactivity towards iodoacetamide. Analysis of high-resolution tandem MS spectra localized these modifications and corroborated the proposed structure (Figure 3.8). Additional evidence for the structure is provided by the following observations: (i) trypsin cuts very slowly at Lys2 near the first crosslink, presumably due to steric congestion (for structure, see Figure 3.6c), (ii) Arg12-related tryptic fragments were not detected (because the scissile bond was transformed to a thiazoline), and (iii) treatment with endoproteinase GluC did not result in digestion at Glu23 (Figure 3.9). Other low intensity signals consistent with a tri-thiazoline/mono-sactionine and tetra-thiazoline species were also detected, suggesting the hybrid pathway does not necessarily follow a strict processing order (Figure 3.6c).

Based on the results of Hyp2.1 and 2.2 (Figures 3.6 and 3.7), a change in the sequence context of a modifiable residue can change which RiPP enzyme will act at that position. Alteration of surrounding residues has been previously shown to enhance or diminish the processing rate of cyclodehydratases.^{25, 52, 53} The use of the SboA core sequence initially provided an advantage to AlbA. However, the processing efficiency of HcaD/F could be enhanced through insertion of a more native-like RCGGC motif, ultimately producing the desired hybrid product.

Because AlbA and HcaD/F were transplanted from unrelated pathways, we could not use evolved biosynthetic controls to ensure that both enzymes act in the desired order to produce a hybrid product. Such controls include using protein-protein interactions that channel the substrate from one enzyme to the next or tuning biosynthetic enzymes to favor core processing only in a specific order (e.g. LanP and PatG proteases preferentially remove recognition sequences after earlier modifications are completed).^{18, 38} In the absence of such controls that are difficult to engineer, a key design feature for controlling hybrid RiPP biosynthetic schemes is using the local sequence context to alter the rate of enzymes that compete for the same residues or install modifications that block the activity of subsequent enzymes.

3.2.3 Thiazoline-lanthipeptide (class II) hybrid

Thus far, three different primary RiPP enzymes have been shown to be tolerant toward substrates with non-native elements, but it has yet to be determined if a single enzyme pair will be generally able to process several different peptide sequences. For convenience, HcaD/F was again chosen for incorporation into another hybrid biosynthetic pathway to investigate this question. The class II lanthipeptide synthetase ProcM was selected as the second enzyme because it is mechanistically unique compared to the class I

NisB/C enzyme and provides another opportunity to test different enzyme types using the hybrid approach. ProcM is known to be highly substrate-tolerant, having 30 different native precursor peptides (ProcAs) and forming a structurally diverse set of lanthipeptides (Figure 3.10).^{28, 29, 54-56} Thus, we envisioned that it would be well suited for producing structurally diverse hybrid RiPPs. Furthermore, use of ProcM would extend the scope of the method because it does not contain an RRE;¹⁶ in fact, the site of leader peptide binding on ProcM is currently not known.

A chimeric leader peptide for ProcM and HcaD/F was designed analogously to the previous hybrids. The minimal HcaA leader peptide was joined with the C-terminal portion of ProcA2.8 because the majority of the ProcA N-terminus is dispensable for processing.²⁸ Accordingly, the final 19 residues of the ProcA2.8 leader peptide were placed after the first 25 residues of the HcaA leader region (Figure 3.11a). As first trial, the native core sequence of ProcA2.8 was used to create Hyp3.1, which contained two Cys. Co-expression of Hyp3.1, HcaD/F, and ProcM (Table 3.1) produced a peptide with two Lan; no thiazolines were detected (Figure 3.12). Thus, as with the sactipeptide hybrid Hyp2.1, use of a native ProcA core peptide in the chimeric peptide resulted in ProcM outcompeting HcaD/F. Rather than improving HcaD/F catalysis by altering the local sequence of the core peptide, we decided to investigate a different control mechanism. We placed the gene for Hyp3.1 on the same plasmid as that for HcaD/F and lowered the copy number of the plasmid encoding ProcM to tune the activity in the desired direction (Table 3.1). This arrangement afforded a hybrid product with one thiazoline, as indicated by mild acid treatment, and one Lan (Figure 3.12). Analysis of MS/MS fragmentation indicated that the thiazoline was located at the N-terminal Cys while the Lan was at the C-terminus (Figure 3.13). It appeared, however, that the thiazoline at core position 3 (residue 13 of the AspN-digested peptide) inhibited dehydration at Ser9, indicating that thiazolines may occasionally interfere with downstream hybrid processing. A low intensity ion consistent with a peptide containing two Lan was also observed, indicating the hybrid was not formed in quantitative yield. Substitution of the acceptor Ser9 to Ala in Hyp3.2 eliminated this side product (Figures 3.14 and 3.15). These data demonstrate how modest sequence alterations readily lead to the production of a single hybrid product without resorting to insertion of longer motifs.

The general tolerability of HcaD/F and ProcM encouraged us to design a new hybrid based on a peptide with a completely different core sequence, ProcA3.3 (Figure 3.10). Like Hyp3.1, this hybrid (Hyp4.1) appended the HcaA minimal leader sequence to the C-terminal portion of the ProcA3.3 leader peptide (Figure 3.11a). The native ProcA3.3 core sequence contains two Cys that normally form a dual MeLan “ring-within-a-ring” topology in prochlorosin 3.3 (Figure 3.10). Co-expression of Hyp4.1, ProcM, and HcaD/F resulted in two major products: Hyp4.1a and Hyp4.1b (Figure 3.16). Hyp4.1a contained one Dhb, one thiazoline, and a non-natural MeLan between Thr18 and Cys21 (Figure 3.17). Hyp4.1b

displayed modifications similar to native prochlorosin 3.3, suggesting that ProcM had processed the substrate before HcaD/F.

With HcaD/F and ProcM both able to act upon two different core peptides derived from native ProcA sequences, we next designed additional hybrids to direct enzymatic processing with greater precision. In the product of Hyp4.1, Cys21 forms a MeLan with Thr18, which places the thiazoline at Cys14 outside the MeLan macrocycle (Figure 3.17). We then wondered if we could create a hybrid in which a thiazoline would be within a (Me)Lan macrocycle. A Thr18Ala mutation was introduced (Hyp4.2; Figure 3.11a) to prevent Cys21 from generating a MeLan with Thr18 with the intention that it would then form a larger thioether ring with Thr11. After co-expression, Hyp4.2 was converted into a mixture of modified products (two thiazolines, thiazoline and Lan, or two Lan), indicating that each Cys could be modified by either enzyme (Figure 3.16). Although diversity generating systems can be advantageous, a mixture of products was not our desired outcome, and thus, we aimed to again tune the hybrid pathway to produce a single RiPP product. Accordingly, we designed Hyp4.3 in which we placed Asp before Cys21 to decrease the efficiency of HcaD/F processing at this position (i.e. Met20Asp substitution to Hyp4.2; Figure 3.11a). Expression of Hyp4.3 with HcaD/F and ProcM yielded primarily one product (Hyp4.3a) that contained one Dhb, one thiazoline, and one MeLan based on mild acid hydrolysis and iodoacetamide labeling (Figure 3.11c). MS/MS indicated that Cys21 was converted to MeLan, supporting our prediction that HcaF/D is less efficient in processing Cys with a preceding Asp residue (Figure 3.18; Figure 3.11c). The thiazoline was verified to be within the macrocycle through linearization by thermolysin digestion and subsequent MS/MS analysis (Figure 3.19). The stereochemistry of the MeLan was confirmed to be the native DL configuration by chiral GC/MS (Figure 3.20), providing support that this ring was likely still formed enzymatically.³⁸ The other low intensity ions observed from co-expression with Hyp4.3 appeared to be a mixture of isobaric species (Figure 3.21).

The previous example demonstrates that a thiazoline can be installed within a large, 11-residue MeLan macrocycle. We next investigated whether a thiazoline could be placed within a smaller (Me)Lan macrocycle. For this purpose, we generated another variant of the ProcA2.8 core (Hyp3.3; Figure 3.11a). We introduced an additional Arg-Cys motif that we envisioned would become a thiazoline within a 7-residue Lan macrocycle. Upon co-expression, the modified peptide had three dehydrations (-54 Da), two of which were susceptible to mild acid treatment, indicating the presence of two thiazolines and one Lan (Figure 3.11d). Lack of iodoacetamide labeling further indicated that all Cys were modified, and MS/MS fragmentation data support the proposed modified Hyp3.3 structure (Figure 3.22).

Overall, these experiments with HcaD/F and ProcM suggest that the chimeric leader peptide strategy will be amenable to creating libraries of modified peptides. After plasmid design optimization, ProcM and HcaD/F processed several peptides with altered core sequences and produced different

modified structures with different arrangements of thiazoline heterocycles and (Me)Lan rings. The ability to further tune hybrid biosynthesis toward a single desired product through simple core sequence changes was also demonstrated. HcaD/F-dependent azoline-formation can apparently be inhibited by employing a C-terminal Cys or by using an “Asp-Cys” motif, which reduces residue competition. Conversely, a simple “Arg-Cys” motif was robustly cyclodehydrated in Hyp3.3. This is notable as the earlier examples relied on larger motifs to introduce thiazolines. With the use of minimalistic motifs to tune enzyme activity, and the usage of enzymes with diverse sequence preferences, we anticipate a wide diversity of new-to-nature RiPPs will be accessible through the chimeric leader peptide engineering approach.

3.2.4 Hybrids with secondary tailoring enzymes

Although our chimeric leader peptide strategy was designed for combining leader-dependent “primary” RiPP biosynthesis enzymes, the method should also allow interfacing with secondary tailoring enzymes. Many RiPP pathways have tailoring enzymes that install functional groups critical for bioactivity.¹¹ We chose to combine the D-Ala-forming Dha reductase NpnJ_A from *Nostoc punctiforme* PCC 73102 (Figure 3.1f) with the HcaD/F and ProcM hybrid system.⁵⁷ NpnJ_A appears to prefer hydrophobic flanking residues, so a D2V/T3S double substitution was introduced into the Hyp4.2 peptide to create Hyp4.4 (Figure 3.23a).⁵⁷ Co-expression of Hyp4.4, HcaD/F, and ProcM initially produced three detectable species (Figure 3.23). The most intense ion (labeled as Hyp4.4a) was a hybrid containing one thiazoline, one MeLan, and one Dha (from dehydration of the newly introduced Ser3). Upon treating Hyp4.4a with NpnJ_A and NADPH in vitro, a new +2 Da product resulted (Hyp4.4b; Figure 3.23d), indicative of D-Ala formation. We also successfully combined HcaD/F with MibD, a flavin-dependent enzyme from *Microbispora* sp. 107891,⁵⁸ to produce a C-terminally decarboxylated, linear azoline-containing peptide (Figure 3.24).

These results indicate that our approach for creating hybrid RiPPs can be extended to leader-independent modifications, provided the proper modifiable sequence is present. Even though tailoring enzymes tend to bind directly to the core peptide, inserting short motifs into the core in conjunction with a chimeric leader peptide led here to three distinct posttranslational modifications from unrelated pathways.

3.3 Discussion

Rational design of biosynthetic pathways to produce custom molecules has long been a goal of combinatorial biosynthesis and natural product synthetic biology. Despite progress in understanding natural pathways, the chemical space that can be explored by current engineering techniques is limited compared to what is theoretically possible.^{9, 10} In this work, we demonstrate how a chimeric leader peptide has the potential to unlock the vast chemical space afforded by RiPPs. This concept was inspired

by how natural pathways can combine different enzymes (Figure 3.2) and represents a major step forward in RiPP engineering. Although prior work has demonstrated enzyme swapping within related pathways, such as for nisin and the cyanobactins,^{27, 30, 32-36} this study is to our knowledge the first demonstration of leveraging the recognition sequences and programmability of RiPP enzymes to mix-and-match the primary, class-defining structural features from unrelated classes. In total, four mechanistically unique enzymes were shown to be tolerant to manipulation of their leader peptides and insertion of non-native sequences.

The primary difficulty we encountered was ensuring that the enzymes acted at the desired locations and in the correct order. Being from unrelated pathways, there were no natural biosynthetic checkpoints to guide processing toward a single hybrid product, meaning the enzymes could compete for the same residue or act in a fashion that blocked the other enzyme. However, modest editing of the core sequence was sufficient to exploit the innate selectivity of an enzyme to yield the desired product. Future work with other enzymes that exhibit different substrate preferences may lead to a toolbox of enzymes for use in specific applications. Optimizing expression levels of each enzyme through plasmid design also adds another layer of control. Despite the successful use of these design principles in this work, there are likely some RiPP modifications that are inherently incompatible or too difficult to generally combine given their different structural requirements or limited promiscuity of their enzymes. Nonetheless, based on the enzymes combined here, we predict that many other hybrid RiPP combinations will be feasible.

In summary, chimeric leader peptides appear to be a broadly applicable and effective platform for creating hybrid RiPPs. In the cases shown here, the method afforded significant product yields (~1 mg/L of culture for the core peptide without any optimization) and employed a standard co-expression plasmid system. Moreover, RiPP enzymes can be chosen even if the enzymes act on the same residue. Our approach was successful regardless of native leader peptide length (8 vs. 60+), its binding affinity (~5 μ M for ProcA/ProcM vs. ~50 nM for HcaA/HcaF),^{37, 55} how it was bound (RRE or non-RRE binding site), or whether the binding site was previously known. Further, our approach is amenable to the inclusion of leader-independent enzymes. Accordingly, the chimeric leader peptide strategy holds much potential for combinatorial RiPP biosynthesis and opens the door for the generation of additional hybrid RiPP compounds.

3.4 Methods

3.4.1 General methods

Materials were purchased from Fisher Scientific, Gold Biotechnology, or Sigma-Aldrich unless otherwise noted. Tryptone and yeast extract were purchased from Dot Scientific. Restriction enzymes (REs), AspN endoprotease, Q5 polymerase, T4 DNA ligase, and deoxynucleotides (dNTPs) were purchased from New

England Biolabs (NEB). Oligonucleotide primers were synthesized by Integrated DNA Technologies (IDT) and are listed in Table 3.2. DNA spin columns were purchased from Epoch Life Sciences, and DNA sequencing was performed by the Roy J. Carver Biotechnology Center (UIUC). All polymerase chain reactions (PCRs) were conducted on an S1000 thermal cycler (Bio-Rad).

3.4.2 Cloning

Plasmids were constructed using REs and DNA ligase or by Gibson assembly (GA). For RE cloning, the target gene was amplified with Q5 polymerase, and after PCR cleanup using the QIAquick PCR Purification Kit Protocol, the amplicon was digested with restriction enzymes for ligation into a similarly digested and purified vector. For amplifications which gave more than one product, a gel extraction was performed. Ligations were performed with T4 DNA ligase at 23 °C for 1 h (100 ng vector and 3-fold excess insert). For GA, the standard protocol was followed with guidance from the NEBuilder Assembly tool (nebuilder.neb.com). Inserts were generated using Q5 polymerase and purified as above. Vector backbones were linearized using PCR or restriction enzymes and similarly purified.

Genes for the cyclodehydratase (HcaF and HcaD),³⁷ lanthipeptide synthetase (ProcM and NisB/C),⁵⁹ dehydroalanine reductase (NpnJ),⁵⁷ and flavoenzyme decarboxylase (MibD)⁶⁰ were cloned previously from their native organisms. All new plasmids used in this study are listed in Table 3.1 and were constructed using primers listed in Table 3.2. The subtilisin radical SAM gene (AlbA) and its precursor peptide (SboA) were cloned from *Bacillus subtilis* strain 168 genomic DNA.

The hybrid peptides were built from HcaA, SboA, or ProcA precursor peptides through overlap extension PCR and initially cloned into a modified pET28 vector with an N-terminal maltose-binding protein (MBP) affinity tag (pET28-MBP). This intermediate was then used to clone MBP-tagged hybrid peptides into duet vectors or to express the peptide for in vitro activity assays. The NisA-derived hybrid peptide was synthesized as a gblock (IDT) for subsequent cloning and was cloned directly into a duet vector in frame with a His₆ tag. Duet vectors pACYCDuet, pCDFDuet, pETDuet, and pRSFDuet were used for co-expression in *E. coli* (see Table 3.1), as has been well established,⁶¹ and genes were sequentially cloned into each multiple cloning site. To insert the cyclodehydratase HcaF and HcaD genes into a single multiple cloning site, the genes were taken together as a single amplicon from pETDuet-HcaF-HcaD.

3.4.3 Site-directed mutagenesis

Changes to the sequence of hybrid peptides were made with cloned PFU or KOD polymerase (a proofreading DNA polymerase isolated from *Thermococcus kodakaraensis*) and the site-directed mutagenesis (SDM) primers in Table 3.2 following the Agilent QuikChange protocol.

3.4.4 In vivo production of modified hybrid peptides

BL21(RIPL-DE3) chemically competent cells were co-transformed with pairs of duet vectors (Table 3.2) and selected with 100 µg/mL ampicillin for pETDuet, 50 µg/mL kanamycin for pRSFDuet or pET28, and 34 µg/mL chloramphenicol for pACYCDuet. Initial cultures were started from single colonies and grown in 10 mL of Luria–Bertani (LB) broth in 18 mm glass culture tubes at 37 °C supplemented with appropriate antibiotics for several hours, and then used to inoculate 0.5 or 1 L of the same media. Cells were grown to ~0.8 OD₆₀₀, put on ice for 10 min, and induced with 0.6 mM IPTG (isopropyl β-D-1-thiogalactopyranoside) at 22 °C except for co-expressions containing AlbA, which were shaken at 100 rpm in non-baffled flasks to reduce aeration (see Figure 3.5).⁶² After 18 h induction, cells were harvested at 4000 × g for 15 min at 4 °C, washed with TBS (Tris-buffered saline: 25 mM Tris pH 8.0, 150 mM NaCl), and centrifuged again at 4000 × g for 10 min at 4 °C. Cell pellets were stored for up to 1 week at -20 °C.

3.4.5 Protein expression and purification

BL21(DE3-RIPL) cells were transformed with plasmids containing the tagged proteins and selected using appropriate antibiotics. Starter cultures were grown in 10 mL LB with antibiotics and used to inoculate 1 L of similarly prepared media. After reaching an OD₆₀₀ of ~0.8, cells were induced overnight with IPTG. Next, cells were harvested by centrifugation at 4000 × g for 15 min at 4 °C, washed with TBS, and then subjected to centrifugation at 4000 × g for 10 min at 4 °C. Cell pellets were stored for up to 1 week at -20 °C. Proteins used for in vitro assays were purified using affinity chromatography via the MBP or His₆ tag.

MBP-tagged proteins were purified from *E. coli* cells using amylose resin (NEB). The resin was pre-equilibrated with 5 bed volumes of Buffer A [50 mM Tris pH 7.5, 500 mM NaCl, 2.5% (v/v) glycerol, 0.1% (v/v) Triton X-100]. Cell pellets were suspended in Buffer A and sonicated 3 times for 30 seconds at 10-13 W using a Misonix MICROSON XL2000 Cell Disruptor with 10 min incubations at 4 °C between each 30 second sonication period to prevent sample heating. The lysate was clarified by centrifugation at 34000 × g for 1 h on a Sorvall RC 6 plus centrifuge, applied to the equilibrated resin, and washed with 10 bed volumes of Buffer B [50 mM Tris pH 7.5, 400 mM NaCl, 2.5% (v/v) glycerol]. Protein was then eluted using Elution Buffer [50 mM Tris pH 7.5, 300 mM NaCl, 2.5% (v/v) glycerol, 10 mM maltose] and collected into Amicon Ultra 15 mL Filter centrifugal filters [30kDa NMWL (Nominal Molecular Weight Limit)]. The eluted protein was concentrated and a 10-fold buffer exchange was performed into Storage Buffer [50 mM HEPES, 300 mM NaCl, 2.5% (v/v) glycerol] through centrifugation at 4000 × g. The final protein concentration was calculated from the absorbance at 280 nm (NanoDrop 2000 UV-Vis Spectrophotometer) with the predicted protein extinction coefficient (<http://web.expasy.org/protparam/>) and the Bradford assay (Pierce™ Coomassie Protein Assay Kit). All proteins were stored at -80 °C. Note

that after initial column loading, all wash and storage buffers were supplemented with 1 mM tris-(2-carboxylethyl)-phosphine (TCEP).

His₆-tagged proteins were purified identically to the above protocol except the buffer composition differed as follows: Buffer A [50 mM Tris pH 7.5, 500 mM NaCl, 2.5% (v/v) glycerol, 15 mM imidazole], Buffer B [50 mM Tris pH 7.5, 400 mM NaCl, 2.5% (v/v) glycerol, 30 mM imidazole], and Elution Buffer [50 mM Tris pH 7.5, 300 mM NaCl, 2.5% (v/v) glycerol, 250 mM imidazole].

3.4.6 Purification of modified hybrid peptides

All hybrid peptides except Hyp1.1 were MBP-tagged and purified as described above. Typically, >20 mg of MBP-tagged peptide (corresponding to >1 mg of core peptide after affinity tag removal) was obtained from 1 L of culture. Hyp1.1 was expressed as a His₆-tagged peptide and purified by the following procedure. The harvested cells were resuspended in 25 mL of LanA Buffer 1 [6 M guanidine hydrochloride, 20 mM NaH₂PO₄, pH 7.5 at 25 °C, 500 mM NaCl, 0.5 mM imidazole], and lysed by sonication (50% amplitude, 1.0 s pulse, 2.0 s pause, 5 min). The sample was subjected to centrifugation (23,700 × g for 30 min at 4 °C) and the supernatant containing the His₆-tagged peptide was retained for further purification. The His₆-tagged peptide was then purified by immobilized metal affinity chromatography (IMAC) using 2-4 mL of His60 Ni Superflow Resin (Clontech). The supernatant was applied to the column and the column was washed with 2 column volumes of LanA Buffer 1 followed by 2 column volumes of LanA Buffer 2 [4 M guanidine hydrochloride, 20 mM NaH₂PO₄, pH 7.5 at 25 °C, 300 mM NaCl, 30 mM imidazole], and then eluted with 3 column volumes of LanA Elution Buffer [4 M guanidine hydrochloride, 20 mM Tris, pH 7.5 at 25 °C, 100 mM NaCl, 1 M imidazole]. The elution fractions containing the peptide were desalted by Vydac C4 SPE column and lyophilized and stored at -80 °C.

3.4.7 In vitro modification assays

In vitro assays were performed with purified proteins and peptides as described previously.^{37, 56, 57, 63} In general, peptides were supplied at 50 μM. Proteins were present at 5 to 20 μM and allowed to react for 16 h. Reactions were carried out in synthetase buffer [50 mM Tris pH 7.5, 125 mM NaCl, 20 mM MgCl₂, 1 mM TCEP]. However, assays with the dehydroalanine reductase (NpnJ_A) were performed in 50 mM HEPES pH 7.4 with 1.5 mM NADPH.

3.4.8 MALDI-TOF-MS

Matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS) was used to analyze posttranslational modifications based on their characteristic mass changes and reactivity towards mild acid or iodoacetamide treatment.⁶⁴ Upon in vitro or in vivo modification, peptides were digested with

a commercial protease (AspN, ArgC, proteomics grade trypsin, GluC, or thermolysin) following the manufacturer's protocol. Hydrochloric acid (200 mM) or iodoacetamide (20 mM at 8 pH) were individually added to an aliquot of the digestion and allowed to react for 1 h in the dark at 22 °C. Next, the samples were desalted with a ZipTip (Millipore) according to the manufacturer's instructions and eluted into 3.5 μ L of 70 % MeCN. Then, 0.5 μ L was applied to a MALDI target plate with 1 μ L of matrix and subsequently mixed until crystallized. For the Hyp2.1 and Hyp2.2 peptides, the ideal matrix solution was 20 mg/mL sinapinic acid (SA) with 10 mg/mL 2,5-dihydroxybenzoic acid (DHB) in 70 % MeCN. All other peptides were applied with a matrix solution of 20 mg/mL of α -cyano-4-hydroxycinnamic acid (CHCA) and 10 mg/mL DHB in 60% MeCN. The samples were analyzed using a Bruker Daltonics UltrafleXtreme MALDI-TOF mass spectrometer in reflector/positive mode using the default RP900_4500_Da.par method.

3.4.9 High performance liquid chromatography (HPLC) purification of modified hybrid peptides

A subset of peptides were desalted and purified (>75%) for MS/MS characterization. Hyp3.2 was purified for determination of the stereochemistry of its MeLan (see below). The separations were performed on an Agilent 1200 series HPLC fitted with a 10 \times 250 mm Betasil C18 column (Fisher Scientific). A gradient elution was used with Buffer A (10 mM ammonium bicarbonate) and Buffer B (10 mM ammonium bicarbonate in 90% MeCN) according to the following linear gradient combinations: at t=0 min, 6% B; t=3.0 min, 6% B; t=6.0 min, 15% B; t=21.0 min, 40% B; t=24.0 min, 100% B; t=33.0 min, 100% B; t=37.0 min, 6% B; t=47.0 min, 6% B. Hyp3.x peptides eluted around 11 min whereas Hyp4.x eluted around 16 min. Hyp2.2 was purified using a different linear gradient combination: at t=0 min, 6% B; t=3.0 min, 6% B; t=7.0 min, 23% B; t=23.0 min, 52% B; t=26.0 min, 100% B; t=32.0 min, 100% B; t=37.0 min, 6% B; t=45.0 min, 6% B. Hyp2.2 eluted at 20 min. After elution from the column, the MeCN portion of the solvent was removed in a vacufuge (Eppendorf), and the remaining liquid was lyophilized using a FreeZone 4.5 L benchtop freeze dry system (Labconco).

3.4.10 HRMS and MS/MS analysis

HPLC-purified peptides or peptides desalted using a ZipTip were diluted 1:1 using ESI mix (80% MeCN, 19% water, 1% formic acid). Samples were directly infused into a ThermoFisher Scientific Orbitrap Fusion ESI-MS using an Advion TriVersa Nanomate 100. The MS was calibrated and tuned with Pierce LTQ Velos ESI Positive Ion Calibration Solution (ThermoFisher). The MS was operated using the following parameters: resolution, 100,000; isolation width (MS/MS), 2 m/z; normalized collision energy (MS/MS), 35; activation q value (MS/MS), 0.4; activation time (MS/MS), 30 ms. Fragmentation was performing using collision-induced dissociation (CID) at 35%. Data analysis was conducted using the Qualbrowser application of Xcalibur software (Thermo-Fisher Scientific).

3.4.11 Determination of methyllanthionine stereochemistry

The stereochemistry of the MeLan in Hyp4.3a was determined according to previously reported methods,⁶⁵ with slight modification. Briefly, after HPLC, the peptide sample was dissolved in 3 mL of 6 M DCl in D₂O (used to distinguish non-enzymatic epimerization products upon GC-MS analysis) and placed in a sealed glass pressure tube. This mixture was heated to 110 °C for 24 h without stirring. After being allowed to cool, the solvent was removed *in vacuo*. In a separate round bottom flask, acetyl chloride (1.5 mL) was added dropwise to MeOH (5 mL) at 0 °C with stirring. This solution (4 mL) was then added to the hydrolysate, which was again heated to 110 °C for 1 h. After cooling, the solvent was removed, and the residue was taken up in 3 mL of CH₂Cl₂. The suspension was then cooled to 0 °C and pentafluoropropionic acid (1 mL) was added dropwise. Next, the sample was heated to 110 °C for 30 min and allowed to cool. The solvent was again removed before suspending the derivatized product in MeOH (200 µL) prior to GCMS analysis.

The derivatized samples were analyzed by GC-MS (Agilent HP 6890N) using an Agilent CP-Chirasil-L-Val column (25 m × 0.25 mm × 0.12 µm), with co-injection of previously prepared DL-MeLan and LL-MeLan standards for stereochemical verification.⁶⁶ Samples were applied to the column via split injection and run according to the following temperature method: 150 °C injection for 3 min, then raised to 200 °C by 3 °C /min and held (21.6 min total runtime). The MS was operated using selected ion monitoring at 379 Da for the detection of MeLan isomers.

3.5 Acknowledgements

This work was supported by the U.S. National Institutes of Health (GM097142 to D.A.M. and GM058822 to W.A.V.), the Chemistry-Biology Interface Training Grant Program (2T32 GM070421 to B.J.B.), the Robert C. and Carolyn J. Springborn Endowment (to B.J.B.), and a National Science Foundation Graduate Research Fellowship (DGE-1144245 to B.J.B.). We also thank Dr. Ian R. Bothwell for assistance in performing lanthionine stereochemistry experiments.

3.6 Figures

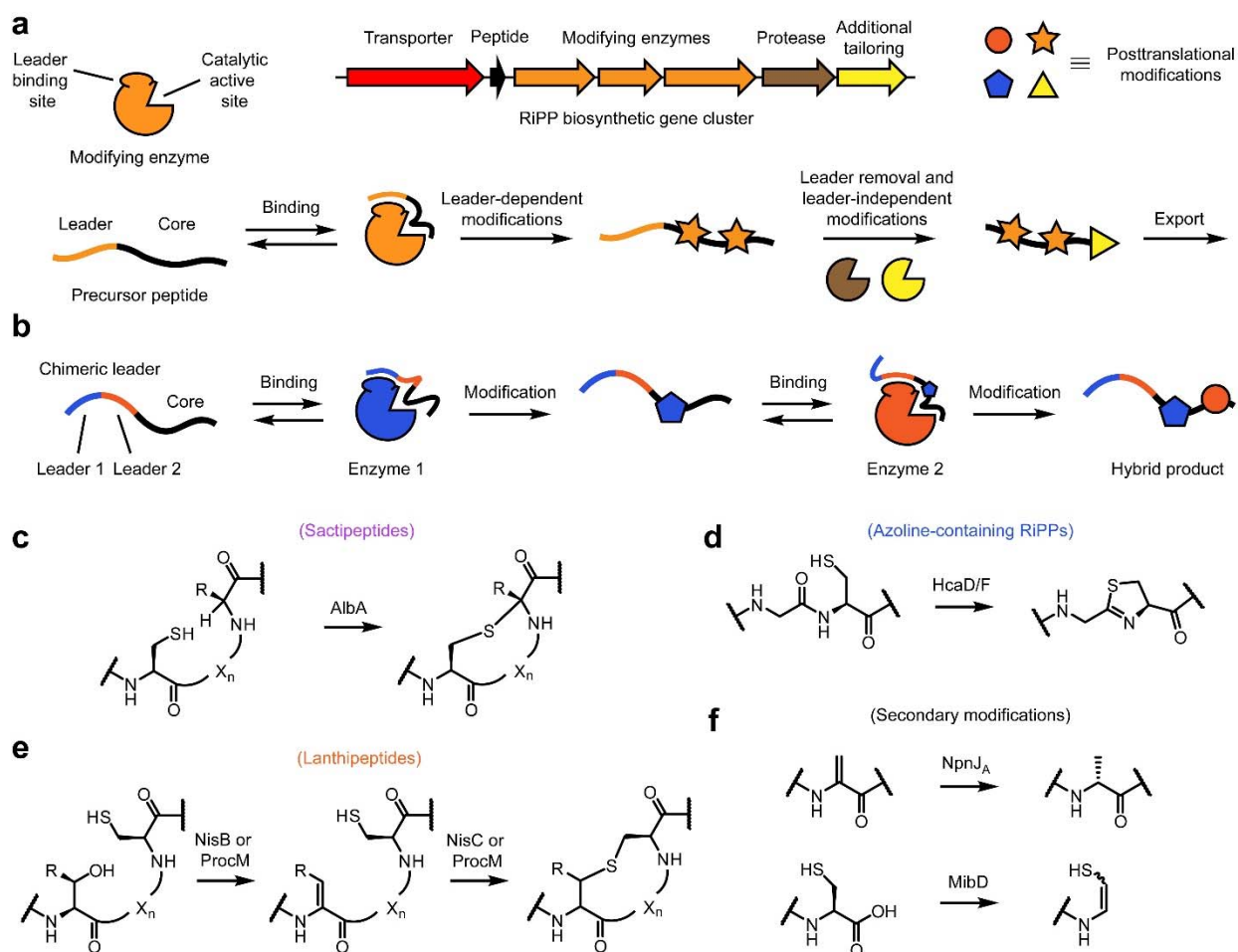


Figure 3.1 Overview of RiPP biosynthesis. (a) A generic RiPP gene cluster and the key maturation steps. (b) Chimeric leader peptide strategy for combining modifying enzymes from unrelated pathways. Examples of primary RiPP modification enzymes combined in this work include: (c) C α -crosslinked sactipeptides, (d) azoline-containing RiPPs and (e) C β -crosslinked lanthipeptides. (f) Examples of secondary RiPP modification enzymes used in this work to install D-Ala residues and decarboxylate C-terminal Cys residues.

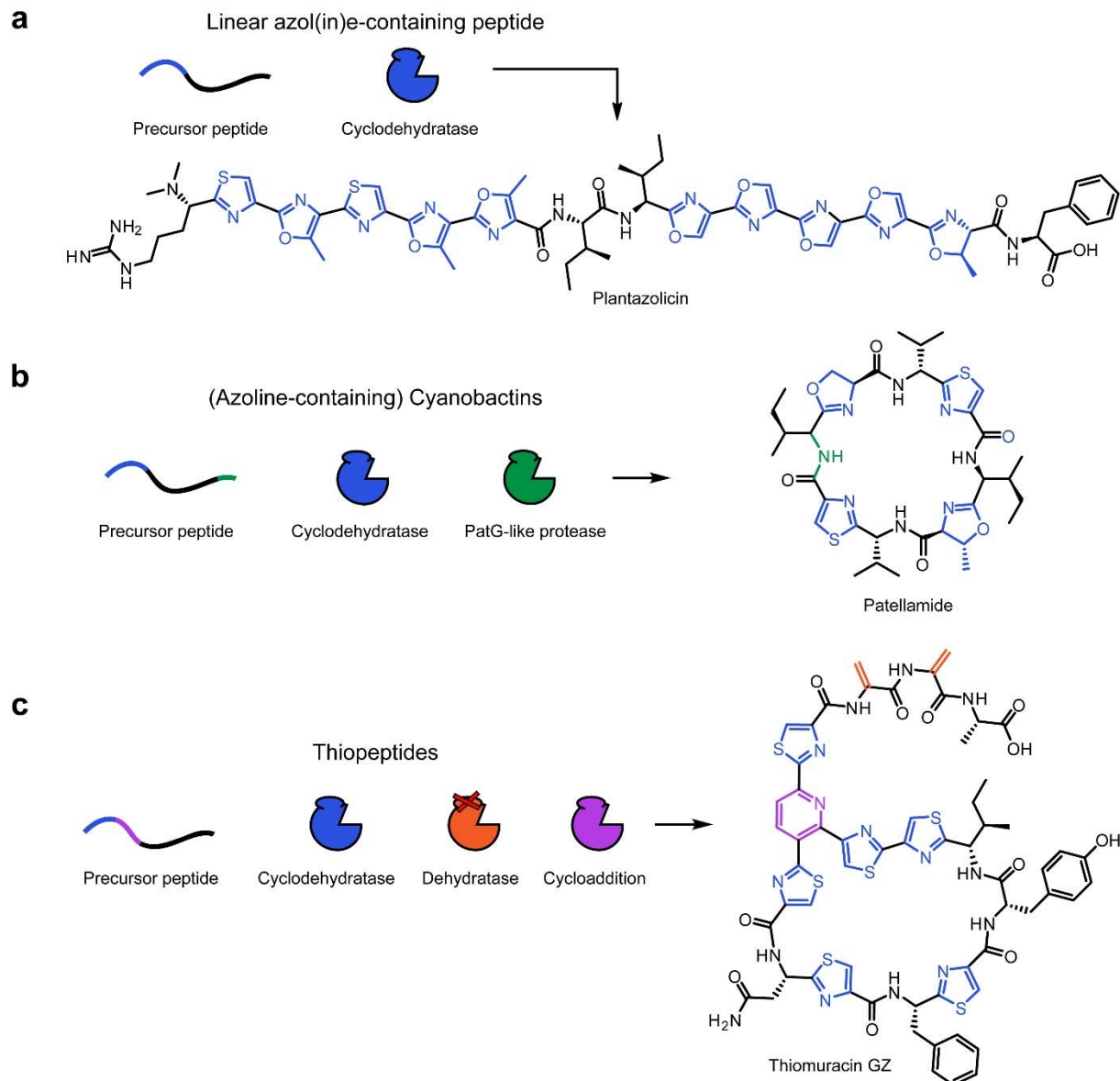


Figure 3.2 Natural “hybrid” RiPP biosynthetic gene clusters. (a) For linear azol(in)e-containing peptides, such as plantazolicin, the cyclodehydratase is the class-defining modifying enzyme. A tailoring enzyme also di-methylates the N-terminus upon leader peptide proteolysis. (b) Cyanobactins are defined by the presence of a PatG-like protease which recognizes a C-terminal motif to cyclize peptides after removal of the N-terminal leader region (by a PatA-like N-terminal protease). However, most cyanobactin gene clusters have acquired a cyclodehydratase and thus install two primary modifications (azolines and N-to-C macrocyclization). (c) Thiopeptides have three major modifications: azol(in)es, dehydrated amino acids, and a 6-membered nitrogen-containing heterocycle (here, pyridine). Because the cyclodehydratase and dehydratase can be found in other RiPP classes, the cycloaddition enzyme is the class-defining feature of

Figure 3.2 (Continued)

thiopeptides. In contrast to RRE-containing and leader peptide-dependent lanthipeptide (LanB) dehydratases,⁶⁷ the related thiopeptide dehydratases, which also contain an RRE, are leader-independent (red “x” in figure).⁶⁸

It is notable that RiPP pathways employ multiple enzymes that are dependent on certain recognition motifs and that these motifs tend to occupy distinct parts of the peptide (see panel b and c).^{24, 27, 68} Leader proteases could also be considered leader-dependent and similarly tend to bind a part of the leader not used by other enzymes.^{24, 27} Nature might have evolved these physically separate binding sites to improve biosynthetic efficiency. Regardless, it appears that the recognition sequences are “plug and play” and simply need to be inserted into an existing peptide for a new enzyme to act on it.

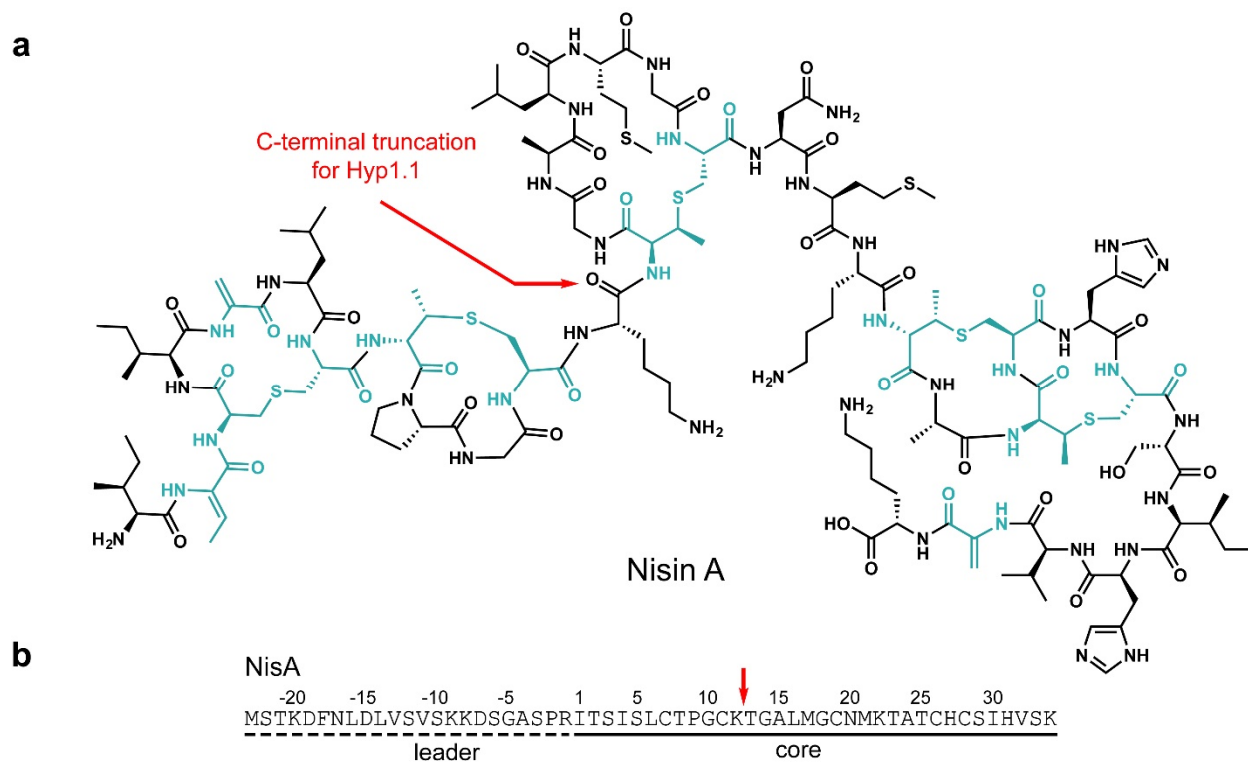


Figure 3.3 Structure and precursor peptide of nisin. (a) The structure of nisin A is shown with a red arrow indicating where Hyp1.1 was truncated to simplify MS/MS analysis of the modified core peptide. (b) The NisA precursor peptide sequence is shown with marked leader and core regions. The C-terminal region of the leader peptide is not required for binding or processing.^{41, 42}

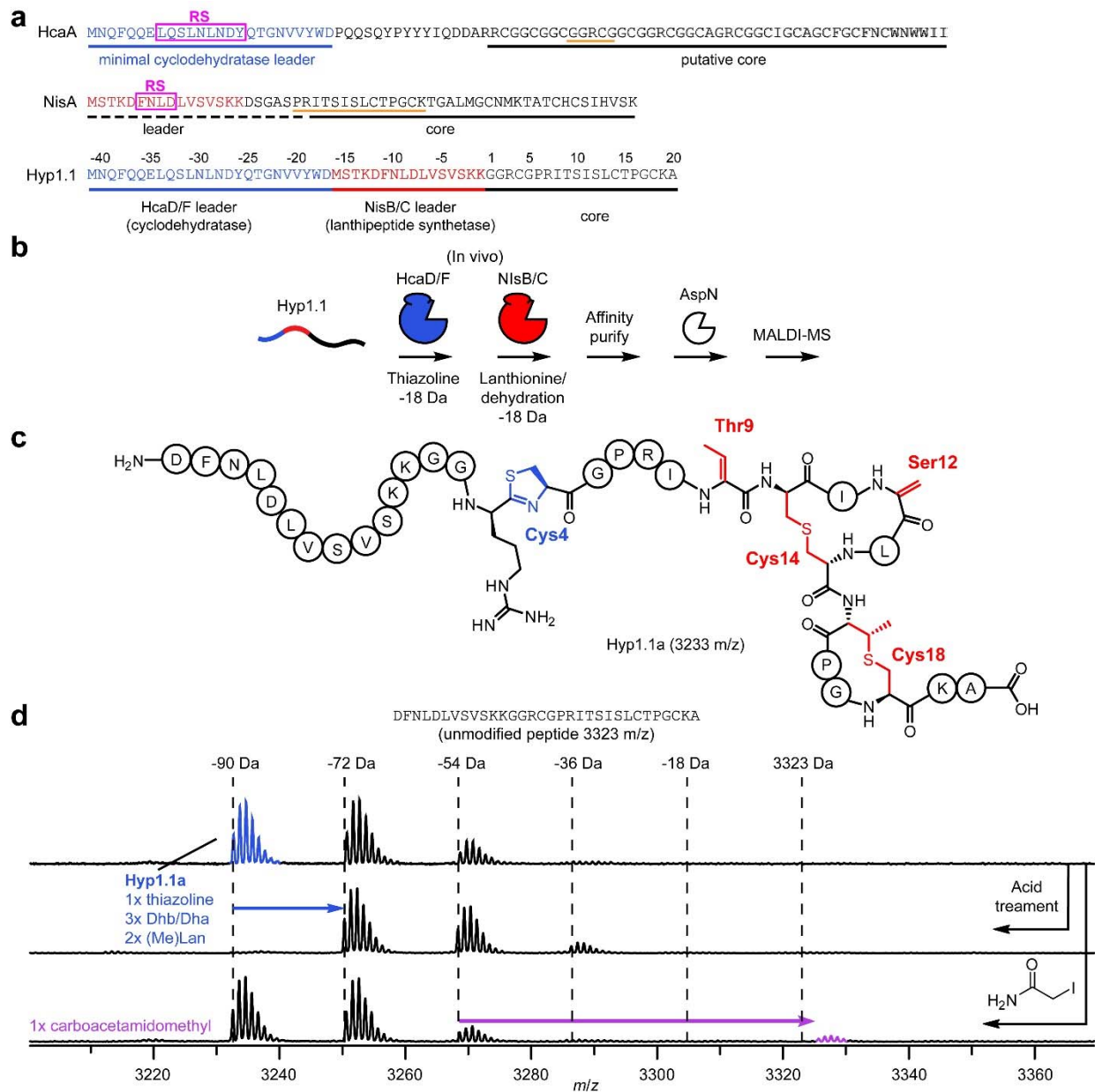


Figure 3.4 Production of a thiazoline-lanthipeptide (class I) hybrid. (a) Design of the Hyp1.1 sequence from HcaA and NisA precursor peptides. Portions of the leader peptides that were combined are colored. The recognition sequences (RSs) bound by the RREs of the cognate enzymes are indicated with a pink box. The orange underlined regions were combined to generate the chimeric core region. (b) Overview of experimental procedure. (c) Structure of thiazoline-lanthipeptide (class I) hybrid Hyp1.1a upon AspN digestion. Thiazolines are blue while dehydrations and (Me)Lan are red. Numbering is based on the core position, not peptide length after digestion. (d) The mass of Hyp1.1a (MALDI-TOF-MS) is consistent with one thiazoline, two dehydrations, and two (Me)Lan. These modifications are supported by acid hydrolysis (blue arrow, +18 Da) and by lack of labeling by iodoacetamide (green arrow, +57 Da).

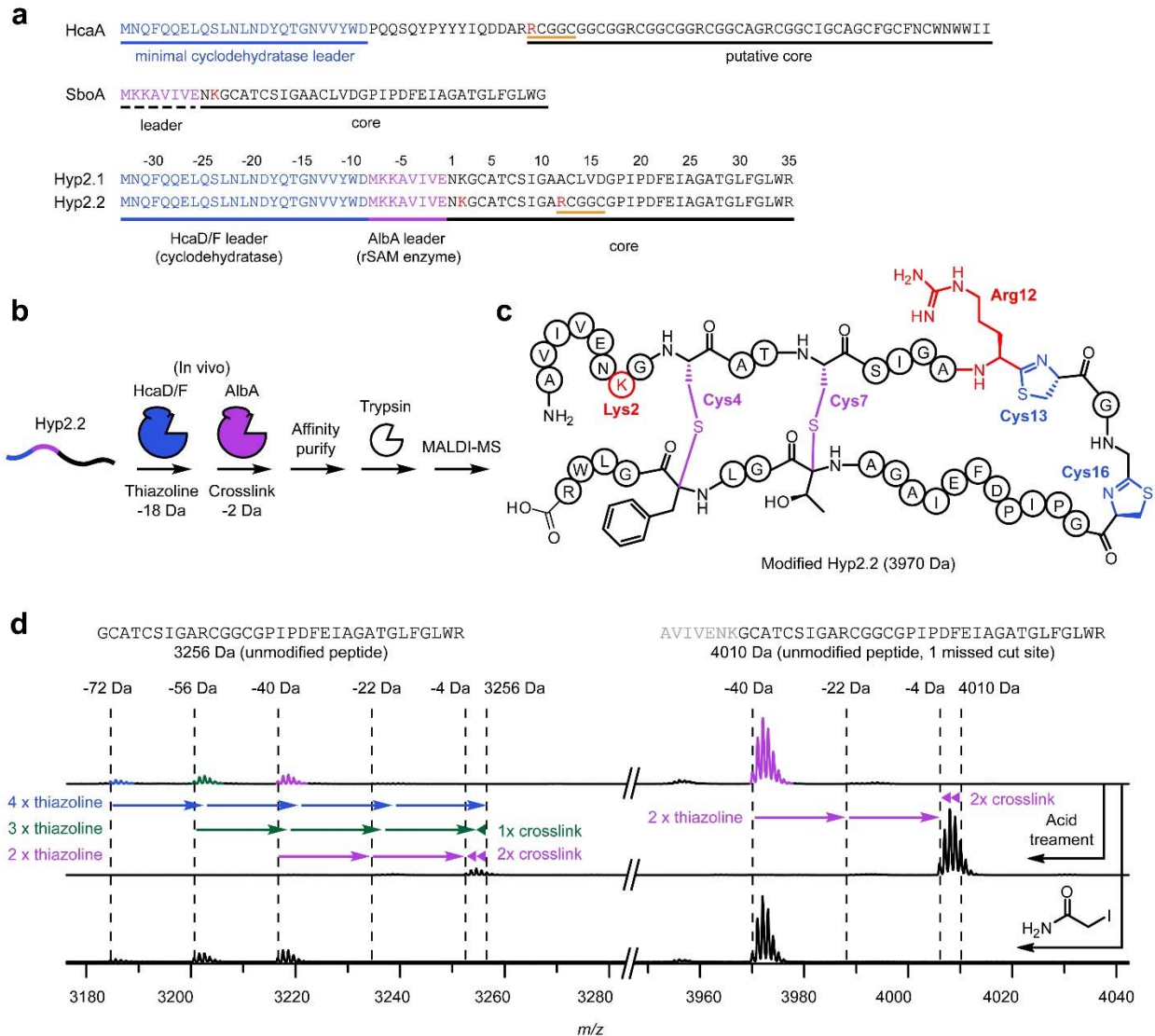


Figure 3.6 Production of a thiazoline-sactipeptide hybrid. (a) Design of hybrid peptides. Orange underlining indicates edited sequences. (b) Overview of the experimental procedure for combination of HcaD/F and AlbA. (c) Structure of thiazoline-sactipeptide hybrid (modified Hyp2.2). Numbering is based on the core position, not peptide length after digestion. (d) The mass of the modified peptide is consistent with two thiazolines and two sactioinone linkages. The two sets of peaks correspond to protease digestion at different sites (gray font). The structure is supported by acid hydrolysis, lack of labeling by iodoacetamide, and resistance to proteolytic digestion (see also Figure S6). Arrows and triangles represent +18 and -2 Da, respectively. Red coloring indicates positions resistant to trypsin digestion.

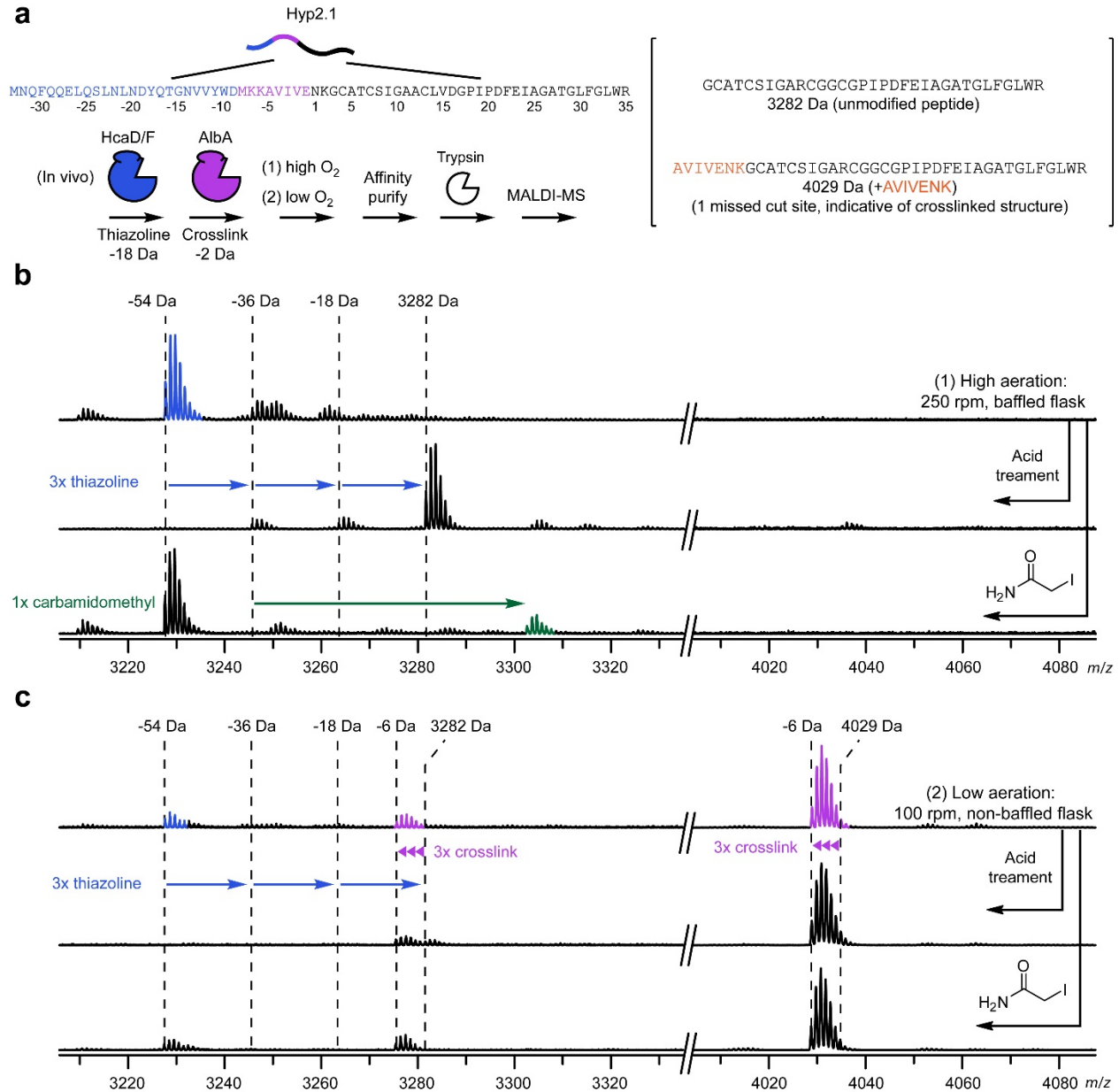


Figure 3.7 Hyp2.1 co-expression with HcaD/F and AlbA at high and low aeration. (a) Overview of experimental procedure. Hyp2.1 was expressed in *E. coli* with HcaD/F (pET-2.1-FD) and AlbA (pACYC-AlbA) at high or low aeration.^{50, 62} (b) Under high aeration (250 rpm, baffled flasks), only thiazoline-containing products were detected. The primary product contained 3 thiazolines with a minor 2 thiazoline-containing species also detected based on acid hydrolysis and iodoacetamide labeling. (c) With low aeration (100 rpm, non-baffled flask), the predominant signal (4023 Da) corresponds to the peptide with three C α -S crosslinks based on lack of iodoacetamide labeling, but a low intensity thiazoline-containing species was observed at lower mass range because trypsin cuts more readily at Lys2 in the non-crosslinked products.

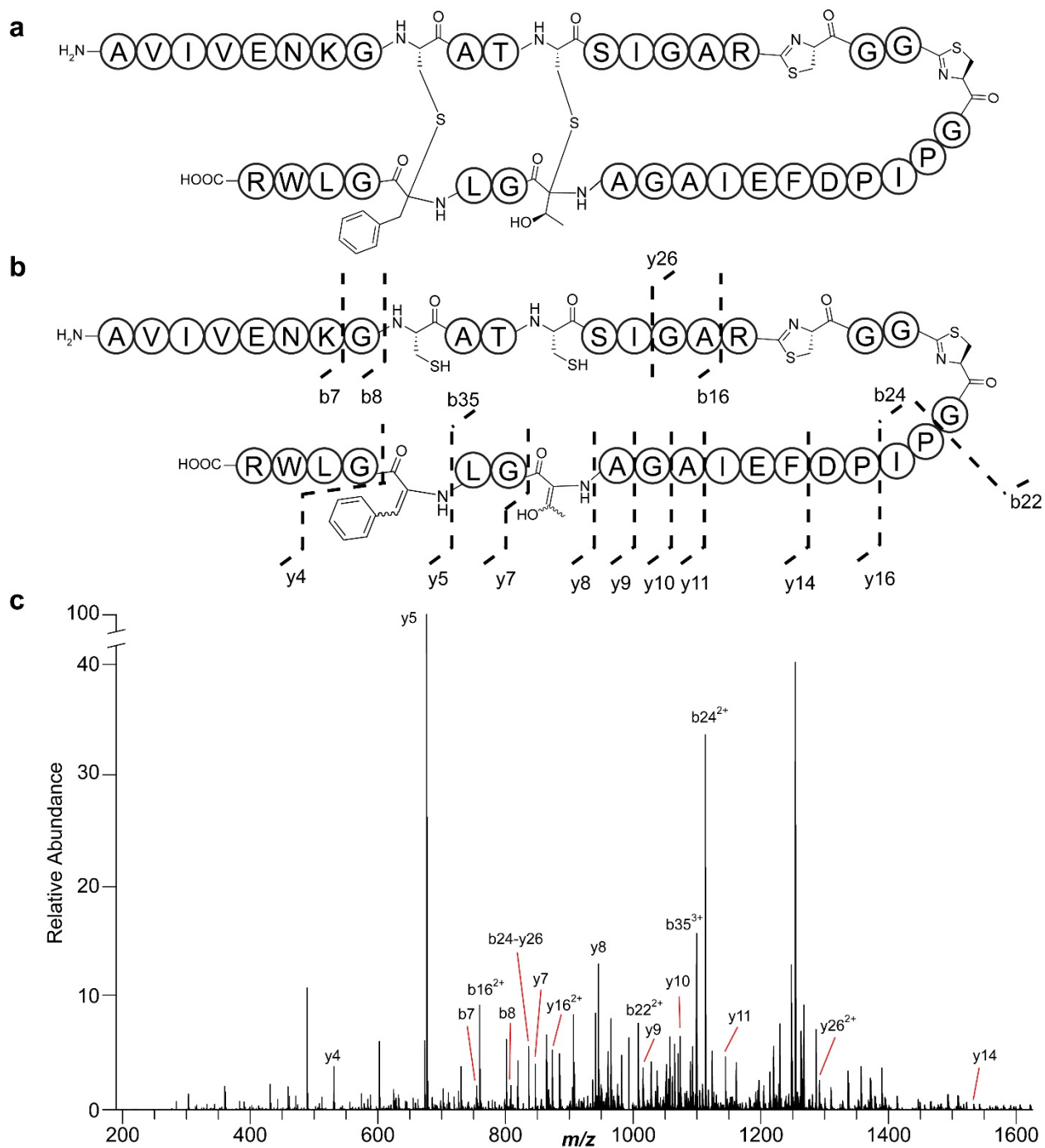


Figure 3.8 MS/MS of thiazoline-modified sactipeptide Hyp2.2. (a) Structure of modified Hyp2.2. (b) Even at low collision energies, thioether crosslinks (thioamidals) are known to undergo elimination to form the corresponding dehydroamino acid.⁵⁰ This enables localization of the crosslinked residues by the characteristic -2 Da at the acceptor amino acid. Fragments identified in panel are mapped onto the structure. (c) MS/MS fragmentation spectrum for Hyp2.2 annotated with identified ions. (d) Table of predicted and observed m/z values for ions observed in panel c (next page).

d

| Ion | Theoretical mass (Da) | Observed mass (Da) | Error (ppm) |
|----------------------|-----------------------|--------------------|-------------|
| b7 ⁺ | 754.4458 | 754.4453 | 0.6 |
| b16 ²⁺ | 759.3762 | 759.3739 | 3.1 |
| b22 ²⁺ | 1007.9575 | 1007.9572 | 0.4 |
| b24 ²⁺ | 1113.0262 | 1113.0250 | 1.1 |
| b24-y26 ⁺ | 836.3654 | 836.3649 | 0.6 |
| b35 ³⁺ | 1098.8557 | 1098.8551 | 0.5 |
| y4 ⁺ | 531.3038 | 531.3035 | 0.5 |
| y5 ⁺ | 676.3566 | 676.3562 | 0.6 |
| y7 ⁺ | 846.4621 | 846.4615 | 0.7 |
| y8 ⁺ | 945.4941 | 945.4935 | 0.7 |
| y9 ⁺ | 1016.5312 | 1016.5304 | 0.8 |
| y10 ⁺ | 1073.5527 | 1073.5522 | 0.5 |
| y11 ⁺ | 1144.5898 | 1144.5888 | 0.9 |
| y14 ⁺ | 1533.7849 | 1533.7836 | 0.8 |
| y16 ²⁺ | 873.4359 | 873.4353 | 0.7 |
| y26 ²⁺ | 1291.1150 | 1291.1135 | 1.1 |

Figure 3.8 (Continued)

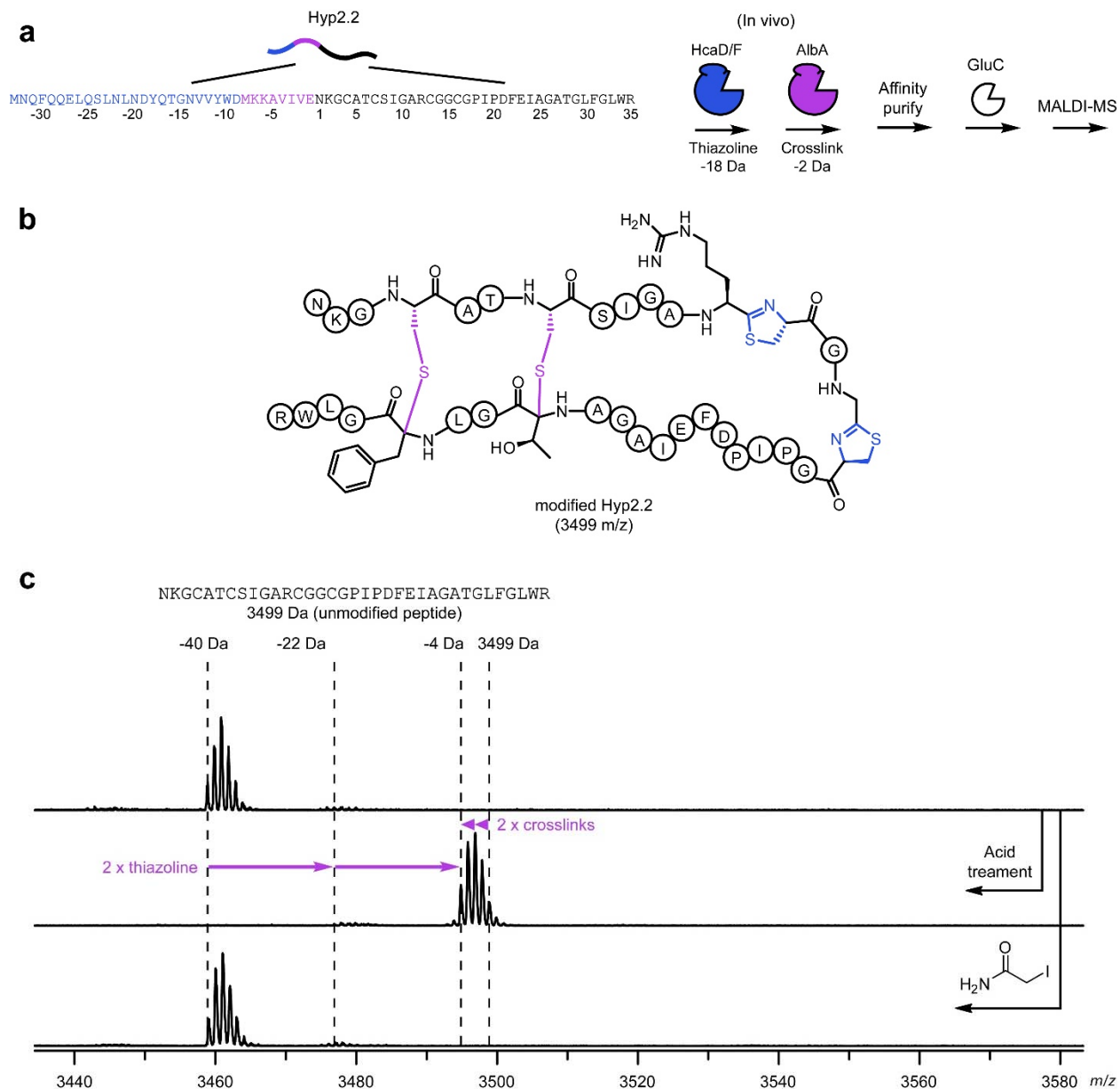


Figure 3.9 GluC digestion of thiazoline-modified and thioether-crosslinked Hyp 2.2. (a) Experimental overview. Hyp2.2 was expressed with HcaD/F (pET-2.2-HcaD/F) and AlbA (pACYC-AlbA) in *E. coli*. (b) Deduced structure of modified Hyp2.2 after treatment with GluC. (c) MALDI-TOF-MS analysis of modified peptide after GluC treatment and chemical derivatization. No digestion occurred at the internal Glu residue of the crosslinked peptide, supporting the proposed topology.

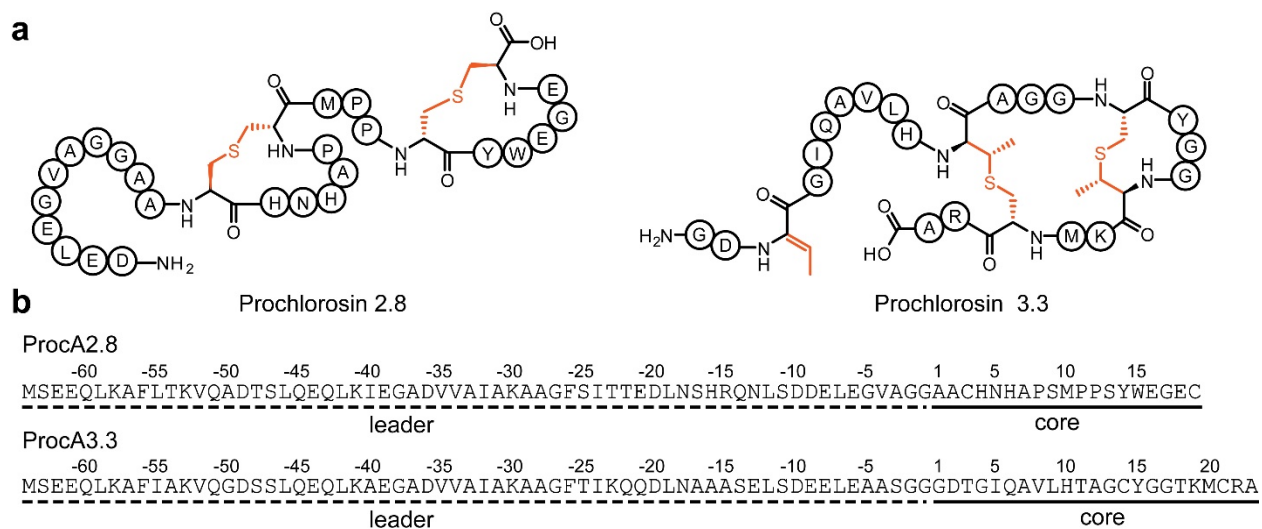


Figure 3.10 Structures and precursor peptides of two prochlorosins. (a) Structure of prochlorosin 2.8 and 3.3 with posttranslational modifications catalyzed by ProcM colored orange. (b) ProcM has over 30 different precursor peptides.²⁹ The sequence of the ProcA2.8 and ProcA3.3 precursor peptides are displayed with their leader and core regions. Only the C-terminal portion of these peptides are required for processing by ProcM,²⁸ and in general LanM enzymes appear to bind the region of the leader peptide closest to the core peptide.⁶⁹

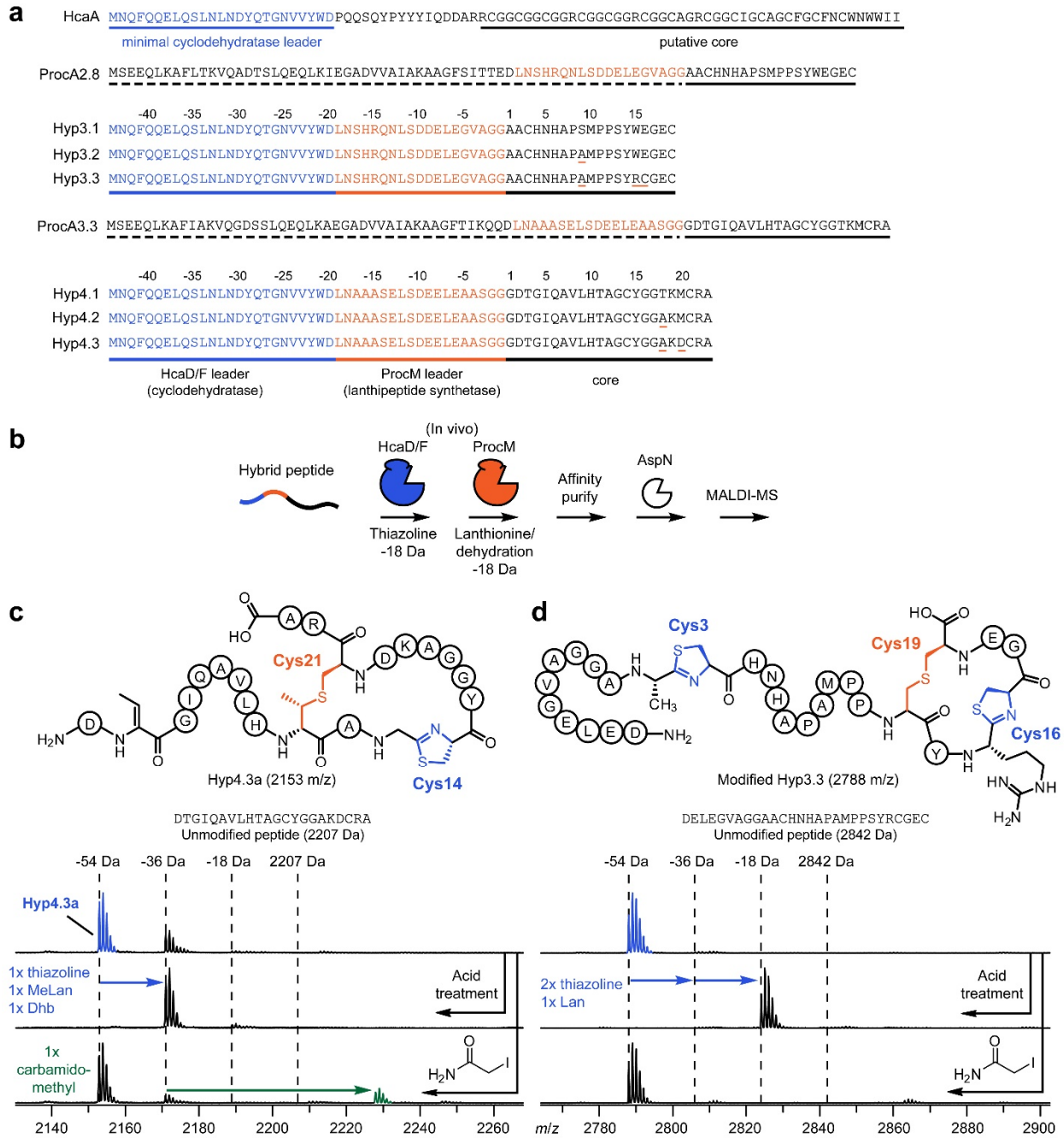


Figure 3.11 Production of hybrid thiazoline-lanthionine (class II) hybrids. (a) Design of hybrid peptides. Orange underlining indicates edited positions of the core peptide. The dashed line indicates the leader peptide of the ProcA2.8 and Proc3.3. (b) Experimental overview for combination of HcaD/F and ProcM. (c) Deduced structure and MALDI-TOF-MS support of posttranslational modifications for Hyp4.3. Numbering is based on the core position, not peptide length after digestion. (d) Same as in panel c but for Hyp3.3.

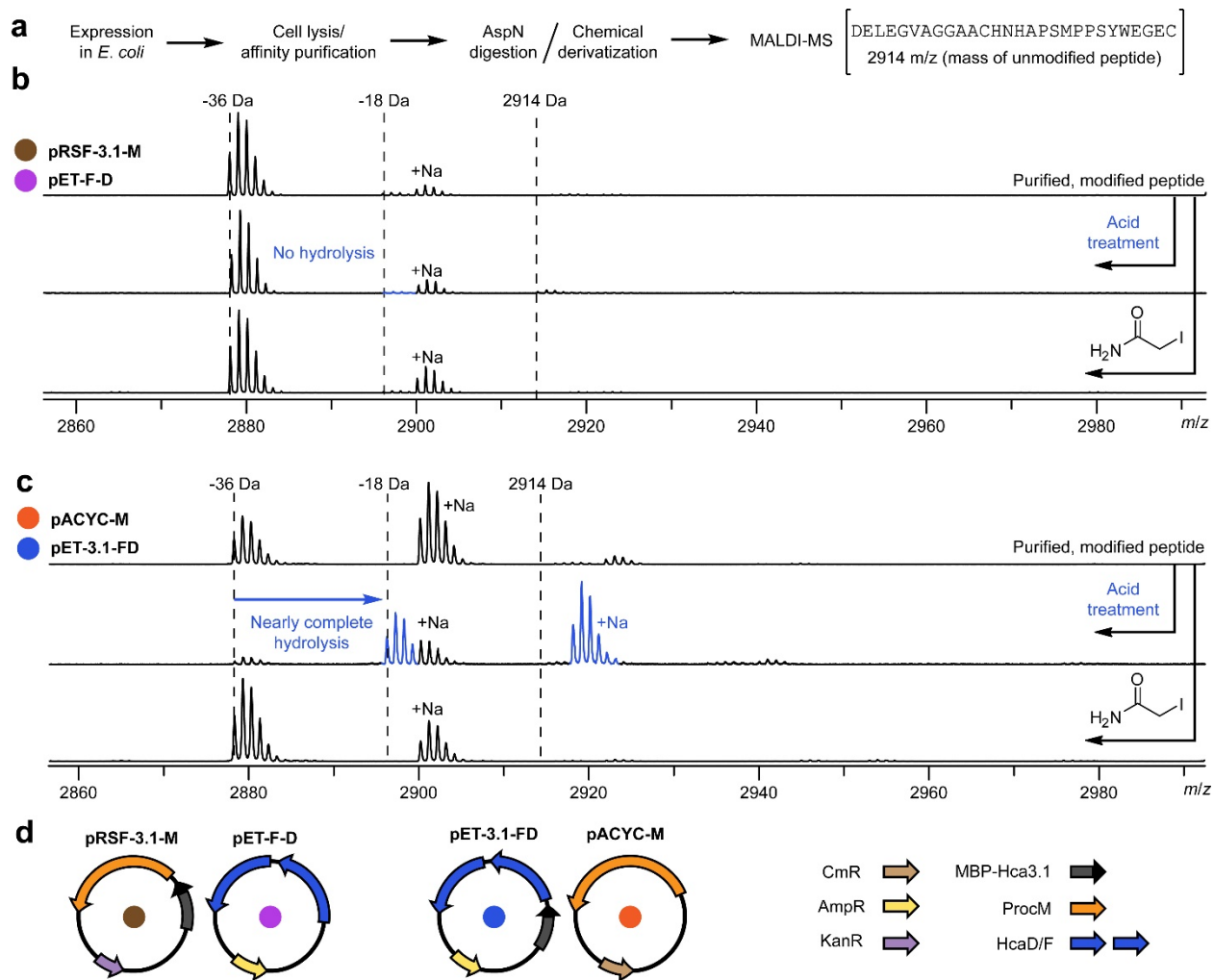


Figure 3.12 Optimization of HcaD/F-ProcM co-expression and plasmid design. (a) Experimental overview. (b) Expression of Hyp3.1 and ProcM on the same plasmid (pRSF-3.1-M) with HcaD/F (pET-F-D) resulted in only lanthionine formation as assessed by MALDI-TOF-MS. (c) Placing ProcM on a lower copy number plasmid (pACYC) and moving HcaD/F to the same plasmid as Hyp3.1 gave primarily a thiazolinone-lanthionine product; however, a minor two lanthionine product was also detected (species at -36 Da that remains after acid treatment). (d) Legend and design of plasmids constructed for this experiment (Table 3.1). CmR, AmpR, and KanR are chloramphenicol, ampicillin, and kanamycin resistance markers.

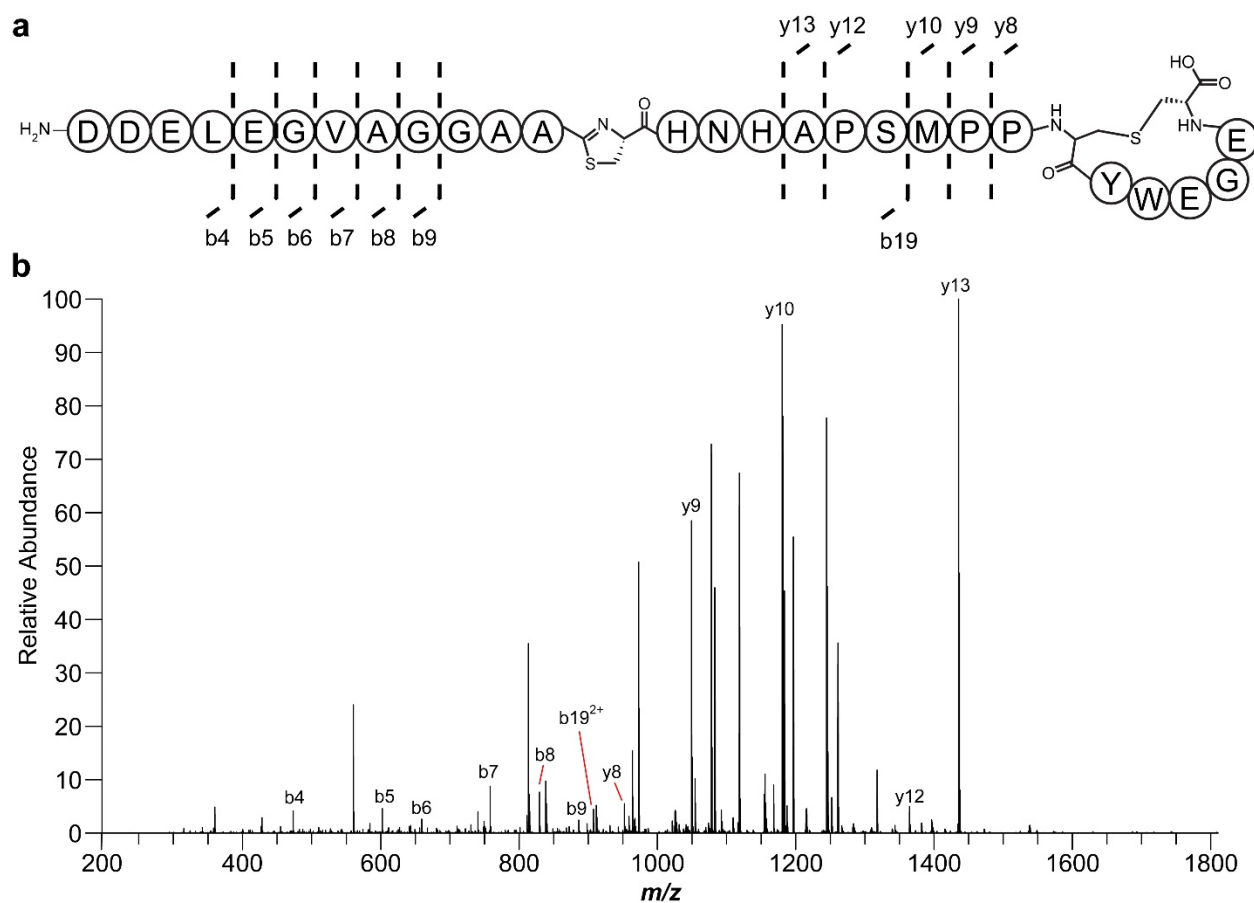


Figure 3.13 MS/MS fragmentation of thiazoline/lanthionine in modified Hyp3.1. (a) Structure of the thiazoline-lanthionine modified Hyp3.1 product. Although Ser9 is normally dehydrated by ProcM, thiazoline formation may inhibit dehydration at this site. Dashed lines indicate identified fragments in the mass spectrum. (b) MS/MS fragmentation spectrum for modified Hyp3.1 annotated with identified ions. (c) Table of predicted and observed m/z values for ions observed in panel b (next page).

C

| Ion | Theoretical mass (Da) | Observed mass (Da) | Error (ppm) |
|-------------------|------------------------------|---------------------------|--------------------|
| b4 ⁺ | 473.1878 | 473.1911 | 6.8 |
| b5 ⁺ | 602.2304 | 602.2346 | 7.0 |
| b6 ⁺ | 659.2519 | 659.2566 | 7.1 |
| b7 ⁺ | 758.3203 | 758.3256 | 7.0 |
| b8 ⁺ | 829.3574 | 829.3631 | 6.8 |
| b9 ⁺ | 886.3789 | 886.3848 | 6.7 |
| b19 ²⁺ | 907.3816 | 907.3879 | 6.9 |
| y8 ⁺ | 952.3505 | 952.3570 | 6.8 |
| y9 ⁺ | 1049.4033 | 1049.4106 | 6.9 |
| y10 ⁺ | 1180.4438 | 1180.4515 | 6.5 |
| y12 ⁺ | 1364.5286 | 1364.5373 | 6.4 |
| y13 ⁺ | 1435.5657 | 1435.5758 | 7.0 |

Figure 3.13 (Continued)

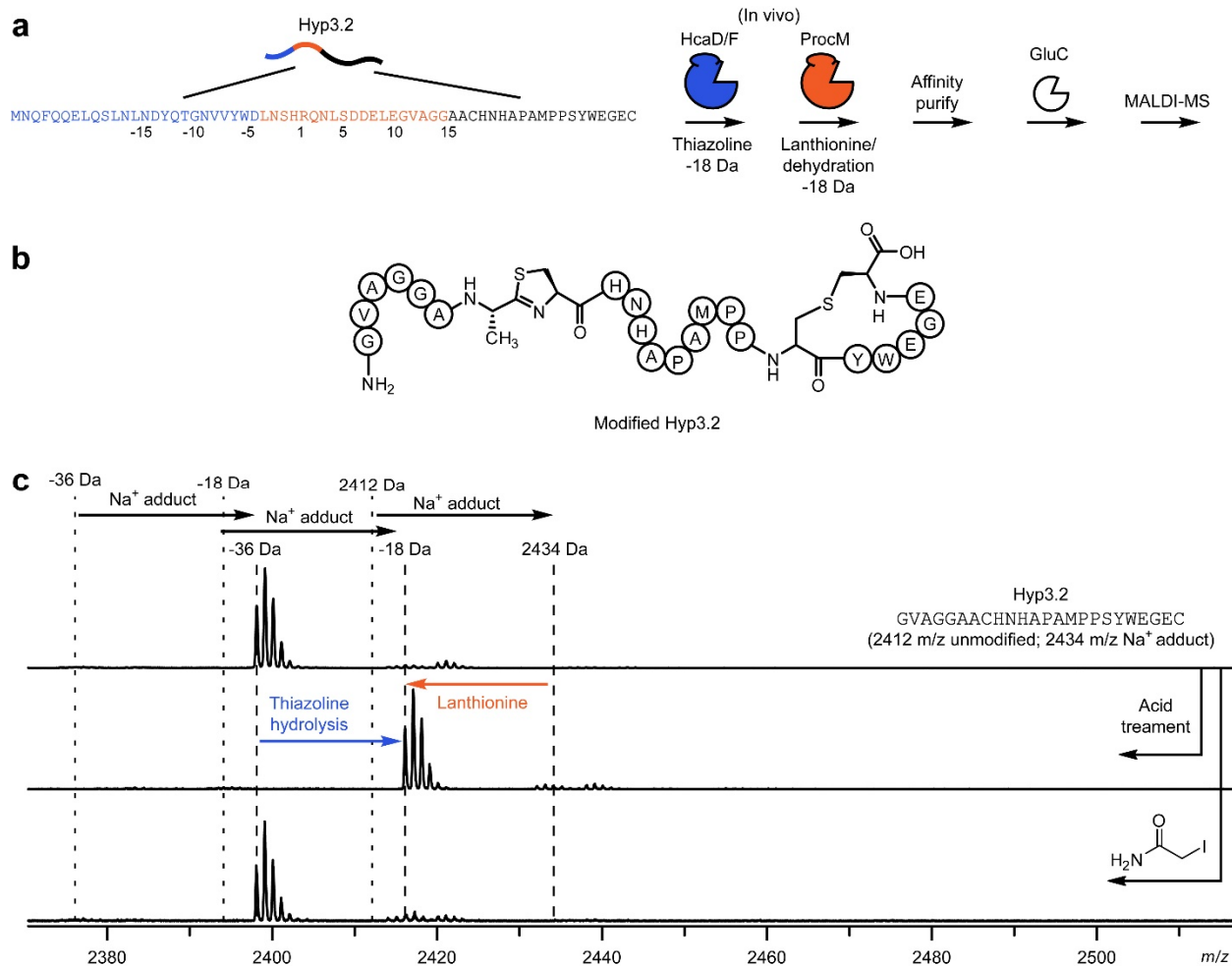


Figure 3.14 MALDI-TOF-MS analysis of thiazoline/lanthionine modified Hyp3.2. Compared to Hyp3.1, Hyp3.2 carries a Ser9Ala substitution. (a) Experimental overview. Hyp3.2 was co-expressed with HcaD/F (pET-3.2-FD) and ProcM (pACYC-M) in *E. coli*. (b) Deduced structure of modified Hyp3.2. (c) MALDI-TOF-MS analysis confirms the presence of a thiazoline based on the +18 Da shift upon mild acid treatment (blue arrow). The lack of iodoacetamide labeling indicates that the other Cys formed a lanthionine (orange arrow). All ions here are sodiated. The lack of GluC digestion at the two Glu near the C-terminus and the MS/MS data in Figure 3.15 support the proposed structure.

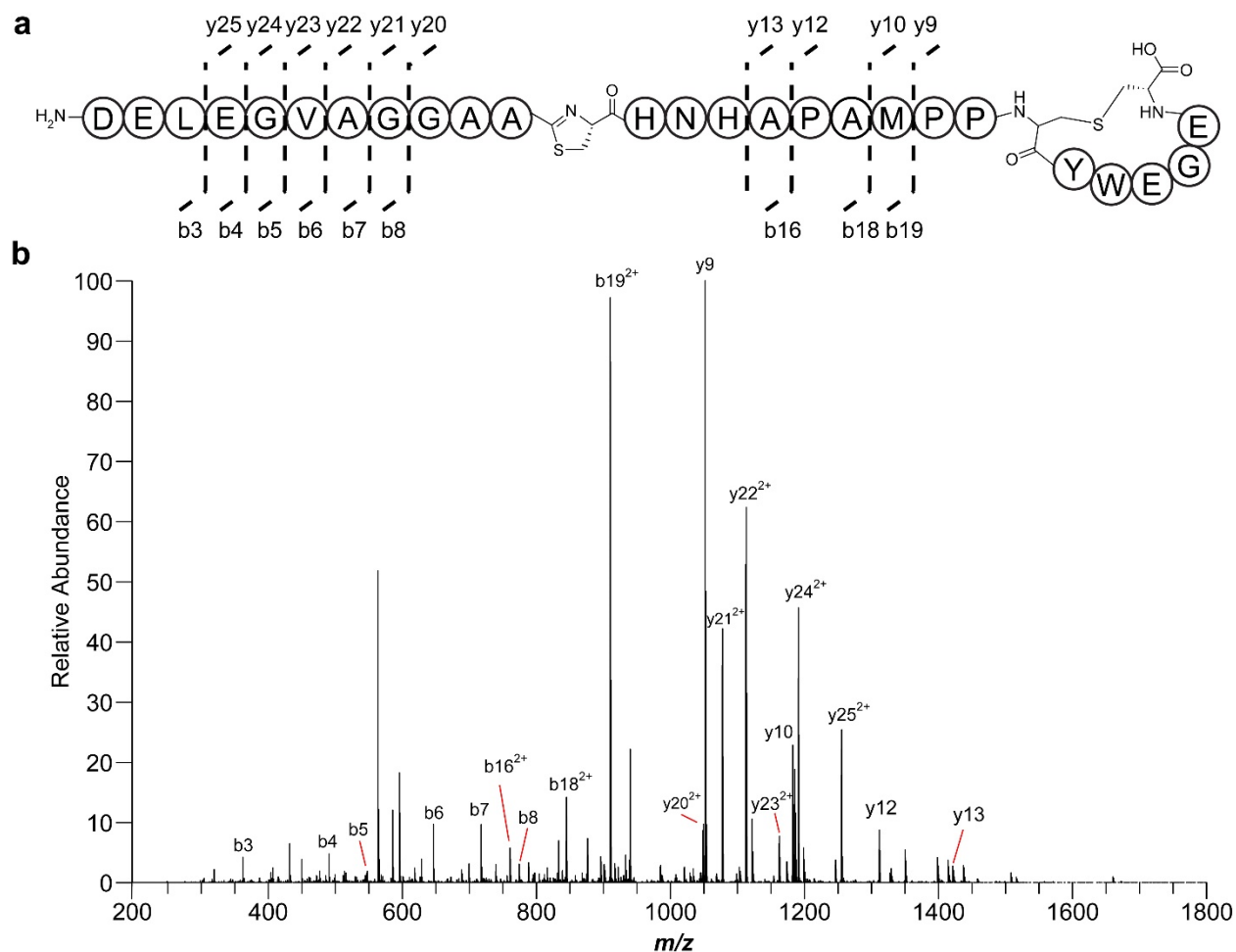


Figure 3.15 MS/MS fragmentation of thiazoline/lanthionine in modified Hyp3.2 digested with AspN. (Previous page) (a) Structure of the thiazoline-lanthionine modified Hyp3.2 product. Dashed lines indicate identified fragments in the mass spectrum. (b) MS/MS fragmentation spectrum for modified Hyp3.2 annotated with identified ions. (c) Table of predicted and observed m/z values found for ions observed in panel b (next page).

C

| Ion | Theoretical mass (Da) | Observed mass (Da) | Error (ppm) |
|-------------------|------------------------------|---------------------------|--------------------|
| b3 ⁺ | 358.1609 | 358.1629 | 5.6 |
| b4 ⁺ | 487.2035 | 487.2062 | 5.7 |
| b5 ⁺ | 544.2249 | 544.2280 | 5.7 |
| b6 ⁺ | 643.2934 | 643.2970 | 5.7 |
| b7 ⁺ | 714.3305 | 714.3345 | 5.7 |
| b8 ⁺ | 771.3519 | 771.3563 | 5.7 |
| b16 ²⁺ | 757.8257 | 757.8298 | 5.4 |
| b18 ²⁺ | 841.8706 | 841.8752 | 5.5 |
| b19 ²⁺ | 907.3909 | 907.3961 | 5.7 |
| y9 ⁺ | 1049.4033 | 1049.4089 | 5.3 |
| y10 ⁺ | 1180.4438 | 1180.4502 | 5.4 |
| y12 ⁺ | 1348.5337 | 1348.5415 | 5.8 |
| y13 ⁺ | 1419.5708 | 1419.5781 | 5.2 |
| y20 ²⁺ | 1046.4166 | 1046.4223 | 5.5 |
| y21 ²⁺ | 1074.9273 | 1074.9330 | 5.3 |
| y22 ²⁺ | 1110.4459 | 1110.4517 | 5.3 |
| y23 ²⁺ | 1159.9800 | 1159.9862 | 5.3 |
| y24 ²⁺ | 1188.4908 | 1188.4972 | 5.4 |
| y25 ²⁺ | 1253.0121 | 1253.0194 | 5.8 |

Figure 3.15 (Continued)

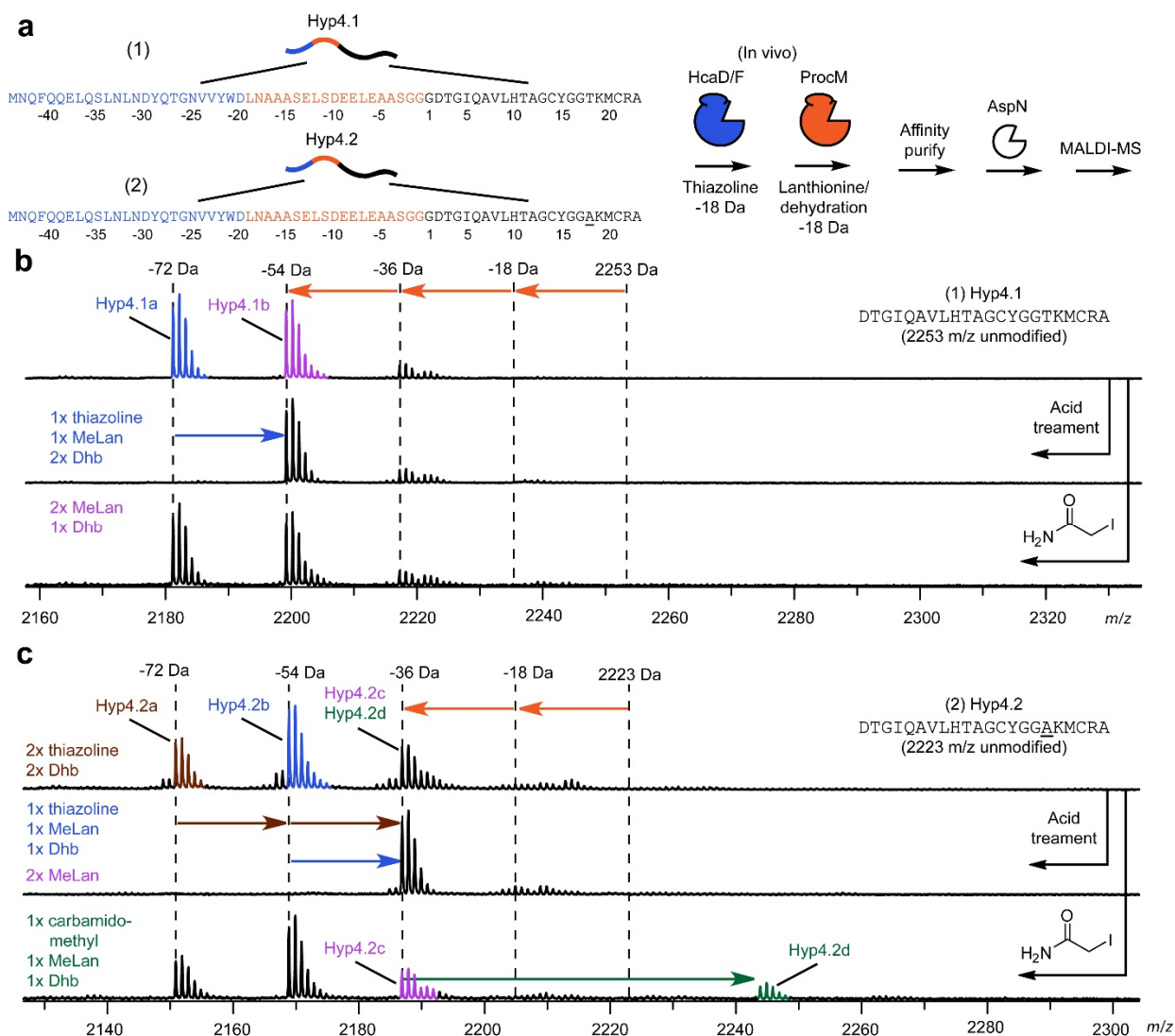


Figure 3.16 Thiiazoline and lanthionine product mixtures produced with Hyp4.1 and Hyp4.2. (a) Experimental overview. (b) When Hyp4.1 was co-expressed with HcaD/F (pET-4.1-FD) and ProcM (pACYC-M), two main products were detected. Hyp4.1a has one thiiazoline and lanthionine (see Figure 3.19 for MS/MS), but the lack of mild acid hydrolysis and lack of iodoacetamide labeling suggested that Hyp4.1b displayed modifications akin to native prochlorosin 3.3. (c) With the Thr18Ala substitution in Hyp4.2 (underlined, panel a), co-expression of Hyp4.2 with HcaD/F (pET-4.2-FD) and ProcM (pACYC-M) produced multiple products indicating there was little preference for the modification order. Hyp4.2a contained two thiazolines, Hyp4.2b contained one thiiazoline, while the other products were not susceptible to mild acid hydrolysis. Hyp4.2c likely contains two lanthionines while Hyp4.2d displayed a free Cys based on iodoacetamide labeling. Blue/brown arrows indicate acid hydrolysis (+18 Da). Orange arrows indicate Dhb/MeLan formation (-18 Da). Green arrow indicates carbamidomethylation upon reaction with iodoacetamide (+57 Da).

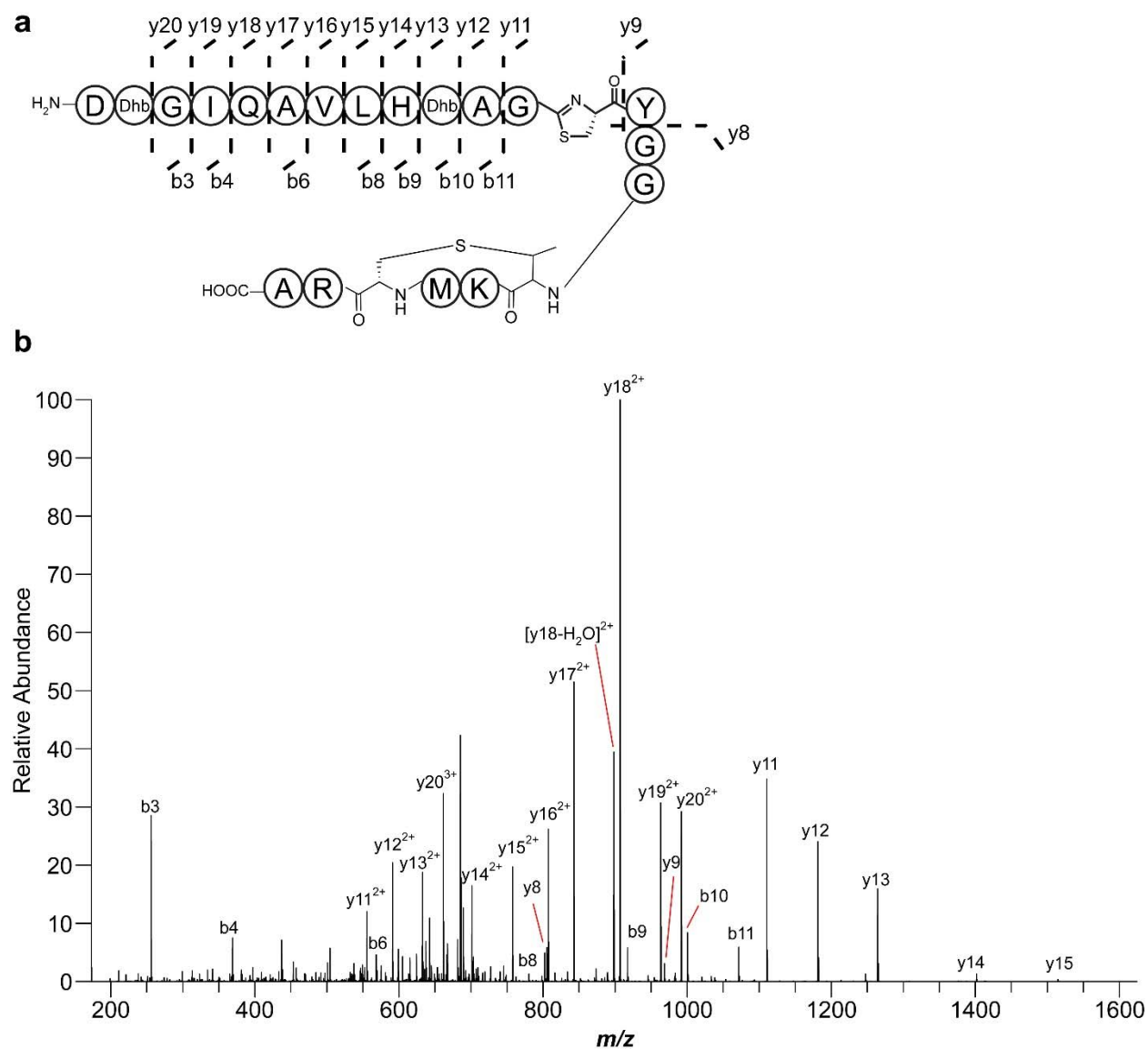


Figure 3.17 MS/MS fragmentation of modified Hyp4.1a. (a) Modified structure of the Dhb, thiazoline, and lanthionine modified Hyp4.1a. Dashed lines indicate identified fragments in the mass spectrum. (b) MS/MS fragmentation spectrum for modified Hyp4.1a annotated with identified ions. (c) Table of predicted and observed m/z values for ions in panel b (next page).

C

| Ion | Theoretical mass (Da) | Observed mass (Da) | Error (ppm) |
|--------------------------------------|-----------------------|--------------------|-------------|
| b3 ⁺ | 256.0928 | 256.0940 | 4.6 |
| b4 ⁺ | 369.1769 | 369.1785 | 4.4 |
| b6 ⁺ | 568.2726 | 568.2741 | 4.3 |
| b8 ⁺ | 780.4250 | 780.4285 | 4.4 |
| b9 ⁺ | 917.4839 | 917.4879 | 4.4 |
| b10 ⁺ | 1000.5255 | 1000.5211 | 4.4 |
| b11 ⁺ | 1071.5582 | 1071.5628 | 4.3 |
| y11 ⁺ | 1110.4642 | 1110.4692 | 4.5 |
| y11 ²⁺ | 555.7357 | 555.7383 | 4.6 |
| y12 ⁺ | 1181.5013 | 1181.5063 | 4.3 |
| y12 ²⁺ | 591.2543 | 591.2570 | 4.7 |
| y13 ⁺ | 1264.5384 | 1264.5439 | 4.3 |
| y13 ²⁺ | 632.7728 | 632.7758 | 4.7 |
| y14 ⁺ | 1401.5973 | 1401.6038 | 4.6 |
| y14 ²⁺ | 701.3023 | 701.3056 | 4.7 |
| y15 ⁺ | 1514.6814 | 1514.6889 | 5.0 |
| y15 ²⁺ | 757.8443 | 757.8479 | 4.7 |
| y16 ²⁺ | 807.3785 | 807.3821 | 4.5 |
| y17 ²⁺ | 842.8971 | 842.9009 | 4.5 |
| y18 ²⁺ | 906.9264 | 906.9303 | 4.4 |
| [y18-H ₂ O] ²⁺ | 897.9211 | 897.9251 | 4.5 |
| y19 ²⁺ | 963.4684 | 963.4727 | 4.5 |
| y20 ²⁺ | 991.9791 | 991.9836 | 4.5 |
| y20 ³⁺ | 661.6552 | 661.6584 | 4.8 |

Figure 3.17 (Continued)

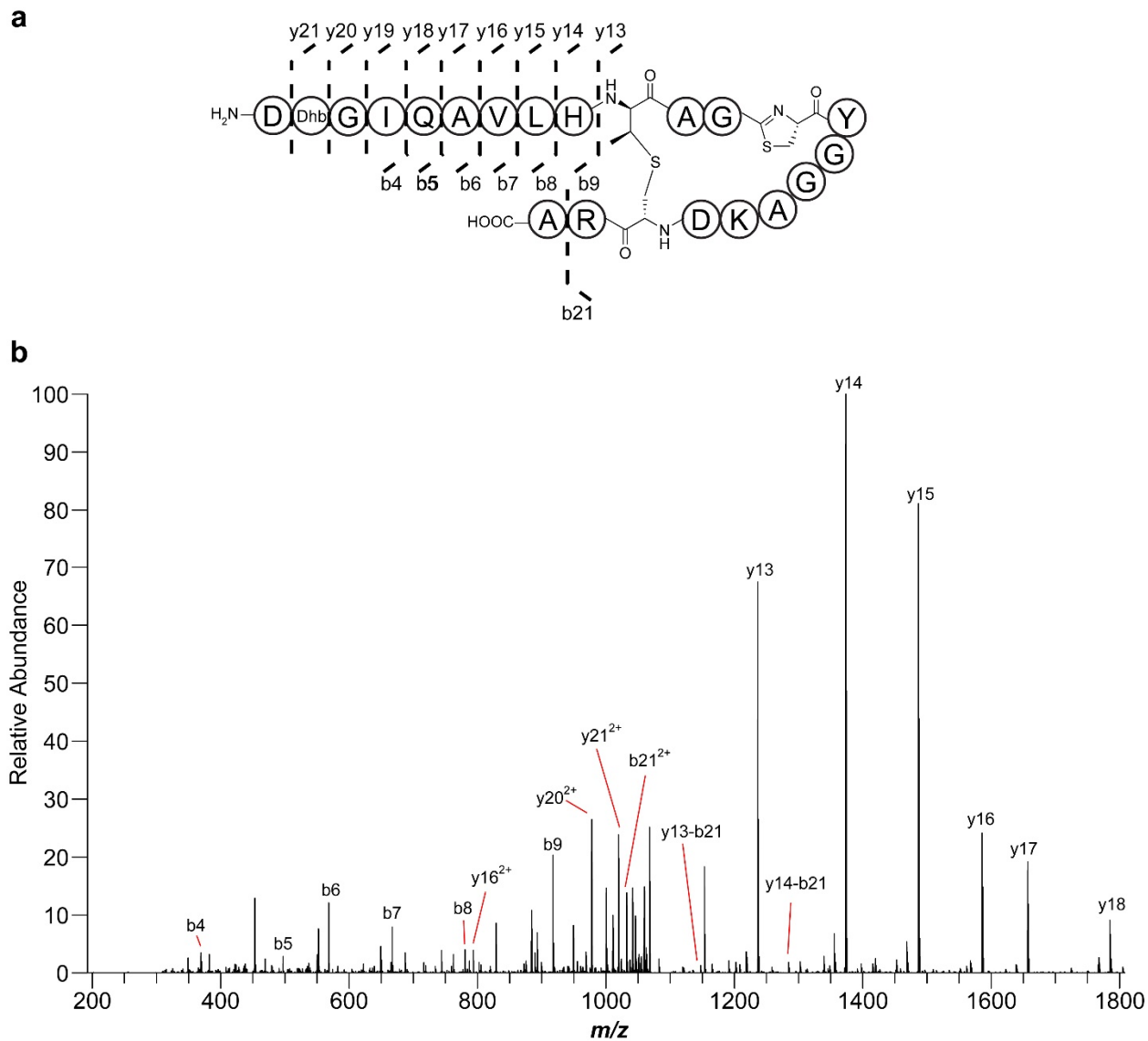


Figure 3.18 MS/MS fragmentation of modified Hyp4.3a. (a) Structure of the Dhb, thiazoline, and lanthionine modified Hyp4.3. Dashed lines indicate identified fragments in the mass spectrum. (b) MS/MS fragmentation spectrum for modified Hyp4.3 annotated with identified ions. (c) Table of predicted and observed m/z values for ions in panel b (next page).

C

| Ion | Theoretical mass (Da) | Observed mass (Da) | Error (ppb) |
|-------------------|------------------------------|---------------------------|--------------------|
| b3 ⁺ | 256.0928 | 256.0927 | 312.4 |
| b4 ⁺ | 369.1769 | 369.1767 | 541.7 |
| b5 ⁺ | 497.2354 | 497.2353 | 281.6 |
| b6 ⁺ | 568.2726 | 568.2725 | 176.0 |
| b7 ⁺ | 667.3410 | 667.3408 | 269.7 |
| b9 ⁺ | 917.4839 | 917.4834 | 566.8 |
| b10 ⁺ | 1000.5211 | 1000.5206 | 479.8 |
| b11 ⁺ | 1071.5582 | 1071.5576 | 559.9 |
| y4 ⁺ | 462.1952 | 462.1950 | 454.4 |
| y5 ⁺ | 590.2901 | 590.2900 | 288.0 |
| y8 ⁺ | 775.3702 | 775.3698 | 451.4 |
| y11 ⁺ | 1080.4536 | 1080.4530 | 536.8 |
| y11 ²⁺ | 540.7304 | 540.7304 | 147.9 |
| y12 ⁺ | 1151.4907 | 1151.4899 | 677.4 |
| y12 ²⁺ | 576.2490 | 576.2488 | 329.7 |
| y13 ⁺ | 1234.5278 | 1234.5279 | 48.6 |
| y13 ²⁺ | 617.7676 | 617.7675 | 161.9 |
| y14 ⁺ | 1371.5867 | 1371.5861 | 444.7 |
| y14 ²⁺ | 686.2970 | 686.2968 | 320.6 |
| y15 ⁺ | 1484.6708 | 1484.6703 | 107.8 |
| y15 ²⁺ | 742.8390 | 742.8389 | 175.0 |
| y16 ²⁺ | 792.3732 | 792.3730 | 290.3 |
| y17 ²⁺ | 827.8918 | 827.8915 | 374.4 |
| y18 ²⁺ | 891.9211 | 891.9206 | 583.0 |
| y19 ²⁺ | 948.4631 | 948.4626 | 601.0 |
| y20 ²⁺ | 976.9739 | 976.9736 | 245.7 |
| y20 ³⁺ | 651.6517 | 651.6516 | 168.8 |

Figure 3.18 (Continued)

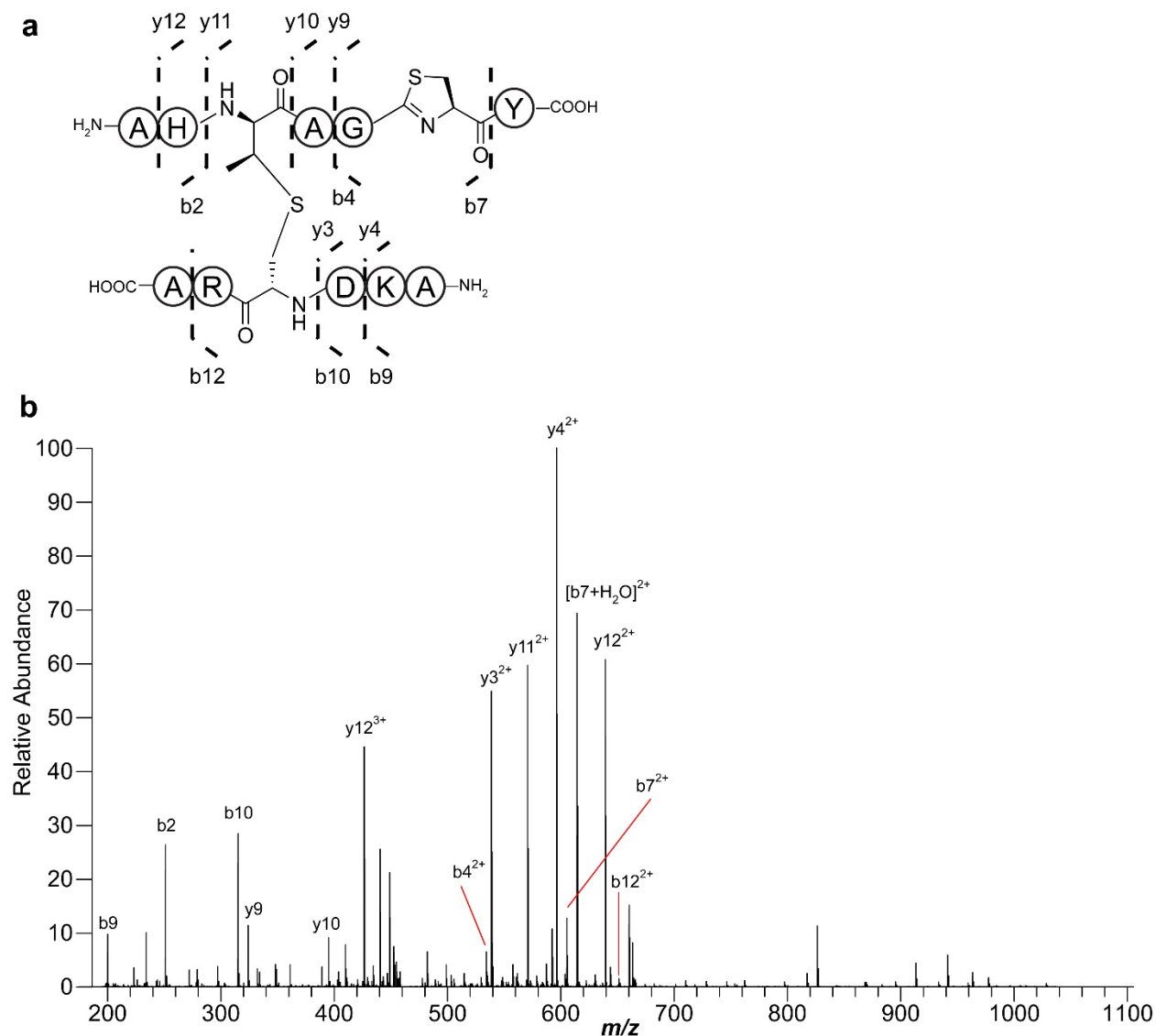


Figure 3.19 MS/MS fragmentation of modified Hyp4.3a after thermolysin digestion. (a) Structure of the lanthionine ring fragment of Hyp4.3 after digestion by thermolysin. Dashed lines indicate identified fragments in the mass spectrum. (b) MS/MS fragmentation spectrum for lanthionine ring of Hyp4.3 after digestion and HPLC-purification annotated with identified ions. (c) Table of predicted and observed m/z values of ions seen in panel b (next page).

C

| Ion | Theoretical mass (Da) | Observed mass (Da) | Error (ppm) |
|-------------------------------------|-----------------------|--------------------|-------------|
| b2 ⁺ | 251.1503 | 251.1516 | 5.4 |
| b4 ²⁺ | 534.2744 | 534.2773 | 5.4 |
| b7 ²⁺ | 605.2844 | 605.2876 | 5.2 |
| [b7+H ₂ O] ²⁺ | 614.2897 | 614.2929 | 5.2 |
| b9 ⁺ | 200.1394 | 200.1405 | 5.7 |
| b10 ⁺ | 315.1663 | 315.1679 | 5.1 |
| b12 ²⁺ | 651.2975 | 651.3010 | 5.3 |
| y3 ²⁺ | 538.7419 | 538.7448 | 5.4 |
| y4 ²⁺ | 596.2553 | 596.2586 | 5.4 |
| y9 ⁺ | 324.1013 | 324.1028 | 4.9 |
| y10 ⁺ | 395.1384 | 395.1403 | 5.0 |
| y11 ²⁺ | 570.7499 | 570.7528 | 5.1 |
| y12 ²⁺ | 639.2793 | 639.2827 | 5.3 |
| y12 ³⁺ | 426.5220 | 426.5242 | 5.1 |

Figure 3.19 (Continued)

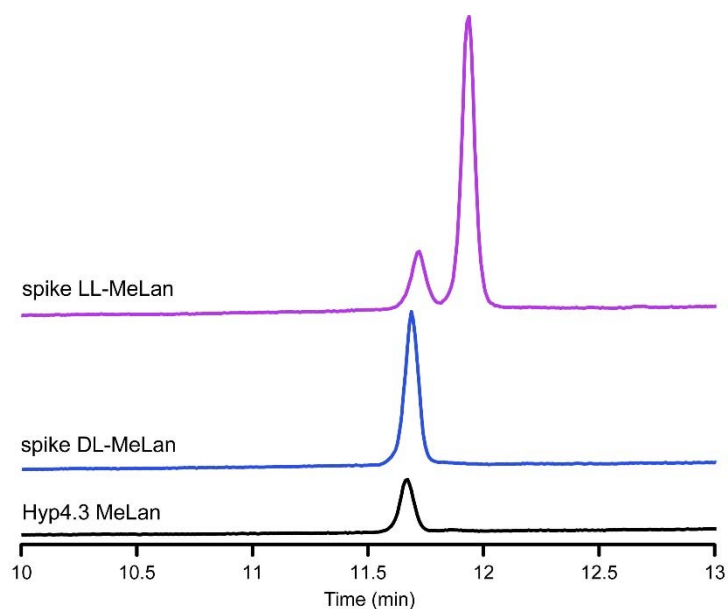


Figure 3.20 GC-MS analysis of Hyp3.3 MeLan. After hydrolysis and derivatization, the MeLan of Hyp3.3 co-eluted when spiked with authentic DL-MeLan but not with LL-MeLan, indicating that the stereochemistry of the lanthionine in Hyp3.3 is DL as in the native prochlorosins.⁶⁵

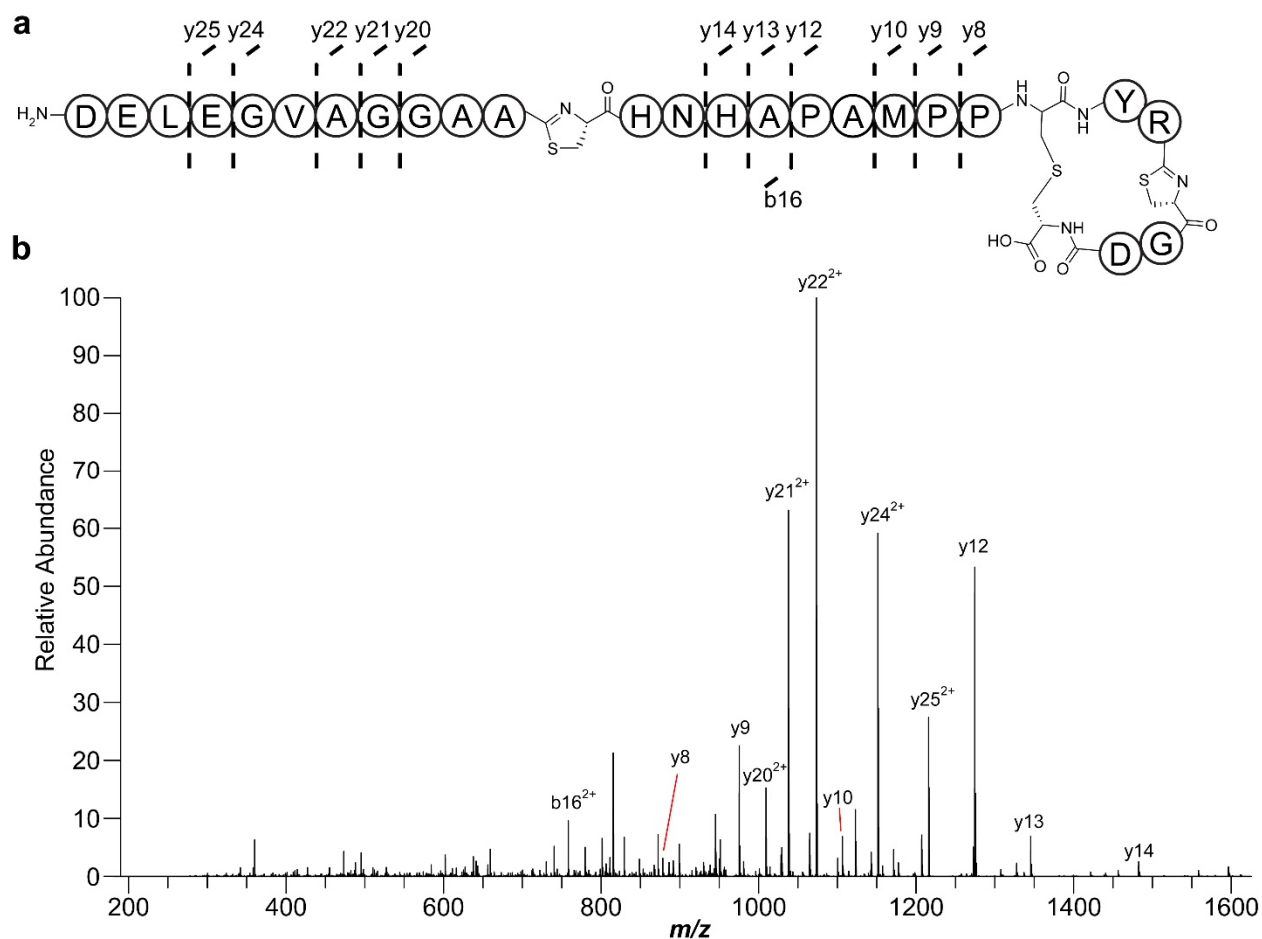


Figure 3.21 MS/MS fragmentation of Hyp4.3 minor products. (a) Structure of two minor products (azoline or Dhb) after expression of Hyp4.3 with HcaD/F (pET-4.3-FD) and ProcM (pACYC-M). A third product, a species containing two lanthionines, also formed (from a Michael-like addition within in the Dhb product) based on MALDI-TOF-MS analysis (Figure 3.6c), but did not undergo fragmentation, likely owing to the lanthionines. Dashed lines indicate identified fragments. (b) Annotated MS/MS spectrum for minor products. (c) Table of predicted and observed m/z values for ions from panel b. Red (azoline) and blue (Dhb) colors indicate the two species (next page)

C

| Azoline species | | | |
|------------------------|------------------------------|---------------------------|--------------------|
| Ion | Theoretical mass (Da) | Observed mass (Da) | Error (ppm) |
| b3 ⁺ | 274.1034 | 274.1045 | 4.3 |
| b4 ⁺ | 387.1874 | 387.1890 | 4.0 |
| b5 ⁺ | 515.2460 | 515.2483 | 4.4 |
| b6 ⁺ | 586.2831 | 586.2856 | 4.2 |
| b7 ⁺ | 685.3515 | 685.3543 | 4.0 |
| b8 ⁺ | 798.4356 | 798.4389 | 4.1 |
| b9 ⁺ | 935.4945 | 935.4983 | 4.1 |
| y13 ²⁺ | 618.7661 | 618.7687 | 4.3 |
| y14 ²⁺ | 687.2955 | 687.2984 | 4.1 |
| y15 ²⁺ | 743.8376 | 743.8407 | 4.2 |
| y16 ²⁺ | 793.3718 | 793.3749 | 4.0 |
| y17 ²⁺ | 828.8903 | 828.8936 | 4.0 |
| y18 ²⁺ | 892.9196 | 892.9230 | 3.8 |
| y20 ²⁺ | 977.9724 | 977.9762 | 3.9 |
| y21 ³⁺ | 685.9999 | 686.0027 | 4.1 |

| Dhb species | | | |
|--------------------|------------------------------|---------------------------|--------------------|
| Ion | Theoretical mass (Da) | Observed mass (Da) | Error (ppm) |
| b3 ⁺ | 256.0928 | 256.0939 | 4.1 |
| b4 ⁺ | 369.1769 | 369.1784 | 4.2 |
| b5 ⁺ | 497.2354 | 497.2374 | 3.9 |
| b6 ⁺ | 568.2726 | 568.2748 | 4.0 |
| b7 ⁺ | 667.3410 | 667.3437 | 4.1 |
| b8 ⁺ | 780.4250 | 780.4283 | 4.2 |
| b9 ⁺ | 917.4839 | 917.4877 | 4.1 |
| y13 ²⁺ | 627.7713 | 627.7740 | 4.2 |
| y14 ²⁺ | 696.3008 | 696.3041 | 4.7 |
| y15 ²⁺ | 752.8428 | 752.8457 | 3.8 |
| y16 ²⁺ | 802.3770 | 802.3799 | 3.6 |
| y17 ²⁺ | 837.8956 | 837.8989 | 4.0 |
| y18 ²⁺ | 901.9249 | 901.9284 | 3.9 |
| y20 ²⁺ | 986.9776 | 986.9814 | 3.8 |
| y21 ³⁺ | 685.9999 | 686.0027 | 4.1 |

Figure 3.21 (Continued)

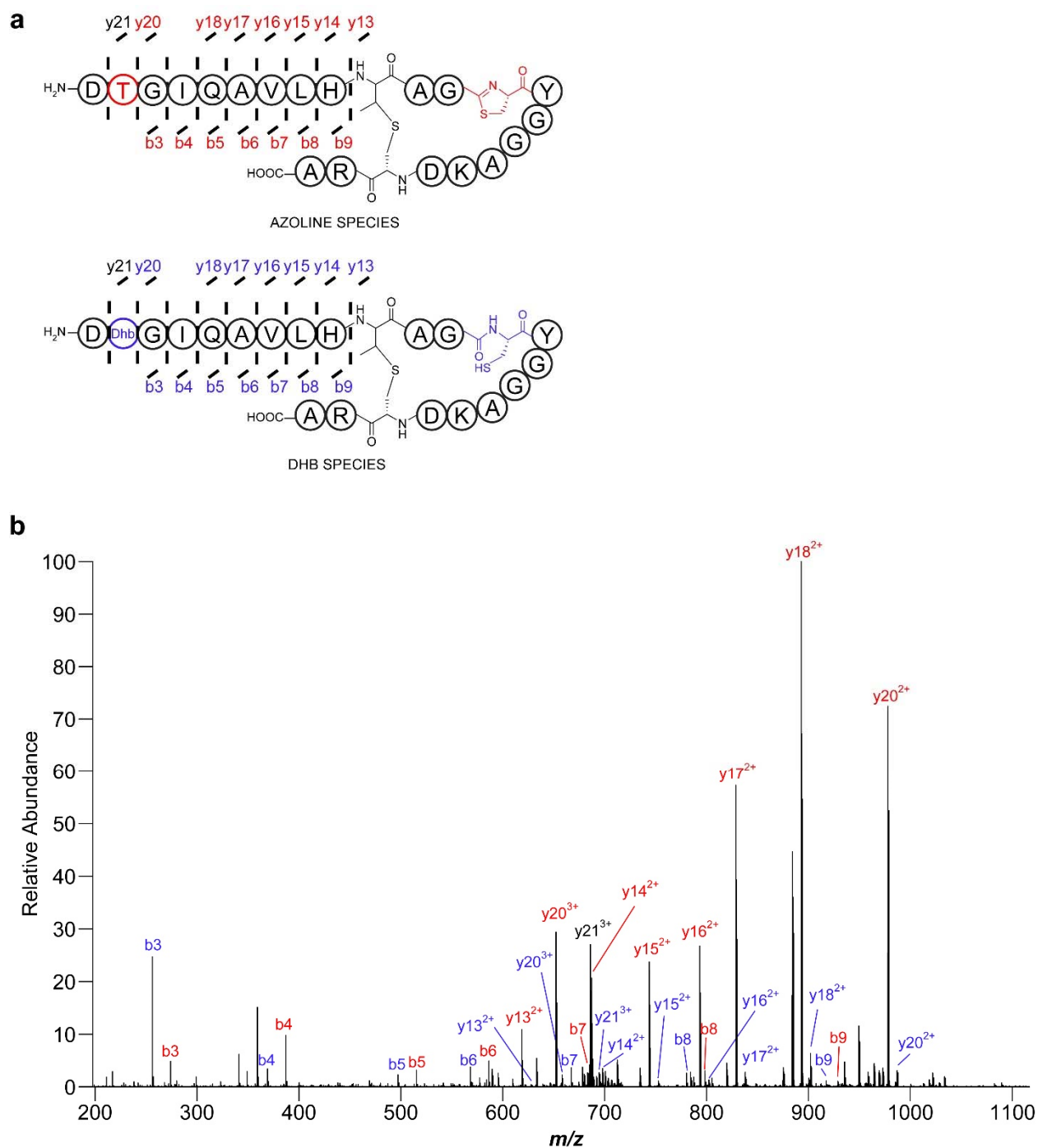


Figure 3.22 MS/MS fragmentation of thiazoline/lanthionine modified Hyp3.3. (a) Structure of the modified Hyp3.3 product containing two thiazolines and one lanthionine. Dashed lines indicate identified fragments in the mass spectrum. (b) MS/MS fragmentation spectrum for modified Hyp3.3 annotated with identified ions. (c) Table of predicted and observed m/z values for ions observed in panel b (next page).

C

| Ion | Theoretical mass (Da) | Observed mass (Da) | Error (ppm) |
|-------------------|------------------------------|---------------------------|--------------------|
| b16 ²⁺ | 757.8257 | 757.8303 | 6.0 |
| y9 ⁺ | 975.3811 | 975.3856 | 4.6 |
| y10 ⁺ | 1106.4216 | 1106.4267 | 4.6 |
| y12 ⁺ | 1274.5115 | 1274.5176 | 4.8 |
| y13 ⁺ | 1345.5486 | 1345.5554 | 5.0 |
| y14 ⁺ | 1482.6075 | 1482.6150 | 5.1 |
| y20 ²⁺ | 1009.4055 | 1009.4103 | 4.8 |
| y21 ²⁺ | 1037.9162 | 1037.9214 | 5.0 |
| y22 ²⁺ | 1073.4348 | 1073.4398 | 4.7 |
| y24 ²⁺ | 1151.4797 | 1151.4851 | 4.7 |
| y25 ²⁺ | 1216.0010 | 1216.0067 | 4.7 |

Figure 3.22 (Continued)

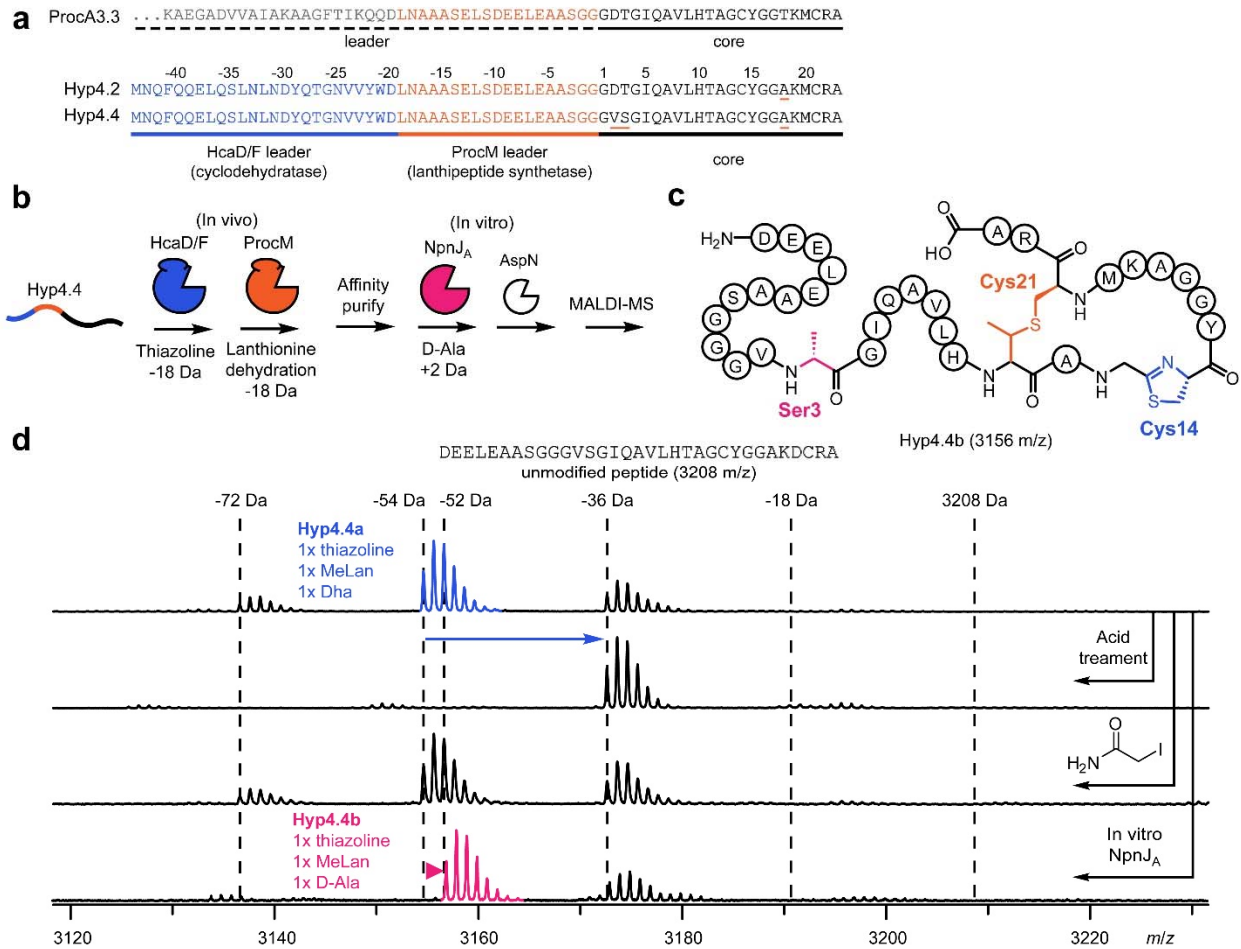


Figure 3.23 Combination of two primary and one secondary RiPP modifications. (a) Design of hybrid peptide. Orange underlining highlights changed residues. (b) Experimental overview for combining HcaF/D, ProcM, and NpnJ_A. (c) Deduced structure of Hyp4.4b. (d) MALDI-TOF-MS analysis of Hyp4.4 products. The initially formed hybrid has a Dha, thiazoline (arrow indicates hydrolysis) and MeLan. The formation of D-Ala from Dha is indicated by the +2 Da shift (triangle) of the Hyp4.4a species upon addition of NpnJ_A. The weak ion labeled as -72 Da is likely a peptide that underwent two cyclodehydrations and two dehydrations.

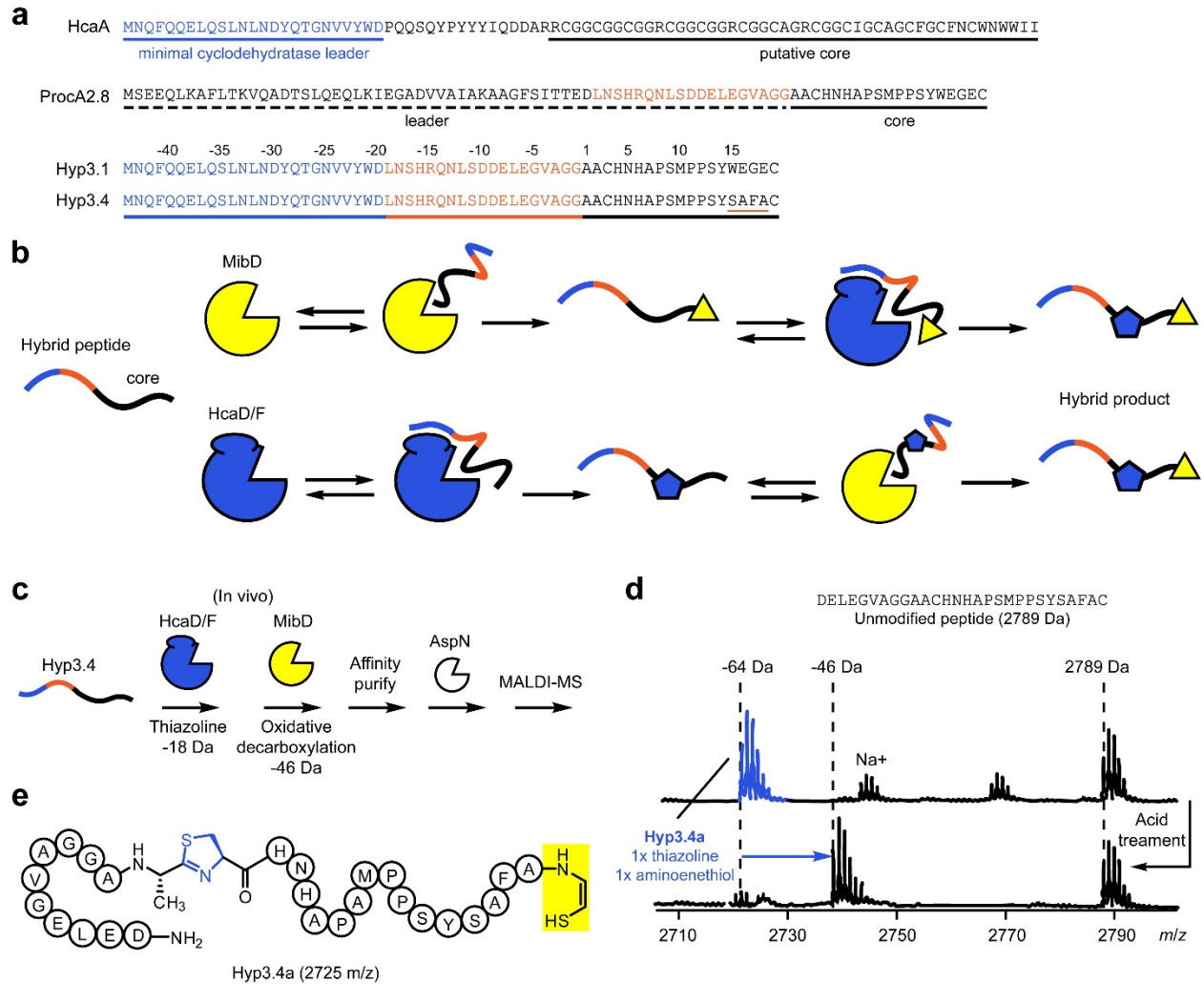


Figure 3.24 Design and production of decarboxylated linear azol(in)e-containing peptide. (a) A hybrid peptide was designed to enable modification by HcaD/F and the leader-independent tailoring enzyme MibD from *Microbispora* sp. 107891. MibD is a LanD enzyme and carries out oxidative decarboxylation of C-terminal Cys to the corresponding aminoenethiolate, and its preferred core sequence is S-(F/W/N)-(N/C)-S-(F/Y/W/N)-C-C.⁶³ This motif was mimicked by changing the C-terminus of the peptide to SAFAC to create Hyp3.4. Although the ProcM leader region was present, ProcM was not used in this experiment. (b) Anticipated hybrid biosynthesis. As both thiazoline and decarboxylation are smaller modifications occurring on different ends of the core peptide, the enzymes can presumably act in any order. (c) Overview of experiment. Hyp3.4 was co-expressed with HcaD/F (pET-3.4-FD) and MibD in pCDFDuet (see plasmid list in Table 3.1). The peptide was purified, digested with AspN, and analyzed by MALDI-TOF-MS. (d) The mass spectrum shows a several ions, but the most intense peak (labeled Hyp3.4a) corresponded to the desired hybrid product. (e) Deduced structure of the hybrid Hyp3.4a product after AspN digestion.

3.7 Tables

Table 3.1 New plasmids generated for this study and their use in combinations. Hybrid peptides (Hyp) are listed in Figure 3.4, 3.6, 3.11, 3.23 and 3.24. HcaF-HcaD comprise a two-component, thiazoline-forming cyclodehydratase, NisB and NisC form a class I lanthipeptide synthetase, while ProcM is a class II lanthipeptide synthetase. Constructs for the tailoring enzymes MibD and NpnJ_A are reported elsewhere.⁵⁷,⁶³ The combinations of plasmids used for co-expression and the figure where the results are presented are also listed. Abbreviations: No., number; MCS, multiple cloning site; MBP, maltose-binding protein.

| No. | Name | Plasmid backbone (copy number ⁷⁰) | Gene(s) MCS1; MCS2 | Co-expression of plasmids; Figure for result |
|-----|------------|---|-----------------------------------|--|
| 1 | pACYC-B-C | pACYCDuet (10-12) | NisB; NisC | 1+2+3; Fig. 3.4 |
| 2 | pRSF-F-D | pRSFDuet (>100) | HcaF; HcaD | 1+2+3; Fig. 3.4 |
| 3 | pCDF-1.1 | pCDFDuet (20-40) | His ₆ -Hyp1.1; (empty) | 1+2+3; Fig. 3.4 |
| 4 | pACYC-AlbA | pACYCDuet (10-12) | (empty); AlbA | See below |
| 5 | pET-2.1-FD | pETDuet (~40) | MBP-Hyp2.1; HcaF-HcaD | 5+4; Fig. 3.7 |
| 6 | pET-2.2-FD | pETDuet (~40) | MBP-Hyp2.2; HcaF-HcaD | 6+4; Fig. 3.6 |
| 7 | pET-3.1-FD | pETDuet (~40) | MBP-Hyp3.1; HcaF-HcaD | 7+16; Fig. 3.7 |
| 8 | pET-3.2-FD | pETDuet (~40) | MBP-Hyp3.2; HcaF-HcaD | 8+16; Fig. 3.14 |
| 9 | pET-3.3-FD | pETDuet (~40) | MBP-Hyp3.3; HcaF-HcaD | 9+16; Fig. 3.11 |
| 10 | pET-3.4-FD | pETDuet (~40) | MBP-Hyp3.4; HcaF-HcaD | 10+11; Fig. 3.24 |
| 11 | pCDF-MibD | pCDFDuet (20-40) | MibD; (empty) | 10+11; Fig. 3.24 |
| 12 | pET-4.1-FD | pETDuet (~40) | MBP-Hyp4.1; HcaF-HcaD | 12+16; Fig. 3.16 |
| 13 | pET-4.2-FD | pETDuet (~40) | MBP-Hyp4.2; HcaF-HcaD | 13+16; Fig. S12 |
| 14 | pET-4.3-FD | pETDuet (~40) | MBP-Hyp4.3; HcaF-HcaD | 14+16; Fig. 4 |
| 15 | pET-4.4-FD | pETDuet (~40) | MBP-Hyp4.4; HcaF-HcaD | 15+16; Fig 5 |
| 16 | pACYC-M | pETDuet (~40) | (empty); ProcM | See above |
| 17 | pRSF-3.1-M | pRSFDuet (>100) | MBP-Hyp3.1; ProcM | 17+18; Fig. S8 |
| 18 | pET-F-D | pETDuet (~40) | HcaF; HcaD | 17+18; Fig. S8 |

Table 3.2 Nucleotide sequences. Restriction enzyme sites are underlined. F, forward; R, reverse. SDM, site-directed mutagenesis.

| Primer Name | Primer sequence (5' to 3') | Purpose |
|---------------------|---|---|
| AlbA F | AAAAGATATCAATGTTTATAGAGCAGATGTTCCATTTATTAATGAAAG | RE cloning of pACYC-AlbA (EcoRV) |
| AlbA R | AAAGACGTCCTAAATAAGCTGGACCACGCTCTTCT | RE cloning of pACYC-AlbA (AatII) |
| HcaD F | AAAAGATCTAATGACTCAAATATATTACTTATAGGGGATGGC | RE cloning of pET-F-D (BglII) |
| HcaD R | AAAAGGTACCCTATGAAACGGGTGCGGATG | RE cloning of pET-F-D (KpnI) |
| HcaF F | AAAGGATCCAATGAGTAACCTTCCAGCACATGC | RE cloning of pET-F-D (BamHI) |
| HcaF R | AAAGCGGCCGCTCATAGACCGTCCCTCTTTGC | RE cloning of pET-F-D (NotI) |
| MBP-Hyp3.1 (pRSF) F | ATCACCACAGCCAGGATCCGAATTCATGAAAATCGAAGAAGGTAAACTGG | Gibson assembly of pRSF-3.1-ProcM (EcoRI) |
| MBP-Hyp3.1 (pRSF) R | GTTCTGACTTAAGCATTATGCGGCCGCTTAGCACTCACCTCCCAAT | Gibson assembly of pRSF-3.1-ProcM (NotI) |
| MBP-Hyp3.1 (pET) F | CATCACCACAGCCAGGATCCGAAAATCGAAGAAGGTAAACTGGTAATC | Gibson assembly of pET-3.1-FD (BamHI) |
| MBP-Hyp3.1 (pET) R | CGACTTAAGCATTATGCGGCCGCTTAGCACTCACCTCCCAATAG | Gibson assembly of pET-3.1-FD (NotI) |
| HcaF-HcaD F | CATATGGCAGATCTCAATTGGATATCAATGAGTAACCTTCCAGCACATG | Gibson assembly of pET-Hyp-FD (EcoRV) |
| HcaF-HcaD R | CCGACTCGAGGGTACCGACGCTCTATGAAACGGGTGCGG | Gibson assembly of pET-Hyp-FD (AatII) |
| MBP-Hyp4.1 F | CATCACCACAGCCAGGATCCGAAAATCGAAGAAGGTAAACTGGTAATC | Gibson assembly of pET-4.1-FD (BamHI) |
| MBP-Hyp4.1 R | TTAAGCATTATGCGGCCGCAAGCTTCTATGCGCGGCACATTTTG | Gibson assembly of pET-4.1-FD (HindIII) |
| MBP-Hyp2.1 F | ACCATCATCACCACAGCCAGGATCCGAAAATCGAAGAAGGTAAACTGGTAATC | Gibson assembly of pET-2.1-FD (BamHI) |
| MBP-Hyp2.1 R | CGACTTAAGCATTATGCGGCCGCTTAACGCCATAGACCGAATAGACC | Gibson assembly of pET-2.1-FD (NotI) |
| HcaA leader F | AAAGGATCCATGAATCAGTTTCAACAAGAACTAC | RE cloning of MBP-tagged Hyp3.1 (BamHI) |
| HcaA leader (3.1) R | GCGATGAGAGTTTAGATCCCAATACACAACATTACCAGT | overlaps ProcA2.8 |
| ProcA2.8 F | GTTGTGTATTGGGATCTAAACTCTCATCGCCAAAATCTG | overlaps HcaA leader |
| ProcA2.8 R | AAAAGCTTTTAGCACTCACCTCCCAATAGG | RE cloning of MBP-tagged Hyp3.1 (HindIII) |
| HcaA leader F | AAAGGATCCATGAATCAGTTTCAACAAGAACTAC | RE cloning of MBP-tagged Hyp4.1 (BamHI) |
| HcaA leader (4.1) R | AGCGGCAGCATTAAGATCCCAATACACAACATTACCAGT | overlaps ProcA3.3 |
| ProcA3.3 F | GTTGTGTATTGGGATCTTAATGCTGCCGCTTCTG | overlaps HcaA leader |
| ProcA3.3 R | AAAAGCTTCTATGCGCGGCACATTTTGG | RE cloning of MBP-tagged Hyp4.1 (HindIII) |
| HcaA leader F | AAAGGATCCATGAATCAGTTTCAACAAGAACTAC | RE cloning of MBP-tagged Hyp2.1 (BamHI) |
| HcaA leader (2.1) R | GACAGTTTTTTTCATATCCCAATACACAACATTACCAG | overlaps SboA |
| SboA F | GTTGTGTATTGGGATATGAAAAAGCTGTCAATTGTAGAAAAC | overlaps HcaA leader |
| SboA R | AAAGCGGCCGCTTACCCCATAGACCGAATAGACCTGTTGC | RE cloning of MBP-tagged Hyp2.1 (NotI) |
| SboA BamHI F | AAAGGATCCATGAAAAAGCTGTCAATTGTAGAAAACAAAGG | RE cloning |
| SboA NotI R | AAAGCGGCCGCTTACCCCATAGACCGAATAGACCTGTTGC | RE cloning |

Table 3.2 (Continued)

| | | |
|----------------|--|-----|
| Hyp2.1 F | ATTTCGGTCTATGGCGTTAAGCGGCCGCACTCGAGCAC | SDM |
| Hyp2.1 R | CGGCCGCTTAACGCCATAGACCGAATAGACCTGTTGC | SDM |
| Hyp2.2 F | ACATGCTCGATCGGAGCCCCGTTGCGGTGGATGTGGTCTATCCCTGATTTTCAAATTGC | SDM |
| Hyp2.2 R | ATTTCAAATCAGGGATAGGACCACATCCACCGCAACGGGCTCCGATCGAGCATGTTGC | SDM |
| Hyp3.2 F | AACCATGCTCCAGCTATGCCTCCATCCTATTGGGA | SDM |
| Hyp3.2 R | GGAGGCATAGCTGGAGCATGGTTATGACAGGCCG | SDM |
| Hyp 3.3 F | GCCTCCATCCTATCGTTGTGGTGAGTGCTAAGCGGCCGATAATGC | SDM |
| Hyp 3.3 R | GCTTAGCACTCACCACAACGATAGGATGGAGGCATAGCTGGAGCATGG | SDM |
| Hyp4.2 F | ACGGCGGGGCCAAAATGTGCGCGCATAG | SDM |
| Hyp4.2 R | CACATTTTGGCCCCCGTAACATCCAGCCG | SDM |
| Hyp4.3 F | GCGGGGCCAAAGACTGCCGCGCATAGAAGCTTGCGGCCGCAT | SDM |
| Hyp4.3 R | TATGCGCGCAGTCTTTGGCCCCGCGTAACATCCAGCCGTG | SDM |
| Hyp4.4 F | CGGCGGAGGCGTTTCCGGCATCCAAGCGGTGCTGCAC | SDM |
| Hyp4.4 R | CTTGGATGCCGGAACGCCTCCGCCGTAGCGGCTTCC | SDM |
| Hyp1.1 Trunc F | GTATTTTCGCTATGTACACCCGGTTGTAAAGCGTAAGGAGCTCTGTAAGGTTGTAACATGA | SDM |
| Hyp1.1 Trunc R | TCATGTTACAACCTTACAGAGCTCCTTACGCTTTTACAACCGGTTGATACATAGCGAAATAC | SDM |
| Hyp1.1 F | GATTTGGTATCTGTTTTCGAAGAAAGATTACGTTGCGGTCCACGCATTACAAGTATTTTC | SDM |
| Hyp1.1 R | GAAATACTTGTAAATGCGTGGACCGCAACGTGAATCTTCTTCGAAACAGATACCAAATC | SDM |
| Hyp4.5 F | AGCGCGTTTGCCTGCTAAGCGGCCGATAATGCTTAAGTCGAACAGAAAGTAATC | SDM |
| Hyp4.5 R | GCCTGTCATAACCATGCTCCATCTATGCCTCCATCCTATAGCGGCTTTGCGTGC | SDM |

3.8 References

- (1) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* **2016**, *79* (3), 629-661.
- (2) Ganesan, A. The Impact of Natural Products Upon Modern Drug Discovery. *Curr. Opin. Chem. Biol.* **2008**, *12* (3), 306-317.
- (3) Katz, L.; Baltz, R. H. Natural Product Discovery: Past, Present, and Future. *J. Ind. Microbiol. Biotechnol.* **2016**, *43* (2-3), 155-176.
- (4) Breitling, R.; Takano, E. Synthetic Biology of Natural Products. *Cold Spring Harb. Perspect. Biol.* **2016**, *8* (10).
- (5) King, J. R.; Edgar, S.; Qiao, K.; Stephanopoulos, G. Accessing Nature's Diversity through Metabolic Engineering and Synthetic Biology. *F1000Res.* **2016**, *5*.

- (6) Smanski, M. J.; Zhou, H.; Claesen, J.; Shen, B.; Fischbach, M. A.; Voigt, C. A. Synthetic Biology to Access and Expand Nature's Chemical Diversity. *Nat. Rev. Microbiol.* **2016**, *14* (3), 135-149.
- (7) Leadlay, P. F. Combinatorial Approaches to Polyketide Biosynthesis. *Curr. Opin. Chem. Biol.* **1997**, *1* (2), 162-168.
- (8) Weissman, K. J.; Leadlay, P. F. Combinatorial Biosynthesis of Reduced Polyketides. *Nat. Rev. Microbiol.* **2005**, *3* (12), 925-936.
- (9) Weissman, K. J. Genetic Engineering of Modular PKSs: From Combinatorial Biosynthesis to Synthetic Biology. *Nat. Prod. Rep.* **2016**, *33* (2), 203-230.
- (10) Winn, M.; Fyans, J. K.; Zhuo, Y.; Micklefield, J. Recent Advances in Engineering Nonribosomal Peptide Assembly Lines. *Nat. Prod. Rep.* **2016**, *33* (2), 317-347.
- (11) Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J., et al. Ribosomally Synthesized and Post-Translationally Modified Peptide Natural Products: Overview and Recommendations for a Universal Nomenclature. *Nat. Prod. Rep.* **2013**, *30* (1), 108-160.
- (12) Oman, T. J.; van der Donk, W. A. Follow the Leader: The Use of Leader Peptides to Guide Natural Product Biosynthesis. *Nat. Chem. Biol.* **2010**, *6* (1), 9-18.
- (13) Walsh, C. T. Blurring the Lines between Ribosomal and Nonribosomal Peptide Scaffolds. *ACS Chem. Biol.* **2014**, *9* (8), 1653-1661.
- (14) McIntosh, J. A.; Donia, M. S.; Schmidt, E. W. Ribosomal Peptide Natural Products: Bridging the Ribosomal and Nonribosomal Worlds. *Nat. Prod. Rep.* **2009**, *26* (4), 537-559.
- (15) Burkhardt, 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.
- (16) Dong, S. H.; Tang, W.; Lukk, T.; Yu, Y.; Nair, S. K.; van der Donk, W. A. The Enterococcal Cytolysin Synthetase Has an Unanticipated Lipid Kinase Fold. *eLife* **2015**, *4*.
- (17) Li, K.; Concurso, H. L.; Li, G.; Ding, Y.; Bruner, S. D. Structural Basis for Precursor Protein-Directed Ribosomal Peptide Macrocyclization. *Nat. Chem. Biol.* **2016**, *12* (11), 973-979.
- (18) Koehnke, J.; Bent, A.; Houssen, W. E.; Zollman, D.; Morawitz, F.; Shirran, S.; Vendome, J.; Nneoyiegbe, A. F.; Trembleau, L.; Botting, C. H., et al. The Mechanism of Patellamide Macrocyclization Revealed by the Characterization of the Patg Macrocyclase Domain. *Nat. Struct. Mol. Biol.* **2012**, *19* (8), 767-772.
- (19) Islam, M. R.; Shioya, K.; Nagao, J.; Nishie, M.; Jikuya, H.; Zendo, T.; Nakayama, J.; Sonomoto, K. Evaluation of Essential and Variable Residues of Nukacin Isk-1 by NNK Scanning. *Mol. Microbiol.* **2009**, *72* (6), 1438-1447.

- (20) Pavlova, O.; Mukhopadhyay, J.; Sineva, E.; Ebright, R. H.; Severinov, K. Systematic Structure-Activity Analysis of Microcin J25. *J. Biol. Chem.* **2008**, *283* (37), 25589-25595.
- (21) Just-Baringo, X.; Albericio, F.; Álvarez, M. Thiopeptide Engineering: A Multidisciplinary Effort Towards Future Drugs. *Angew. Chem., Int. Ed.* **2014**, *53* (26), 6602-6616.
- (22) Maksimov, M. O.; Pan, S. J.; James Link, A. Lasso Peptides: Structure, Function, Biosynthesis, and Engineering. *Nat. Prod. Rep.* **2012**, *29* (9), 996-1006.
- (23) Field, D.; Cotter, P. D.; Ross, R. P.; Hill, C. Bioengineering of the Model Lantibiotic Nisin. *Bioengineered* **2015**, *6* (4), 187-192.
- (24) Yang, X.; van der Donk, W. A. Ribosomally Synthesized and Post-Translationally Modified Peptide Natural Products: New Insights into the Role of Leader and Core Peptides During Biosynthesis. *Chem. Eur. J.* **2013**, *19* (24), 7662-7677.
- (25) Ruffner, D. E.; Schmidt, E. W.; Heemstra, J. R. Assessing the Combinatorial Potential of the Ripp Cyanobactin Tru Pathway. *ACS Synth. Biol.* **2015**, *4* (4), 482-492.
- (26) Donia, M. S.; Hathaway, B. J.; Sudek, S.; Haygood, M. G.; Rosovitz, M. J.; Ravel, J.; Schmidt, E. W. Natural Combinatorial Peptide Libraries in Cyanobacterial Symbionts of Marine Ascidians. *Nat. Chem. Biol.* **2006**, *2* (12), 729-735.
- (27) Sardar, D.; Pierce, E.; McIntosh, J. A.; Schmidt, E. W. Recognition Sequences and Substrate Evolution in Cyanobactin Biosynthesis. *ACS Synth. Biol.* **2015**, *4* (2), 167-176.
- (28) Zhang, Q.; Yang, X.; Wang, H.; van der Donk, W. A. High Divergence of the Precursor Peptides in Combinatorial Lanthipeptide Biosynthesis. *ACS Chem. Biol.* **2014**, *9* (11), 2686-2694.
- (29) Li, B.; Sher, D.; Kelly, L.; Shi, Y.; Huang, K.; Knerr, P. J.; Joewono, I.; Rusch, D.; Chisholm, S. W.; van der Donk, W. A. Catalytic Promiscuity in the Biosynthesis of Cyclic Peptide Secondary Metabolites in Planktonic Marine Cyanobacteria. *Proc. Natl. Acad. Sci. USA* **2010**, *107* (23), 10430-10435.
- (30) Sardar, D.; Schmidt, E. W. Combinatorial Biosynthesis of Ripples: Docking with Marine Life. *Curr. Opin. Chem. Biol.* **2016**, *31*, 15-21.
- (31) Burkhart, B. J.; Schwalen, C. J.; Mann, G.; Naismith, J. H.; Mitchell, D. A. YcaO-Dependent Posttranslational Amide Activation: Biosynthesis, Structure, and Function. *Chem. Rev.* **2017**, *117* (8), 5389-5456.
- (32) van Heel, A. J.; Mu, D.; Montalban-Lopez, M.; Hendriks, D.; Kuipers, O. P. Designing and Producing Modified, New-to-Nature Peptides with Antimicrobial Activity by Use of a Combination of Various Lantibiotic Modification Enzymes. *ACS Synth. Biol.* **2013**, *2* (7), 397-404.

- (33) Zhang, Q.; van der Donk, W. A. Catalytic Promiscuity of a Bacterial A-N-Methyltransferase. *FEBS Lett.* **2012**, *586* (19), 3391-3397.
- (34) Sardar, D.; Lin, Z.; Schmidt, E. W. Modularity of Ripp Enzymes Enables Designed Synthesis of Decorated Peptides. *Chem. Biol.* **2015**, *22* (7), 907-916.
- (35) Zhao, X.; Kuipers, O. P. Identification and Classification of Known and Putative Antimicrobial Compounds Produced by a Wide Variety of Bacillales Species. *BMC genomics* **2016**, *17*, 882.
- (36) Hao, Y.; Pierce, E.; Roe, D.; Morita, M.; McIntosh, J. A.; Agarwal, V.; Cheatham, T. E., 3rd; Schmidt, E. W.; Nair, S. K. Molecular Basis for the Broad Substrate Selectivity of a Peptide Prenyltransferase. *Proc. Natl. Acad. Sci. USA* **2016**, *113* (49), 14037-14042.
- (37) Dunbar, K. L.; Tietz, J. I.; Cox, C. L.; Burkhart, B. J.; Mitchell, D. A. Identification of an Auxiliary Leader Peptide-Binding Protein Required for Azoline Formation in Ribosomal Natural Products. *J. Am. Chem. Soc.* **2015**, *137* (24), 7672-7677.
- (38) Repka, L. M.; Chekan, J. R.; Nair, S. K.; van der Donk, W. A. Mechanistic Understanding of Lanthipeptide Biosynthetic Enzymes. *Chem. Rev.* **2017**, *117* (8), 5457-5520.
- (39) Garg, N.; Salazar-Ocampo, L. M.; van der Donk, W. A. In Vitro Activity of the Nisin Dehydratase NisB. *Proc. Natl. Acad. Sci. USA* **2013**, *110* (18), 7258-7263.
- (40) Ortega, M. A.; Hao, Y.; Zhang, Q.; Walker, M. C.; van der Donk, W. A.; Nair, S. K. Structure and Mechanism of the Trna-Dependent Lantibiotic Dehydratase NisB. *Nature* **2015**, *517* (7535), 509-512.
- (41) Khusainov, R.; Moll, G. N.; Kuipers, O. P. Identification of Distinct Nisin Leader Peptide Regions That Determine Interactions with the Modification Enzymes NisB and NisC. *FEBS open bio* **2013**, *3*, 237-242.
- (42) Plat, A.; Kluskens, L. D.; Kuipers, A.; Rink, R.; Moll, G. N. Requirements of the Engineered Leader Peptide of Nisin for Inducing Modification, Export, and Cleavage. *Appl. Environ. Microbiol.* **2011**, *77* (2), 604-611.
- (43) Chatterjee, C.; Patton, G. C.; Cooper, L.; Paul, M.; van der Donk, W. A. Engineering Dehydro Amino Acids and Thioethers into Peptides Using Lacticin 481 Synthetase. *Chem. Biol.* **2006**, *13* (10), 1109-1117.
- (44) Goto, Y.; Ito, Y.; Kato, Y.; Tsunoda, S.; Suga, H. One-Pot Synthesis of Azoline-Containing Peptides in a Cell-Free Translation System Integrated with a Posttranslational Cyclodehydratase. *Chem. Biol.* **2014**, *21* (6), 766-774.
- (45) Rink, R.; Wierenga, J.; Kuipers, A.; Kluskens, L. D.; Driessen, A. J. M.; Kuipers, O. P.; Moll, G. N. Production of Dehydroamino Acid-Containing Peptides by *Lactococcus Lactis*. *Appl. Environ. Microbiol.* **2007**, *73* (6), 1792-1796.

- (46) Lubelski, J.; Overkamp, W.; Kluskens, L. D.; Moll, G. N.; Kuipers, O. P. Influence of Shifting Positions of Ser, Thr, and Cys Residues in Prenisin on the Efficiency of Modification Reactions and on the Antimicrobial Activities of the Modified Prepeptides. *Appl. Environ. Microbiol.* **2008**, *74* (15), 4680-4685.
- (47) McKinnie, S. M. K.; Ross, A. C.; Little, M. J.; Vederas, J. C. The Solid Phase Supported Peptide Synthesis of Analogues of the Lantibiotic Lactocin S. *MedChemComm* **2012**, *3* (8), 971-975.
- (48) Zhang, F.; Kelly, W. L. Saturation Mutagenesis of Tsra Ala4 Unveils a Highly Mutable Residue of Thiostrepton A. *ACS Chem. Biol.* **2015**, *10* (4), 998-1009.
- (49) Fluhe, L.; Knappe, T. A.; Gattner, M. J.; Schafer, A.; Burghaus, O.; Linne, U.; Marahiel, M. A. The Radical SAM Enzyme AlbA Catalyzes Thioether Bond Formation in Subtilosin A. *Nat. Chem. Biol.* **2012**, *8* (4), 350-357.
- (50) Himes, P. M.; Allen, S. E.; Hwang, S.; Bowers, A. A. Production of sactipeptides in Escherichia Coli: Probing the Substrate Promiscuity of Subtilosin A Biosynthesis. *ACS Chem. Biol.* **2016**, *11* (6), 1737-1744.
- (51) Fluhe, L.; Marahiel, M. A. Radical S-Adenosylmethionine Enzyme Catalyzed Thioether Bond Formation in Sactipeptide Biosynthesis. *Curr. Opin. Chem. Biol.* **2013**, *17* (4), 605-12.
- (52) Melby, J. O.; Dunbar, K. L.; Trinh, N. Q.; Mitchell, D. A. Selectivity, Directionality, and Promiscuity in Peptide Processing from a Bacillus Sp. Al Hakam Cyclodehydratase. *J. Am. Chem. Soc.* **2012**, *134* (11), 5309-5316.
- (53) Tang, W.; Jimenez-Oses, G.; Houk, K. N.; van der Donk, W. A. Substrate Control in Stereoselective Lanthionine Biosynthesis. *Nat. Chem.* **2015**, *7* (1), 57-64.
- (54) Yu, Y.; Mukherjee, S.; van der Donk, W. A. Product Formation by the Promiscuous Lanthipeptide Synthetase ProcM Is under Kinetic Control. *J. Am. Chem. Soc.* **2015**, *137* (15), 5140-5148.
- (55) Thibodeaux, C. J.; Ha, T.; van der Donk, W. A. A Price to Pay for Relaxed Substrate Specificity: A Comparative Kinetic Analysis of the Class II Lanthipeptide Synthetases ProcM and HalM2. *J. Am. Chem. Soc.* **2014**, *136* (50), 17513-17529.
- (56) Mukherjee, S.; van der Donk, W. A. Mechanistic Studies on the Substrate-Tolerant Lanthipeptide Synthetase ProcM. *J. Am. Chem. Soc.* **2014**, *136* (29), 10450-10459.
- (57) Yang, X.; van der Donk, W. A. Post-Translational Introduction of D-Alanine into Ribosomally Synthesized Peptides by the Dehydroalanine Reductase Npnj. *J. Am. Chem. Soc.* **2015**, *137* (39), 12426-12429.

- (58) Ortega, M. A.; Hao, Y.; Walker, M. C.; Donadio, S.; Sosio, M.; Nair, S. K.; van der Donk, W. A. Structure and Trna Specificity of MibB, a Lantibiotic Dehydratase from Actinobacteria Involved in NAI-107 Biosynthesis. *Cell Chem. Biol.* **2016**, *23* (3), 370-80.
- (59) Shi, Y.; Yang, X.; Garg, N.; van der Donk, W. A. Production of Lantipeptides in Escherichia Coli. *J. Am. Chem. Soc.* **2011**, *133* (8), 2338-2341.
- (60) Ortega, M. A.; Cogan, D. P.; Mukherjee, S.; Garg, N.; Li, B.; Thibodeaux, G. N.; Maffioli, S. I.; Donadio, S.; Sosio, M.; Escano, J., et al. Two Flavoenzymes Catalyze the Post-Translational Generation of 5-Chlorotryptophan and 2-Aminovinyl-Cysteine During NAI-107 Biosynthesis. *ACS Chem Biol* **2017**, *12* (2), 548-557.
- (61) Tolia, N. H.; Joshua-Tor, L. Strategies for Protein Coexpression in Escherichia Coli. *Nat. Methods* **2006**, *3* (1), 55-64.
- (62) McDaniel, L. E.; Bailey, E. G. Effect of Shaking Speed and Type of Closure on Shake Flask Cultures. *Appl. Microbiol.* **1969**, *17* (2), 286-290.
- (63) Ortega, M. A.; Cogan, D. P.; Mukherjee, S.; Garg, N.; Li, B.; Thibodeaux, G. N.; Maffioli, S. I.; Donadio, S.; Sosio, M.; Escano, J., et al. Two Flavoenzymes Catalyze the Post-Translational Generation of 5-Chlorotryptophan and 2-Aminovinyl-Cysteine During NAI-107 Biosynthesis. *ACS Chem. Biol.* **2017**, *12* (2), 548-557.
- (64) Dunbar, K. L.; Mitchell, D. A. Insights into the Mechanism of Peptide Cyclodehydrations Achieved through the Chemoenzymatic Generation of Amide Derivatives. *J. Am. Chem. Soc.* **2013**, *135* (23), 8692-8701.
- (65) Tang, W.; van der Donk, W. A. Structural Characterization of Four Prochlorosins: A Novel Class of Lantipeptides Produced by Planktonic Marine Cyanobacteria. *Biochemistry* **2012**, *51* (21), 4271-4279.
- (66) Liu, W.; Chan, A. S.; Liu, H.; Cochrane, S. A.; Vederas, J. C. Solid Supported Chemical Syntheses of Both Components of the Lantibiotic Lacticin 3147. *J. Am. Chem. Soc.* **2011**, *133* (36), 14216-14219.
- (67) Wieckowski, B. M.; Hegemann, J. D.; Mielcarek, A.; Boss, L.; Burghaus, O.; Marahiel, M. A. The Pqqd Homologous Domain of the Radical Sam Enzyme Thnb Is Required for Thioether Bond Formation During Thurincin H Maturation. *FEBS Lett.* **2015**, *589* (15), 1802-1806.
- (68) Zhang, Z.; Hudson, G. A.; Mahanta, N.; Tietz, J. I.; van der Donk, W. A.; Mitchell, D. A. Biosynthetic Timing and Substrate Specificity for the Thiopeptide Thiomuracin. *J. Am. Chem. Soc.* **2016**, *138* (48), 15511-15514.

- (69) Xie, L.; Miller, L. M.; Chatterjee, C.; Averin, O.; Kelleher, N. L.; van der Donk, W. A. Lacticin 481: In Vitro Reconstitution of Lantibiotic Synthetase Activity. *Science* **2004**, *303* (5658), 679-681.
- (70) Held, D.; Yaeger, K.; Novy, R. New Coexpression Vectors for Expanded Compatibilities in E. Coli. *InNovations* **2003**, *18*, 4-6.

Appendix A: PDF Reprints of other coauthored publications

A.1 Structural and functional insight into an unexpectedly selective *N*-methyltransferase involved in plantazolicin biosynthesis

Reproduced with permission from Lee, J.; Hao, Y.; Blair, P.M.; Melby, J.O.; Agarwal, V.; Burkhart, B.J.; Nair, S.K.; Mitchell, D.A. Structural and functional insight into an unexpectedly selective *N*-methyltransferase involved in plantazolicin biosynthesis. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 12954-12959. PMID: PMC3740862. Copyright 2013 National Academy of Sciences.

I used PyMOL to analyze the structure of two methyltransferases, BamL and BpumL, and to model peptide substrates. This analysis revealed potential constraints for substrate recognition and provided insight into the narrow substrate scope of these two enzymes. I created Figure S5 and generated mutant proteins (BamL T38A, T38F, L132A, L132F, L162A, and L162F).

Structural and functional insight into an unexpectedly selective *N*-methyltransferase involved in plantazolicin biosynthesis

Jaeheon Lee^{a,1}, Yue Hao^{a,b,1}, Patricia M. Blair^c, Joel O. Melby^{a,c}, Vinayak Agarwal^{a,d}, Brandon J. Burkhart^{a,c}, Satish K. Nair^{a,b,d,2}, and Douglas A. Mitchell^{a,c,e,2}

^aInstitute for Genomic Biology, Departments of ^bBiochemistry, ^cChemistry, and ^dMicrobiology, and ^eCenter for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801

Edited by Gregory A. Petsko, Brandeis University, Waltham, MA, and approved June 28, 2013 (received for review April 1, 2013)

Plantazolicin (PZN), a polyheterocyclic, *N,N*'-dimethylarginine-containing antibiotic, harbors remarkably specific bactericidal activity toward strains of *Bacillus anthracis*, the causative agent of anthrax. Previous studies demonstrated that genetic deletion of the *S*-adenosyl-L-methionine-dependent methyltransferase from the PZN biosynthetic gene cluster results in the formation of desmethylPZN, which is devoid of antibiotic activity. Here we describe the in vitro reconstitution, mutational analysis, and X-ray crystallographic structure of the PZN methyltransferase. Unlike all other known small molecule methyltransferases, which act upon diverse substrates in vitro, the PZN methyltransferase is uncharacteristically limited in substrate scope and functions only on desmethylPZN and close derivatives. The crystal structures of two related PZN methyltransferases, solved to 1.75 Å (*Bacillus amyloliquefaciens*) and 2.0 Å (*Bacillus pumilus*), reveal a deep, narrow cavity, putatively functioning as the binding site for desmethylPZN. The narrowness of this cavity provides a framework for understanding the molecular basis of the extreme substrate selectivity. Analysis of a panel of point mutations to the methyltransferase from *B. amyloliquefaciens* allowed the identification of residues of structural and catalytic importance. These findings further our understanding of one set of orthologous enzymes involved in thiazole/oxazole-modified microcin biosynthesis, a rapidly growing sector of natural products research.

enzymology | mutagenesis | RIPP natural product

Plantazolicin (PZN) is a poly-azol(in)e-containing molecule of ribosomal origin from the plant-growth promoting bacterium, *Bacillus amyloliquefaciens* FZB42 (1-3). PZN exhibits selective bactericidal activity toward *Bacillus anthracis* (3). All of the genes required for PZN production, immunity, and export cluster within a 10-kb region of the FZB42 genome (Fig. 1*A*). Genome mining has identified highly similar PZN biosynthetic gene clusters in *Bacillus pumilus*, *Clavibacter michiganensis* subsp. *sepedonicus*, *Corynebacterium urealyticum*, and *Brevibacterium linens* (3). PZN is biosynthesized from a 41-residue, inactive precursor peptide (Fig. 1*A*). Distinguishing chemical features of PZN are the two contiguous poly-azol(in)e moieties, which like all thiazole/oxazole-modified microcin (TOMM) natural products, originate from Cys and Ser/Thr residues on the C-terminal region of the precursor peptide (4-6). During heterocycle formation, a cyclodehydratase first converts Cys and Ser/Thr to thiazoline and (methyl)oxazoline, respectively. This ATP-dependent transformation formally removes water from the preceding amide bond (7-10). Subsequent dehydrogenation yields the aromatic thiazole and (methyl)oxazole (11). During PZN maturation, all 10 Cys and Ser/Thr residues within the C-terminal core region are cyclized, yielding 9 azole heterocycles and 1 methylloxazoline (Fig. 1*B*). Further modification includes leader peptide proteolysis and methylation to yield the final metabolite. The PZN methyltransferase dimethylates the N terminus (Arg) and is *S*-adenosyl-L-methionine (SAM)-dependent. The N-terminal methylation of

ribosomal peptides from bacteria is exceptionally rare. With respect to natural products, the only compounds that we are aware of that undergo N-terminal dimethylation besides PZN are the linaridin antibiotics (e.g., cypemycin, grisemycin), the best studied of which contain *N,N*'-dimethylAla (12-14).

Through genetic manipulation and alteration of cultivation conditions, PZN biosynthetic intermediates have been intercepted, permitting investigation into the details of downstream tailoring reactions and the bioactivity of partially processed substrates (3). After genetic deletion of the methyltransferase from *B. amyloliquefaciens*, desmethylPZN was isolated in roughly equivalent yield (1). DesmethylPZN was found to be devoid of antibiotic activity, as was also shown for desmethylcypemycin (12). As with PZN, the corresponding methyltransferase for cypemycin (CypM) tailoring was confirmed through deletion studies (12) and has recently been characterized in vitro (14). Although CypM and the PZN methyltransferase catalyze similar reactions, they do not share significant amino acid sequence similarity outside of the predicted SAM-binding sites. In this work, we report the in vitro reconstitution of the PZN methyltransferase, which exhibits a striking selectivity for polyheterocyclized substrates. We have also determined the high-resolution cocrystal structures of the PZN methyltransferases from *B. amyloliquefaciens* (BamL) and *B. pumilus* (BpumL), each bound to *S*-adenosyl-L-homocysteine (SAH), and have carried out detailed characterization of site-specific variants of BamL. These studies provide a molecular rationale for the unexpected selectivity for an otherwise broadly acting superfamily of catalysts.

Results

Enzymatic Conversion of DesmethylPZN to PZN. To assess enzymatic activity, we cloned, expressed, and purified BamL as a fusion with maltose-binding protein (MBP). MBP-BamL was then added to reactions containing desmethylPZN, SAM, and a nucleosidase that converts the SAH by-product to adenine and *S*-ribosyl-L-homocysteine, minimizing potential product inhibition (Pfs) (15). Reactions were initiated by the addition of SAM, quenched at desired time points, and analyzed by MALDI-MS (Fig. 2). In the presence of desmethylPZN and SAM, MBP-BamL

Author contributions: J.L., Y.H., S.K.N., and D.A.M. designed research; J.L., Y.H., P.M.B., J.O.M., V.A., and B.J.B. performed research; P.M.B. contributed new reagents/analytic tools; J.L., Y.H., P.M.B., J.O.M., V.A., B.J.B., S.K.N., and D.A.M. analyzed data; and J.L., S.K.N., and D.A.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 4KVZ and 4KWC).

¹J.L. and Y.H. contributed equally to this work.

²To whom correspondence may be addressed. E-mail: douglasm@illinois.edu or s-nair@life.illinois.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1306101110/-DCSupplemental.

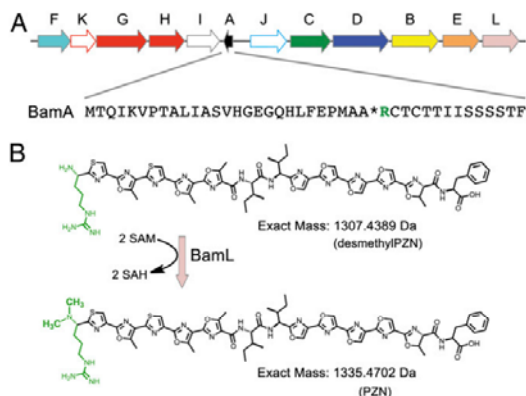


Fig. 1. PZN biosynthesis. (A) The PZN biosynthetic gene cluster with single-letter codes above each ORF. The amino acid sequence of the PZN precursor peptide (designated as A) is shown. *, leader peptide cleavage site; B, dehydrogenase; C/D, cyclodehydratase; E, putative leader peptidase; F, putative immunity protein; G/H, ABC transporters; green, site of dimethylation; I/J, unknown function; K, transcriptional regulator; L, methyltransferase. (B) The BamL and SAM-dependent conversion of desmethylPZN to PZN with the N-terminal Arg residues shown in green. Note that the most C-terminal heterocycle is a methyloxazoline, which can be selectively hydrolyzed to yield Thr.

catalyzed the formation of PZN (m/z 1336). Omitting either SAM or BamL resulted in no product formation. Identical results were achieved with the BpumL methyltransferase (from *B. pumilus*), another bona fide producer of PZN (3) (Fig. 2). Removal of the MBP tag from BamL led to a lower level of PZN formation, suggesting that MBP does not interfere with methyltransferase activity (SI Appendix, Table S1). The higher level of PZN formation with MBP was likely due to increased stability conferred by the tag (16). Using the MALDI-MS end point assay, we further evaluated the activity of MBP-BamL by testing the effect of pH, buffer, and a variety of other additives (SI Appendix, Table S1). Omission of Pfs from the reaction led to only a modest decrease (~25%) in product formation. Using isothermal titration calorimetry (ITC), we obtained a dissociation constant between SAM and BamL of $7.5 \pm 0.8 \mu\text{M}$ (SI Appendix, Fig. S1). In accord with lack of SAH by-product inhibition, identical ITC runs with SAH generated unusual titration curves that could not be mathematically fit to any reasonable binding model. To confirm the site of BamL dimethylation, collision-induced dissociation (CID) spectra were obtained using Fourier transform MS. A previously characterized diagnostic fragment ion, m/z 1277.4291 Da, was observed, indicating that the site of methylation was on the expected N-terminal α amine (SI Appendix, Fig. S2) (3).

Despite the variety of in vitro reconstitution reactions screened (SI Appendix, Table S1), monomethylPZN was never detected (m/z 1322). Even at shorter time points, when consumption of desmethylPZN was incomplete, monomethylPZN was not observed. In an additional effort to detect this species, we performed BamL reactions under pseudosingle turnover conditions using variable concentrations of BamL, desmethylPZN, and SAM (SI Appendix, Fig. S3). Under no condition was monomethylPZN (m/z 1322) detected. However, when all reaction components were supplied at 100 μM , we observed a minor peak consistent with hydrolyzed (+18 Da) monomethylPZN (m/z 1340). Unfortunately, the low signal-to-noise ratio did not allow for structural confirmation.

DesmethylPZN Substrate Analogs. Because of the hydrolytic instability of PZN, we observed by MALDI-MS hydrolyzed

desmethylPZN (m/z 1326) and hydrolyzed PZN (m/z 1354) in all in vitro reactions (Fig. 2 and SI Appendix, Fig. S3). These species arise from the hydrolysis of the sole methyloxazoline ring. We have previously reported on the lability of this heterocycle, which reinstates the most C-terminal Thr (Fig. 1) (3). To establish whether BamL could directly accept hydrolyzed desmethylPZN (m/z 1326) as a substrate to produce hydrolyzed PZN (m/z 1354), we subjected desmethylPZN to conditions that yielded quantitative conversion to hydrolyzed desmethylPZN for BamL substrate testing (SI Appendix, Materials and Methods). Analysis of reaction products run under identical conditions showed that BamL processed hydrolyzed desmethylPZN with efficiency approximately equal to that of desmethylPZN (Figs. 2 and 3A and B). From earlier work, we noted that another desmethylPZN variant harboring two methyloxazolines (dihydrodesmethylPZN, m/z 1310) was readily accessible by using oxygen-saturated cultivation (3). Because of separation difficulties, we tested a mixture of m/z 1308, 1310, 1326, and 1328 (desmethylPZN analog mixture) for processing by BamL. Analysis of reaction products by MALDI-MS showed the presence of PZN (m/z 1336), dihydroPZN (m/z 1338, note isotopic ratio), hydrolyzed PZN (m/z 1354), and hydrolyzed dihydroPZN (m/z 1356, Fig. 3C). These data indicate BamL successfully processed dihydrodesmethylPZN.

Unusual Substrate Selectivity of BamL. To further explore substrate permissiveness, we tested bradykinin and a synthetic peptide termed “RPG” (sequence given in Fig. 3) as BamL substrates. To a first approximation, bradykinin and the RPG peptide mimic desmethylPZN. Both harbor an N-terminal Arg, are similar in size to desmethylPZN, and contain Pro where several azole heterocycles are found in desmethylPZN (Fig. 3A). As with Pro, azoles are five-membered, nitrogen-containing heterocycles that serve as peptide conformational restraints. Although not a perfect match for an azole, the ability of Pro to structurally and functionally substitute in TOMM natural products has been documented (17). Upon treatment of bradykinin and the RPG

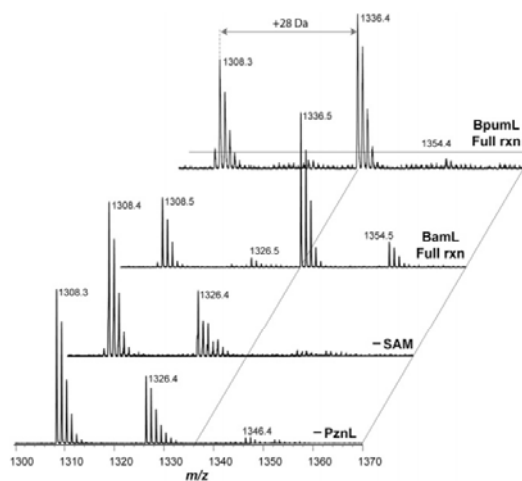


Fig. 2. Activity of purified methyltransferases. MALDI-MS was used to monitor the conversion of desmethylPZN (m/z 1308) to PZN (m/z 1336). The m/z 1326 and 1354 species represent hydrolyzed desmethylPZN and PZN, respectively. Samples containing all reaction components used either the methyltransferase from BamL or BpumL. The SAM cosubstrate and BamL enzyme were omitted from the samples labeled -SAM and -BamL, respectively. PznL, PZN methyltransferase (either BamL or BpumL); rxn, reaction.

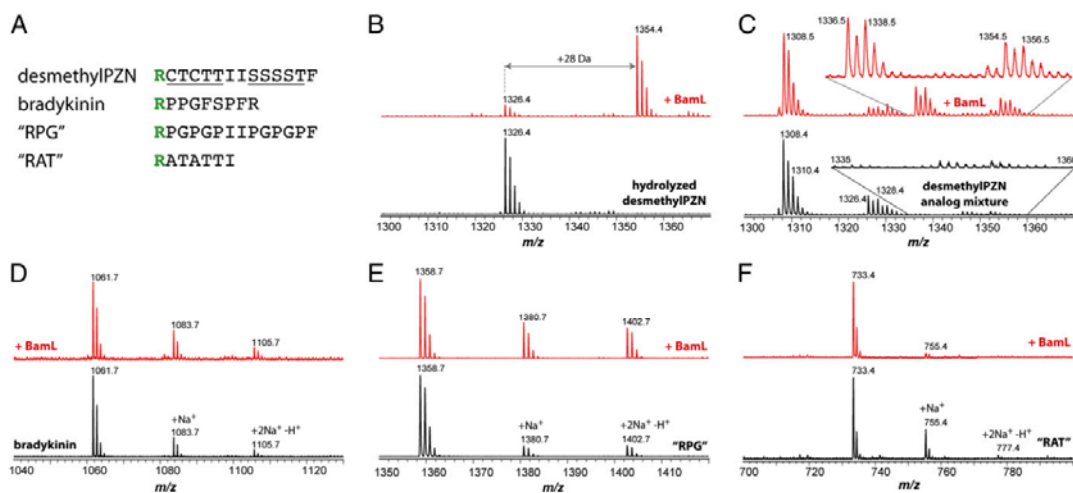


Fig. 3. Substrate tolerance of BamL. MALDI-MS was used to assess the ability of BamL to process substrates other than desmethylPZN. (A) Subset of peptides tested as substrates with the N-terminal Arg shown in bold. Heterocyclized residues of desmethylPZN are underlined. The most C-terminal heterocycle is reverted to Thr in hydrolyzed desmethylPZN. (B–F) Bottom spectra, unreacted peptides; top spectra, after reaction with BamL. (B) Hydrolyzed desmethylPZN (m/z 1326). (C) DesmethylPZN analog mixture that contained desmethylPZN (m/z 1308), dihydrodesmethylPZN (m/z 1310), and the hydrolyzed forms (m/z 1326 and 1328, respectively). (D) Bradykinin (m/z 1061). (E) “RPG” peptide (m/z 1358). (F) “RAT” peptide (m/z 733).

peptide with BamL and SAM, no reaction products were detected, even after increasing the enzyme concentration and allowing for extended reaction times (Fig. 3 *D* and *E*). Similarly, adrenocorticotrophic hormone peptide 18–39, with an N-terminal amino acid sequence of RPV, was not processed by BamL. Because Pro is an imperfect match for an azole, we also tested whether BamL would process an additional synthetic peptide “RAT” (sequence given in Fig. 3), which contained the left segment of the native PZN precursor sequence except that the Cys residues were replaced with Ala. Again, no reaction product was observed, even after increasing the enzyme concentration to near single turnover levels (Fig. 3*F*). Two tetrapeptides, RAAA and RGGG, were also evaluated as BamL substrates, but no reaction products were detected. As a final test, BamL accepted Arg amide (carboxylate replaced with a neutral amide) as a substrate, albeit a very inefficient one. After increased reaction time with elevated substrate concentration, high-resolution MS identified a mass consistent with dimethylArg amide (experimental 202.1677 Da, theoretical 202.1688 Da, error 4.5 ppm). The very low intensity of this peak suggested that <5% of the starting material was converted. These results suggested that BamL does not simply recognize peptides with an N-terminal Arg and furthermore is intolerant to substrates not containing desmethylPZN-like (i.e., polyheterocyclic) functionality. In contrast to the specificity exhibited by BamL, other characterized SAM-dependent methyltransferases acting on small molecules and peptides have much broader substrate scopes. For example, the methyltransferases involved in rebeccamycin (18, 19), dehydrophos (20), aminocoumarin (21), and CypM (14) biosynthesis methylate a wide range of substrates.

Structure Determination of Two PZN Methyltransferases. To provide insight into the unexpected lack of activity toward non-heterocyclized peptide substrates, we determined the crystal structures of BamL (1.75 Å resolution) and BpumL (2.0 Å resolution), each in complex with SAH. Initial crystallographic phases were determined at 3.2 Å resolution using SeMet-labeled crystals of BpumL, followed by model building and refinement

against higher resolution data until convergence. Phases for BamL were determined by molecular replacement using a partial model of BpumL. Relevant data collection and refinement statistics are given in *SI Appendix, Table S2*. The BamL and BpumL structures consist of a core Rossmann-fold domain composed of seven β -strands surrounded by six α -helices, with an overall architecture similar to other methyltransferases (Fig. 4 *A* and *B*). As expected from the 48% sequence identity between BamL and BpumL, the structures are superimposable with an rmsd of 1.1 Å over 256 aligned α carbons. A DALI search against the Protein Data Bank (PDB) identifies the closest structural homologs as the bacterial Hen1 methyltransferase (PDB ID code, 3JWG; Z-score, 16.8; rmsd, 2.5 Å over 174 aligned α carbons) and the hypothetical bacterial YecO protein (PDB ID code, 1IM8; Z-score, 16.7; rmsd, 2.9 Å over 194 aligned α carbons). The conservation in structure is restricted to the core Rossmann fold domain and persists despite minimal similarities (roughly 10–13% identity) across the primary structure.

Unlike the structures of typical small molecule methyltransferases, the architectures of BamL and BpumL are minimalistic and lack additional secondary structure decorations. One molecule of SAH is bound across the interior of BamL with the adenine ring enclosed in a cavity defined along the sides by S92, F137, and a loop encompassing residues S112–A114 across the top. The side chain of D91 is within interaction distance with the 2' hydroxyl of the ribose (Fig. 4*C*). BamL residues R42 and H131 provide additional contacts with SAH via the homocysteine carboxylate. Notably, the trajectory of the homocysteine moiety of SAH is not collinear with the adenine but is deflected toward the center of the polypeptide by F21 and a loop encompassing G68–Q71. This would position the electrophilic methyl group of SAM at the base of a long tunnel running to the surface of the protein. The walls of this tunnel are defined by a number of hydrophobic residues including F21, Y33, T38, L132, L162, Y182, and L183. These residues are strictly conserved between BamL and BpumL, with the exception of L183, which is I181 in BpumL (*SI Appendix, Fig. S4*). Because the depth of this tunnel is of sufficient length to accommodate the first four to five residues

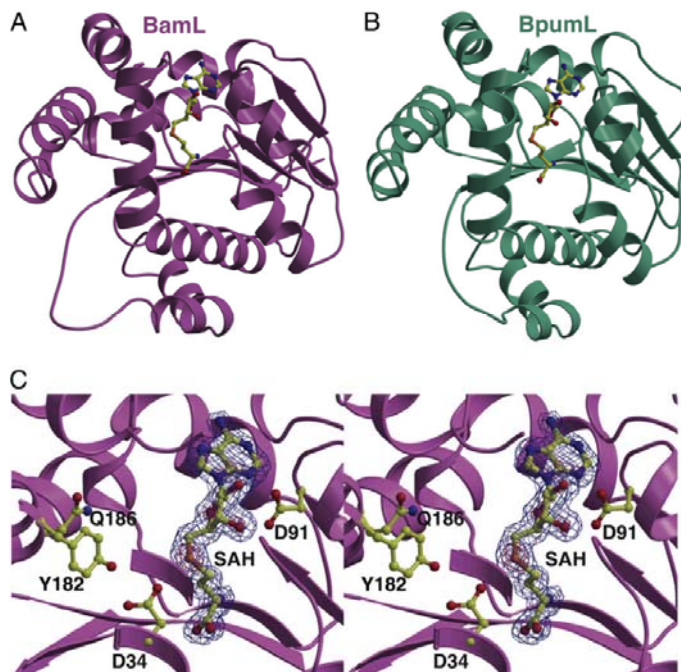


Fig. 4. Ribbon diagrams derived from the X-ray crystal structures of (A) BamL and (B) BpumL, with the bound SAH ligand shown in yellow ball-and-stick representation. (C) Electron density maps calculated using Fourier coefficients ($F_{obs} - F_{calc}$) with phases derived from the refined 1.75 Å resolution structure of BamL calculated with the coordinates for SAH omitted before one round of crystallographic refinement. The map is contoured at 2.6 σ (blue mesh) and 8 σ (red mesh) and the final refined coordinates are superimposed. The coordinates for SAH and polar active site residues that are important in catalysis are shown as yellow (ball-and-stick).

of desmethylPZN, we presume that it defines the substrate-binding pocket (*SI Appendix, Fig. S5*). At the base of the tunnel, a shorter, perpendicular cavity is found. This shorter cavity is immediately adjacent to the presumptive position of the electrophilic methyl group of SAM and is of suitable size to accommodate the N-terminal Arg side chain of (desmethyl)PZN. Residues D34, Y182, and Q186 of BamL presumably stabilize the N-terminal amine of the substrate (Fig. 4C). Evidence for the role of these residues in substrate engagement is borne out by the mutational analyses described in the following section.

Selection and Enzymatic Activity of BamL Mutants. To better understand catalysis, we aligned the amino acid sequence of BamL with a series of well-characterized methyltransferases. With the exception of identifying residues that comprise the SAM-binding pocket, this exercise shed little light on what residues may be important for BamL catalysis because of the poor conservation across the methyltransferase superfamily (22). Fortunately, sequence alignment of the five known PZN methyltransferases (including BpumL) identified in our earlier work (3) revealed a manageable number of conserved residues with which to begin the enzymological dissection of BamL (*SI Appendix, Fig. S4*). We thus targeted eight residues of BamL to be replaced with Ala (*SI Appendix, Table S3*). We then analyzed the structures of BamL and BpumL cocrystallized with SAH and mutated seven additional positions hypothesized to be critical either for substrate binding (desmethylPZN and SAM) or catalysis. At four of these locations, the wild-type (WT) residue was replaced with Ala and an additional alternative residue, ultimately yielding a database of 19 mutant proteins (Table 1).

To establish if any point mutant reduced enzymatic activity or structural instability, we first expressed and purified all 19 mutant BamL proteins from *Escherichia coli* under identical conditions (*SI Appendix, Materials and Methods*). BamL W20A, F21A, D91A,

and D126A underwent extensive proteolytic degradation and were low yielding, indicating that these positions play an important role in structural stability (*SI Appendix, Fig. S6*). In an effort to obtain values for the kinetic constants k_{cat} and K_m for the mutant panel, we attempted quantitative liquid chromatography MS, coupled-enzymatic assays, and tritiated SAM assays. Despite many attempts, a number of problems arose that limited our ability to determine the desired kinetic parameters. As an alternative, we used internally calibrated MALDI-MS to compare the ion intensities of the desmethylPZN starting material with the PZN product under three reaction conditions (Table 1). In what is referred to as reaction condition A [10 μ M MBP-BamL, 10 μ M Pfs (SAH nucleosidase), 50 μ M desmethylPZN, and 3 mM SAM for 16 h at 37 °C], all tested mutants except for BamL T38A, T38F, L132A, L132F, Y182F, and S190A showed impaired activity. In a more stringent test of activity, the concentrations of MBP-BamL, Pfs, and desmethylPZN were all lowered 10-fold, and the reaction time was limited to 1 h at 22 °C (referred to as reaction condition B). Under condition B, we only observed product formation from BamL T38A and S190A. Importantly, the SAM concentration was intentionally kept high (3 mM) in reaction condition B. To evaluate if a BamL point mutant affected the ability to properly handle SAM, we examined a subset of the mutants in reaction condition C. Here, the desmethylPZN concentration was returned to 50 μ M and the SAM concentration was reduced to 500 μ M (other variables were the same as in condition B). Upon comparison with condition B, detectable product formation was resurrected from reactions with BamL T38F, L132A, L162A, and Y182F (Table 1). This tripartite assay allowed for the generalized assessment of which residues were required for proper handling of the desmethylPZN and SAM cosubstrates.

Detection of Monoalkylated PZN Analogs. Although our earlier attempts to identify monomethylPZN as a BamL reaction

Table 1. Relative enzymatic activity of BamL mutants

| BamL | Cond. A conv. | Cond. B conv. | Cond. C conv. | BamL | Cond. A conv. | Cond. B conv. | Cond. C conv. |
|--------------------|---------------|--------------------|--------------------|-------|---------------|--------------------|--------------------|
| WT | 1.00 ± 0.16* | 1.00 ± 0.15* | 1.00 ± 0.08* | L132F | 1.00 ± 0.06 | <0.05 [†] | ND |
| W20A [‡] | ND | ND | ND | D161A | 0.16 ± 0.01 | <0.05 [†] | ND |
| F21A [‡] | ND | ND | ND | L162A | 0.65 ± 0.11 | <0.05 [†] | 0.24 ± 0.05 |
| D34A | 0.62 ± 0.08 | <0.05 [†] | ND | L162F | 0.76 ± 0.13 | <0.05 [†] | ND |
| T38A | 1.00 ± 0.03 | 0.59 ± 0.10 | 0.37 ± 0.12 | R164A | 0.43 ± 0.01 | <0.05 [†] | ND |
| T38F | 1.00 ± 0.04 | <0.05 [†] | 0.25 ± 0.03 | Y182A | 0.23 ± 0.08 | <0.05 [†] | ND |
| R42A | 0.40 ± 0.07 | <0.05 [†] | <0.05 [†] | Y182F | 0.96 ± 0.03 | <0.05 [†] | 0.37 ± 0.02 |
| D91A [‡] | ND | ND | ND | D185A | 0.77 ± 0.06 | <0.05 [†] | ND |
| D126A [‡] | ND | ND | ND | Q186A | 0.45 ± 0.07 | <0.05 [†] | <0.05 [†] |
| L132A | 0.92 ± 0.08 | <0.05 [†] | 0.59 ± 0.10 | S190A | 1.00 ± 0.07 | 0.74 ± 0.32 | ND |

Relative activity was determined by comparing the ion intensities of the starting material to product by MALDI-TOF-MS. Results derive from two independent enzyme preparations and at least two independent reactions per prep. Values are normalized to that obtained for WT enzyme. Cond. A: MBP-BamL (10 μM), Pfs (10 μM), desmethylPZN (50 μM), and SAM (3 mM) were added to Tris buffer (50 mM, pH 7.8) and allowed to proceed for 16 h at 37 °C. Cond. B: MBP-BamL (1 μM), Pfs (1 μM), desmethylPZN (5 μM), and SAM (3 mM) were added to Tris buffer (50 mM, pH 7.8) and allowed to proceed for 1 h at 22 °C. Cond. C: MBP-BamL (1 μM), Pfs (1 μM), desmethylPZN (50 μM), and SAM (500 μM) were added to Tris buffer (50 mM, pH 7.8) and allowed to proceed for 1 h at 22 °C. Cond., condition; conv., fraction converted; ND, not determined.

*Error is reported as SD of the mean ($n = 4$).

[†]The detection limit for product formation is estimated to be ~5% of that found in the WT reaction.

[‡]Several point mutants were extensively degraded during heterologous expression in *E. coli* (SI Appendix, Materials and Methods); thus, their activities were ND.

intermediate were unsuccessful, we observed a strong indication that this species was formed upon analysis of the BamL Y182F reaction using condition A (SI Appendix, Fig. S7). After supplying WT BamL to the Y182F reaction mixture containing m/z 1308, 1322, and 1336, all species converged to m/z 1336 (PZN). Convergence provided definitive evidence that m/z 1322 was indeed monomethylPZN, an on-pathway reaction intermediate as expected.

A close inspection of the SAM-binding pocket within BamL/BpumL suggested that slightly larger alkyl substituents might be tolerated (Fig. 4 and SI Appendix, Fig. S5). To evaluate if an ethyl group could be transferred from a SAM analog bearing an electrophilic ethyl group, rather than the naturally occurring methyl substituent, we synthesized “ethyl SAM” using an established procedure (23). Enzymatic reactions were carried out as performed previously. Analysis of the reaction product by MALDI-MS revealed the transfer of a single ethyl group to desmethylPZN, yielding monoethylPZN (SI Appendix, Fig. S7). This variant is isobaric with native PZN (m/z 1336), which bears two methyl groups. Subjecting this species to CID gave a diagnostic fragment ion of m/z 1277, which localized the ethyl modification to the N terminus. Intriguingly, despite increased reaction times, addition of fresh enzyme, and the use of either ethyl SAM or natural SAM, the corresponding dialkylated PZN derivative was not observed. This suggests that there are insurmountable steric barriers to the formation of monomethyl-monoethylPZN (m/z 1350) and diethylPZN (m/z 1364).

Discussion

In this work, we successfully reconstituted the in vitro activity of the methyltransferase involved in PZN biosynthesis (Fig. 2). We discovered that, in contrast to other small-molecule methyltransferases, BamL is highly specific for substrates closely related to desmethylPZN. Indeed, a variety of peptidic substrates containing N-terminal Arg residues, with and without conformationally restrictive Pro residues included, all failed to be processed (Fig. 3). The molecular basis for this unusual selectivity is that the PZN-binding cleft is sufficiently narrow that substrates lacking contiguous polyheterocyclic structures are simply too wide to be accommodated (Fig. 4 and SI Appendix, Fig. S5). Heterocyclization substantially reduces the width of a peptide as the carbonyl oxygens are removed and the β -nucleophilic side chains

(Cys, Ser/Thr) are backbone-cyclized. Moreover, contiguously heterocyclic peptides can more readily adopt the narrow, planar conformations required to reach the active site (Fig. 1 and SI Appendix, Fig. S5). We posit that the lack of activity toward the tetrapeptide substrates bearing minimal side chains (RAAA and RGGG) could arise from these and related features. Contrary to RAAA/RGGG, desmethylPZN can likely pass through the tunnel “pinch point” without adopting an energetically unfavorable conformation or having the enzyme undergo a major structural rearrangement. DesmethylPZN is also considerably more hydrophobic than the tetrapeptides, as evidenced by reverse-phase C₁₈ HPLC elution (desmethylPZN, 72% MeOH; RAAA/RGGG, <15% MeOH for both). Moreover, desmethylPZN is structurally preorganized in a linear form; thus, the entropic penalty for enzyme binding should be vastly lower than for the more conformationally flexible tetrapeptides.

To establish if PZN methyltransferases harbor a narrow substrate-binding cleft, we solved the high-resolution crystal structures of two family members (BamL and BpumL), each in complex with SAH (Fig. 4). Although the overall fold of BamL/BpumL was characteristic of other known methyltransferases, a remarkable feature was a deep tunnel (14–15 Å) running from the surface of the protein toward the active site. Several generally conserved residues within the PZN methyltransferases define the walls of this tunnel, and an orthogonal cleft that we propose stabilizes the guanidinium group of PZN. Although attempts to crystallize BamL/BpumL with (desmethyl)PZN were unsuccessful, analysis of the available structures prompted us to postulate that the tunnel binds the five N-terminal residues of PZN (SI Appendix, Fig. S5). To experimentally support this model, a series of bioinformatics- and structure-guided point mutations were introduced into BamL. The resultant methyltransferase activities were measured in a three-tiered assay, alerting us to residues of potential importance in substrate binding, enzymatic catalysis, and protein structural stability (Table 1). Four BamL mutants (W20A, F21A, D91A, and D126A) were extensively degraded during *E. coli* heterologous expression and their activities were not determined (Table 1 and SI Appendix, Fig. S6). For mutants less prone to degradation, we first used an excess of enzyme, an elevated concentration of both substrates (SAM and desmethylPZN), and an extended reaction time to detect any trace of activity (condition A). The mutations tested that compose

the putative PZN substrate-binding tunnel were BamL D34A, T38A/F, L132A/F, L162A, Y182A/F, and Q186A. Of these, D34A, L162A, and Q186A exhibited decreased activity, and Y182A activity was severely impaired, likely because of structural instability (Table 1 and *SI Appendix*, Fig. S6). Under condition B, where the enzyme concentration, reaction time, and desmethylPZN concentration were greatly decreased, the only tunnel mutation that gave detectable product formation was T38A. To support that the tunnel mutants were indeed involved in PZN binding, the desmethylPZN concentration was raised and the SAM concentration was lowered (condition C). Product formation was reinstated for all tested mutations (i.e., T38F, L132A, L162A, and Y182F) except for R42A and Q186A, the latter of which is in close proximity to two structurally stabilizing positions (W20 and F21).

Using ITC, we measured a dissociation constant of 7.5 ± 0.8 μ M between BamL and SAM (*SI Appendix*, Fig. S1). In accord with the crystal structures, the three-tiered enzymatic assay supports critical roles for BamL R42 and D91 in binding SAM. As mentioned previously, enzymatic activity was not resurrected for R42A upon changing from condition B to condition C, indicative of an interaction with SAM (Table 1). The structure reveals that the side chain of BamL R42 is engaged in an ionic interaction with the carboxylate moiety of SAH. Further, the carboxylate side chain of BamL D91 makes several contacts with the SAH ribose hydroxyl groups, which upon mutation to Ala yielded degraded protein.

To be effective catalysts, BamL and other PZN methyltransferases need to precisely position the SAM methyl donor and acceptor (*SI Appendix*, Fig. S8). Given that amines tend to be protonated at physiological pH, BamL might be expected to include an active site base. Although the pK_a of this group could be dramatically perturbed in the interior of an enzyme, negating the need for an active site base, our data suggest that BamL D34 and Y182 may be directly involved in catalysis. Although residual activity remains with the point mutants D34A and Y182F (Table 1), the side chains of these residues are in close proximity to each other and the thioether sulfur of SAH (Fig. 4). Considering our

unsuccessful attempts to identify monomethylPZN as a reaction intermediate with WT BamL (*SI Appendix*, Fig. S3), it was gratifying to observe this intermediate with BamL Y182F (*SI Appendix*, Fig. S7). This result, taken together with the data in Table 1, suggests severely impaired catalysis. It is notable that two members of the PZN methyltransferase family, from *C. urealyticum* and *B. linens*, naturally harbor a Phe at this position (BamL Y182F) but a Tyr at BamL F21 (*SI Appendix*, Fig. S4). As with position 182, residue 21 is in close proximity to D34 and the thioether sulfur atom of SAH.

The results presented here have revealed that the enzyme responsible for converting desmethylPZN to PZN exhibits an unusually selective substrate scope in vitro. Given that PZN is a highly discriminating antibiotic with known activity only toward *B. anthracis*, the causative agent of anthrax, the characterization of the responsible biosynthetic enzymes remains an important undertaking. Our study lays a foundation for future work wherein the rational engineering of PZN variants bearing unnatural modifications can be produced for the purposes of establishing structure-activity relationships.

Materials and Methods

Detailed methods are provided in the *SI Appendix, Materials and Methods*. Briefly, these cover the experimental procedures followed for the preparation of starting materials, such as isolation of desmethylPZN, cloning, expression, and purification of methyltransferase enzymes, and the generation of site-directed mutants. Also given are the conditions for methyltransferase reconstitution, MS, protein crystallization, structure determination, protein analysis, and ITC.

ACKNOWLEDGMENTS. We thank John Cronan for the Pfs (SAH nucleosidase) expression plasmid and Xinyun Cao for her technical assistance. We thank members of the D.A.M. laboratory for critically reviewing this manuscript. The Bruker UltrafleXtreme MALDI mass spectrometer was purchased in part with National Center for Research Resources' National Institutes of Health (NIH) Grant S10 RR027109. Use of the Advanced Photon Source, operated for the US Department of Energy by the Argonne National Laboratory, was supported by Contract DE-AC02-06CH11357. This work was supported in part by the institutional funds provided by the University of Illinois, the NIH Director's New Innovator Award Program Grant DP2 OD008463 (to D.A.M.), and the Robert C. and Carolyn J. Springborn Endowment (to B.J.B.).

- Scholz R, et al. (2011) Plantazolicin, a novel microcin B17/streptolysin S-like natural product from *Bacillus amyloliquefaciens* FZB42. *J Bacteriol* 193(1):215–224.
- Kalyon B, et al. (2011) Plantazolicin A and B: Structure elucidation of ribosomally synthesized thiazole/oxazole peptides from *Bacillus amyloliquefaciens* FZB42. *Org Lett* 13(12):2996–2999.
- Molohon KJ, et al. (2011) Structure determination and interception of biosynthetic intermediates for the plantazolicin class of highly discriminating antibiotics. *ACS Chem Biol* 6(12):1307–1313.
- Lee SW, et al. (2008) Discovery of a widely distributed toxin biosynthetic gene cluster. *Proc Natl Acad Sci USA* 105(15):5879–5884.
- Arnison PG, et al. (2013) Ribosomally synthesized and post-translationally modified peptide natural products: Overview and recommendations for a universal nomenclature. *Nat Prod Rep* 30(1):108–160.
- Melby JO, Nard NJ, Mitchell DA (2011) Thiazole/oxazole-modified microcins: Complex natural products from ribosomal templates. *Curr Opin Chem Biol* 15(3):369–378.
- McIntosh JA, Schmidt EW (2010) Marine molecular machines: Heterocyclization in cyanobactin biosynthesis. *ChemBioChem* 11(10):1413–1421.
- Milne JC, Eliot AC, Kelleher NL, Walsh CT (1998) ATP/GTP hydrolysis is required for oxazole and thiazole biosynthesis in the peptide antibiotic microcin B17. *Biochemistry* 37(38):13250–13261.
- Dunbar KL, Melby JO, Mitchell DA (2012) YcaO domains use ATP to activate amide backbones during peptide cyclodehydrations. *Nat Chem Biol* 8(6):569–575.
- Melby JO, Dunbar KL, Trinh NQ, Mitchell DA (2012) Selectivity, directionality, and promiscuity in peptide processing from a *Bacillus* sp. AI Hakam cyclodehydratase. *J Am Chem Soc* 134(11):5309–5316.
- Milne JC, et al. (1999) Cofactor requirements and reconstitution of microcin B17 synthetase: A multienzyme complex that catalyzes the formation of oxazoles and thiazoles in the antibiotic microcin B17. *Biochemistry* 38(15):4768–4781.
- Claesen J, Bibb M (2010) Genome mining and genetic analysis of cypemycin biosynthesis reveal an unusual class of posttranslationally modified peptides. *Proc Natl Acad Sci USA* 107(37):16297–16302.
- Claesen J, Bibb MJ (2011) Biosynthesis and regulation of grisemycin, a new member of the linaridin family of ribosomally synthesized peptides produced by *Streptomyces griseus* IFO 13350. *J Bacteriol* 193(10):2510–2516.
- Zhang Q, van der Donk WA (2012) Catalytic promiscuity of a bacterial α -N-methyltransferase. *FEBS Lett* 586(19):3391–3397.
- Hendricks CL, Ross JR, Pichersky E, Noel JP, Zhou ZS (2004) An enzyme-coupled colorimetric assay for S-adenosylmethionine-dependent methyltransferases. *Anal Biochem* 326(1):100–105.
- Kapust RB, Waugh DS (1999) Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci* 8(8):1668–1674.
- Mitchell DA, et al. (2009) Structural and functional dissection of the heterocyclic peptide cytotoxin streptolysin S. *J Biol Chem* 284(19):13004–13012.
- Zhang C, et al. (2006) RebG- and RebM-catalyzed indolocarbazole diversification. *ChemBioChem* 7(5):795–804.
- Zhang C, Weller RL, Thorson JS, Rajski SR (2006) Natural product diversification using a non-natural cofactor analogue of S-adenosyl-L-methionine. *J Am Chem Soc* 128(9):2760–2761.
- Lee JH, et al. (2010) Characterization and structure of Dhpl, a phosphonate O-methyltransferase involved in dehydrophos biosynthesis. *Proc Natl Acad Sci USA* 107(41):17557–17562.
- Anderle C, et al. (2007) Improved mutasynthetic approaches for the production of modified aminocoumarin antibiotics. *Chem Biol* 14(8):955–967.
- Martin JL, McMillan FM (2002) SAM (dependent) I AM: The S-adenosylmethionine-dependent methyltransferase fold. *Curr Opin Struct Biol* 12(6):783–793.
- Islam K, Zheng W, Yu H, Deng H, Luo M (2011) Expanding cofactor repertoire of protein lysine methyltransferase for substrate labeling. *ACS Chem Biol* 6(7):679–684.

A.2 Discovery of a new ATP-binding motif involved in peptidic azoline biosynthesis

Reproduced with permission from Dunbar, K.L.; Chekan, J.R.; Cox, C.L.; Burkhart, B.J.; Nair, S.K.; Mitchell, D.A. Discovery of a new ATP-binding motif involved in peptidic azoline biosynthesis. *Nat. Chem. Biol.* **2014**, *10*, 823-829. PMID: PMC4167974. Copyright 2014 Nature Publishing Group.

I discovered that BalhC was necessary and sufficient for leader peptide binding whereas BalhD did not significantly interact with the leader peptide. I generated Figure 4c.

Discovery of a new ATP-binding motif involved in peptidic azoline biosynthesis

Kyle L Dunbar^{1,2,6}, Jonathan R Chekan^{3,6}, Courtney L Cox^{2,4}, Brandon J Burkhardt^{1,2}, Satish K Nair^{2,3,5*} & Douglas A Mitchell^{1,2,4*}

Despite intensive research, the cyclodehydratase responsible for azoline biogenesis in thiazole/oxazole-modified microcin (TOMM) natural products remains enigmatic. The collaboration of two proteins, C and D, is required for cyclodehydration. The C protein is homologous to E1 ubiquitin-activating enzymes, whereas the D protein is within the YcaO superfamily. Recent studies have demonstrated that TOMM YcaOs phosphorylate amide carbonyl oxygens to facilitate azoline formation. Here we report the X-ray crystal structure of an uncharacterized YcaO from *Escherichia coli* (Ec-YcaO). Ec-YcaO harbors an unprecedented fold and ATP-binding motif. This motif is conserved among TOMM YcaOs and is required for cyclodehydration. Furthermore, we demonstrate that the C protein regulates substrate binding and catalysis and that the proline-rich C terminus of the D protein is involved in C protein recognition and catalysis. This study identifies the YcaO active site and paves the way for the characterization of the numerous YcaO domains not associated with TOMM biosynthesis.

The YcaO family of proteins currently comprises nearly 5,000 members distributed across the bacterial and archaeal domains. Disparate functions have been ascribed to members of this family, which is sometimes referred to as DUF181 (DUF, domain of unknown function). In *E. coli*, the deletion or overexpression of the eponymous YcaO protein (Ec-YcaO; Fig. 1a) suggested that it potentiates the methylation of ribosomal protein S12 and influenced biofilm formation, respectively^{1,2}. However, a molecular explanation for these observations is currently unavailable. Another YcaO-associated activity is the ATP-dependent cyclodehydration of serine, threonine and cysteine residues to azoline heterocycles, which is the defining modification of TOMM natural products (Fig. 1a,b)³. TOMMs display diverse structures and activities^{3,4}, with some implicated in bacterial pathogenesis⁵, making the ~1,000 bioinformatically identifiable TOMM YcaO proteins noteworthy members of the larger superfamily.

Although the TOMM YcaO domain was first implicated in cyclodehydration reactions in the mid-1990s⁶, its exact role remains unclear^{7,8}. The function of the TOMM YcaO (D protein) is intimately linked to members of the E1 ubiquitin-activating enzyme family (C protein) found in canonical TOMM biosynthetic clusters^{6,9,10}. Underscoring this linked function, roughly half of known TOMM clusters express C and D as a single polypeptide^{3,10}. Studies on both fused and unfused cyclodehydratases have demonstrated that these domains are necessary and sufficient for TOMM azoline formation^{7,11}. Consequently, the C-D complex is referred to as the TOMM cyclodehydratase (or alternatively, heterocyclase^{8,10}). As early studies on the cyclodehydratase were unable to observe activity from either protein in isolation^{6,9,12}, the respective contributions of C and D were inferred by bioinformatics. Given the ATP dependence of the reaction¹³ and the homology of C to members of the E1 ubiquitin-activating superfamily¹⁰, which includes other ATP-using enzymes (for example, MccB, ThiF and MoeB)¹⁴, it was assumed that C was responsible for cyclodehydration, whereas the uncharacterized YcaO (D protein) played a regulatory or scaffolding role^{9,10}.

In 2012, research from our group challenged these assignments with the characterization of the TOMM cyclodehydratase from *Bacillus* sp. Al Hakam (Balh; Fig. 1a)⁷. This YcaO protein (BalhD) displayed ATP-dependent cyclodehydratase activity in the absence of the cognate C protein (BalhC); however, BalhC potentiated cyclodehydration by nearly 1,000-fold. Considering that the C protein had been implicated in precursor peptide recognition in streptolysin S biosynthesis¹⁵ and that BalhC dictated the regio- and chemoselectivity of the Balh cyclodehydratase¹⁶, we hypothesized that the YcaO contained the active site residues, whereas the C protein was responsible for binding the peptide substrate. Although YcaO proteins lack recognizable ATP-binding motifs, the presence of one or more YcaOs in the bottromycin and trifolixatin biosynthetic clusters, which lack recognizable C proteins, supports these functional assignments (Supplementary Results, Supplementary Fig. 1)^{17–21}.

Although the above studies assigned a putative activity to TOMM YcaOs, a molecular understanding of cyclodehydratase catalysis remained elusive. Recently, the X-ray crystal structure of a fused cyclodehydratase was reported (TruD; Protein Data Bank (PDB) code 4BS9), providing what is to our knowledge the first structural glimpse of a TOMM cyclodehydratase⁸. The C domain adopted the expected E1 fold, whereas the YcaO fold was unique. As the structure lacked both the ATP and peptide substrates, no information regarding substrate engagement and catalysis could be gleaned. However, the lack of structural homology of YcaO to known ATP-binding proteins led to the reassertion that the C domain was responsible for ATP binding and carbonyl activation, whereas the YcaO domain catalyzed the requisite nucleophilic attack⁸.

Here, we report the structure of a non-TOMM YcaO from *E. coli* in various nucleotide-bound and nucleotide-free forms and demonstrate that the most conserved residues in YcaOs comprise a previously uncharacterized ATP-binding motif. We show that these ATP-binding residues are critical for catalysis in TOMM YcaOs using BalhD as a model cyclodehydratase. Further, we identify the active site of TOMM cyclodehydratases and demonstrate that the

¹Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA. ²Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA. ³Department of Biochemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA. ⁴Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA. ⁵Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA. ⁶These authors contributed equally to this work. *e-mail: douglasm@illinois.edu or snair@uiuc.edu



conserved, proline-rich C termini are involved in active site organization and C protein binding. Our results strongly support a model where ATP use is a universal feature of YcaOs (TOMM and non-TOMM) and where TOMM C proteins recognize the peptide substrate and potentiate the activity of the cognate YcaO.

RESULTS

E. coli YcaO hydrolyzes ATP to AMP and PP_i

With the ATP-dependent cyclodehydratase activity of BalhD previously established⁷, we attempted to locate the ATP-binding site in BalhD by 8-azido-ATP cross-linking; however, these experiments were unsuccessful. Furthermore, the TruD crystal structure did not reveal an obvious ATP-binding site⁸, and BalhD was refractory to numerous crystallization attempts. We reasoned that because TOMM YcaOs evolved to interact with their cognate C proteins, working with a non-TOMM (with no C protein partner) might alleviate the previously encountered challenges. The local genomic environment of Ec-YcaO does not contain an E1 homolog (i.e., TOMM C protein; Fig. 1a), and Ec-YcaO is not known to interact with an E1 homolog, making it an attractive candidate for structural and biochemical characterization. Ec-YcaO was cloned into a tobacco etch virus (TEV) protease-cleavable maltose-binding protein (MBP)-fusion vector and expressed in *E. coli* (Supplementary Fig. 2). Although the function of Ec-YcaO was unknown, characterized cyclodehydratases hydrolyze ATP in the absence of peptide substrates⁷. Consequently, we measured the ATPase activity of Ec-YcaO using an established purine nucleoside phosphorylase assay²². This assay revealed that Ec-YcaO indeed hydrolyzed ATP, preferentially generating AMP and PP_i (Supplementary Fig. 3). Although ATP hydrolysis was slow, perhaps because the native substrate was not present, Ec-YcaO displayed a K_m for ATP of ~80 μM (Supplementary Fig. 3), comparable to that of several characterized cyclodehydratases^{7,11,13}.

Crystallization of Ec-YcaO

The structure of nucleotide-free Ec-YcaO (containing a mercurial salt for phasing) was determined to a Bragg limit of 2.63 Å and revealed a circularly symmetric homodimer in the asymmetric unit (Supplementary Table 2). The overall structure consists of an N-terminal YcaO domain of ~400 residues and a 150-residue C-terminal domain resembling a tetratricopeptide repeat that mediates dimerization. A structure-based comparison against the PDB revealed similarity solely with TruD, the only other solved YcaO structure (r.m.s. deviation of 3.1 Å over 279 aligned Cα atoms)²³, confirming that YcaOs constitute a new structural fold (Supplementary Fig. 4).

To identify the ATP-binding site, we determined the structure of Ec-YcaO in complex with multiple nucleotides. Co-crystallization of Ec-YcaO with ATP produced an AMP-bound structure (2.25 Å), suggesting that *in situ* hydrolysis had occurred (Fig. 2a). To clarify the residues involved in ATP binding, we also determined the co-structure of Ec-YcaO with αβ-methyleneadenosine 5'-triphosphate (AMPCPP, a nonhydrolyzable ATP analog). These three structures facilitated the characterization of the ATP-binding site in the YcaO superfamily.

Structural characterization of ATP-binding in Ec-YcaO

Analysis of the 2.25-Å resolution AMP-bound and the 3.29-Å resolution AMPCPP-bound co-crystal structures revealed that the adenine ring is recognized by Glu191

and Asn187 via interactions through the N7 nitrogen and between Ser16 and the exocyclic N6 (Fig. 2b). Additionally, Lys9 resides above one face of the adenine ring, whereas Ala70, Ser71 and Gly74 are found within an α-helix that extends below the adenine and ribose rings, forming a hydrophobic surface. The ribose within the AMP- and AMPCPP-bound structures is oriented perpendicular to the adenine ring, with Ser184 and Glu78 coordinating the 2'- and 3'-hydroxyls, respectively (Fig. 2c,d). Although Ser71 coordinates the α-phosphate in both structures, Arg286 coordinates to the α-phosphate in only the AMP-bound form (Fig. 2c,d). To our surprise, two Mg²⁺ ions are found in the nucleotide-binding pocket in both structures. In the AMP structure, Glu199 and Glu78 ligate one Mg²⁺ ion, and Glu290 and Glu75 bind the second (Fig. 2c). The Mg²⁺ ions are coordinated in a similar fashion in the AMPCPP structure with the subtle difference that Glu202, rather than Glu75, coordinated the second Mg²⁺ (Fig. 2d). This slight change in the coordination of the second Mg²⁺ ion positions the metal ions on opposite sides of the β- and γ-phosphates. Furthermore, Arg203 coordinates the γ-phosphate of AMPCPP (Fig. 2d). The interactions between Ec-YcaO and AMPCPP are summarized in Figure 2e.

The ATP-binding site is conserved in TOMM YcaOs

Using the nucleotide-bound structures of Ec-YcaO, we established the conservation of the ATP-binding residues across the superfamily. First, we generated a Cytoscape sequence similarity network²⁴ of all of the YcaO members in InterPro (IPR003776)²⁵. During assembly, redundant sequences were removed, leaving ~2,000 sequences in the network. Although the sequence of the TOMM precursor peptide dictates the structure of the natural product, there is also a strong correlation between TOMM structure and the sequence similarity of the cognate YcaO³. For the network, all of the YcaO sequences were manually annotated on the basis of neighboring genes. YcaOs were categorized as being involved in TOMM biosynthesis if there was a gene encoding a recognizable C protein in the local region (~10 kb on either side of the *ycaO* gene) or if the protein had an experimentally verified link to a known TOMM (for example, bottromycin^{17–20} and trifolitin²¹). The remaining YcaOs were separated into two other categories, non-TOMM YcaOs (for example, Ec-YcaO) and TfuA-associated non-TOMM YcaOs. The latter were found within 10 kb of a gene encoding for the protein TfuA, which is implicated in trifolitin biosynthesis²⁶. Whenever possible, TOMM YcaOs were further subdivided by expected structural class. On the basis of these classifications, we determined that an expectation value of 10⁻⁶⁰ gave an optimal separation of YcaO sequences into isofunctional clusters (Supplementary Fig. 5).

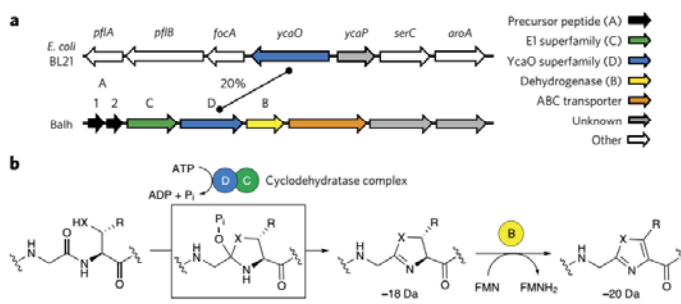


Figure 1 | YcaO gene clusters and characterized roles of YcaO proteins. (a) The local genomic environment for Ec-YcaO and BalhD is depicted along with the percentage amino acid identity for the YcaOs. Gene assignments are shown. **(b)** Azole heterocycles in TOMM natural products are installed by the successive action of a cyclodehydratase (C and D proteins) and a flavin mononucleotide (FMN)-dependent dehydrogenase (B protein). The cumulative mass change for each step is shown below the modification. X = S or O; R = H or CH₃.

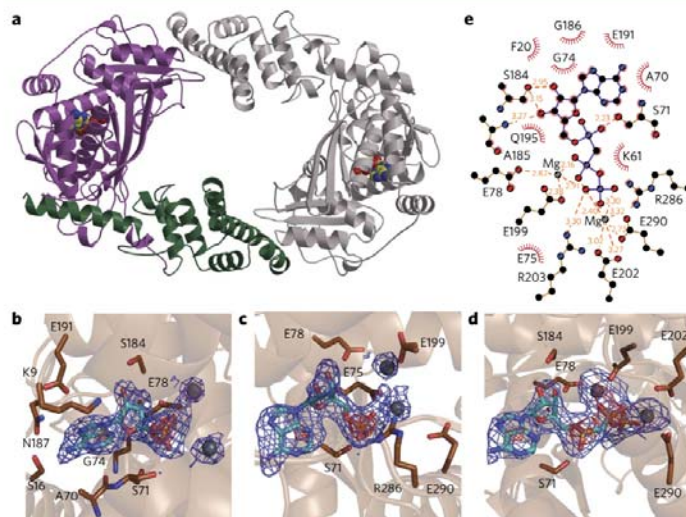


Figure 2 | Structure of Ec-YcaO and ATP-binding residues of Ec-YcaO. (a) Structure of the Ec-YcaO homodimer with one monomer in purple (ATP-binding domain) and green (tetratricopeptide repeat domain) and the other monomer in gray. AMP is shown as spheres in both monomers. (b,c) Orthogonal views of the Ec-YcaO AMP-bound structure in the vicinity of the active site, showing residues responsible for adenine and ribose binding (b) and Mg^{2+} and P_i binding (c). AMP and Mg^{2+} ions are cyan and gray, respectively. The superimposed difference Fourier maps are contoured at levels of 2.0σ (blue) and 8.0σ (red). (d) AMP (cyan) and Mg^{2+} (gray) bound in the Ec-YcaO active site coordinates with superimposed difference Fourier maps contoured at levels of 2.0σ (blue) and 6.0σ (red). Residues responsible for P_i, ribose and Mg^{2+} binding are indicated. (e) A ligand interaction diagram for the AMP-bound structure is shown. Putative hydrogen bonds are shown in orange with distances indicated, and red arcs denote hydrophobic interactions. Due to slight differences in residue orientation in the monomer subunits, only a subset of the interactions is displayed for clarity.

Using the sequence similarity network as a guide, 349 of the ~2,000 members in the nonredundant network were selected from across all of the clusters, and a maximum likelihood tree was generated (Supplementary Fig. 6). Among these 349 were all of the singletons, defined as divergent family members not grouping with any other YcaO at an e-value of 10^{-50} . Using this diversity-maximized tree, a sequence logo for each of the regions involved in ATP binding was generated using WebLogo (Fig. 3a)²⁷. The logos clearly demonstrate that the ATP-binding pocket is highly conserved in the YcaO family across all three groups (i.e., TOMM, non-TOMM and TfuA-associated non-TOMM). The ATP-binding residues were found to be the most conserved feature in the YcaO superfamily (Fig. 3b). This is in stark contrast to TOMM C proteins, which lack the ATP-binding residues conserved in all of the characterized non-TOMM E1 ubiquitin-activating superfamily members¹⁴ (Supplementary Fig. 7). Furthermore, the conservation in the ATP-binding residues is maintained in all of the characterized TOMM YcaOs (Supplementary Figs. 8 and 9), suggesting that the previously reported carbonyl activation mechanism is likely to be a universal biosynthetic feature^{7,16}.

The Ec-YcaO ATP-binding site is vital for BalhD activity

Because the native substrate of Ec-YcaO is unknown, we validated the ATP-binding residues by conducting structure-function studies on BalhD. An alignment of BalhD and Ec-YcaO permitted the mapping of the nucleotide- and Mg^{2+} -binding residues onto BalhD (Supplementary Figs. 4 and 9). Subsequently, an alanine

mutagenesis scan was performed on the polar residues of BalhD predicted to bind ATP. Every mutation was well tolerated in terms of protein yield and stability (Supplementary Fig. 2). The effect on heterocycle formation on BalhA1 (the peptide substrate) by the mutant BalhD proteins, in the presence of BalhC, was monitored in a 16-h endpoint assay (Supplementary Fig. 10). Of the 11 mutated residues in the ATP-binding pocket, four were able to convert BalhA1 to the previously reported penta-azoline species⁷, three showed intermediate levels of processing (two to four heterocycles), and the remaining four generated no heterocyclic products within the limit of detection (Table 1). To quantify the effect of each mutation to BalhC and BalhD activity, the rate of ATP hydrolysis was monitored using the K_M concentration of BalhA1 (15 μ M) and a concentration of ATP that would be saturating for wild-type BalhD (3 mM). Mutants unable to cyclize BalhA1, even after extended reaction times, displayed no detectable ATP hydrolysis over the assay background (Supplementary Fig. 11). Likewise, mutants that installed five azolines on BalhA1 in the endpoint assay had the highest ATP hydrolysis rates. These data are congruent with our earlier work showing that ATP hydrolysis is tightly coupled to heterocycle formation⁷. The YcaO mutations examined here did not appear to disrupt this feature of TOMM cyclodehydration.

Although mutation of the BalhD ATP-binding pocket reduced cyclodehydratase activity, an alternative interpretation of the above data could be that these mutations interfered with the association of BalhC and BalhD. A unique feature of the Balh cyclodehydratase

is that BalhD is catalytically active in the absence of BalhC⁷. This permitted the use of BalhD-only activity measurements to determine whether the alanine mutations affected the intrinsic cyclodehydratase activity. Heterocycle formation endpoint assays (16 h) were again conducted, but this time with 50 μ M BalhA1 and 25 μ M BalhD mutant to account for the expected ~1,000-fold drop in catalytic activity in the absence of BalhC⁷. The resultant mass spectra confirmed that the decrease in cyclodehydration arose from a perturbation in BalhD activity (Table 1 and Supplementary Fig. 12).

For all of the BalhD mutants with measureable cyclodehydratase activity, we obtained the Michaelis-Menten kinetic parameters for BalhA1 and ATP (Table 1 and Supplementary Fig. 13). Every mutation negatively affected the observed k_{cat} (k_{obs}), indicating that the selected residues were of catalytic importance. Apart from K281A, and to a lesser extent S72A, all of the ATP-binding site mutations of BalhD substantially increased the K_M for ATP. In contrast, the only mutant in this series to substantially raise the K_M for BalhA1 was R198A (Table 1).

Four BalhD mutants (i.e., BalhD^{E76A}, BalhD^{E79A}, BalhD^{E194A} or BalhD^{E197A}) did not exhibit detectable cyclodehydratase activity. Potential explanations include an inability to bind the substrates (BalhA1 and/or ATP) or hydrolyze ATP or a structural perturbation with these mutants. Previous work demonstrated that BalhC and BalhD hydrolyze ATP slowly in the absence of BalhA1 (ref. 7). In reactions with wild-type BalhD, addition of BalhC potentiates the rate of ATP hydrolysis by 2.5-fold over an additive rate of both proteins; however, when the four inactive mutants were assayed, no

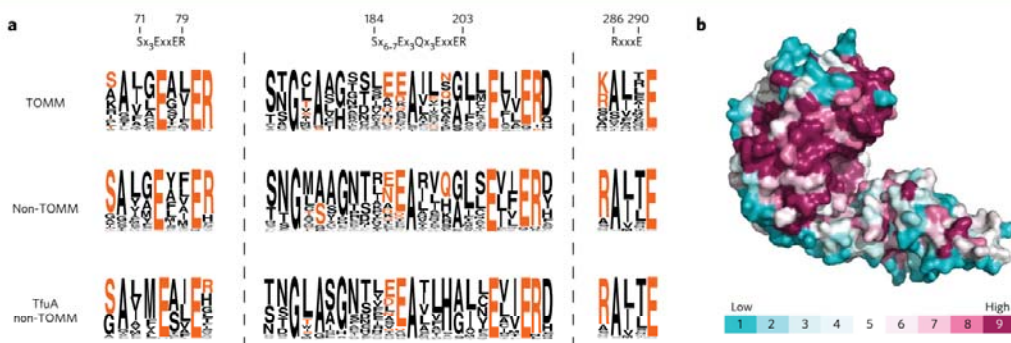


Figure 3 | Conservation of the Ec-YcaO ATP-binding residues in the superfamily. (a) WebLogo frequency plots for the ATP- and Mg^{2+} -binding residues of YcaO domains for each subclass (TOMM, non-TOMM and TfuA non-TOMM). Owing to the high level of diversity in the sequences, WebLogos for the N-terminal ATP-binding residues were not generated. The ATP-binding motif identified in Ec-YcaO is shown above each of the ATP-binding regions with the number representing the residue in Ec-YcaO and conserved residues in orange. (b) YcaO superfamily sequence conservation was mapped onto the structure of Ec-YcaO, which highlights strong conservation of the ATP-binding region.

potentiation was observed (Supplementary Fig. 14). This suggested that the lack of BalhD activity was due to a structural perturbation or the inability to bind or hydrolyze ATP. Unfortunately, attempts to directly measure ATP binding or a secondary structure perturbation of BalhD with isothermal titration calorimetry or CD spectroscopy, respectively, were problematic owing to the solubility characteristics of BalhD. However, we reasoned that the latter could be assayed indirectly by monitoring the interaction between BalhC and a mutant BalhD through a competition assay and size-exclusion chromatography. Although all of the BalhD mutants were able to associate with the BalhC (Supplementary Fig. 15), BalhD^{E76A} and BalhD^{E194A} did so with reduced affinity, suggesting that these mutations affected the BalhC-BalhD interaction surface. Conversely, the wild type-like affinity that BalhD^{E197A} and BalhD^{E197A} displayed for BalhC suggested that these mutants were inactive owing to an inability to bind or hydrolyze ATP.

The BalhD C terminus affects BalhC binding and catalysis

In addition to the conserved ATP-binding site, TOMM YcaO proteins have a highly conserved, proline-rich C terminus. In the most pronounced cases, the final five residues of the YcaO are PXPXP (Supplementary Fig. 16)⁹. The proline-rich C terminus is not conserved in non-TOMM or TfuA-associated YcaO domains

(Supplementary Fig. 16), implicating the motif in either C protein recognition or cyclodehydratase activity. This hypothesis is supported by the observation that the C terminus of TruD is in close proximity to the YcaO ATP-binding site and is surface-accessible (Supplementary Fig. 16). We first interrogated the importance of this motif by truncating five residues from the BalhD C terminus. This minor perturbation abolished the catalytic activity of BalhC and BalhD (Table 2 and Supplementary Fig. 17). Removing the C-terminal three residues of BalhD produced an identical result, and removal of the C-terminal residue of BalhD (BalhD^{P429*}, where the asterisk represents a stop codon) decreased activity by > 100-fold (Table 2 and Supplementary Fig. 17). Similarly, extending the C terminus by a single amino acid (BalhD PXPXP), or deleting two amino acids upstream of the PXPXP motif (BalhD Δ 418-419; Δ 2 AA), also resulted in inactive cyclodehydratases (Table 2 and Supplementary Fig. 17).

To establish whether altering the BalhD C terminus affected the interaction with BalhC, we assessed the ability of BalhC to potentiate the background ATPase activity of BalhD mutants lacking detectable activity. Potentiation was not observed in any case (Supplementary Fig. 18), indicating that mutants of the BalhD PXPXP motif had lost the ability to bind or hydrolyze ATP or to bind

Table 1 | Mutations to the ATP-binding pocket of BalhD decrease cyclodehydratase activity.

| Mutation | Ring formation ^a | | BalhA1 kinetics | | | ATP kinetics | | |
|----------|-----------------------------|----------|--------------------------------|-------------------------------|--|--------------------------------|-------------------------------|--|
| | CD | D only | k_{obs} (min ⁻¹) | K_M (μ M) ^b | k_{obs}/K_M (M ⁻¹ s ⁻¹) | k_{obs} (min ⁻¹) | K_M (μ M) ^b | k_{obs}/K_M (M ⁻¹ s ⁻¹) |
| WT | 5; 100% | 2-3; 45% | 12.9 ± 0.4 | 16 ± 2 | 13,000 | 12.2 ± 0.3 | 240 ± 20 | 850 |
| S72A | 5; 100% | 2; 40% | 8.3 ± 0.2 | 11 ± 1 | 12,500 | 8.0 ± 0.4 | 360 ± 60 | 370 |
| E76A | 0; 0% | 0; 0% | ND | ND | ND | ND | ND | ND |
| E79A | 0; 0% | 0; 0% | ND | ND | ND | ND | ND | ND |
| R80A | 2-4; 63% | 0-1; 10% | 0.79 ± 0.05 | 16 ± 5 | 823 | 0.88 ± 0.02 | 620 ± 50 | 24 |
| Q186A | 5; 100% | 0-2; 10% | 3.9 ± 0.1 | 12 ± 1 | 5,400 | 7.5 ± 0.3 | 2,500 ± 200 | 50 |
| N190A | 5; 100% | 0-2; 15% | 5.1 ± 0.2 | 27 ± 3 | 3,150 | 4.7 ± 0.2 | 920 ± 110 | 85 |
| E194A | 0; 0% | 0; 0% | ND | ND | ND | ND | ND | ND |
| E197A | 0; 0% | 0; 0% | ND | ND | ND | ND | ND | ND |
| R198A | 2-4; 40% | 0; 0% | 3.3 ± 0.2 | 50 ± 8 | 1,100 | 2.4 ± 0.2 | 760 ± 65 | 52 |
| K281A | 5; 100% | 2-3; 45% | 7.9 ± 0.2 | 16 ± 2 | 8,200 | 6.1 ± 0.1 | 230 ± 20 | 440 |
| E286A | 2-4; 64% | 0; 0% | 0.49 ± 0.03 | 27 ± 3 | 272 | 0.42 ± 0.01 | 1,200 ± 100 | 6 |

^a% processing = (5-P₅ + 4-P₄ + 3-P₃ + 2-P₂ + 1-P₁)/(5); where P_x is the percentage of the substrate with x number of azolines. The number of heterocycles formed in the assay is listed in parentheses.

^bApparent K_M . Error on the Michaelis-Menten parameters represents the s.d. from the regression analysis. ND, not determined.

Table 2 | Mutations to the C terminus of BalhD disrupt catalysis.

| Mutation | Ring formation ^a | | BalhA1 kinetics | | | ATP kinetics | | |
|---------------|-----------------------------|----------|---------------------------------------|--------------------------------------|---|---------------------------------------|--------------------------------------|---|
| | CD | D only | k_{obs} (min ⁻¹) | K_M (μM) ^b | k_{obs}/K_M (M ⁻¹ s ⁻¹) | k_{obs} (min ⁻¹) | K_M (μM) ^b | k_{obs}/K_M (M ⁻¹ s ⁻¹) |
| WT | 5; 100% | 1-3; 45% | 12.9 ± 0.4 | 16 ± 2 | 13,000 | 12.2 ± 0.3 | 240 ± 20 | 850 |
| P425* | 0; 0% | 0; 0% | ND | ND | ND | ND | ND | ND |
| P427* | 0; 0% | 0; 0% | ND | ND | ND | ND | ND | ND |
| P429* | 1-2; 36% | 0; 0% | 0.19 ± 0.03 | 80 ± 20 | 31 | 0.21 ± 0.05 | 110 ± 10 | 32 |
| $\Delta 2$ AA | 0; 0% | 0; 0% | ND | ND | ND | ND | ND | ND |
| PxPxPG | 0; 0% | 0; 0% | ND | ND | ND | ND | ND | ND |
| H426A | 0-2; 25% | 0; 0% | ND | ND | ND | ND | ND | ND |
| P427G | 3-5; 80% | 0; 0% | 0.43 ± 0.04 | 35 ± 8 | 204 | 0.41 ± 0.02 | 700 ± 100 | 10 |
| F428A | 3-4; 80% | 0; 0% | 0.70 ± 0.02 | 22 ± 2 | 530 | 0.69 ± 0.03 | 480 ± 60 | 24 |
| P429G | 4; 80% | 0-2; 6% | 4.7 ± 0.1 | 14 ± 1 | 5,600 | 5.0 ± 0.1 | 75 ± 4 | 1,100 |
| GxGxG | 0-3; 45% | 0; 0% | ND | ND | ND | ND | ND | ND |

^a% processing = (5P₅ + 4P₄ + 3P₃ + 2P₂ + 1P₁)/(5); where P_x is the percentage of the substrate with x number of azolines. The number of heterocycles formed in the assay is listed in parentheses.

^bApparent K_M . Asterisk indicates stop codon. Error on the Michaelis-Menten parameters represents the s.d. from the regression analysis. ND, not indicated.

BalhC. We next assessed the ability of each PxPxP mutant to bind BalhC. Using a combination of size-exclusion chromatography and a competition assay, all of the truncations to the BalhD PxPxP motif were shown to have decreased affinity to BalhC (Supplementary Fig. 19). Moreover, the order of heterocycle formation was dysregulated in BalhD^{P429*}, reminiscent of wild-type BalhD reactions lacking BalhC (Supplementary Fig. 20)¹⁶.

Intrigued by the loss of activity observed upon extending or truncating the C terminus, we next investigated the importance of the amino acid composition of the BalhD PxPxP motif (PHPPF₄₂₆). As with the truncations, any mutation to the five C-terminal residues of BalhD decreased cyclodehydratase activity (Table 2 and Supplementary Fig. 21). The decrease in activity ranged from ~2.5-fold (BalhD^{P429G}) to 100-fold (BalhD^{H426A}), with severity diminishing the closer the mutation was to the C terminus. This result was consistent with the observation that the C-terminal residues of TruD are located in a channel leading to the active site. As with the PxPxP truncations, every mutant tested, apart from BalhD^{F428A}, displayed a decreased affinity for BalhC (Supplementary Fig. 22). Consistent with this observation, mutation of the terminal amino acid (P429G) resulted in an aberrant order of heterocycle formation (Supplementary Fig. 23). Increasing the flexibility of the PHPPF motif by substituting it with GHGFG, yielded an inactive cyclodehydratase (Table 2 and Supplementary Fig. 21).

Although these results implicated the C terminus of BalhD in BalhC recognition, a decrease in BalhC affinity could not explain the data in its entirety. For example, both BalhD^{P425*} and BalhD^{P429*} displayed reduced interactions with BalhC, but only BalhD^{P425*} was catalytically inactive (Table 2 and Supplementary Fig. 19). Furthermore, BalhD PxPxPG showed a wild-type level of interaction with BalhC despite being catalytically inactive. Given these results, we tested the activity of each mutant in the absence of BalhC. Analogous to the mutations to the ATP-binding pocket, mutations to the PxPxP motif affected the intrinsic activity of BalhD (Supplementary Fig. 24). For all tractable BalhD mutants, BalhA1 and ATP Michaelis-Menten kinetic curves were obtained for the mutant BalhC–BalhD complexes. Although the largest effects were on k_{obs} , the mutations also affected the K_M for ATP and BalhA1 (Table 2).

Unlike the ATP-binding mutants, the changes to K_M for the two substrates were similar in the PxPxP mutants, suggesting that the C terminus is involved in active site organization and catalysis, not substrate binding. Furthermore, the importance of the YcaO C terminus seems to be general for TOMM biosynthesis, given that the cyclodehydratase activity of McbD (microcin B17 YcaO protein) was also abolished when the C terminus was truncated (Supplementary Fig. 25).

BalhC regulates BalhD ATP-binding and catalysis

To further characterize the C terminus of BalhD, we generated a derivative containing a C-terminal His₆ tag (PA₄LEH₆, where P is the last residue of wild-type BalhD). As expected from earlier experiments with BalhD PxPxPG, addition of the longer tag abolished heterocycle formation (Supplementary Fig. 26). Although heterocycle formation is stoichiometric with ATP hydrolysis for wild-type cyclodehydratase⁷, the C-terminal His₆-tagged BalhD displayed robust ATPase activity, irrespective of the presence of BalhA1 (Fig. 4a). Moreover, this high level of unproductive ATP hydrolysis was potentiated by the addition of BalhC to the same extent observed with wild-type BalhD (2.5-fold increase; see Supplementary Fig. 14), indicating that the His₆ tag did not interfere with BalhC recognition (Fig. 4a). With a BalhD derivative displaying robust BalhC-independent ATPase activity in hand, we evaluated the role of BalhC on ATP use by BalhD by obtaining ATP Michaelis-Menten kinetics parameters for BalhD-A₄LEH₆ alone and in complex with BalhC (Fig. 4b).

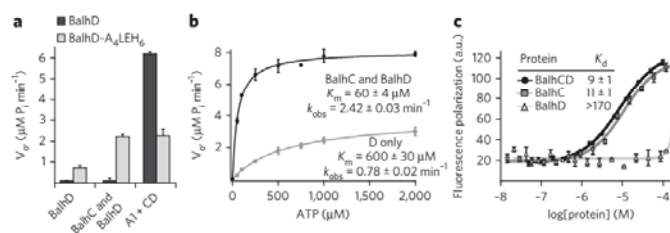


Figure 4 | BalhC modulates ATP binding and hydrolysis by BalhD and is responsible for leader peptide binding. (a) ATP hydrolysis rates measured by the PNP assay. (b) ATP kinetic curves for BalhD-A₄LEH₆ with and without BalhC. Error bars represent the s.d. from the mean ($n = 3$), and error on the Michaelis-Menten parameters represents the s.d. from the regression analysis. (c) A fluorescent polarization curve for fluorescein-labeled BalhA1 leader peptide recognition by BalhC and BalhD is displayed. Error bars represent the s.e.m. of three independent titrations. Errors on the K_D values represent the error from curve fitting. a.u., arbitrary units.

These data indicate that the addition of BalhC modulates BalhD activity by increasing the k_{obs} and decreasing the K_{M} for ATP.

TOMM C proteins provide leader peptide binding

Precursor peptide recognition in the majority of ribosomally synthesized post-translationally modified peptides occurs in a bipartite fashion. An N-terminal sequence (leader peptide) serves as the recognition sequence by the modification enzymes, and the C-terminal sequence (core peptide) contains the sites of post-translational modification^{4,28}. We previously demonstrated that the regioselectivity and directionality of BalhD azoline formation is dependent on BalhC¹⁶. Thus, we hypothesized that BalhC was responsible for presenting the core peptide to BalhD, most likely by engaging the leader peptide. However, the identification of BalhD mutants with a perturbed K_{M} for BalhA1 suggested that the YcaO domain might have a role in substrate recognition. To assess the role of the C and D proteins in peptide substrate recognition, a fluorescein-labeled BalhA1 leader peptide was used to monitor binding to BalhC and BalhD by fluorescence polarization. Owing to a very weak, potentially nonspecific, association with BalhD, the K_{d} toward the BalhA1 leader peptide could not be obtained. In contrast, BalhC displayed a K_{d} of 11 μM (Fig. 4c), near the previously measured K_{M} for BalhA1 of 16 μM ²⁹. Moreover, the addition of BalhD did not significantly alter BalhC's affinity for the BalhA1 leader peptide ($P > 0.05$). Consequently, these data support a model where (i) BalhD does not engage the leader region of BalhA1 and (ii) the elevation in K_{M} value of select BalhD mutants for BalhA1 is due to a decreased affinity for the core region of BalhA1.

DISCUSSION

We have discovered that Ec-YcaO contains a new ATP-binding fold. Given the steric and electrostatic complimentary requirements for binding ATP, the YcaO strategy is reminiscent of that of other structurally characterized ATP-binding proteins. For example, the Lys- α -helix 'sandwich' involved in adenine recognition is similar to the conserved arginine and glycine motif found in class I amino acyl tRNA synthetases³⁰. Furthermore, select members of the ATP-grasp and PurM families have been shown to bind ATP through the use of multiple divalent cations^{31,32}; however, in these proteins, the Mg^{2+} ions are coordinated to all three phosphates and not just the β - and γ -phosphates. As these similarities in ATP binding occur despite a lack of structural and primary homology between YcaO and all of the other known ATP-binding proteins, this represents an example of convergent evolution in ATP-binding domains.

The ATP-binding residues are the most highly conserved motifs in the YcaO superfamily and, appropriately, represent a prominent signature for the hidden Markov model that bioinformatically defines the YcaO family (IPR003776). Our extensive bioinformatics analysis, X-ray crystallographic data on Ec-YcaO and biochemical characterization of BalhD confirm that ATP use is a conserved feature in the superfamily. In spite of the low level of overall similarity between Ec-YcaO and BalhD, we were able to demonstrate that the YcaO ATP-binding motif was critical for cyclodehydratase activity. Although the mutations affected BalhD activity to differing extents, the impact of mutating a particular residue on Balh cyclodehydratase activity was proportional to the level of conservation within the YcaO family (Fig. 3a).

During our analysis of the YcaO ATP-binding motif, we observed a marked difference between the TOMM and non-TOMM YcaO domains. TOMM YcaOs (D proteins) almost invariably harbor proline-rich C-termini, with PxxP most often serving as the terminal five residues of the protein. Though the widespread nature of the PxxP motif had been previously recognized⁹, before this work, it was unclear whether this motif had any role in TOMM biogenesis. Our data indicate that the C terminus of TOMM YcaOs assists in both C protein recognition and cyclodehydration. It is rare for the

C terminus of an enzyme to be important for catalysis^{33–37}. In fact, the terminal regions are often highly sequence-variable within a protein family. Notably, the C-terminal proline content of a YcaO has powerful predictive value. If present, the YcaO is quite probably involved in TOMM biosynthesis. This tentative assignment can be confidently made even without knowledge of the flanking genes. As such, we hypothesize that a subset of the 249 (~8%) non-TOMM YcaOs that contain a proline-rich C terminus may actually be stand-alone TOMM YcaOs (akin to the bottromycin YcaOs).

Previously, BalhC was shown to potentiate BalhD via an unknown mechanism⁷. The current study indicates that this potentiation occurs via two distinct mechanisms. First, the serendipitous identification of a C-terminal His-tagged construct of BalhD with robust ATP hydrolysis (BalhD-A₄LEH₂) allowed us to show that the presence of BalhC increases k_{obs} and lowers the K_{M} for ATP. Although our data suggest that C protein potentiation occurs via allosteric activation, follow-up studies will be required to validate this hypothesis. Second, our data demonstrate that BalhC is responsible for binding the leader peptide of BalhA1, thus efficiently bringing the substrate in close proximity to the BalhD active site. This result is in accord with a previous study implicating SacG in leader peptide recognition during streptolysin S biosynthesis¹⁵. In further support of a general role for TOMM C proteins in peptide substrate binding, the 'C portion' (homologous to E1 ubiquitin-activating enzymes) of TruD has an MccB-like N-terminal 'peptide clamp'⁸, which is responsible for leader peptide binding in microcin C7 biosynthesis³⁸. Combined with the fact that TOMM C proteins lack the ATP-binding site that is conserved in all of the characterized non-TOMM E1 ubiquitin-activating family members, all lines of evidence suggest that TOMM C proteins engage the leader peptide while simultaneously potentiating the carbonyl activation chemistry of their cognate YcaO domain (D protein; Supplementary Fig. 27).

In light of this, it is not yet clear how stand-alone TOMM YcaO proteins (i.e., for bottromycin and trifolixotoxin production) perform cyclodehydrations in the absence of a C protein. Given the diversity between these stand-alone and canonical (C protein-containing) TOMM YcaOs, we envision that multiple solutions to the substrate recognition problem could exist. For example, it is possible that these biosynthetic pathways use an unidentified companion protein to bind the precursor peptide. Alternatively, these YcaO proteins may have evolved to bind a specific motif within the core peptide and modify the substrate a single time. Of note is the fact that the bottromycin and trifolixotoxin stand-alone YcaO domains are each predicted to install a single heterocycle^{17–21}. This is in stark contrast to canonical TOMMs that process a wide array of core peptides, often at numerous locations^{3,29,39}. Such promiscuity is common in other ribosomally synthesized post-translationally modified peptides and most likely accounts for the existence of leader peptides (i.e., the modification enzymes can be specific for motifs within the leader peptide but promiscuous on the core once the enzyme-substrate complex is formed)^{4,28}. Further work, including the reconstitution of a stand-alone YcaO, will be required to test these claims.

The capacity to bind ATP (or possibly other nucleotide triphosphates) seems to be ubiquitous in the YcaO superfamily, but it remains unclear whether the TOMM cyclodehydratase-like direct activation of carbonyls is a universal feature. It is intriguing that YcaOs have recently been implicated in the formation of thioamides⁴⁰ and macroamide rings^{17–20}, as both of these modifications could conceivably occur through carbonyl activation. In addition to providing major insight into the mechanics of TOMM cyclodehydration, the results presented here provide an initial framework to explore the elusive functions of the 4,000 uncharacterized non-TOMM YcaOs.

Received 14 March 2014; accepted 19 June 2014;
published online 17 August 2014

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. PDB. Coordinates for apo-YcaO, AMP-bound YcaO and AMPCPP-bound YcaO structures were deposited under accession codes 4Q84, 4Q86 and 4Q85, respectively.

References

- Strader, M.B. *et al.* A proteomic and transcriptomic approach reveals new insight into β -methylthiolation of *Escherichia coli* ribosomal protein S12. *Mol. Cell. Proteomics* **10**, M110005199 (2011).
- Tenorio, E. *et al.* Systematic characterization of *Escherichia coli* genes/ORFs affecting biofilm formation. *FEMS Microbiol. Lett.* **225**, 107–114 (2003).
- Melby, J.O., Nard, N.J. & Mitchell, D.A. Thiazole/oxazole-modified microcins: complex natural products from ribosomal templates. *Curr. Opin. Chem. Biol.* **15**, 369–378 (2011).
- Arnison, P.G. *et al.* Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **30**, 108–160 (2013).
- Molloy, E.M., Cotter, P.D., Hill, C., Mitchell, D.A. & Ross, R.P. Streptolysin S-like virulence factors: the continuing saga. *Nat. Rev. Microbiol.* **9**, 670–681 (2011).
- Li, Y.M., Milne, J.C., Madison, L.L., Kolter, R. & Walsh, C.T. From peptide precursors to oxazole and thiazole-containing peptide antibiotics: microcin B17 synthase. *Science* **274**, 1188–1193 (1996).
- Dunbar, K.L., Melby, J.O. & Mitchell, D.A. YcaO domains use ATP to activate amide backbones during peptide cyclodehydrations. *Nat. Chem. Biol.* **8**, 569–575 (2012).
- Koehnke, J. *et al.* The cyanobactin heterocyclase enzyme: a processive adenylase that operates with a defined order of reaction. *Angew. Chem. Int. Ed. Engl.* **52**, 13991–13996 (2013).
- Lee, S.W. *et al.* Discovery of a widely distributed toxin biosynthetic gene cluster. *Proc. Natl. Acad. Sci. USA* **105**, 5879–5884 (2008).
- Schmidt, E.W. *et al.* Patellamide A and C biosynthesis by a microcin-like pathway in *Prochloron didemni*, the cyanobacterial symbiont of *Lissoclinum patella*. *Proc. Natl. Acad. Sci. USA* **102**, 7315–7320 (2005).
- McIntosh, J.A. & Schmidt, E.W. Marine molecular machines: heterocyclization in cyanobactin biosynthesis. *ChemBioChem* **11**, 1413–1421 (2010).
- Milne, J.C. *et al.* Cofactor requirements and reconstitution of microcin B17 synthase: a multienzyme complex that catalyzes the formation of oxazoles and thiazoles in the antibiotic microcin B17. *Biochemistry* **38**, 4768–4781 (1999).
- Milne, J.C., Eliot, A.C., Kelleher, N.L. & Walsh, C.T. ATP/GTP hydrolysis is required for oxazole and thiazole biosynthesis in the peptide antibiotic microcin B17. *Biochemistry* **37**, 13250–13261 (1998).
- Schulman, B.A. & Harper, J.W. Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. *Nat. Rev. Mol. Cell Biol.* **10**, 319–331 (2009).
- Mitchell, D.A. *et al.* Structural and functional dissection of the heterocyclic peptide cytotoxin streptolysin S. *J. Biol. Chem.* **284**, 13004–13012 (2009).
- Dunbar, K.L. & Mitchell, D.A. Insights into the mechanism of peptide cyclodehydrations achieved through the chemoenzymatic generation of amide derivatives. *J. Am. Chem. Soc.* **135**, 8692–8701 (2013).
- Huo, L., Rachid, S., Stadler, M., Wenzel, S.C. & Muller, R. Synthetic biotechnology to study and engineer ribosomal bottromycin biosynthesis. *Chem. Biol.* **19**, 1278–1287 (2012).
- Hou, Y. *et al.* Structure and biosynthesis of the antibiotic bottromycin D. *Org. Lett.* **14**, 5050–5053 (2012).
- Gomez-Escribano, J.P., Song, L., Bibb, M.J. & Challis, G.L. Posttranslational β -methylation and macrolactamidation in the biosynthesis of the bottromycin complex of ribosomal peptide antibiotics. *Chem. Sci.* **3**, 3522–3525 (2012).
- Crone, W.J.K., Leeper, F.J. & Truman, A.W. Identification and characterization of the gene cluster for the anti-MRSA antibiotic bottromycin: expanding the biosynthetic diversity of ribosomal peptides. *Chem. Sci.* **3**, 3516–3521 (2012).
- Breil, B.T., Ludden, P.W. & Triplett, E.W. DNA sequence and mutational analysis of genes involved in the production and resistance of the antibiotic peptide trifolitoxin. *J. Bacteriol.* **175**, 3693–3702 (1993).
- Webb, M.R. A continuous spectrophotometric assay for inorganic phosphate and for measuring phosphate release kinetics in biological systems. *Proc. Natl. Acad. Sci. USA* **89**, 4884–4887 (1992).
- Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).
- Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
- Hunter, S. *et al.* InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res.* **40**, D306–D312 (2012).
- Breil, B., Borneman, J. & Triplett, E.W. A newly discovered gene, *tfuA*, involved in the production of the ribosomally synthesized peptide antibiotic trifolitoxin. *J. Bacteriol.* **178**, 4150–4156 (1996).
- Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
- Oman, T.J. & van der Donk, W.A. Follow the leader: the use of leader peptides to guide natural product biosynthesis. *Nat. Chem. Biol.* **6**, 9–18 (2010).
- Melby, J.O., Dunbar, K.L., Trinh, N.Q. & Mitchell, D.A. Selectivity, directionality, and promiscuity in peptide processing from a *Bacillus* sp. Al Hakam cyclodehydratase. *J. Am. Chem. Soc.* **134**, 5309–5316 (2012).
- Denessiouk, K.A. & Johnson, M.S. When fold is not important: a common structural framework for adenine and AMP binding in 12 unrelated protein families. *Proteins* **38**, 310–326 (2000).
- Zhang, Y., Morar, M. & Ealick, S.E. Structural biology of the purine biosynthetic pathway. *Cell. Mol. Life Sci.* **65**, 3699–3724 (2008).
- Fawaz, M.V., Topper, M.E. & Firestone, S.M. The ATP-grasp enzymes. *Bioorg. Chem.* **39**, 185–191 (2011).
- Rakus, J.F. *et al.* Evolution of enzymatic activities in the enolase superfamily: d-mannonate dehydratase from *Novosphingobium aromaticivorans*. *Biochemistry* **46**, 12896–12908 (2007).
- Chen, W., Biswas, T., Porter, V.R., Tsodikov, O.V. & Garneau-Tsodikova, S. Unusual rigors of acetyltransferase Eis, a cause of drug resistance in XDR-TB. *Proc. Natl. Acad. Sci. USA* **108**, 9804–9808 (2011).
- Selvy, P.E., Lavrier, R.R., Lindsley, C.W. & Brown, H.A. Phospholipase D: enzymology, functionality, and chemical modulation. *Chem. Rev.* **111**, 6064–6119 (2011).
- Bhatnagar, R.S., Futterer, K., Waksman, G. & Gordon, J.I. The structure of myristoyl-CoA:protein N-myristoyltransferase. *Biochim. Biophys. Acta* **1441**, 162–172 (1999).
- Climie, S.C., Carreras, C.W. & Santi, D.V. Complete replacement set of amino acids at the C terminus of thymidylate synthase: quantitative structure-activity relationship of mutants of an enzyme. *Biochemistry* **31**, 6032–6038 (1992).
- Regni, C.A. *et al.* How the MccB bacterial ancestor of ubiquitin E1 initiates biosynthesis of the microcin C7 antibiotic. *EMBO J.* **28**, 1953–1964 (2009).
- Donia, M.S. *et al.* Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians. *Nat. Chem. Biol.* **2**, 729–735 (2006).
- Izawa, M., Kawasaki, T. & Hayakawa, Y. Cloning and heterologous expression of the thioviridamide biosynthesis gene cluster from *Streptomyces olivoviridis*. *Appl. Environ. Microbiol.* **79**, 7110–7113 (2013).

Acknowledgments

We are grateful to C. Deane and K. Taylor for the generation of select BalhD mutants and J. Melby for assistance with collecting MS/MS data. This work was supported by the US National Institutes of Health (NIH) (1R01 GM097142 to D.A.M., 1R01 GM102602 to S.K.N. and 2T32 GM070421 to K.L.D., B.J.B. and J.R.C.). Additional support was from the Harold R. Snyder Fellowship (University of Illinois at Urbana-Champaign (UIUC) Department of Chemistry to K.L.D.), the Robert C. and Carolyn J. Springborn Endowment (UIUC Department of Chemistry to B.J.B.), the National Science Foundation Graduate Research Fellowship (DGE-1144245 to B.J.B.) and the University of Illinois Distinguished Fellowship (UIUC Graduate College to J.R.C.) The Bruker UltrafleXtreme MALDI TOF/TOF mass spectrometer was purchased in part with a grant from the NIH–National Center for Research Resources (S10 RR027109 A).

Author contributions

Experiments were designed by D.A.M., S.K.N., K.L.D., J.R.C., C.L.C. and B.J.B. and were performed by K.L.D., J.R.C., C.L.C. and B.J.B. The manuscript was written by D.A.M., K.L.D. and J.R.C. with critical editorial input from S.K.N., C.L.C. and B.J.B. The overall study was conceived and managed by D.A.M. with S.K.N. overseeing all aspects of protein structure determination.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Correspondence and requests for materials should be addressed to S.K.N. or D.A.M.

A.3 Identification of an auxiliary leader peptide-binding protein required for azoline formation in ribosomal natural products

Reproduced with permission from Dunbar, K.L.; Tietz, J.I.; Cox, C.L.; Burkhart, B.J. & Mitchell, D.A. Identification of an auxiliary leader peptide-binding protein required for azoline formation in ribosomal natural products. *J. Am. Chem. Soc.* **2015**, *137*, 7672-7677. PMID: PMC4481143. Copyright 2015 American Chemical Society.

I demonstrated the role of HcaF in leader peptide binding, showed that HcaD does not interact with the leader peptide, and performed binding studies with leader peptide mutants to reveal the contribution of different residues to binding and the minimal peptide length. I produced Table 1 and Figure 3 and S15.

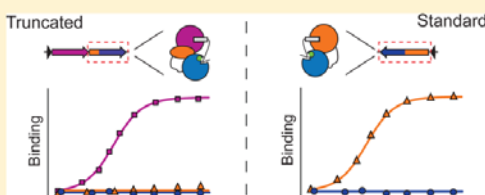
Identification of an Auxiliary Leader Peptide-Binding Protein Required for Azoline Formation in Ribosomal Natural Products

Kyle L. Dunbar,^{†,‡} Jonathan I. Tietz,[†] Courtney L. Cox,^{‡,§} Brandon J. Burkhardt,^{†,‡} and Douglas A. Mitchell^{*,†,‡,§}

[†]Department of Chemistry, [‡]Institute for Genomic Biology, [§]Department of Microbiology, University of Illinois at Urbana–Champaign, Urbana, Illinois 61801, United States

Supporting Information

ABSTRACT: Thiazole/oxazole-modified microcins (TOMMs) are a class of post-translationally modified peptide natural products bearing azole and azoline heterocycles. The first step in heterocycle formation is carried out by a two component cyclodehydratase comprised of an E1 ubiquitin-activating and a YcaO superfamily member. Recent studies have demonstrated that the YcaO domain is responsible for cyclodehydration, while the TOMM E1 homologue is responsible for peptide recognition during azoline formation. Although all characterized TOMM biosynthetic clusters contain this canonical TOMM E1 homologue (C domain), we also identified a second, highly divergent E1 superfamily member, annotated as an Ocin-ThiF-like protein (F protein), associated with more than 300 TOMM biosynthetic clusters. Here we describe the *in vitro* reconstitution of a novel TOMM cyclodehydratase from such a cluster and demonstrate that this auxiliary protein is required for cyclodehydration. Using a combination of biophysical techniques, we demonstrate that the F protein, rather than the C domain, is responsible for engaging the peptide substrate. The C domain instead appears to serve as a scaffolding protein, bringing the catalytic YcaO domain and the peptide binding Ocin-ThiF-like protein into proximity. Our findings provide an updated biosynthetic framework that provides a foundation for the characterization and reconstitution of approximately 25% of bioinformatically identifiable TOMM synthetases.



INTRODUCTION

In thiazole/oxazole-modified microcin (TOMM) natural products, thiazoline and (methyl)oxazoline heterocycles are synthesized by a two-component cyclodehydratase composed of a ubiquitin-activating enzyme E1 homologue (C domain) and a member of the YcaO superfamily (D domain).^{1,2} In approximately 50% of cases, the genes encoding the C and D domains are fused and are expressed as a single polypeptide (fused cyclodehydratase).¹ In many TOMMs, the azolines are oxidized to azoles by a flavin mononucleotide (FMN)-dependent dehydrogenase (B domain).^{1,2} Previous research has demonstrated that the TOMM D domain catalyzes the ATP-dependent cyclodehydration reaction, while the TOMM C domain recognizes the peptide substrate through specific motifs within the N-terminal leader peptide and regulates D domain activity.^{3–5} However, the biophysical basis for leader peptide binding and activity potentiation by the C domain has yet to be elucidated.

Nearly all members of the *Bacillus cereus* group (including *B. anthracis*) harbor an uncharacterized TOMM biosynthetic cluster, previously designated as a “heterocycloanthracin” (HCA) biosynthetic cluster (Figure 1).^{6,7} Recently, the first report regarding the isolation of a HCA natural product has emerged;⁸ however, the structure of the compound was not described. Here we report the reconstitution of the fused

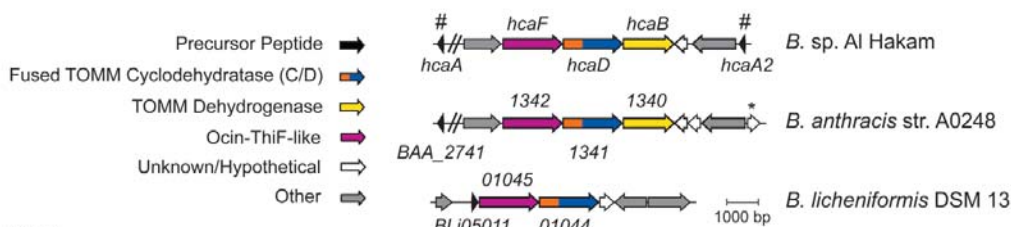
TOMM cyclodehydratase from the *Bacillus* sp. Al Hakam HCA cluster and demonstrate that cyclodehydratase activity is dependent on the presence of an uncharacterized protein (annotated as an Ocin-ThiF-like domain based on weak similarity to E1 superfamily members) found in all HCA clusters.⁶ We show that the C domain of the fused HCA cyclodehydratase is N-terminally truncated compared to characterized TOMM C domains and is unable to bind the precursor peptide. Furthermore, we establish that the previously uncharacterized Ocin-ThiF-like protein is responsible for leader peptide binding. A bioinformatic survey of all currently identifiable TOMM clusters revealed that approximately 25% of biosynthetic gene clusters contain a similarly truncated TOMM C domain and that, in nearly every case, an Ocin-ThiF-like protein is found in the gene cluster.

RESULTS AND DISCUSSION

The genes corresponding to the *Bacillus* sp. Al Hakam HCA dehydrogenase (Figure 1; HcaB) and fused cyclodehydratase (Figure 1; HcaD) and two potential precursor peptides (HcaA and HcaA2; Figures 1 and S1; Supporting Information) were expressed in *E. coli* with N-terminal His₆ or maltose-binding

Received: November 13, 2014

Published: May 29, 2015



HcaA

MNQFQQELQSLNLDYQTGNVVYWDPPQSQPYPPYYIQDDARRCGGCGGCGGRCGGCGGRCGGCAGRCGGCVGCAGCFGCFNCWNWWII

HcaA2

MDLPACDAEANWVAFQISCANVSSLDSQARALYIVALMENSWSSSETAGFGSFGSCGSSGRWARRPAHIYFTYFQ

Figure 1. HCA biosynthetic clusters. Open reading frame (ORF) diagrams for three representative HCA biosynthetic clusters are displayed along with putative precursor peptides from the *Bacillus* sp. Al Hakam cluster. #, ORF not annotated in GenBank. An unannotated homologue of *hcaA2* is present in *B. anthracis* within the ORF marked by an asterisk. Cyclizable residues are colored orange. The putative core peptide for HcaA is underlined.

protein (MBP) tags, respectively. While HcaD expressed well in *E. coli*, HcaB was heavily proteolyzed and purified without the requisite FMN cofactor (Figure S2). Attempts to obtain a full-length, holo HcaB from orthologous biosynthetic clusters were equally unsuccessful (Figure S2). When HcaA was treated with HcaD in the presence of ATP, no modification was observed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS), even at an enzyme to substrate ratio of 1:2 (Figure 2). Given that a subset of HCA

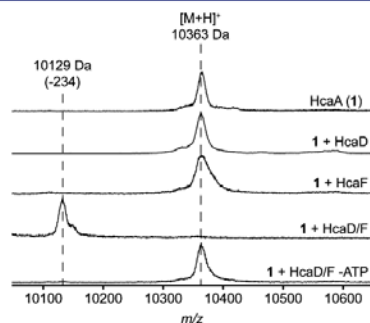


Figure 2. HcaF is required for cyclodehydratase activity. A MALDI-TOF mass spectral overlay for HcaA processing is displayed. The mass of the starting material and the fully cyclodehydrated species are displayed along with the mass shift relative to the unmodified peptide.

clusters lack a B domain,⁶ and previous work with fused cyclodehydratases has demonstrated that azoline formation does not depend on the presence of a dehydrogenase even when it is present in the cluster,^{9–11} we reasoned that HcaD activity could be dependent upon another, unidentified protein. Bioinformatic analysis identified a potential candidate, annotated as an Ocini-ThiF-like protein (hereafter referred to as a TOMM F protein), found adjacent to the cyclodehydratase in all HCA clusters (Figure 1; HcaF). When HcaF was expressed as a His₆-tagged fusion in *E. coli* and added to reactions containing HcaD, 13 azoline heterocycles (resulting in a -234 Da mass shift) were installed on HcaA (Figure 2). No modification was observed with HcaF alone, indicating that

both HcaD and HcaF are required for cyclodehydration (Figure 2). Using iodoacetamide labeling, these modifications were localized to the 13 cysteines in the peptide (Figure S3). Heterocycle formation was ATP- and YcaO-dependent, as omission of ATP or mutagenesis of the ATP-binding site in the D (YcaO) domain of HcaD abolished activity (Figure 2 and Figure S4). Attempts to modify HcaA2 were unsuccessful, suggesting that this peptide is either not a substrate for the HCA cyclodehydratase or that an additional protein is required for modification (Figure S5). As HcaA2 is encoded within another gene (Figure 1),⁷ we believe that the former hypothesis is more probable.

Consistent with other TOMM cyclodehydratases, HcaD and HcaF hydrolyzed ATP to ADP and phosphate (Figure S6 and S7).^{9,12,13} As work with previous TOMM cyclodehydratases has demonstrated that ATP hydrolysis is tightly coupled to heterocycle formation,¹² we attempted to obtain kinetic parameters for HcaA modification using a previously described purine nucleoside phosphorylase (PNP)-dependent phosphate detection assay.^{7,12,14} However, initial attempts to determine the rate of HcaA processing were stymied by the high nonspecific ATPase activity in HcaA control reactions. We reasoned that the high cysteine content of HcaA might account for the ATPase activity, and thus we generated truncations of HcaA lacking the C-terminal 36 (HcaA-C53*^{*}; * denotes stop codon) and 29 amino acids (HcaA-C60*^{*}; see Figure S1). Both truncated peptides displayed negligible ATPase activity and were fully processed when treated with HcaD and HcaF in the presence of ATP (Figure S3). The initial rates for each peptide demonstrated that both HcaD and HcaF were required for maximal ATP consumption and that both substrates were modified at similar rates (Figure S7). Moreover, control reactions revealed that removal of the N-terminal MBP-tag from the precursor peptide was not a prerequisite for processing (Figure S7). Consistent with previously characterized TOMM cyclodehydratases, ATP hydrolysis occurs even in the absence of substrate (Figure S6 and S7).¹² Unlike the previously characterized enzymes, the rate of background ATP hydrolysis is ~25% the rate of the HcaA-saturated cyclodehydratase. Combined with the low apparent K_M for the HcaA derivatives and the slow rate of substrate processing, the determination of kinetic parameters was difficult. However, k_{obs} and apparent K_m can be approximated as $1.5 \pm 0.1 \text{ min}^{-1}$ and 5

$\pm 1 \mu\text{M}$, respectively (Figure S7). These values are similar to the parameters obtained for the fused cyclodehydratase TruD (2.6 min^{-1} , $1 \mu\text{M}$).⁹

Intrigued by the ability of the HCA cyclodehydratase to modify truncated analogs of HcaA, we sought to determine the order of cyclodehydration on the native substrate. However, a time course experiment performed with HcaA indicated that the enzyme processed the substrate with a high level of processivity, as partially cyclized species were not observed by MALDI-TOF-MS (Figure S8). When the reaction time course was repeated with HcaA-C60*, a similar result was obtained (Figure S8). As previously characterized TOMM cyclodehydratases have modified their substrates in a distributive fashion,^{7,10,15} the processive modification of HcaA by HcaD is noteworthy.

Although studies performed on other TOMM synthetases have demonstrated that cyclodehydratase activity can be dependent on the presence of the TOMM B protein (dehydrogenase),^{16–18} no other components have been shown to be required for azoline formation. As such, the absolute requirement for the TOMM F protein was unexpected. TOMM C domains have been shown to be responsible for binding the leader peptide.^{3,5} Although the peptide-binding site has not been identified, recognition has been proposed to be accomplished by an N-terminal “peptide clamp” domain akin to that found in the microcin C7 maturation protein MccB.^{3,19} This assignment is bolstered by a recent structure of the nisin dehydratase, NisB, bound to its leader peptide, which demonstrated that a homologous domain is responsible for precursor peptide recognition in the biosynthesis of type I lantipeptides.²⁰ Comparing HcaD to other fused TOMM cyclodehydratases, we noticed that the enzyme was significantly shorter than previously characterized members and appeared to lack the putative “peptide clamp” (Figure S9). As such, we hypothesized that HcaF was involved in leader peptide recognition.

Using size-exclusion chromatography, we first established the oligomeric state of HcaD and HcaF. Individually, both proteins had retention volumes consistent with a monomer (Figure S10). When HcaD and HcaF were analyzed as a 1:1 mixture, a new peak was formed with a retention volume consistent with a 1:1 complex (Figure S10). When MBP-HcaA leader peptide (HcaA-LP; Figure S1) was added to samples containing HcaD, HcaF, or both proteins, a new peak was observed only in samples containing HcaF (Figure S10). In order to quantify the strength of this interaction, an MBP-free fluorescein-tagged version of HcaA-LP was generated, and binding to HcaD and HcaF was assessed by fluorescence polarization (FP). As before, HcaA-LP binding was only observed when HcaF was present (Figure 3), and the addition of HcaD did not affect binding. Importantly, the tight interaction between HcaF and HcaA obtained in this assay was consistent with the high level of processivity observed. Indeed, the large difference between the apparent K_M ($5 \mu\text{M}$) and K_d (37 nM) for HcaA indicates that k_{off} must be slower than k_{cat} , a requirement for processivity.

Based on the primary sequence of HcaF, it was unclear which region of the protein engaged the leader peptide. Although the protein displays weak similarity to E1 superfamily members (see Figure S11), attempts to locate a MccB-like clamp or a NisB-like peptide-binding domain in HcaF using BLAST were unsuccessful. This is not surprising given the high level divergence seen with these domains.^{17,20} Further study of

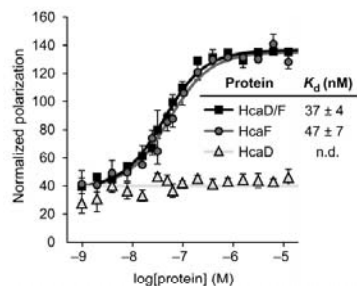


Figure 3. HcaF mediates leader peptide binding. Fluorescence polarization binding experiment with FITC-labeled HcaA-LP and HcaD, HcaF, or HcaD/F. Error bars represent standard deviation from the mean ($n = 3$). Curves were fit according to the methods. Error on K_d values is standard error of the mean determined from curve fitting; n.d., not detected.

HcaF will be required to identify the leader peptide-binding site.

Saturation transfer difference (STD) NMR spectroscopy can be used to detect binding between a protein and ligand;²¹ in certain cases, this has been used for elucidating binding epitopes.²² We performed ^1H STD-NMR to reveal the regions of HcaA-LP contacted by HcaF. Due to the solution instability of HcaA-LP under the conditions used for NMR (aggregation was seen after several hours, limiting the scope of available experiments) and the large number of overlapping, similar resonances (the peptide is replete with Asn, Gln, and Asp residues, for instance, which were difficult to differentiate in the absence of isotope labeling and multidimensional experiments), only partial assignment of the resonances was possible by ^1H and ^1H - ^1H TOCSY (Table S3 and Figures S12 and S13). Nevertheless, resonances from a number of unique residues distributed throughout the peptide were assignable and used to assess contact with HcaF (Figure S14). These data implicated Met1 and Phe4 in binding and demonstrated that the assignable residues in the C-terminus of HcaA-LP did not closely contact HcaF. In addition to these unique residues, several Leu and Tyr residues appeared to be strongly binding; however, due to sequence redundancy, the exact positions could not be unambiguously assigned.

In order to resolve this ambiguity and to corroborate the STD-NMR data, alanine point substitutions were made to HcaA-LP, and recognition by HcaF was assessed using a competition-based FP assay (Table 1 and Figure S15). Consistent with the NMR data, mutation of either Met1 or Phe4 resulted in a 4- and 10-fold reduction in binding affinity, respectively. While all of the Leu mutations resulted in a 100-fold decrease in binding affinity, Tyr16 was the only Tyr residue (of six) that appeared to be critical for recognition. Combined with the STD-NMR data, these results suggest that HcaA-LP recognition occurs primarily through the N-terminal portion of the sequence. In order to test this further, a series of C-terminal truncations were made to the HcaA-LP and binding was assayed by competition FP (Table 1 and Figure S15). Truncations of HcaA-LP of up to 15 residues were tolerated without a decrease in binding affinity; however, removal of an additional five residues severely reduced binding. Combined, our data demonstrate that HcaF engages HcaA-LP through the N-terminal portion of the peptide and that hydrophobic

Table 1. Competition FP Results for HcaA-LP Mutations^a

| HcaA-LP Mutant | IC ₅₀ (nM) | K _i (nM) |
|----------------|-----------------------|---------------------|
| WT | 310 ± 40 | 25 ± 3 |
| M1A | 1200 ± 400 | 180 ± 50 |
| F4A | 3500 ± 700 | 570 ± 110 |
| L8A | ~25,000 | ~4200 |
| L11A | >50,000 | >8500 |
| L13A | >50,000 | >8500 |
| Y16A | >50,000 | >8500 |
| Y23A | 320 ± 60 | 27 ± 5 |
| Y31A | 330 ± 50 | 28 ± 4 |
| Y33A | 380 ± 50 | 37 ± 5 |
| Y34A | 420 ± 40 | 44 ± 4 |
| Y35A | 290 ± 50 | 21 ± 4 |
| V21* | ~25,000 | ~4200 |
| P26* | 340 ± 30 | 30 ± 3 |
| P32* | 370 ± 50 | 35 ± 5 |
| I36* | 370 ± 80 | 35 ± 8 |

^aError on IC₅₀ values is the standard error of the mean determined from curve fitting. Competitive K_i values were calculated from the experimentally-determined IC₅₀. Percent errors from the IC₅₀ measurements were used to obtain the error for the calculated K_i values. An asterisk denotes a stop codon.

residues throughout this region are important for recognition. Notably, previous studies with the TOMM biosynthetic enzymes involved in streptolysin S and microcin B17 biosynthesis have demonstrated that leader peptide is primarily engaged through a FxxxB (B = V, I, or L) motif.^{5,23} This suggests that a conserved strategy for peptide recognition has been propagated through at least part of the TOMM family.

Previous work with TOMM and lanthipeptide biosynthetic enzymes has demonstrated that select members have the ability to process substrates lacking leader peptides, albeit with reduced efficiency.^{11,24–27} The decrease in processing efficiency can often be rescued in part by the addition of the leader peptide *in trans*. These studies have demonstrated that, in addition to providing substrate recognition, leader peptides activate their cognate biosynthetic machineries, presumably by biasing the enzymes to an active conformation. In order to determine if the HCA cyclodehydratase displays similar leader peptide-independent activity, HcaA derivatives lacking the leader peptide (HcaA-Core and HcaA-Core-CS3*⁸; Figure S1) were cloned and expressed as MBP-fusions. Following TEV protease cleavage to remove the MBP-tag, peptides were treated with HcaD and HcaF. Although both substrates were processed by the HCA cyclodehydratase, the efficiency of processing was severely reduced relative to the full-length variants (Figures 4 and S16). Addition of HPLC-purified HcaA-LP to reactions increased the efficiency of processing, suggesting that leader peptide binding activates the HCA cyclodehydratase. Consistent with the important role that the leader peptide plays in substrate binding, modification of the core peptides occurred in a distributive, not processive, fashion regardless of the addition of the leader peptide *in trans* (Figures 4 and S16).

As with reactions performed with full-length HcaA, controls lacking either HcaD or HcaF displayed no processing (Figures 4 and S16). This result suggests that HcaF potentiates HcaD activity in addition to binding HcaA-LP. Although our current data do not reveal the underlying cause of this potentiation, two possibilities include the presentation of the core peptide to the

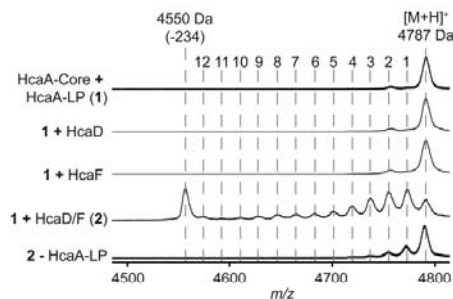


Figure 4. HcaA-LP activates the HCA cyclodehydratase. MALDI-TOF mass spectral overlay for HcaA-Core modification by HcaD/F is displayed. The mass of the starting material and the fully cyclodehydrated species are displayed along with the mass shift from the unmodified peptide. Dashed vertical guides are provided to indicate the expected masses for all partially cyclodehydrated species.

HcaD active site and the allosteric regulation of HcaD. Additional experiments will be required to unravel this phenomenon.

With the major roles for HcaF and the D (YcaO) domain of HcaD determined, we next attempted to determine the role of the C domain of HcaD. In order to accomplish this, we expressed each domain of HcaD (C domain: HcaD-E1; D domain: HcaD-YcaO) individually as an N-terminal His₆-tag fusion (Figure S2). Although the individual domains were inactive alone or with HcaF, a reaction containing all three components displayed cyclodehydratase activity (Figure S17). Based on this *in trans* activity, we reasoned that the interactions between the three components were still intact when HcaD was split into HcaD-E1 and HcaD-YcaO. Repeating the analytical size-exclusion chromatography experiment with the individual domains of HcaD clearly demonstrated that the C domain of HcaD was responsible for binding HcaF and the D domain of HcaD. This result is consistent with the naturally truncated C domain of HcaD serving as a scaffolding protein between the catalytic D domain and the leader peptide-binding, cyclodehydratase-activating TOMM F protein (Figures 5A and S17).

At the time of writing, there are over 300 bioinformatically identifiable TOMM F proteins in UniProt²⁸ distributed almost exclusively in HCA and thiopeptide biosynthetic clusters (Figures 5B, S18, and S19). Consistent with the function described above, 97% of TOMM F proteins are found in clusters containing N-terminally truncated fused TOMM cyclodehydratases (Figure S9). Although the sequence of the TOMM F protein is highly variable (Figure S11), the strong correlation between the presence of an F protein and the absence of the MccB-like peptide clamp on the C domain (Figure S9) suggests leader peptide binding will be a ubiquitous feature of this protein family.

CONCLUSIONS

We have reconstituted the activity of a novel TOMM cyclodehydratase found in nearly all members of the *B. cereus* group and have demonstrated that cyclodehydration is dependent on an uncharacterized biosynthetic protein, which we call the TOMM F protein. Through diverse biophysical assays, we have demonstrated that the TOMM F protein, rather than the C domain of the cyclodehydratase, is responsible for

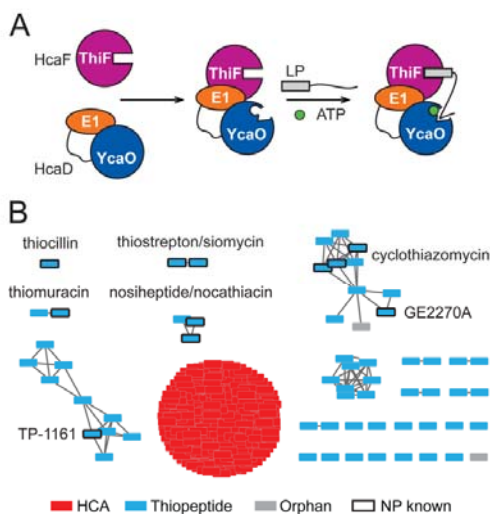


Figure 5. Model for TOMM F protein-dependent cyclodehydratases and distribution of TOMM F proteins. (A) A model for cyclodehydration in clusters containing an N-terminally truncated fused cyclodehydratase and a TOMM F protein is displayed. (B) A sequence similarity network of all bioinformatically identifiable TOMM F proteins is displayed. Each node represents a unique TOMM F protein. Lines connecting nodes share a BLAST Expect value (E -value) $< 10^{-50}$. Nodes are colored based on the class of natural product the cluster is predicted to produce. Nodes are bounded by a black box if the TOMM F protein is found in a cluster for which a natural product (NP) has been structurally characterized.

leader peptide recognition and cyclodehydratase activation. Our data are consistent with the C domain functioning as a scaffold to bring the catalytic D domain of the TOMM cyclodehydratase into contact with the substrate-binding TOMM F protein. Based on our bioinformatic analysis, we predict that the TOMM F protein will be required for the *in vitro* and *in vivo* biosynthesis of natural products from over 300 TOMM clusters (~25% of known TOMM clusters). Indeed, previous work with the nosiheptide (a thiopeptide) biosynthetic cluster demonstrated that the TOMM F protein was required for biosynthesis *in vivo*.²⁹ Although a biosynthetic role could not be provided at that time, we propose that the protein is required for precursor peptide recognition and cyclodehydratase activation. Additional studies will be necessary to confirm this hypothesis and to identify the leader peptide-binding domain in the TOMM F protein.

■ ASSOCIATED CONTENT

Supporting Information

Experimental methods and supporting figures. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.5b04682.

■ AUTHOR INFORMATION

Corresponding Author

*dougasm@illinois.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We are grateful to J. Melby for cloning HcaA2, N. Trinh for cloning HcaA, L. Zhu for assistance with the STD-NMR, and E. Molloy for the critical review of the manuscript. This work was supported by the US National Institutes of Health (NIH) (1R01 GM097142 to D.A.M. and 2T32 GM070421 to K.L.D., J.I.T., and B.J.B.). Additional support was from the Harold R. Snyder Fellowship (UIUC-Dept. of Chemistry to K.L.D.), the Robert C. and Carolyn J. Springborn Endowment (UIUC-Dept. of Chemistry to J.I.T. and B.J.B.), and the National Science Foundation Graduate Research Fellowship (DGE-1144245 to B.J.B.). The Bruker UltrafleXtreme MALDI TOF/TOF mass spectrometer was purchased in part with a grant from the National Center for Research Resources, NIH (S10 RR027109 A). A portion of the NMR data was collected in the IGB Core on a 600 MHz NMR funded by NIH grant number S10-RR028833.

■ REFERENCES

- (1) Melby, J. O.; Nard, N. J.; Mitchell, D. A. *Curr. Opin. Chem. Biol.* **2011**, *15*, 369.
- (2) Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J.; Cotter, P. D.; Craik, D. J.; Dawson, M.; Dittmann, E.; Donadio, S.; Dorrestein, P. C.; Entian, K. D.; Fischbach, M. A.; Garavelli, J. S.; Goransson, U.; Gruber, C. W.; Haft, D. H.; Hemscheidt, T. K.; Hertweck, C.; Hill, C.; Horswill, A. R.; Jaspars, M.; Kelly, W. L.; Klinman, J. P.; Kuipers, O. P.; Link, A. J.; Liu, W.; Marahiel, M. A.; Mitchell, D. A.; Moll, G. N.; Moore, B. S.; Muller, R.; Nair, S. K.; Nes, I. F.; Norris, G. E.; Olivera, B. M.; Onaka, H.; Patchett, M. L.; Piel, J.; Reaney, M. J. T.; Rebuffat, S.; Ross, R. P.; Sahl, H. G.; Schmidt, E. W.; Selsted, M. E.; Severinov, K.; Shen, B.; Sivonen, K.; Smith, L.; Stein, T.; Sussmuth, R. D.; Tagg, J. R.; Tang, G. L.; Truman, A. W.; Vederas, J. C.; Walsh, C. T.; Walton, J. D.; Wenzel, S. C.; Willey, J. M.; van der Donk, W. A. *Nat. Prod. Rep.* **2013**, *30*, 108.
- (3) Dunbar, K. L.; Chekan, J. R.; Cox, C. L.; Burkhart, B. J.; Nair, S. K.; Mitchell, D. A. *Nat. Chem. Biol.* **2014**, *10*, 823.
- (4) Dunbar, K. L.; Mitchell, D. A. *ACS Chem. Biol.* **2013**, *8*, 473.
- (5) Mitchell, D. A.; Lee, S. W.; Pence, M. A.; Markley, A. L.; Limm, J. D.; Nizet, V.; Dixon, J. E. *J. Biol. Chem.* **2009**, *284*, 13004.
- (6) Haft, D. H. *Biol. Direct* **2009**, *4*, 15.
- (7) Melby, J. O.; Dunbar, K. L.; Trinh, N. Q.; Mitchell, D. A. *J. Am. Chem. Soc.* **2012**, *134*, 5309.
- (8) Chopra, L.; Singh, G.; Choudhary, V.; Sahoo, D. K. *Appl. Environ. Microbiol.* **2014**, *80*, 2981.
- (9) McIntosh, J. A.; Schmidt, E. W. *ChemBioChem* **2010**, *11*, 1413.
- (10) Koehnke, J.; Bent, A. F.; Zollman, D.; Smith, K.; Houssen, W. E.; Zhu, X.; Mann, G.; Lebl, T.; Scharff, R.; Shirran, S.; Botting, C. H.; Jaspars, M.; Schwarz-Linek, U.; Naismith, J. H. *Angew. Chem., Int. Ed. Engl.* **2013**, *52*, 13991.
- (11) Goto, Y.; Ito, Y.; Kato, Y.; Tsunoda, S.; Suga, H. *Chem. Biol.* **2014**, *21*, 766.
- (12) Dunbar, K. L.; Melby, J. O.; Mitchell, D. A. *Nat. Chem. Biol.* **2012**, *8*, 569.
- (13) Milne, J. C.; Eliot, A. C.; Kelleher, N. L.; Walsh, C. T. *Biochemistry* **1998**, *37*, 13250.
- (14) Webb, M. R. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 4884.
- (15) Kelleher, N. L.; Hendrickson, C. L.; Walsh, C. T. *Biochemistry* **1999**, *38*, 15623.
- (16) Li, Y. M.; Milne, J. C.; Madison, L. L.; Kolter, R.; Walsh, C. T. *Science* **1996**, *274*, 1188.
- (17) Lee, S. W.; Mitchell, D. A.; Markley, A. L.; Hensler, M. E.; Gonzalez, D.; Wohlrab, A.; Dorrestein, P. C.; Nizet, V.; Dixon, J. E. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5879.
- (18) Gonzalez, D. J.; Lee, S. W.; Hensler, M. E.; Markley, A. L.; Dahesh, S.; Mitchell, D. A.; Bandeira, N.; Nizet, V.; Dixon, J. E.; Dorrestein, P. C. *J. Biol. Chem.* **2010**, *285*, 28220.

- (19) Regni, C. A.; Roush, R. F.; Miller, D. J.; Nourse, A.; Walsh, C. T.; Schulman, B. A. *EMBO J.* **2009**, *28*, 1953.
- (20) Ortega, M. A.; Hao, Y.; Zhang, Q.; Walker, M. C.; van der Donk, W. A.; Nair, S. K. *Nature* **2015**, *517*, 509.
- (21) Wagstaff, J. L.; Taylor, S. L.; Howard, M. J. *Mol. Biosyst.* **2013**, *9*, 571.
- (22) Hurtado-Gomez, E.; Abian, O.; Munoz, F. J.; Hernaiz, M. J.; Velazquez-Campoy, A.; Neira, J. L. *Biophys. J.* **2008**, *95*, 1336.
- (23) Roy, R. S.; Kim, S.; Baleja, J. D.; Walsh, C. T. *Chem. Biol.* **1998**, *5*, 217.
- (24) Dunbar, K. L.; Mitchell, D. A. *J. Am. Chem. Soc.* **2013**, *135*, 8692.
- (25) Khusainov, R.; Kuipers, O. P. *ChemBioChem* **2012**, *13*, 2433.
- (26) Levensgood, M. R.; Patton, G. C.; van der Donk, W. A. *J. Am. Chem. Soc.* **2007**, *129*, 10314.
- (27) Oman, T. J.; Knerr, P. J.; Bindman, N. A.; Velasquez, J. E.; van der Donk, W. A. *J. Am. Chem. Soc.* **2012**, *134*, 6952.
- (28) UniProt Consortium. *Nucleic Acids Res.* **2014**, *42*, D191.
- (29) Yu, Y.; Duan, L.; Zhang, Q.; Liao, R.; Ding, Y.; Pan, H.; Wendt-Pienkowski, E.; Tang, G.; Shen, B.; Liu, W. *ACS Chem. Biol.* **2009**, *4*, 855.

A.4 *In vitro* biosynthesis and substrate tolerance of the plantazolicin family of natural products

Reproduced with permission from Deane, C.D.; Burkhart, B.J.; Blair, P.M.; Tietz, J.I.; Lin, A.; Mitchell, D.A. *In vitro* biosynthesis and substrate tolerance of the plantazolicin family of natural products. *ACS Chem. Biol.* **2016**, *11*, 2232-2243. PMID: PMC4992447. Copyright 2016 American Chemical Society.

I performed experiments to investigate leader peptide recognition by the PZN synthetase and homologous proteins from the Bpum, Cms, and Blin biosynthetic gene clusters. From these results, I authored a section of this paper entitled “Leader Peptide Recognition and Binding by the PZN Synthetase.” I also assisted in the overall editing and writing of the paper and generated Table S1 and S2 and Figure S3, 5, S6, S40, and S41.

In Vitro Biosynthesis and Substrate Tolerance of the Plantazolicin Family of Natural Products

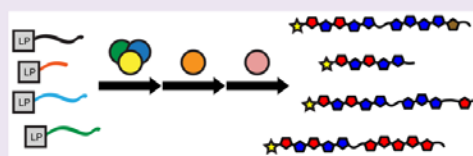
Caitlin D. Deane,^{†,‡} Brandon J. Burkhardt,^{†,‡} Patricia M. Blair,[†] Jonathan I. Tietz,[†] Alice Lin,[‡] and Douglas A. Mitchell^{*,†,‡,§}

[†]Department of Chemistry, [‡]Carl R. Woese Institute for Genomic Biology, and [§]Department of Microbiology, University of Illinois at Urbana—Champaign, Urbana, Illinois, United States

Supporting Information

ABSTRACT: Plantazolicin (PZN) is a ribosomally synthesized and post-translationally modified peptide (RiPP) natural product that exhibits extraordinarily narrow-spectrum antibacterial activity toward the causative agent of anthrax, *Bacillus anthracis*. During PZN biosynthesis, a cyclodehydratase catalyzes cyclization of cysteine, serine, and threonine residues in the PZN precursor peptide (BamA) to azolines. Subsequently, a dehydrogenase oxidizes most of these azolines

to thiazoles and (methyl)oxazoles. The final biosynthetic steps consist of leader peptide removal and dimethylation of the nascent *N*-terminus. Using a heterologously expressed and purified heterocycle synthetase, the BamA peptide was processed *in vitro* concordant with the pattern of post-translational modification found in the naturally occurring compound. Using a suite of BamA-derived peptides, including amino acid substitutions as well as contracted and expanded substrate variants, the substrate tolerance of the heterocycle synthetase was elucidated *in vitro*, and the residues crucial for leader peptide binding were identified. Despite increased promiscuity compared to what was previously observed during heterologous production in *E. coli*, the synthetase retained exquisite selectivity in cyclization of unnatural peptides only at positions which correspond to those cyclized in the natural product. A cleavage site was subsequently introduced to facilitate leader peptide removal, yielding mature PZN variants after enzymatic or chemical dimethylation. In addition, we report the isolation and characterization of two novel PZN-like natural products that were predicted from genome sequences but whose production had not yet been observed.



Natural products are a rich source of chemically and mechanistically diverse antibiotics.¹ Within this broad space, the ribosomally synthesized and post-translationally modified peptide (RiPP) natural products comprise a class garnering increased attention.² Plantazolicin (PZN) is an anti-*B. anthracis* RiPP produced by *Bacillus velezensis* (formerly *Bacillus amyloliquefaciens*) FZB42^{3–6} and a member of a subclass of RiPPs termed the linear azol(in)e-containing peptides (LAPs).^{7,8} LAPs, thiopeptides, and azol(in)e-containing cyanobactins can also collectively be referred to as thiazole/oxazole-modified microcins (TOMMs).² During TOMM biosynthesis, a trimeric heterocycle synthetase (BCD) converts select Cys, Ser, and Thr residues in the C-terminal (core) region of the precursor peptide to thiazole, oxazole, and methyloxazol(in)e moieties (Supporting Information Figure 1a).⁷ The D-protein (a member of the YcaO superfamily), in partnership with the C-protein, first installs azolines through an ATP-mediated cyclodehydration which proceeds through a phosphorylated hemiothoamide,^{9,10} and the B-protein [flavin mononucleotide (FMN)-dependent dehydrogenase] then oxidizes select azolines to azoles.¹¹ The C-protein serves to bind the *N*-terminal (leader) region of the precursor peptide, enhancing the rate of processing.^{12–14} Following heterocyclization, the biosynthesis of PZN is completed by removal of the leader peptide, likely by a type II CaaX protease,^{15,16} and

dimethylation of the new *N*-terminus by an *S*-adenosylmethionine (SAM)-dependent methyltransferase.^{17,18}

Though originating from a peptide, the structure of mature PZN after post-translational modification is remarkably non-peptide-like (Figure 1a), with two sets of five contiguous heterocycles conferring rigidity and hydrophobicity.^{4,5} In addition to the PZN biosynthetic gene cluster (BGC) found in *B. velezensis*, homologous gene clusters have been identified in a number of other bacterial strains (Figure 1b, Supporting Information Figure 1),⁸ although only *Bacillus pumilus* has been demonstrated to produce PZN.⁴ The putative precursor peptides associated with these BGCs exhibit a highly conserved core region, which would result in similar groupings of contiguous heterocycles if indeed post-translationally modified like PZN (Figure 1c, Supporting Information Figure 1). The unique structure of PZN endows it with an extremely narrow spectrum of activity and a membrane-targeting mode of action unlike other known antibiotics.¹⁹ PZN exhibits bactericidal activity against *B. anthracis*, the causative agent of anthrax, with a minimum inhibitory concentration (MIC) of 1–2 $\mu\text{g/mL}$, while not killing other species, even closely related ones.^{4,19} It

Received: April 28, 2016

Accepted: June 1, 2016

Published: June 1, 2016

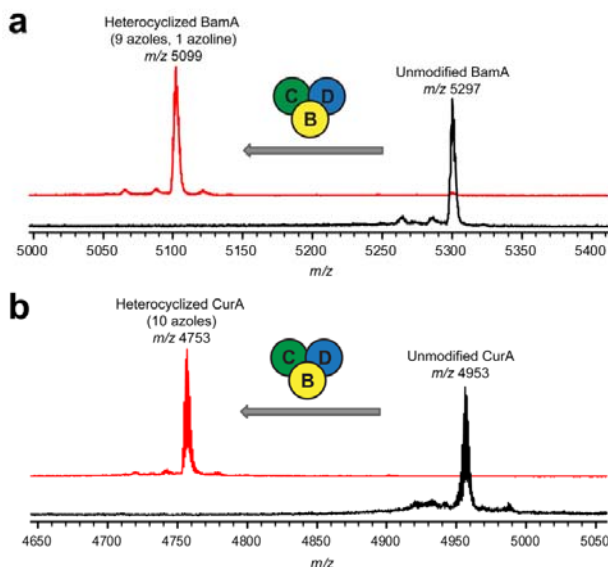


Figure 2. (a) Reaction of the PZN precursor peptide BamA with the *Bam* heterocycle synthetase *in vitro* results in the installation of nine azoles and one azoline, consistent with PZN produced from the native host. (b) Reaction of the precursor peptide CurA with the *Cur* heterocycle synthetase *in vitro* results in the installation of 10 azoles.

(MBP) and affinity purified (Supporting Information Figure 2). Unfortunately, all attempts to express MBP-BamC yielded only degraded protein. A codon-optimized version of the gene for *E. coli* expression (BamC_{opt}; Supporting Methods) was synthesized and heterologously expressed to yield a modest amount of enzyme (~2 mg/L; Supporting Information Figure 2), but size-exclusion chromatography indicated that BamC_{opt} did not form an observable complex with BamB and BamD, which could impair the activity of the synthetase (Supporting Information Figure 3a). In order to obtain more significant quantities of properly functioning protein, we elected to substitute BamC with a closely related homologue, which is an approach that has been successful for other heterocycle synthetases.^{42,43} The C-protein from the PZN BGC in *Bacillus pumilus* (BpumC, accession number WP_041117410; Figure 1b) shares 91% amino acid similarity (63% identity) with BamC⁴ and, when expressed as an MBP fusion protein in *E. coli*, yielded over 90 mg/L following affinity purification (Supporting Information Figure 2). In contrast to BamC_{opt}, BpumC formed a complex with BamB and BamD as assessed by size exclusion chromatography (Supporting Information Figure 3b). Gratifyingly, when these MBP-tagged proteins (BamB, BpumC, and BamD) were combined with MBP-BamA, TEV protease (to remove the MBP tags), and ATP, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) indicated a mass loss consistent with the formation of nine azoles and one azoline (−198 Da) onto the precursor peptide (Figure 2a). Subsequent Fourier transform ion cyclotron resonance tandem mass spectrometry (FT-ICR-MS/MS) was used to localize these modifications, which was identical to the *bona fide* natural product (Supporting Information Figure 4). While BamC_{opt} did not appear to form a stable complex with other components of the synthetase

(Supporting Information Figure 3a), overnight reactions with BamC_{opt} did form the expected −198 Da product (Supporting Information Figure 5). Ultimately, the more robust protein yield and activity of BpumC warranted its usage in all subsequent assays.

Bolstered by our success with *in vitro* reconstitution of the *Bam* heterocycle synthetase, we also set about reconstituting the homologous synthetase from *Corynebacterium urealyticum* (*Cur*, Figure 1b).^{4,5} Size-exclusion chromatography indicated that CurB, CurC, and CurD formed a complex (Supporting Information Figure 6), and combination of the MBP-tagged synthetase proteins with ATP, TEV, and MBP-CurA resulted in a mass loss of 200 Da for CurA, consistent with the installation of 10 azoles (Figure 2b).

Although some TOMM cyclodehydratase enzymes are capable of activity in the absence of their cognate dehydrogenase, or even in the absence of their cognate C-protein,^{9,43} this was not the case for either the *Bam* or *Cur* cyclodehydratases, as the exclusion of any one component of the synthetase from either BGC abolished activity (Supporting Information Figure 7). For these synthetases, all three components of the heterocycle synthetase were required for activity, as has been demonstrated for most other studied TOMM clusters, including microcin B17 and streptolysin S.^{7,42,44} Thus, in order to reveal the order of biosynthetic events, the enzymatic activities of the cyclodehydratase and dehydrogenase needed to be disentangled. Toward this goal, we prepared two inactive forms of the dehydrogenase, BamB-Y206A and BamB-R93A. BamB-Y206A contains an alanine substitution at a key active site residue,^{11,45} while BamB-R93A no longer copurified with FMN upon heterologous expression in *E. coli* but still associated with BpumC/BamD by size-exclusion chromatography (Supporting Information Figure 3c). When MBP-BamA

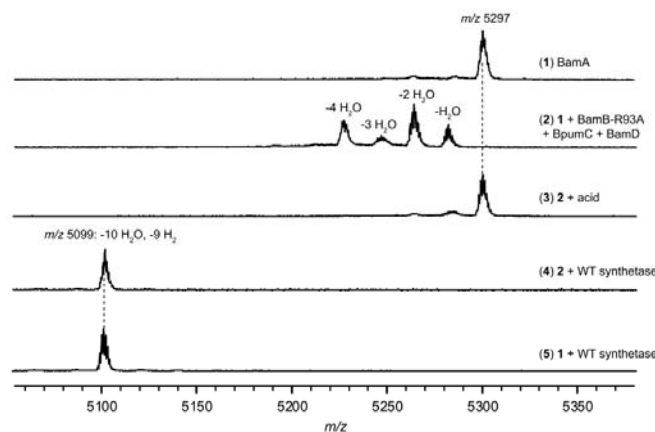


Figure 3. Reaction of BamA (1) with a heterocycle synthetase containing the inactive BamB-R93A results in the installation of up to four azolines, without their oxidation to azoles (2). Subsequent treatment of this species mixture with acid hydrolyzes the azolines to their original amino acids (3), whereas treatment with a fully wild-type synthetase yields the full complement of nine azoles and one azoline (4), comparable to treatment of unmodified BamA with wild-type synthetase (5). Substitution of the catalytically inactive BamB-Y206A for BamB-R93A yields equivalent results.

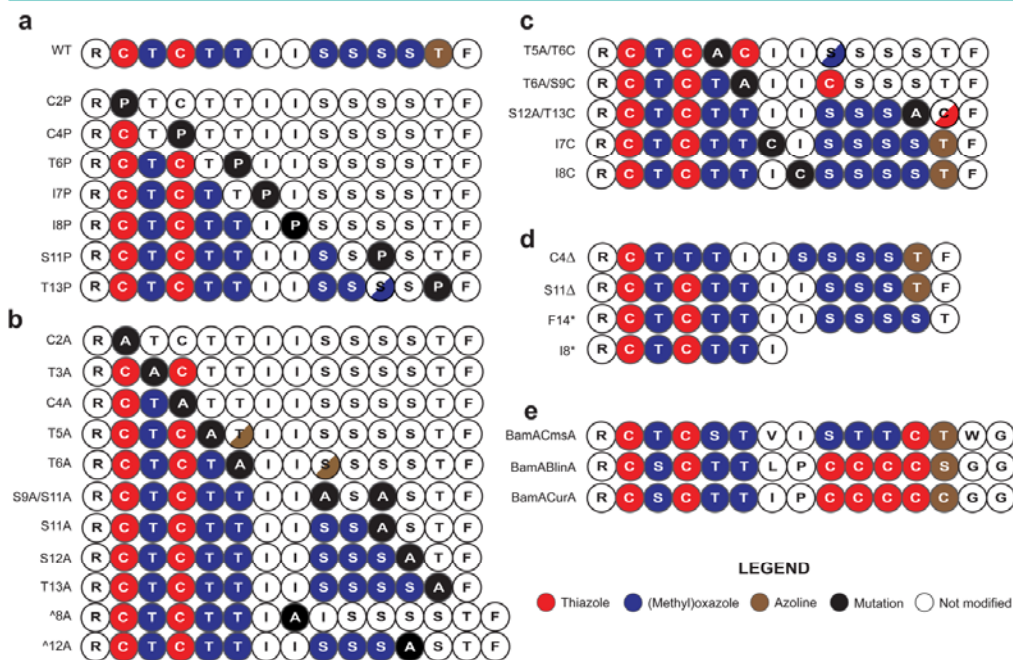


Figure 4. Processing of mutant BamA peptides by the Bam synthetase *in vitro*. Mutated and uncyclized residues are shown in solid black circles; half-filled circles indicate a mixture of species. (a) Mutation of most residues to Pro prevents cyclization both C-terminal and in the -1 position to the Pro. Wild-type BamA (WT) is shown for comparison. (b) Mutation of a normally cyclized position to Ala, or insertion of Ala between two wild-type residues, generally prevents cyclization C-terminal to the Ala. (c) A Cys residue in the next heterocyclizable position relative to Ala enables cyclodehydration of that residue, though Cys in a position not normally cyclized remains unmodified. (d) Deletion of a residue does not affect processing by the synthetase, though one residue must remain uncyclized at the C-terminus. (e) The Bam synthetase *in vitro* is capable of processing substrates with the core peptides of other members of the PZN family.

was treated with BpumC, BamD, and either BamB-R93A or BamB-Y206A, the cyclodehydratase installed at most four azoline heterocycles, which were localized to the four most N-terminal positions in the core peptide (Figure 3, Supporting Information Figure 8). Subsequent treatment of this species with synthetase containing an active dehydrogenase yielded a fully modified substrate (nine azoles and one azoline). While the four-azoline species proved to be a competent *in vitro* intermediate, it is not necessarily a natural, on-pathway biosynthetic intermediate. Nevertheless, one of these four azolines is most likely the first to be installed during the normal course of biosynthesis. All attempts to isolate additional intermediates from the Bam or Cur synthetase reactions were unsuccessful due to the apparent high processivity of the enzymes (Supporting Information Figure 9). However, the observation that the cyclodehydratase cannot install all 10 azolines in the absence of an active dehydrogenase is strongly suggestive that during the normal reaction course, azole formation begins at the N-terminal residues prior to completion of azoline formation at the C-terminal residues.

Substrate Tolerance of the PZN Synthetase *in Vitro*.

With the PZN heterocycle synthetase reconstituted *in vitro*, we next turned our attention to determining the parameters of substrate tolerance for these enzymes with variant precursor peptides. MBP-tagged precursors with core region substitutions were generated using site-directed mutagenesis and *E. coli* heterologous expression. After treating the panel of purified variant peptides with the heterocycle synthetase *in vitro*, the extent of processing was determined by MALDI-TOF-MS, with subsequent FT-ICR-MS/MS to localize the sites of post-translational modification (Figure 4, Supporting Information Figures 10–36). Overall, the ability of the synthetase to process variant substrates was increased over what was previously observed in *E. coli*,⁴¹ presumably due to a greater level of control over the reaction conditions (e.g., higher concentrations of reaction components are possible *in vitro*). Nonetheless, many of the variant peptides remained incompletely processed, reaffirming that the PZN cyclodehydratase is remarkably substrate selective.

Initially, we examined the limits of substrate scope for the Bam synthetase by creating a series of variant peptides wherein a Cys, Ser, or Thr residue was replaced with Pro, which has been used as a crude steric mimic for an azoline in previous studies.^{43,46} While the Bam synthetase was able to install heterocycles in the expected positions N-terminal to the interrogated Pro residue, neither the residue proximal to Pro nor any residues following Pro were cyclized (Figure 4a, Supporting Information Figures 10–15). Presumably, the conformational restriction imparted on the peptide backbone by the presence of Pro in the +1 position prevented the Bam cyclodehydratase from installing a heterocycle at that position, perhaps due to an inability to properly fit the substrate in the active site or some other catalytic deficiency. While such an intolerance for Pro in the +1 position is consistent with previous studies using the cyclodehydratase from *Bacillus* sp. Al Hakam,⁴³ the Cur cyclodehydratase, among other TOMM natural products,^{47,48} is capable of accepting Pro in the –1 position (Figure 1c, Figure 2b).

Similar N-terminal but not C-terminal processing was also observed for a number of Ala substitutions, though the Bam cyclodehydratase was able to cyclize residues with Ala in the +1 position (Figure 4b, Supporting Information Figures 16–25). In a few cases involving Ala near the middle of the core peptide,

the cyclodehydratase was capable of installing an azoline in the next heterocyclizable position after the Ala, but this was always a minor species, perhaps due in part to hydrolysis of the azoline. Dehydrogenation of this azoline to the azole was never observed in the position following Ala, nor was cyclodehydration observed in more than one position following Ala. This abortive trend was also observed in cases where an Ala residue was inserted between two wild-type residues, including between the two central Ile residues in the core peptide (^8A, ^12A; ^ indicates insertion). Together, these results further support the hypothesis that heterocyclization during PZN biosynthesis proceeds in an N- to C-terminal direction.

The single notable exception to this trend of Ala preventing significant heterocyclization of C-terminal residues was in the case of the BamA-T3A substrate, for which the synthetase fully cyclized both Cys2 and Cys4 to thiazoles (Figure 4b, Supporting Information Figure 16). Further heterocyclization C-terminal to Cys4 was not observed, consistent with the other Ala-substituted precursor peptides. It is possible that the increased nucleophilicity of the Cys side chain relative to that of Ser/Thr, plus the more electrochemically favorable oxidation to thiazole relative to oxazole, were responsible for the ability of the synthetase to overcome the introduction of Ala in the preceding (–1) position. To test this hypothesis, three double substitutions were constructed and assayed for modification in which the Ala residue was followed by Cys in the next position normally heterocyclized: T5A/T6C, T6A/S9C, and S12A/T13C. In each of these cases, the synthetase was able to convert the unnatural Cys to a thiazole, though some Cys remained unprocessed in the S12A/T13C peptide (Figure 4c, Supporting Information Figures 26–28). Together, these results demonstrate that the increased reactivity of Cys, in essence, chemically rescues processing when the activity of the synthetase would normally be halted by Ala. However, when Cys was placed at positions not heterocyclized in the wild-type PZN (i.e., I7C and I8C), susceptibility to iodoacetamide labeling demonstrated that these Cys residues remained unmodified, whereas other positions of BamA were processed as in the wild-type peptide (Figure 4c, Supporting Information Figure 29). Such exquisite site-selectivity is consistent with that previously observed in *E. coli*.⁴¹

To further examine the effects of disrupting the natural arrangement of heterocyclizable residues, several precursors with contracted core peptides were constructed and assayed for modification by the Bam synthetase. Shortening the precursor peptide by one residue in the center of either block of five heterocyclizable residues (C4Δ, S11Δ; Δ indicates deletion) did not significantly disrupt processing, as assessed by the MALDI-TOF-MS end point assay, resulting in a maximum of nine heterocycles installed (Figure 4d, Supporting Information Figures 30 and 31). Likewise, truncation of the precursor peptide to remove the C-terminal half of the core peptide (I8*; * indicates a stop codon) also led to the installation of five heterocycles (Figure 4d, Supporting Information Figure 32). Finally, removal of the C-terminal Phe residue (F14*) prevented heterocyclization only at the preceding residue, Thr13 (Figure 4d, Supporting Information Figure 33), indicating that the Bam synthetase is unable to install a C-terminal heterocycle.

In addition to the precursors with one or two substituted residues, three potential substrates with more extensive alterations were also assayed. These chimeric peptides consisted of the leader peptide from BamA and the core peptides from

three other BGCs in the PZN family. These peptides were previously assayed in *E. coli*, where no post-translational modifications were observed.⁴¹ However, the reaction conditions using purified enzymes now enabled modification of all three chimeric peptides (Figure 4e, Supporting Information Figures 34–36). The ability of the synthetase to modify the BamA-CurA and BamA-BlnA chimeras is especially notable, as each contains an internal Pro residue. As previously demonstrated in this work, the presence of a Pro residue in the BamA core peptide prevented heterocyclization in positions C-terminal to the Pro (Figure 4a). Yet in the case of BamA-CurA and BamA-BlnA, Pro8 is followed immediately by several Cys residues. We hypothesize that the increased nucleophilicity of the Cys side chain again may be chemically rescuing an otherwise unprocessed substrate, as was the case for the Ala/Cys double mutants (Figure 4c). The processing of BamA-CurA also demonstrates site-selectivity on the part of the PZN dehydrogenase, as it did not oxidize the penultimate position to the thiazole, whereas CurB oxidized all azolines in CurA to the corresponding azoles (Figure 2b, Supporting Information Figure 37).

In addition to the heterocyclizable residues, the other key position in BamA at which substitution was not tolerated by the heterocycle synthetase in *E. coli* was Arg1.⁴¹ Previously, it was unclear whether the intolerance toward even the most conservative substitution (Arg to Lys) at this position was due to selectivity by the cyclodehydratase or prevention of leader peptide cleavage. With the Bam synthetase reconstituted *in vitro*, its tolerance for alternative residues at the N-terminus of the core peptide could now be assessed in the absence of the leader peptidase. *In vitro*, the R1K construct was accepted by the Bam synthetase, though not robustly, as indicated by the presence of a number of incompletely cyclized species (Supporting Information Figure 38a). In addition to the “conservative” R1K substitution, a variety of other alterations to this position were also assayed. The Bam synthetase was revealed to be more permissive to small residues (R1G, R1S) than to bulky (R1I, R1F) or negatively charged (R1D) residues, though only in the case of R1K could the synthetase install the full complement of 10 heterocycles (Supporting Information Figure 38a). The initial enzymatic rate was also determined for the synthetase in the presence of each of these Arg1 mutants, and those substrates which permitted the installation of more heterocycles also generally had greater initial velocities (Supporting Information Figure 38b). The Bam synthetase thus appeared to strongly favor Arg1.

In addition to determining the relative rates of processing for the Arg1 variants, several other precursor peptides were kinetically assayed using the Bam synthetase (Table 1). The

observed V_0 values for these substrate variants were relatively unperturbed regardless of the number of heterocycles installed or the substituted position in the core. Such similarity suggested that the first cyclodehydration event (Cys2) is the rate-determining step, and this result offers further support that modification proceeds in an N- to C-terminal direction. Indeed, many data reported herein would be difficult to reconcile if an alternative processing order was operating. Meanwhile, in contrast to the similar V_0 values, the apparent K_m values varied greatly. Although it can be difficult to interpret kinetic parameters for a multistep reaction where substrate binding is primarily dictated at a location distal to the active site (i.e., the C-protein engages the leader peptide, *vide infra*),¹² lower apparent K_m values correlated with an increased total number of modifications, such as with the native substrate and T13P (Figure 4a). Such a trend suggests that substrates receiving more modifications interact more productively with the active site, which is expected for any substrate with a lower K_m . The higher apparent K_m for the T3A and T6A variants also suggests that these Thr residues enhance productive interactions with the enzyme, perhaps through the β -methyl substituent on their side chains.

Leader Peptide Recognition and Binding by the PZN Synthetase. In order to more fully understand the substrate scope of the heterocycle synthetases, we next investigated the contribution of the leader peptide toward substrate binding and processing. Previous work has shown that CurC, and not CurD, interacts with a CurA leader peptide, exemplifying the previously described role for the C-protein component of the heterocycle synthetase in leader peptide binding via the RiPP recognition element (RRE) domain.¹² Building on these previous results, we examined the interaction between an N-terminally FITC-labeled BamA leader peptide (FITC-BamA-LP, Supporting Information Table 1) and the synthetase proteins by fluorescence polarization (FP). Unexpectedly, BpumC on its own was incapable of avid FITC-BamA-LP binding, with an extrapolated K_d of approximately 100 μ M (Figure 5). In the presence of BamD, the K_d decreased to $20 \pm 2 \mu$ M, though BamD alone did not substantially bind the leader peptide (Figure 5), indicating that BpumC was indeed the primary engaging protein, in line with previous results. The surprisingly weak binding by BpumC alone was not a result of the noncognate nature of the BamA/BpumC interaction, since the binding of FITC-BpumA-LP (the cognate sequence) to BpumC gave comparable results ($K_d \sim 100 \mu$ M, extrapolated; Figure 5). In light of the observation that the full, trimeric PZN heterocycle synthetase (BamB/BpumC/BamD) was required to modify BamA (Supporting Information Figure 7), we hypothesized that BpumC requires its associated partners for efficient leader peptide binding. To test this, BamB-R93A, which lacks the FMN cofactor, was employed so that the intrinsic fluorescence of the flavin would not interfere in the FP assay. When BamB-R93A, BpumC, and BamD were combined to form the complete heterocycle synthetase, leader peptide binding was considerably enhanced ($K_d = 1 \pm 0.1 \mu$ M, Figure 5). This result is in stark contrast to several previously studied TOMM synthetases, where the C-protein, which contains the ~ 90 residue RRE, was sufficient for effective peptide binding.^{12,46} We hypothesize that some C-proteins must leverage their association with other synthetase components to induce structural rearrangements that potentiate leader peptide binding.⁴⁹

Table 1. Kinetics of Select BamA Substrates with the Bam Synthetase

| mutant | V_0 (μ M P _i min ⁻¹) | apparent K_m (μ M) | heterocycles installed |
|-------------|--|---------------------------|------------------------|
| WT | 1.76 \pm 0.09 | 3.22 \pm 1.08 | 10 |
| T3A | 1.71 \pm 0.13 | 26.8 \pm 5.9 | 2 |
| T6A | 2.74 \pm 0.25 | 55.5 \pm 11.3 | 4 |
| I8P | 2.35 \pm 0.12 | 7.61 \pm 2.28 | 5 |
| S12A | 2.35 \pm 0.19 | 8.36 \pm 2.86 | 8 |
| T13P | 2.48 \pm 0.09 | 5.73 \pm 1.26 | 8 |
| C4 Δ | 1.88 \pm 0.22 | 9.37 \pm 4.3 | 9 |
| BamA-CurA | 1.78 \pm 0.12 | 12.3 \pm 2.8 | 10 |

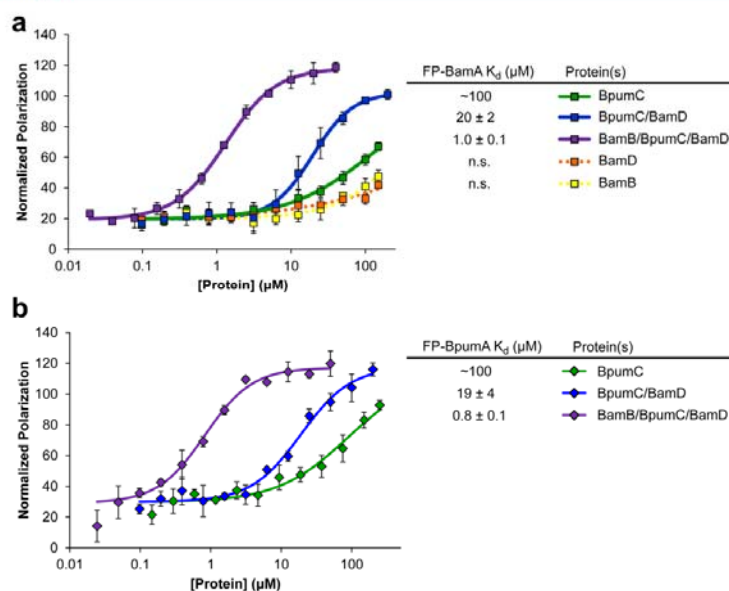


Figure 5. Fluorescence polarization (FP) binding assay with BamB-R93A/BpumC/BamD and FITC-labeled BamA leader peptide (LP) (a) or FITC-labeled-BpumA-LP (b). K_d values not shown are $>150 \mu\text{M}$. Error bars represent SD of at least three independent measurements. Errors on K_d values are given as the SEM from regression analysis. For binding curves that do not reach saturation, K_d values were extrapolated.

To investigate the importance of various precursor peptide regions for substrate binding, we constructed a series of C-terminal truncations (Table 2). The results indicated that not only did the core region not detectably contribute to the overall binding of the peptide, but also that the first 24 residues of the precursor peptide were minimally necessary to maintain wild-type binding affinity. Next, Ala scanning of BamA was performed to identify individual residues of the leader peptide (designated by negative numbers) responsible for interactions with the synthetase. As measured by a competition FP assay (Table 2), BamA and BpumA had nearly equal IC_{50} values for the synthetase, which were also similar to the K_d of their fluorescent derivatives. Overall, most of the single Ala substitutions had only modest effects on the binding of the peptide, although P(-21)A and L(-18)A had a much more substantial effect, which was borne out by kinetic assays (Table 2).

To ensure that these observations were not *in vitro* artifacts, a previously described *E. coli* heterologous production system was used to assess processing of the BamA variants *in vivo*.⁴¹ In this assay, the production of the mature PZN-I7V variant (a known, well-tolerated core peptide substitution) from a plasmid-borne copy of the *bamA* gene containing a leader peptide substitution was compared to the production of the wild-type PZN from a fosmid-borne copy, as measured by MALDI-TOF-MS (Table 2; Supporting Information Figure 39). Overall, these *in vivo* results corroborated those obtained *in vitro*, indicating that two residues, Pro(-21) and Leu(-18), are primarily responsible for the peptide's interaction with the synthetase in what appears to be a $_{-21}\text{PXXL}_{-18}$ recognition motif. Interestingly, in other precursor peptides within the PZN family, this Pro is highly conserved, whereas the Leu is less so

Table 2. Binding and Processing of Leader Peptide Mutants

| peptide | IC_{50} (μM) | <i>in vitro</i> (%) ^b | <i>in vivo</i> ^c |
|--------------------------|------------------------------------|----------------------------------|-----------------------------|
| BamA-WT | 2.8 ± 0.3 | 100 ± 9.8 | +++ |
| BamA-T(-26)A | 5.4 ± 0.6 | 108 ± 6.7 | +++ |
| BamA-Q(-25)A | 4.4 ± 0.6 | 94 ± 9.8 | ++ |
| BamA-I(-24)A | 5.5 ± 0.7 | 117 ± 5.4 | + |
| BamA-K(-23)A | 5.9 ± 0.6 | 89 ± 5.1 | ++ |
| BamA-V(-22)A | 4.8 ± 0.3 | 102 ± 1.0 | + |
| BamA-P(-21)A | >100 | 54 ± 8.6 | — ^d |
| BamA-T(-20)A | 5.5 ± 0.7 | 95 ± 2.7 | ++ |
| BamA-L(-18)A | >100 | 63 ± 2.9 | — |
| BamA-I(-17)A | 2.7 ± 0.3 | 99 ± 7.1 | ++ |
| BamA-S(-15)A | 9 ± 2 | 100 ± 1.3 | + |
| BamA-V(-14)A | 8.2 ± 0.9 | 110 ± 11.7 | + |
| BamA-H(-13)A | 13 ± 3 | 106 ± 1.4 | + |
| BamA-G(-12)* | 58 ± 8 | n.d. | n.d. |
| BamA-Q(-9)* | 13.5 ± 2.7 | n.d. | n.d. |
| BamA-F(-6)* | 6.8 ± 1 | n.d. | n.d. |
| BamA-M(-3)* | 2.4 ± 0.3 | n.d. | n.d. |
| BpumA-WT | 1.9 ± 0.1 | 105 ± 5.3 | + |
| CurA-WT | 22 ± 4 | 108 ± 4 | n.d. ^e |
| CurA-N(-26) ^g | 24 ± 4 | 100 ± 9 | n.d. |
| CurA-P(-23)A | >150 | 7.1 ± 0.7 | n.d. |
| CurA-P(-22)T | 53 ± 9 | 105 ± 1 | n.d. |
| CurA-V(-20)A | >150 | 44.4 ± 0.3 | n.d. |

^aCurA-N(-26)S represents leader sequence of BlinA. ^bNormalized to wild-type BamA or CurA, respectively. ^cDefined relative to that of the plasmid-borne wild-type BamA: 100–75% (+++), 75–25% (++), 25–5% (+). ^dLimit of detection estimated to be 5% based on requiring that the signal be 3 times greater than baseline noise. ^en.d., not determined.

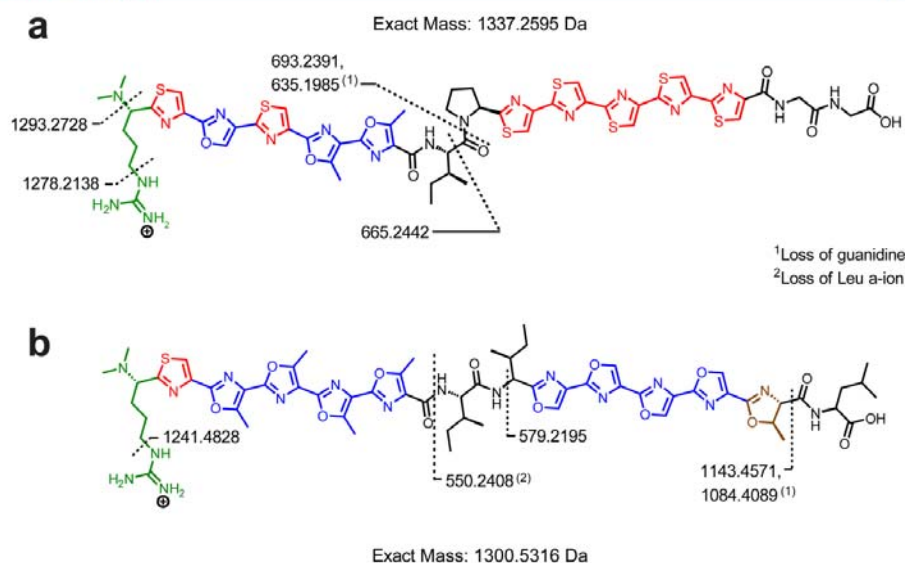


Figure 6. Proposed structures for the novel natural products coryneazolicin (CZN; a) and badiazolicin (BZN; b), as supported by FT-ICR-MS/MS (see also Supporting Figures 45 and 48).

(Supporting Information Figure 1c). Furthermore, these results indicate that the conserved residues at the C-terminus of the leader region (FEPXAA), which are likely part of a conserved protease recognition motif, contribute only modestly to the overall peptide affinity.

Select CurA peptides with substitutions at the equivalent positions in the leader peptide were also assayed for binding and processing by the Cur synthetase *in vitro* (Table 2). In line with the conserved nature of the position (Supporting Information Figure 1c), Pro(-23) in CurA was crucial for binding, but not the neighboring nonconserved Pro(-22). Unsurprisingly, the binding of CurA-N(-26)S, whose leader peptide is identical to that of BlinA,¹² was virtually no different from that of wild-type CurA. Taken together with the leader peptide binding experiments using the Bam synthetase, these results identify the critical residues that mediate recognition by the PZN heterocycle synthetases.

Beyond being necessary for leader peptide binding, we next sought to determine if these key residues were sufficient for binding by assaying a variety of precursor peptides from the PZN family with noncognate synthetases. In the FP assay, the Bam synthetase exhibited weakened binding to CmsA compared to BamA or BpumA and none to BlinA, while CurC (sufficient for precursor peptide binding) exhibited binding only to the highly similar BlinA and CmsC only to CmsA (Supporting Information Table 2). These observations were borne out by a MALDI-TOF-MS processing assay, which demonstrated that the Bam synthetase was capable of installing ten heterocycles on either BamA or BpumA, but only six on the less-similar CmsA, and none on either BlinA or CurA (Supporting Information Figure 40). Likewise, the Cur synthetase installed heterocycles only on CurA or BlinA, and not on BamA, BpumA, or CmsA (Supporting Information Figure 40). As the C-protein provides the main contact for

binding to the leader peptide, it was feasible that substituting the C-protein in these synthetase complexes could enable binding, and thus processing, of noncognate precursor peptides. However, these "hybrid synthetases" were not active, likely due in part to the inability to form productive complexes between a C-protein and its noncognate partners (Supporting Information Figure 41). Additional attempts to confer binding of noncognate substrates through simply swapping the RRE domains of these C-proteins were unsuccessful (data not shown).

These noncognate experiments demonstrated that relatively few conserved leader peptide residues [i.e., Pro(-21) in BamA] are necessary but not sufficient for modification. These results also suggest that synthetase recognition is encoded across the length of the peptide, even though affinity is primarily localized to a few "hot spots."⁵⁰ Even though most single substitutions had a minor effect, multiple residue differences in the leader peptide (e.g., between BamA and CurA) clearly prevented noncognate processing, as the Bam synthetase was able to fully process noncognate core peptides when its cognate leader was present (Figure 4e). Interestingly, CmsA does not contain the important Leu(-18) relative to BamA but can still be partially processed by the Bam synthetase, likely due to the overall greater similarity between leader peptides for CmsA and BamA, compared to BamA and CurA (Supporting Information Figure 40). Overall, the ability of these various synthetases to process noncognate substrates appears to correlate with evolutionary distance.

Production of Novel PZN Analogs. Having investigated the substrate tolerance of the Bam synthetase, it was of interest whether these variant precursor peptides could be used to generate mature PZN analogs. Following heterocyclization, the next biosynthetic step is removal of the leader peptide, which is putatively performed by the protease encoded in the BGC (*bamE*, Figure 1b).^{15,16} As a predicted type II CaaX protease,

BamE is thought to be an integral membrane protein;⁵¹ thus, we treated BamA that had been heterocyclized *in vitro* with a crude preparation of *B. velezensis* membranes. Unfortunately, no proteolytic product was observed (data not shown). As an alternative approach to removing the *N*-terminal leader for the *in vitro* production of PZN analogs, we introduced a cleavage site for a soluble, commercial reagent, as has been previously successful for the production of other RiPPs.^{52–55} Using site-directed mutagenesis, Ala(-1) was changed to one of several other residues: Lys (trypsin), Phe (chymotrypsin), Glu (endoproteinase GluC), and Met (cyanogen bromide, CNBr). A(-1)M was the best-tolerated substitution in combination with enzymatic heterocyclization, as determined by MALDI-TOF-MS (Supporting Information Figure 42a). Subsequent treatment of the heterocycle-containing peptide with CNBr in the presence of formic acid^{56,57} yielded a species whose mass was consistent with that of hydrolyzed desmethylPZN, indicating cleavage of the leader peptide (Supporting Information Figure 42b). Unfortunately, the acidic conditions necessary for the CNBr reaction also hydrolyzed the methyloxazoline normally present in PZN (Figure 1a).

The A(-1)M substitution was introduced into select BamA variants, in addition to CurA, to enable chemical removal of the leader peptide. Following CNBr treatment, these peptides were then subjected to *N*-terminal dimethylation by the SAM-dependent methyltransferase involved in PZN biosynthesis (BamL), whose activity had previously been reconstituted *in vitro*.^{16–18} BamL dimethylated each of these PZN variants, including the CurA variant, as determined by MALDI-TOF-MS (Supporting Information Figure 43). Though the methyltransferase has significant substrate selectivity owing to its narrow substrate-binding channel,¹⁷ all of the PZN variants used in these reactions contained both the *N*-terminal Arg and at least one azole, which is minimally necessary for robust methyltransferase activity. Although enzymatic dimethylation was successful, more robust modification was achieved by reductive alkylation with formaldehyde and borane-pyridine (Supporting Information Figure 44).⁵⁸ Regardless of the method of dimethylation, this sequence of reactions enabled the production of a new collection of PZN variants beyond what was previously achievable *in vivo*.⁴¹

Through the system of treating CurA-A(-1)M with the Cur synthetase, CNBr-based removal of the leader peptide, and reductive alkylation of the *N*-terminus, we obtained sufficient quantities of the PZN-like natural product from *C. urealyticum* to enable FT-ICR-MS/MS analysis of the purified compound, which corroborated the proposed structure containing 10 azoles and *N*-terminal dimethylation (Figure 6a, Supporting Information Figure 45). ¹H NMR also established the presence of sharp singlets consistent with azole formation (Supporting Information Figure 46). Similar to the name “plantazolicin,” we have named this anticipated natural product “coryneazolicin” (CZN). Our production of CZN demonstrates the use of *in vitro* biosynthesis to generate a predicted natural compound whose production has not been observed from the native producer. To establish if CZN is naturally produced will require further investigation.

After obtaining CZN, we next turned our attention to recently identified PZN-like BGCs from other bacteria (Supporting Information Figure 1). Several of these BGCs encode precursor peptides with substitutions in the core region compared to BamA, the products of which could aid in further determination of the structure–activity relationships of the

PZN class of natural products. Although the unannotated precursor peptide from *Bacillusadius*, BadaA, lies in a different location relative to other genes in the cluster (Figure 1b), it only differs from BamA by two residues in the core region (Figure 1c). Methanolic extracts of *B.adius* contained an *m/z* 1300 species consistent with the expected mass of the precursor peptide with PZN-like modifications. Furthermore, a less intense species was found with *m/z* 1318, consistent with a PZN-like compound where the methyloxazoline moiety was hydrolyzed (Supporting Information Figure 47).³ Purification of “badiazolicin” (BZN) enabled FT-ICR-MS/MS analysis, which supported a PZN-like structure containing nine azoles, one azoline, and *N*-terminal dimethylation (Figure 6b, Supporting Information Figure 48). After quantification by ¹H NMR, BZN exhibited a minimum inhibitory concentration (MIC) of 2 μg/mL against *B. anthracis* by microbroth dilution assay, but no growth inhibitory activity toward *B. subtilis* (MIC > 16 μg/mL) or methicillin-resistant *S. aureus* (MIC > 32 μg/mL).

In conclusion, we have demonstrated the *in vitro* reconstitution of two heterocycle synthetases involved in the biosynthesis of PZN by *B. velezensis* and CZN by *C. urealyticum*. Overall, the synthetases *in vitro* afforded a greater degree of biosynthetic insight than previous work with PZN biosynthesis in cells, enabling characterization of synthetase complex assembly and substrate binding properties. Using the Bam heterocycle synthetase, we have explored substrate tolerance in the PZN precursor peptide and proposed that cyclodehydration occurs in an *N*-terminal to *C*-terminal direction. These synthetases were then used to achieve the total *in vitro* biosynthesis of a number of PZN variants, culminating in the isolation of two novel natural products, CZN and BZN.

■ METHODS

***In Vitro* Synthetase Assays.** Heterocycle synthetase reactions to assess substrate tolerance or necessity of individual components contained 100 μM MBP-tagged precursor peptide, 10 μM of each of the pertinent MBP-tagged synthetase enzymes, and 40 μg/mL TEV protease in synthetase buffer [50 mM Tris (pH 7.5), 125 mM NaCl, 20 mM MgCl₂, 10 mM dithiothreitol, and 3 mM ATP].⁹ Reactions were carried out in a 100 μL volume for 18 h at 22 °C, at which point they were quenched by the addition of 100 μL of MeCN to precipitate the large proteins. After centrifugation for 5 min at 16 000g, the supernatant was either dried by speedvac or diluted to 12.5% MeCN by the addition of water to facilitate desalting before analysis by mass spectrometry.

For synthetase reactions where isolation of intermediates was attempted, synthetase enzymes were pretreated with TEV to remove MBP and present at 2 μM in the reaction. Other components were present at the concentrations listed above, and the reaction was quenched by the addition of 100 μM MeCN at 10 and 30 min.

Mass Spectrometry. For MALDI-MS, peptide samples were desalted by C18 ZipTip according to manufacturer instructions and analyzed using a Bruker Daltonics UltrafleXtreme MALDI-TOF/TOF instrument operating in reflector/positive mode with α -cyano-4-hydroxycinnamic acid as the matrix.

For high-resolution mass spectrometry and MS/MS, desalted peptides were resuspended in 80% (v/v) MeCN and 1% AcOH and centrifuged at 11 000g for 5 min prior to analysis by direct infusion Fourier transform mass spectrometry (FT-MS). An Advion NanoMate 100 was used to directly infuse samples to an LTQ-FTMS/MS (ThermoFisher) operating at 11 T. The MS was calibrated weekly, following the manufacturer's instructions, and tuned daily using Pierce LTQ Velos ESI Positive Ion Calibration Solution (ThermoFisher). Spectra were collected with a resolution of 100 000. Ions were selected for ion trap fragmentation or FT-MS/MS fragmentation based on

signal intensity, and spectra were collected using the following parameters: isolation width of 5 m/z , normalized collision energy of 35, activation q value of 0.4, and activation time of 30 ms. Data analysis was performed using the Qualbrowser application within Thermo-Fisher Xcalibur v 2.2.

In Vitro Kinetics Assay. Peptide processing kinetics were measured using a previously described purine nucleoside phosphorylase (PNP)-coupled assay to detect ADP production from ATP hydrolysis.^{9,59} Because ATP consumption is 1:1 with heterocycle formation, this assay provides a measure of the rate of heterocycle formation. In general MBP-BamB, MBP-BpumC, and MBP-BamD were pretreated with TEV protease at RT to cleave the tag, but for the Cur synthetase proteins, TEV treatment was performed at 4 °C because extended incubation at RT caused protein precipitation. After adding the enzyme cleavage mixture to a cuvette for a final concentration of 5 μ M BCD, enzymatic reactions were initiated by the addition of a peptide substrate mixed with synthetase buffer, 200 μ M 7-methyl-6-mercapto-7-methylpurineriboside (Berry and Associates), and 0.2 U of PNP at 22 °C. Maximal initial rates were determined by averaging the slope over the first 2 min of the reaction. The rate of processing for leader peptide mutants was determined with 30 μ M BamA/BpumA or 25 μ M for CurA and their respective synthetase complexes, whereas the BamA-R1X mutants were tested at 100 μ M. Kinetic parameters for the Bam/Bpum complex were determined using variable substrate concentrations and were calculated using the nonlinear Michaelis–Menten fit with OriginPro9.1 (OriginLab). All experiments were performed in triplicate.

Fluorescence Polarization Binding Assay. The interaction between FITC-conjugated precursor peptides and different synthetase proteins was measured with a previously employed fluorescence polarization (FP) assay.¹⁵ In brief, the assay was performed in nonbinding-surface, 384-black-well polystyrene microplates (Corning) with serial dilutions of the indicated protein(s) and 25 nM of the indicated fluorescent peptide in a storage buffer. After 30 min of incubation at RT, the dilutions were measured using a FilterMax F5 multimode microplate reader (Molecular Devices) with λ_{ex} = 485 nm and λ_{em} = 538 nm. All assays were performed in triplicate, and K_d values were determined from a nonlinear dose–response curve in OriginPro9.1 (OriginLab). Error bars represent standard error of the mean.

To assess the interaction of nonfluorescent peptides with the PZN synthetase, a competition FP binding assay was used as previously described.⁴⁹ MBP-tagged precursor peptides were serially diluted and mixed with 25 nM of a fluorescent peptide and synthetase component(s). The BamA/BpumA peptides were competed with FITC-BamA-LP to bind to 3 μ M BamB-R93A, BpumC, and BamD, whereas the CurA mutants were competed with FITC-CurA-LP to bind 25 μ M CurC. IC₅₀ values were determined from the 50% inhibition point calculated using a dose response curve with Origin Pro 9.1 (OriginLab).

Leader Peptide Cleavage. Dried precursor peptide (after synthetase reaction and MeCN precipitation) was resuspended in 60% formic acid and 1–2 small crystals of cyanogen bromide (CNBr) were added. The reaction was allowed to proceed for 4 h at 22 °C, at which point it was quenched by the addition of 5 volumes of water and desalted by ZipTip, then dried by speedvac.

Methylation. Enzymatic dimethylation using BamL was performed using <50 μ M peptide (exact concentration not known), 10 μ M MBP-BamL, 10 μ M Pfs, and 3 mM S-adenosylmethionine in 50 mM Tris (pH 8), for 18 h at 37 °C, as described previously.¹⁷ Chemical dimethylation was performed using <1 mM peptide (exact concentration not known), 200 mM formaldehyde, and 300 mM borane-pyridine in 10 mM NH₄CO₃ with 50% MeOH, for 18 h at 22 °C.⁵⁸

Production of CZN. Synthetase reactions using MBP-tagged CurA, CurB, CurC, and CurD were performed as described above, with the exceptions that 250 μ M MBP-CurA was used, the buffer contained 6 mM ATP, and reactions were carried out in replicates of 200 μ L each. After precipitation with MeCN and centrifugation, the supernatants were dried by speedvac, then resuspended and pooled in

60% formic acid (v/v) for CNBr digestion as described above. After the CNBr reaction was quenched with water, the solvent was removed by blowing air. The dried reaction mixture was then resuspended in water to remove water-soluble components and centrifuged for 10 min at 16 000g, after which the supernatant was removed and the remaining solids resuspended in MeOH to extract CZN. After centrifugation for 10 min at 16 000g, the MeOH supernatant (containing crude CZN) was removed and used in the chemical methylation reaction as described above.

This crude CZN solution in 50% MeOH (v/v) was reverse phase purified using a Thermo BETASIL C18 column (250 mm \times 10 mm; pore size, 100 Å; particle size, 5 μ m) at a flow rate of 4 mL min⁻¹. A gradient of 40–95% MeOH with 10 mM NH₄HCO₃ in the aqueous phase over 41 min was used. The fractions containing CZN (as monitored by A₂₅₄ and later verified by MALDI-MS) were collected into 20 mL borosilicate vials, and the solvent was removed *in vacuo*.

Production of BZN. *Bacillus badius* ATCC14574 was grown in Luria–Bertani (LB) broth (10 mL/18 mm glass culture tube) at 37 °C overnight. Sterilized aluminum trays (16–7/16" \times 11–5/8" cake pans) containing nutrient agar (1 L/tray) were inoculated with 3 mL of overnight culture and incubated for 48 h at 37 °C. Cells were harvested with a razor blade and Tris-buffered saline (TBS; 160 mL/tray) and pelleted by centrifugation (4000g, 20 min, 10 °C). The supernatant was decanted, and the pellets were stored at –20 °C until extraction.

Crude BZN was extracted by resuspending the pellets in MeOH (160 mL/tray) through vortex agitation, and the resuspended cells were equilibrated for 4 h at 22 °C on a shaking platform. The supernatant was retained after centrifugation (4000g, 20 min, 10 °C), vacuum filtered with Whatman filter paper, and concentrated by rotary evaporation. Lyophilization of the extract yielded a yellow solid (200 mg/tray). The crude material was dissolved in 50% MeCN, where the sample separated into two layers. Both layers were added to Celite, which was dried by rotary evaporation. The dried Celite was packed into a cartridge for solid loading of BZN onto a RediSep RP 130g C18 column (Teledyne Isco) for purification by MPLC using a Combiflash system (30–95% MeCN/10 mM aqueous NH₄HCO₃ over 16 column volumes). The fractions containing BZN were pooled, concentrated by rotary evaporation, and lyophilized to dryness. BZN was then reverse phase purified using the same method as for CZN, except that a gradient of 85–95% MeOH with 10 mM aqueous NH₄HCO₃ over 20 min was used. The fractions containing BZN were collected in 20 mL borosilicate vials and the solvent removed *in vacuo*. The isolated yield of HPLC-purified BZN was 0.7–0.8 μ g/L of nutrient agar.

NMR. Samples were dissolved in 600 μ L of CD₃OD (99.95 atom % D, Cambridge Isotope Laboratories). NMR spectra were obtained with an Agilent VNMRS 750 MHz narrow bore magnet spectrometer equipped with a 5 mm triple resonance (¹H–¹³C–¹⁵N) triaxial gradient probe and pulse-shaping capabilities. Samples were held at 298 K during acquisition. Standard Varian pulse sequences were used; a relaxation delay (d1) of 28 s was used during acquisition for quantitative NMR (qNMR), and 90° pulse widths were calibrated and used (8.50 μ s for CZN acquisition; 8.75 μ s for BZN acquisition). A total of 320 transients were recorded for BZN; 2048 transients were recorded for CZN. Apodization (0.4 Hz line-broadening), phase correction, integration, integral normalization, and automated baseline correction were applied prior to qNMR processing. Spectra were recorded with VNMRJ 4.2 software, and data were processed using MestReNova 8.1.1. Chemical shifts (δ , ppm) were referenced internally to the solvent peak (methanol). For quantitative NMR, the probe was calibrated using the qEstimate tool VNMRJ on the standard ³¹P sample (48.5 mM triphenylphosphate in CDCl₃, Varian part #: 00-968120-97).

Antibacterial Activity Assays. The concentration of BZN was calculated by qNMR, which has been used for accurate quantification of microgram quantities of natural products.⁶⁰ The native qNMR functionality within VNMRJ 4.2 was used to quantify BZN based on the average integration of its azole peaks, which occur as distinct singlets in the aromatic region. Determination of MIC values for BZN was performed as described previously.¹⁹

■ ASSOCIATED CONTENT

■ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscchembio.6b00369.

Supporting methods, figures, and tables (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Phone: 1-217-333-1345. Fax: 1-217-333-0508. E-mail: douglasm@illinois.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank A. Maniak for assistance with cloning and L. Zhu (NMR Laboratory, University of Illinois) for assistance with NMR. This work was supported in part by the U.S. National Institutes of Health (NIH, 1R01 GM097142 to D.A.M.), the NIH Director's New Innovator Award Program (DP2 OD008463 to D.A.M.), the Robert C. and Carolyn J. Springborn Endowment (to C.D.D., B.J.B., and J.L.T.), the James R. Beck Graduate Fellowship from the Department of Chemistry at the University of Illinois (to C.D.D.), a National Science Foundation Graduate Research Fellowship (DGE-1144245 to B.J.B.), and ACS Division of Medicinal Chemistry Predoctoral Fellowships (to P.M.B. and J.L.T.). The Bruker UltrafleXtreme MALDI TOF/TOF mass spectrometer was purchased in part with a grant from the National Center for Research Resources, NIH (S10 RR027109 A).

■ REFERENCES

- (1) Newman, D. J., and Cragg, G. M. (2016) Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* 79, 629–661.
- (2) Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., Camarero, J. A., Campopiano, D. J., Challis, G. L., Clardy, J., Cotter, P. D., Craik, D. J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P. C., Entian, K. D., Fischbach, M. A., Garavelli, J. S., Goransson, U., Gruber, C. W., Haft, D. H., Hemscheidt, T. K., Hertweck, C., Hill, C., Horswill, A. R., Jaspars, M., Kelly, W. L., Klinman, J. P., Kuipers, O. P., Link, A. J., Liu, W., Marahiel, M. A., Mitchell, D. A., Moll, G. N., Moore, B. S., Muller, R., Nair, S. K., Nes, I. F., Norris, G. E., Olivera, B. M., Onaka, H., Patchett, M. L., Piel, J., Reaney, M. J., Rebuffat, S., Ross, R. P., Sahl, H. G., Schmidt, E. W., Selsted, M. E., Severinov, K., Shen, B., Sivonen, K., Smith, L., Stein, T., Sussmuth, R. D., Tagg, J. R., Tang, G. L., Truman, A. W., Vederas, J. C., Walsh, C. T., Walton, J. D., Wenzel, S. C., Willey, J. M., and van der Donk, W. A. (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* 30, 108–160.
- (3) Scholz, R., Molohon, K. J., Nachtigall, J., Vater, J., Markley, A. L., Sussmuth, R. D., Mitchell, D. A., and Borriss, R. (2011) Plantazolicin, a novel microcin B17/streptolysin S-like natural product from *Bacillus amyloliquefaciens* FZB42. *J. Bacteriol.* 193, 215–224.
- (4) Molohon, K. J., Melby, J. O., Lee, J., Evans, B. S., Dunbar, K. L., Bumpus, S. B., Kelleher, N. L., and Mitchell, D. A. (2011) Structure determination and interception of biosynthetic intermediates for the plantazolicin class of highly discriminating antibiotics. *ACS Chem. Biol.* 6, 1307–1313.
- (5) Kalyon, B., Helaly, S. E., Scholz, R., Nachtigall, J., Vater, J., Borriss, R., and Sussmuth, R. D. (2011) Plantazolicin A and B: structure elucidation of ribosomally synthesized thiazole/oxazole peptides from *Bacillus amyloliquefaciens* FZB42. *Org. Lett.* 13, 2996–2999.
- (6) Dunlap, C. A., Kim, S. J., Kwon, S. W., and Rooney, A. P. (2015) *Bacillus velezensis* is not a later heterotypic synonym of *Bacillus amyloliquefaciens*; *Bacillus methylotrophicus*, *Bacillus amyloliquefaciens* subsp. *plantarum* and '*Bacillus oryzicola*' are later heterotypic synonyms of *Bacillus velezensis* based on phylogenomics. *Int. J. Syst. Evol. Microbiol.* 66, 1212.
- (7) Lee, S. W., Mitchell, D. A., Markley, A. L., Hensler, M. E., Gonzalez, D., Wohlrab, A., Dorrestein, P. C., Nizet, V., and Dixon, J. E. (2008) Discovery of a widely distributed toxin biosynthetic gene cluster. *Proc. Natl. Acad. Sci. U. S. A.* 105, 5879–5884.
- (8) Cox, C. L., Doroghazi, J. R., and Mitchell, D. A. (2015) The genomic landscape of ribosomal peptides containing thiazole and oxazole heterocycles. *BMC Genomics* 16, 778.
- (9) Dunbar, K. L., Melby, J. O., and Mitchell, D. A. (2012) YcaO domains use ATP to activate amide backbones during peptide cyclodehydrations. *Nat. Chem. Biol.* 8, 569–575.
- (10) Dunbar, K. L., Chekan, J. R., Cox, C. L., Burkhart, B. J., Nair, S. K., and Mitchell, D. A. (2014) Discovery of a new ATP-binding motif involved in peptidic azoline biosynthesis. *Nat. Chem. Biol.* 10, 823–829.
- (11) Melby, J. O., Li, X., and Mitchell, D. A. (2014) Orchestration of enzymatic processing by thiazole/oxazole-modified microcin dehydrogenases. *Biochemistry* 53, 413–422.
- (12) Burkhart, B. J., Hudson, G. A., Dunbar, K. L., and Mitchell, D. A. (2015) A prevalent peptide-binding domain guides ribosomal natural product biosynthesis. *Nat. Chem. Biol.* 11, 564–570.
- (13) Koehnke, J., Mann, G., Bent, A. F., Ludewig, H., Shirran, S., Botting, C., Lebl, T., Houssen, W. E., Jaspars, M., and Naismith, J. H. (2015) Structural analysis of leader peptide binding enables leader-free cyanobactin processing. *Nat. Chem. Biol.* 11, 558–563.
- (14) Oman, T. J., and van der Donk, W. A. (2010) Follow the leader: the use of leader peptides to guide natural product biosynthesis. *Nat. Chem. Biol.* 6, 9–18.
- (15) Maxson, T., Deane, C. D., Molloy, E. M., Cox, C. L., Markley, A. L., Lee, S. W., and Mitchell, D. A. (2015) HIV protease inhibitors block streptolysin S production. *ACS Chem. Biol.* 10, 1217–1226.
- (16) Pei, J., and Grishin, N. V. (2001) Type II CAAX prenyl endopeptidases belong to a novel superfamily of putative membrane-bound metalloproteases. *Trends Biochem. Sci.* 26, 275–277.
- (17) Lee, J., Hao, Y., Blair, P. M., Melby, J. O., Agarwal, V., Burkhart, B. J., Nair, S. K., and Mitchell, D. A. (2013) Structural and functional insight into an unexpectedly selective N-methyltransferase involved in plantazolicin biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* 110, 12954–12959.
- (18) Piwowarska, N. A., Banala, S., Overkleeft, H. S., and Sussmuth, R. D. (2013) Arg-Thz is a minimal substrate for the N(alpha),N(alpha)-arginyl methyltransferase involved in the biosynthesis of plantazolicin. *Chem. Commun.* 49, 10703–10705.
- (19) Molohon, K. J., Blair, P. M., Park, S., Doroghazi, J. R., Maxson, T., Hershfield, J. R., Flatt, K. M., Schroeder, N. E., Ha, T., and Mitchell, D. A. (2016) Plantazolicin is an ultranarrow-spectrum antibiotic that targets the *Bacillus anthracis* membrane. *ACS Infect. Dis.* 2, 207–220.
- (20) Appleyard, A. N., Choi, S., Read, D. M., Lightfoot, A., Boakes, S., Hoffmann, A., Chopra, I., Bierbaum, G., Rudd, B. A., Dawson, M. J., and Cortes, J. (2009) Dissecting structural and functional diversity of the lantibiotic mersacidin. *Chem. Biol.* 16, 490–498.
- (21) Caetano, T., Krawczyk, J. M., Mosker, E., Sussmuth, R. D., and Mendo, S. (2011) Heterologous expression, biosynthesis, and mutagenesis of type II lantibiotics from *Bacillus licheniformis* in *Escherichia coli*. *Chem. Biol.* 18, 90–100.
- (22) Molloy, E. M., Ross, R. P., and Hill, C. (2012) 'Bac' to the future: bioengineering lantibiotics for designer purposes. *Biochem. Soc. Trans.* 40, 1492–1497.
- (23) Boakes, S., Ayala, T., Herman, M., Appleyard, A. N., Dawson, M. J., and Cortes, J. (2012) Generation of an actagardine A variant library through saturation mutagenesis. *Appl. Microbiol. Biotechnol.* 95, 1509–1517.

- (24) van Heel, A. J., Mu, D., Montalban-Lopez, M., Hendriks, D., and Kuipers, O. P. (2013) Designing and producing modified, new-to-nature peptides with antimicrobial activity by use of a combination of various lantibiotic modification enzymes. *ACS Synth. Biol.* 2, 397–404.
- (25) Field, D., Molloy, E. M., Iancu, C., Draper, L. A., O' Connor, P. M., Cotter, P. D., Hill, C., and Ross, R. P. (2013) Saturation mutagenesis of selected residues of the alpha-peptide of the lantibiotic lactacin 3147 yields a derivative with enhanced antimicrobial activity. *Microb. Biotechnol.* 6, 564–575.
- (26) Molloy, E. M., Field, D., O'Connor, P. M., Cotter, P. D., Hill, C., and Ross, R. P. (2013) Saturation mutagenesis of lysine 12 leads to the identification of derivatives of nisin A with enhanced antimicrobial activity. *PLoS One* 8, e58530.
- (27) Pau, S. J., and Link, A. J. (2011) Sequence diversity in the lasso peptide framework: discovery of functional microcin J25 variants with multiple amino acid substitutions. *J. Am. Chem. Soc.* 133, 5016–5023.
- (28) Piscotta, F. J., Sharp, J. M., Liu, W. R., and Link, A. J. (2015) Expanding the chemical diversity of lasso peptide MccJ25 with genetically encoded noncanonical amino acids. *Chem. Commun.* 51, 409–412.
- (29) Maksimov, M. O., Koos, J. D., Zong, C., Lisko, B., and Link, A. J. (2015) Elucidating the specificity determinants of the AtxE2 lasso peptide isopeptidase. *J. Biol. Chem.* 290, 30806–30812.
- (30) Bowers, A. A., Acker, M. G., Koglin, A., and Walsh, C. T. (2010) Manipulation of thioicillin variants by prepeptide gene replacement: structure, conformation, and activity of heterocycle substitution mutants. *J. Am. Chem. Soc.* 132, 7519–7527.
- (31) Li, C., Zhang, F., and Kelly, W. L. (2011) Heterologous production of thiostrepton A and biosynthetic engineering of thiostrepton analogs. *Mol. Biosyst.* 7, 82–90.
- (32) Li, C., Zhang, F., and Kelly, W. L. (2012) Mutagenesis of the thiostrepton precursor peptide at Thr7 impacts both biosynthesis and function. *Chem. Commun.* 48, 558–560.
- (33) Bowers, A. A., Acker, M. G., Young, T. S., and Walsh, C. T. (2012) Generation of thioicillin ring size variants by prepeptide gene replacement and in vivo processing by *Bacillus cereus*. *J. Am. Chem. Soc.* 134, 10313–10316.
- (34) Zhang, F., and Kelly, W. L. (2015) Saturation mutagenesis of TsrA Ala4 unveils a highly mutable residue of thiostrepton A. *ACS Chem. Biol.* 10, 998–1009.
- (35) Zhang, F., Li, C., and Kelly, W. L. (2016) Thiostrepton variants containing a contracted quinaldic acid macrocycle result from mutagenesis of the second residue. *ACS Chem. Biol.* 11, 415–424.
- (36) Tianero, M. D., Donia, M. S., Young, T. S., Schultz, P. G., and Schmidt, E. W. (2012) Ribosomal route to small-molecule diversity. *J. Am. Chem. Soc.* 134, 418–425.
- (37) Goto, Y., Ito, Y., Kato, Y., Tsunoda, S., and Suga, H. (2014) One-pot synthesis of azoline-containing peptides in a cell-free translation system integrated with a posttranslational cyclodehydratase. *Chem. Biol.* 21, 766–774.
- (38) Houssem, W. E., Bent, A. F., McEwan, A. R., Pieiller, N., Tabudravu, J., Koehnke, J., Mann, G., Adaba, R. I., Thomas, L., Hawas, U. W., Liu, H., Schwarz-Linek, U., Smith, M. C., Naismith, J. H., and Jaspars, M. (2014) An efficient method for the *in vitro* production of azol(in)e-based cyclic peptides. *Angew. Chem., Int. Ed.* 53, 14171–14174.
- (39) Ruffner, D. E., Schmidt, E. W., and Heemstra, J. R. (2014) Assessing the combinatorial potential of the RiPP cyanobactin tru pathway. *ACS Synth. Biol.* 4, 482–492.
- (40) Sardar, D., Lin, Z., and Schmidt, E. W. (2015) Modularity of RiPP enzymes enables designed synthesis of decorated peptides. *Chem. Biol.* 22, 907–916.
- (41) Deane, C. D., Melby, J. O., Molohon, K. J., Susarrey, A. R., and Mitchell, D. A. (2013) Engineering unnatural variants of plantazolicin through codon reprogramming. *ACS Chem. Biol.* 8, 1998–2008.
- (42) Gonzalez, D. J., Lee, S. W., Hensler, M. E., Markley, A. L., Dahesh, S., Mitchell, D. A., Bandeira, N., Nizet, V., Dixon, J. E., and Dorrestein, P. C. (2010) Clostridiolysin S, a post-translationally modified biotoxin from *Clostridium botulinum*. *J. Biol. Chem.* 285, 28220–28228.
- (43) Melby, J. O., Dunbar, K. L., Trinh, N. Q., and Mitchell, D. A. (2012) Selectivity, directionality, and promiscuity in peptide processing from a *Bacillus* sp. Al Hakam cyclodehydratase. *J. Am. Chem. Soc.* 134, 5309–5316.
- (44) Milne, J. C., Roy, R. S., Eliot, A. C., Kelleher, N. L., Wokhlu, A., Nickels, B., and Walsh, C. T. (1999) Cofactor requirements and reconstitution of microcin B17 synthetase: a multienzyme complex that catalyzes the formation of oxazoles and thiazoles in the antibiotic microcin B17. *Biochemistry* 38, 4768–4781.
- (45) Dunbar, K. L., and Mitchell, D. A. (2013) Insights into the mechanism of peptide cyclodehydrations achieved through the chemoenzymatic generation of amide derivatives. *J. Am. Chem. Soc.* 135, 8692–8701.
- (46) Mitchell, D. A., Lee, S. W., Pence, M. A., Markley, A. L., Limm, J. D., Nizet, V., and Dixon, J. E. (2009) Structural and functional dissection of the heterocyclic peptide cytotoxin streptolysin S. *J. Biol. Chem.* 284, 13004–13012.
- (47) Donia, M. S., Ravel, J., and Schmidt, E. W. (2008) A global assembly line for cyanobactins. *Nat. Chem. Biol.* 4, 341–343.
- (48) Morris, R. P., Leeds, J. A., Naegeli, H. U., Oberer, L., Memmert, K., Weber, E., LaMarche, M. J., Parker, C. N., Burrer, N., Esterov, S., Hein, A. E., Schmitt, E. K., and Krastel, P. (2009) Ribosomally synthesized thiopeptide antibiotics targeting elongation factor Tu. *J. Am. Chem. Soc.* 131, 5946–5955.
- (49) Dunbar, K. L., Tietz, J. L., Cox, C. L., Burkhart, B. J., and Mitchell, D. A. (2015) Identification of an auxiliary leader peptide-binding protein required for azoline formation in ribosomal natural products. *J. Am. Chem. Soc.* 137, 7672–7677.
- (50) Clackson, T., and Wells, J. A. (1995) A hot spot of binding energy in a hormone-receptor interface. *Science* 267, 383–386.
- (51) Pryor, E. E., Jr., Horanyi, P. S., Clark, K. M., Fedoriv, N., Connelly, S. M., Koszelak-Rosenblum, M., Zhu, G., Malkowski, M. G., Wiener, M. C., and Dumont, M. E. (2013) Structure of the integral membrane protein CAAX protease Ste24p. *Science* 339, 1600–1604.
- (52) Goto, Y., Li, B., Claesen, J., Shi, Y., Bibb, M. J., and van der Donk, W. A. (2010) Discovery of unique lanthionine synthetases reveals new mechanistic and evolutionary insights. *PLoS Biol.* 8, e1000339.
- (53) Shi, Y., Yang, X., Garg, N., and van der Donk, W. A. (2011) Production of lantipeptides in *Escherichia coli*. *J. Am. Chem. Soc.* 133, 2338–2341.
- (54) Plat, A., Kluskens, L. D., Kuipers, A., Rink, R., and Moll, G. N. (2011) Requirements of the engineered leader peptide of nisin for inducing modification, export, and cleavage. *Appl. Environ. Microbiol.* 77, 604–611.
- (55) Okesli, A., Cooper, L. E., Fogle, E. J., and van der Donk, W. A. (2011) Nine post-translational modifications during the biosynthesis of cinnamycin. *J. Am. Chem. Soc.* 133, 13753–13760.
- (56) Schreiber, J., and Witkop, B. (1964) Reaction of cyanogen bromide with mono- + diamino acids. *J. Am. Chem. Soc.* 86, 2441–2445.
- (57) Slootweg, J. C., Liskamp, R. M., and Rijkers, D. T. (2013) Scalable purification of the lantibiotic nisin and isolation of chemical/enzymatic cleavage fragments suitable for semi-synthesis. *J. Pept. Sci.* 19, 692–699.
- (58) Krusemark, C. J., Frey, B. L., Smith, L. M., and Belshaw, P. J. (2011) Complete chemical modification of amine and acid functional groups of peptides and small proteins. *Methods Mol. Biol.* 753, 77–91.
- (59) Webb, M. R. (1992) A continuous spectrophotometric assay for inorganic phosphate and for measuring phosphate release kinetics in biological systems. *Proc. Natl. Acad. Sci. U. S. A.* 89, 4884–4887.
- (60) Kronic, A., and Orjala, J. (2015) Application of high-field NMR spectroscopy for characterization and quantitation of submilligram quantities of isolated natural products. *Magn. Reson. Chem.* 53, 1043–1050.