© 2013 Yanxiang (Nancy) Shi
EXPLORATION OF THE BIOSYNTHESIS OF LANTHIPEPTIDES

BY

YANXIANG (NANCY) SHI

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, 2013

Urbana, Illinois

Doctoral Committee:

Professor Wilfred A. van der Donk, Chair
Professor Jonathan V. Sweedler
Professor Douglas A. Mitchell
Professor David M. Kranz
Abstract

Natural products play important roles in the survival of organisms, and provide a large pool of candidates for drug discovery and pharmaceutical investigations. A major class of peptide-based natural products are ribosomally synthesized and post-translationally modified (RiPPs). Lanthipeptides, which possess (methyl)lanthionine structures, are a class of intensively studied RiPPs. Based on the biosynthesis enzymes that introduce the thioether motifs, lanthipeptides are classified into four classes. Class I lanthipeptides utilize two distinct enzymes, a dehydratase and a cyclase, while class II lanthipeptides adopt a bifunctional enzyme. Trifunctional enzymes with independent evolution pathways are employed by class III and class IV lanthipeptides. A major subclass of lanthipeptides possess antimicrobial activity, known as lantibiotics with known modes of action including lipid II binding, pore formation, inhibition of spore outgrowth, and phosphatidylethanolamine (PE) binding. Several lanthipeptides display no antimicrobial activities but have other functions important for bacterial life cycles, such as facilitating aerial hyphae formation during sporulation.

In the past, most lanthipeptides were discovered based on their activity mainly through antimicrobial screening campaigns. With the advances in genome sequencing and bioinformatics, a variety of lanthipeptides have been identified by genome mining, presumably with more unexplored functions and molecular targets. Prochlorosins are an unusual group of lanthipeptides, which were discovered through genome mining and have no known antibiotic activities. Further biological function investigations of prochlorosins require pure products in adequate amounts, which was not achievable
through natural product isolation or *in vitro* enzymatic assays. In this thesis, Chapter 2 describes the development of a novel co-expression system to produce prochlorosins by *in vivo* enzymatic modification in *E. coli*, providing prochlorosins with high purity and sufficient quantity for subsequent characterizations. Chapter 3 presents an example of utilizing the co-expression methodology to produce antimicrobial lanthipeptides, with additional modifications besides dehydration and cyclization. The precursor peptide of actagardine, one of the only two lantibiotics that have been evaluated in clinical trials thus far, was fully modified in *E. coli* as presented in Chapter 3. A luciferase-like monooxygenase located in the biosynthetic cluster of actagardine was also reconstituted, both *in vitro* and *in vivo*, to introduce a sulfoxide group onto actagardine and its derivatives. Based on the success of *in vivo* co-expression of lanthipeptides with identified or unidentified activities, Chapter 4 describes efforts towards the incorporation of non-canonical amino acids into lanthipeptides, in order to expand the structural diversity of these natural products, tune their activities, and introduce probes for subsequent functional investigations. The work presented in this thesis provides methods to effectively produce the peptide-based natural products lanthipeptides, in adequate quantities for subsequent investigations including study of substrate peptides and their modification enzymes, peptide structure elucidation, as well as sequence and structure diversification.
For My Mom
Acknowledgement

Placing the acknowledgement in front of all the main texts in the thesis is a bright idea. Without all the people that have supported and helped me, it will be impossible to have this thesis.

At first, I would like to thank my supervisor, Professor Wilfred A. van der Donk. I would like to thank him for accepting me as one of his students five years ago, thank him as a guider in the maze, thank his constructive instructions and generous support along the journey, his encouragement and care during my frustrated times, and all his suggestions and critiques. With his lead, I have learned and grown a lot during the five years. His rigorous scholarship and great breadth of knowledge will lead me for life.

Also, I would like to thank Prof. Jonathan V. Sweedler, Prof. David Kranz, and Prof. Douglas A. Mitchell for being my committee members. Every time talking and having discussion with them, I can always feel the original passion for science. Their enthusiasm and critical thinking will still influence me in the future.

Professors from other departments and institutes also offered tremendous support to my work during the five years. Prof. Sallie W. Chisholm (Massachusetts Institute of Technology) and her student Andres Cubillos-Ruiz have made it a very enjoyable collaboration on the prochlorosin project through the past five years. Prof. Peter Schultz (the Scripps Research Institute) generously provided their engineered plasmids, while Prof. Ryan Mehl (Oregon State University) and Prof. Wenshe Liu (Texas A&M University) not only shared their plasmid material but also their research source in the non-canonical amino acid incorporation project. Meanwhile, Prof. Andrew N. Miller
(UIUC), Prof. Angela D. Kent (UIUC), Prof. Bryan A. White (UIUC) and their students were kind enough to share their equipment, knowledge and resources with me for different projects and trials.

I also want to thank Furong Sun, Dr. Haijun Yao, Dr. Kevin Tucker and the whole SCS mass spectrometry facility for their regular support, thank Dr. Dean Olson and Tracie Hubert for their daily discussions, thank Dr. Alvaro Hernandez and the staff in Biotechnology Center for facilitating the sequencing works, and thank Nancy Holda and Martha Freeland for coordinating the daily affairs in the lab, also as a “lab mom” taking care of us. The chemical biology office and the SCS student office have provided considerate care since day one, and they deserve sincere appreciation.

Without financial support, no research work can be done. Both NIH and Howard Hughes Medical Institute have provided generous funding for my daily work. I would also like to thank Martha Hetzner Nilsen and Dr. Peter V. Hetzner for providing financial support through the Marvel Fellowship, the R. C. Fuson Travel Award and Grad College Conference Travel Award for providing me the opportunity to present my work in the national conference.

Regular and important support came from my dear labmates. We spent the majority of time together, I believe even more than with our roommates or significant others. Dr. Bo Li, Dr. Yuki Goto, and Dr. Trent Oman were mentors showing me all the details of the projects and experiments. Dr. Neha Garg helped me in the purification of ProcM. Xiao Yang dedicated quite a lot in the development of the co-expression system. John Hung, Noah Bindman and Dr. Gabrielle Thibodeaux shared their great ideas and precious experience in the non-canonical amino acid incorporation. Dr. Patrick Knerr instructed
me in SPPS, while Min Zeng helped me to perform western blotting trials. Dr. Qi Zhang helped to add more meanings to some of my “extra” work. Special thanks to Alejandro (Alex) Bueno, who started to work with me when he was still a sophomore. Alex participated in various projects before graduation, having contributed not only good works but also many fun memories. Manuel Ortega, Yi Yu, Kyle Dunbar, and Kaitlin Deane all provided special input during their rotations.

I also want to say special “thank you” to Ayse Okesli. The first day we met was still like yesterday, while all of a sudden we have been together working in one room for five years. She was always there during happy and down time, with encouragements, advice, hugs and stories. We have laughed together, cried together, and grown together, holding each other not only in work but also in daily life.

I feel really grateful to work with so many great people in the past five years. With a variety of expertise, they have given me enormous advice and taught me a variety of new things. Besides aforementioned, all of my labmates including Dr. Leigh Anne Furgerson, Dr. Lisa Cooper, Dr. John Whitteck, Dr. Jin Hee Lee, Dr. Kevin M. Clark, Dr. Ian Gut, Dr. Juan Velasquez, Dr. Nicholas Llewellyn, Dr. Remco Merkx, Dr. Svetlana A. Borisova, Dr. Michael Kuemin, Dr. Heather Cooke, Dr. Seung Young Kim, Dr. Ran Zhang, Dr. Huan Wang, Dr. Rebecca Splain, Dr. Jiangtao Gao, Dr. Despina Bougioukou, Dr. Zedu Huang, Dr. Debapriya Dutta, Spencer Peck, Weixin Tang, Chantal Garcia de Gonzalo, Subha Mukherjee, Xiling Zhao, Nidhi Kakkar, Josh Wagoner, Emily C Ulrich, Lindsey Johnstone, Isabel Neacato, Tong Hee Koh, and Ike Joewono, also people from other labs, such as Ying Xiao (as a very supportive friend too), Dr. Kou-San Ju, Joel
Cioni, Xiaomin Yu, Dr. Christine Herman, and Dr. Dylan Dodd, have made the journey full of colors and excitements.

Last but not least, I would like to thank my mom. She has sacrificed tremendously to raise me, and let me come oversea in pursuit of my advance study. I also want to thank my boyfriend, Xueda Wen. I was blessed to meet him in this campus and we share life with each other. I would also like to thank all the friends I met during the five years, you all make the five years full with love.

Yanxiang (Nancy) Shi

2013.8
TABLE OF CONTENTS

LIST OF FIGURES ........................................................................................................xii

LIST OF TABLES ...........................................................................................................xiv

CHAPTER 1: LANTHIPEPTIDES: BIOSYNTHESIS AND MÓDES OF ACTION .................................................................1
  1.1 INTRODUCTION ........................................................................................................1
  1.2 LANTHIPEPTIDES: DEFINITION AND CLASSIFICATIONS ................................3
  1.3 LANTHIPEPTIDE BIOACTIVITY AND MODES OF ACTIONS .....................10
  1.4 SUMMARY AND OUTLOOK ..................................................................................13
  1.5 REFERENCES ..........................................................................................................15

CHAPTER 2: IN VITRO AND IN VIVO RECONSTITUTION OF THE BIOSYNTHESIS OF PROCHLOROSINS ....................................................23
  2.1 INTRODUCTION ......................................................................................................23
  2.2 RESULTS AND DISCUSSION ..................................................................................27
    2.2.1 ProcM dehydrates and cyclizes ProcAs and their mutants ......................27
    2.2.2 Ring topology determination of modified ProcA core peptides ............30
    2.2.3 One-vector co-expression of His₆-ProcA and ProcM .........................33
    2.2.4 Large scale prochlorosin preparation for bioactivity investigations ..........42
    2.2.5 Incorporation of a non-canonical amino acid into modified ProcA ..................45
  2.3 SUMMARY AND OUTLOOK ..................................................................................48
  2.4 MATERIALS AND METHODS ................................................................................48
    2.4.1 Materials ...........................................................................................................48
    2.4.2 General methods ...............................................................................................49
    2.4.3 Construction of procA2.4/pET15b and procA2.8/pET15b .....................50
    2.4.4 Mutagenesis of ProcA2.1, ProcA2.11 and ProcA3.2 .........................51
    2.4.5 Overexpression and purification of linear precursor peptides of ProcA and ProcA mutants .................................................................53
    2.4.6 Overexpression and purification of His₆-tagged ProcM ......................55
    2.4.7 Activity assays for ProcM ...............................................................................57
    2.4.8 Construction of pRSFDuet-1 derivatives for co-expression of ProcM and ProcA .................................................................57
    2.4.9 Overexpression and purification of His₆-tagged modified ProcA and ProcA mutants .................................................................60
    2.4.10 Protease cleavage ..........................................................................................63
2.4.11 Tandem mass spectrometry analysis of prochlorosins and their derivatives .........................................................64
2.4.12 HPLC purification of prochlorosins modified in E. coli ..........65
2.4.13 Solid phase peptide synthesis (SPPS) of linear ProcA2.11 and ProcA3.2 core peptide derivatives .............................66
2.4.14 Deprotection of solid phase synthesized products .................66
2.4.15 Purification of deprotected SPPS products ..........................67
2.4.16 Incorporation of para-benzoyl phenylalanine (pBpa or B) in ProcA3.2 .................................................................68

2.5 REFERENCES .................................................................................72

CHAPTER 3: HETEROLOGOUS PRODUCTION OF THE LANTIBIOTIC ALA(0)ACTAGARDINE IN ESCHERICHIA COLI .........................................................74
3.1 INTRODUCTION ............................................................................74
3.2 RESULTS AND DISCUSSION ......................................................78
  3.2.1 Expression vector construction for GarA, GarM and GarO ......78
  3.2.2 Modification of GarA and GarA derivatives by GarM in E. coli ...............................................................80
  3.2.3 Ring topology confirmation of modifications on GarA derivatives ..............................................................83
  3.2.4 Modification of GarA and GarA derivatives by GarM and GarO in E. coli ......................................................84
  3.2.5 In vitro reconstitution of GarO ................................................86
  3.2.6 Overexpression and purification of GarM .............................89
3.3 SUMMARY AND OUTLOOK ...............................................................89
3.4 MATERIALS AND METHODS ..........................................................90
  3.4.1 Materials .............................................................................90
  3.4.2 General methods .................................................................90
  3.4.3 Genomic DNA extraction of Actinoplanes garbadinensis ATCC 31049 .................................................................91
  3.4.4 Construction of pRSFDuet-1 derivatives for garM, garA, and garO .................................................................92
  3.4.5 Overexpression and purification of GarM-modified His6-tagged GarA and GarA-V15L/I16V/I19insK ......................95
  3.4.6 Overexpression and purification of His6-tagged GarO ............97
  3.4.7 Proteolytic removal of GarA leader peptide .........................98
  3.4.8 Product evaluation for in vivo coexpression of GarA, GarM and GarO .................................................................99
  3.4.9 HPLC purification of Ala(0)-DAB-Lys ..................................99
  3.4.10 Iodoacetamide treatment of modified His6-GarA and
its derivatives to detect free cysteines ....................................100
3.4.11 GarO enzymatic assays .............................................100
3.4.12 Tandem mass spectrometry analysis of GarO modified
  Ala(0)-DAB-Lys .......................................................101
3.4.13 Bioactivity assay for deoxyactagardine derivatives ..........102
3.4.14 Overexpression and purification of His6-tagged GarM ....103

3.5 REFERENCES ..................................................................105

CHAPTER 4: EFFORTS TOWARDS THE INCORPORATION OF NON-
  CANONICAL AMINO ACID RESIDUES INTO LANTHIPEPTIDES ....108
  4.1 INTRODUCTION ................................................................108
  4.2 RESULTS AND DISCUSSION .............................................113
    4.2.1 Overexpression test with an engineered
      tyrosyl-tRNA synthetase ...........................................113
    4.2.2 Overexpression test with an engineered
      methionyl-tRNA synthetase ......................................114
    4.2.3 Overexpression test with an engineered
      pyrrolysyl-tRNA synthetase ....................................117
  4.3 SUMMARY AND OUTLOOK ..............................................120
  4.4 MATERIALS AND METHODS ..........................................121
    4.4.1 Materials .............................................................121
    4.4.2 General methods ....................................................122
    4.4.3 Construction of procA/procM/pRSFDuet-1 mutants
      containing an amber codon .....................................122
    4.4.4 Incorporation of para-benzoyl phenylalanine
      (pBpa) in ProcA ......................................................123
    4.4.5 Incorporation of H-Hpg-OH using
      methionyl-tRNA synthetase ....................................125
    4.4.6 Construction of nisA/nisB/pRSFDuet-1 mutants and
      nisC/tRNA/pCDFDuet-1 containing an amber codon ....125
    4.4.7 Incorporation of non-canonical amino acids in NisA ....127
  4.5 REFERENCES .............................................................129
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>General biosynthesis scheme of RiPPs</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>Representative structural motifs introduced by post-translational modifications into lanthipeptides</td>
</tr>
<tr>
<td>Figure 1.3</td>
<td>Four classes of lanthipeptide synthetases responsible for the generation of dehydrated amino acid residues and (methyl)lanthionine structures</td>
</tr>
<tr>
<td>Figure 1.4</td>
<td>Class I lanthipeptide biosynthesis pathway as represented by the maturation of nisin</td>
</tr>
<tr>
<td>Figure 1.5</td>
<td>Class II lanthipeptide biosynthesis pathway as represented by the maturation of lacticin 481</td>
</tr>
<tr>
<td>Figure 1.6</td>
<td>Biosynthetic pathway to labionin</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>Amino acid sequence alignment of twenty-nine ProcA precursor peptides</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>In vitro dehydrations of ProcAs by ProcM analyzed by MALDI-ToF-MS</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>In vitro cyclization patterns of ProcAs modified by ProcM shown by tandem mass spectrometry</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>ESI-MS/MS of ProcA3.2TEV modified by ProcM in vitro and treated by TEV protease</td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>MALDI-MS spectra of precursor peptides modified by ProcM in <em>E. coli</em> and treated with a protease to remove the leader peptide</td>
</tr>
<tr>
<td>Figure 2.6</td>
<td>ESI-MSMS of four ProcA peptides processed by ProcM in <em>E. coli</em> and cleaved by commercial proteases</td>
</tr>
<tr>
<td>Figure 2.7</td>
<td>MALDI-MS spectrum of ProcA4.3-G−1R precursor peptides modified by ProcM in <em>E. coli</em> and treated with trypsin to remove part of the leader peptide</td>
</tr>
<tr>
<td>Figure 2.8</td>
<td>MALDI-MS spectra of several precursor peptides modified by ProcM in <em>E. coli</em> and treated with a protease</td>
</tr>
<tr>
<td>Figure 2.9</td>
<td>MALDI-MS spectra of four prochlorosins modified by ProcM in <em>E. coli</em> after commercial protease cleavage and HPLC purification</td>
</tr>
<tr>
<td>Figure 2.10</td>
<td>Analytical RP-HPLC of purified prochlorosins</td>
</tr>
<tr>
<td>Figure 2.11</td>
<td>MALDI-MS spectrum of ProcA3.2TEV with Phe26 substituted by pBpa, modified by ProcM in <em>E. coli</em> and treated with GluC to remove the major part of the leader peptide</td>
</tr>
<tr>
<td>Figure 2.12</td>
<td>Precursor peptide sequences of some of the lanthipeptides used in this study</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Biosynthesis pathway of actagardine and its derivatives, with</td>
</tr>
</tbody>
</table>

xii
sequence alignments of GarA and its homologue, LigA ............75

Figure 3.2 The biosynthetic gene clusters of (A) actagardine and (B) DAB ..........77

Figure 3.3 MALDI-MS spectra of GarA and GarA-V15L/I16V/I19insK modified by GarM in E. coli and treated with trypsin and aminopeptidase A8200 to remove the leader peptide .......................81

Figure 3.4 MALDI-MS spectra of GarM-modified GarA incubated with IAA (blue) and without IAA (red) and treated with trypsin ......................82

Figure 3.5 ESI-MS/MS spectra of GarA-V15L/I16V/I19insK modified in E. coli by GarM and GarO and treated in vitro with trypsin and aminopeptidase A8200 ..................................................83

Figure 3.6 Antimicrobial assays of actagardine derivatives ................................84

Figure 3.7 MALDI-MS spectra of GarA modified by GarM and GarO ...........85

Figure 3.8 MALDI-MS spectrum of GarA modified by GarO from co-expression of garA/garM/garO/pRSFDuet-1 ...............................86

Figure 3.9 MALDI-MS spectrum of modified GarA-V15L/I16V/I19insK from co-expression of garA-V15L/I16V/I19insK/garM/pRSFDuet-1 and 2garO/pCDFDuet-1 ...................................................87

Figure 3.10 Analysis of His6-GarO by gel filtration chromatography ..................88

Figure 4.1 NAAs utilized by the methionyl-tRNA synthetase and prochlorosin sequences .................................................................110

Figure 4.2 Examples of reported NAAs incorporated into proteins by the evolved pyrolysyl-tRNA synthetase-pylT pair ................................111

Figure 4.3 Nisin A structure ......................................................................112

Figure 4.4 MALDI-MS spectra of incorporation of Hpg into (A) ProcA2.8, and (B) Proc3.3-G–1K .................................................................115

Figure 4.5 MALDI-MS spectra of the incorporation of L-Phe(3-CF3)-OH into NisA-I4V/S5B/L6G .................................................................118

Figure 4.6 MALDI-MS spectra of the incorporation of L-Phe(3-CF3)-OH into (A) NisA-K22B and (B) NisA-S29B ........................................119

Figure 4.7 MALDI-MS spectrum of the truncation product from NisA-S29B during co-expression trial to incorporate L-Phe(3-CF3)-OH ........120


**LIST OF TABLES**

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.1</td>
<td></td>
</tr>
<tr>
<td>Samples sent to MIT for bioactivity assays</td>
<td>45</td>
</tr>
<tr>
<td>Table 2.2</td>
<td></td>
</tr>
<tr>
<td>Primer sequences for cloning and mutagenesis of procAs in pET15b</td>
<td>53</td>
</tr>
<tr>
<td>Table 2.3</td>
<td></td>
</tr>
<tr>
<td>Primer sequences for cloning and mutagenesis of procAs for co-expression</td>
<td>59</td>
</tr>
<tr>
<td>Table 2.4</td>
<td></td>
</tr>
<tr>
<td>Plasmid vectors used in Chapter 2</td>
<td>70</td>
</tr>
<tr>
<td>Table 2.5</td>
<td></td>
</tr>
<tr>
<td>Summary of the constructs and purified peptides mentioned in Chapter 2</td>
<td>71</td>
</tr>
<tr>
<td>Table 3.1</td>
<td></td>
</tr>
<tr>
<td>Primer sequences for cloning and mutagenesis of garA, garM and garO</td>
<td>94</td>
</tr>
<tr>
<td>Table 3.2</td>
<td></td>
</tr>
<tr>
<td>Plasmids mentioned in Chapter 3</td>
<td>105</td>
</tr>
<tr>
<td>Table 4.1</td>
<td></td>
</tr>
<tr>
<td>Primer sequences for cloning and mutagenesis of procAs</td>
<td>123</td>
</tr>
<tr>
<td>Table 4.2</td>
<td></td>
</tr>
<tr>
<td>Primer sequences for cloning and mutagenesis of nisA, nisC and pyt</td>
<td>127</td>
</tr>
<tr>
<td>Table 4.3</td>
<td></td>
</tr>
<tr>
<td>Plasmids mentioned in Chapter 4</td>
<td>129</td>
</tr>
</tbody>
</table>
1.1 INTRODUCTION

Natural products are chemicals produced by living organisms for a variety of reasons during their life cycles. These compounds play important roles for the survival of organisms, such as defending against pathogens,\(^1\) inhibiting the growth of competitors,\(^2\) functioning as signaling molecules for communication,\(^3\) or as by-products of metabolism.\(^4\) Because of their diversified activities, natural products have supplied a large pool of candidates for drug discovery,\(^5\) inspiring the development of antimicrobial,\(^6\) anticancer,\(^7\) antimalarial,\(^8\) and anti-HIV drugs.\(^9\) The development of genome sequencing and bioinformatic studies have revealed that a major class of peptide-based natural products are ribosomally synthesized and post-translationally modified that have recently been named RiPPs (as an abbreviation for Ribosomally synthesized and Post-translationally modified Peptides).\(^10\) The biosynthetic pathways of RiPPs are designated as Post-Ribosomal Peptide Synthesis (PRPS), as opposed to Non-Ribosomal Peptide Synthesis (NRPS), which produce natural products through modular assembly.\(^10\)

Extensive modifications have been introduced into RiPPs through PRPS, such as methylation,\(^11\) dehydration,\(^12\) amidation,\(^13\) hetero- or macro-cyclization,\(^14\) glycosylation,\(^15\) epimerization,\(^14\) and aromatization.\(^16\) With the introduction of structures not accessible with the canonical 20 amino acids, these modifications have not only significantly expanded the chemical diversity of the RiPPs and increased their functionality and
bioactivity, but also augmented their stability by restricting their conformations and protecting the peptides against proteolysis.

RiPPs are encoded as a precursor peptide in the genome, typically with a length of 20 – 110 amino acid residues. These precursor peptides generally contain an N-terminal leader peptide and a C-terminal core peptide. The leader peptides serve multiple proposed functions during the maturation of the final products, including: 1) serving as docking molecules recognized by the biosynthetic enzymes to initiate modification, 2) facilitating precursor-peptide folding during the modifications, and 3) protecting the producing strains from their bioactive final products by keeping the modified core peptide in an inert state inside the cells. Some precursor peptides possess an extra short sequence attached N-terminally to the leader peptide known as a signal peptide, directing the precursors to their subcellular locations for certain post-translational modifications. Some precursor peptides append a so-called recognition sequence attached C-terminally

![Diagram](image)

Figure 1.1 General biosynthesis scheme of RiPPs. Most RiPP precursor peptides contain a leader peptide N-terminally attached to the core peptide, except bottromycin for which the leader peptide is attached C-terminally to the core peptide.
to the core peptide for product excision and cyclization (Figure 1.1).\textsuperscript{19} Noteworthy, the precursor peptide of bottromycin lacks an N-terminal leader peptide but contains a long sequence attached C-terminally to the core peptide with functions similar to the leader peptide. Such a C-terminal sequence is designated as a “follower” peptide.\textsuperscript{20}

\subsection*{1.2 LANTHIIPEPTIDES: DEFINITION AND CLASSIFICATIONS}

Lanthipeptides are a class of intensively studied RiPPs, with the name referring to lanthionine-containing peptides (formerly known as lantipeptides).\textsuperscript{21} They are characterized by unusual thioether linkages named lanthionine (Lan) and/or methyllanthionine (MeLan). Lanthipeptides are ribosomally synthesized as a linear precursor peptide generally with an N-terminal leader peptide and a C-terminal core peptide. Biosynthetic enzymes catalyze the dehydration of some or all of the serine (Ser) and threonine (Thr) residues in the core peptide to 2,3-didehydroalanine (Dha) and (Z)-2,3-didehydrobutyrine (Dhb) moieties, respectively. Cysteine residues from the core peptides then undergo intramolecular Michael addition onto these α, β-unsaturated amino acids, forming (methyl)lanthionine rings catalyzed by biosynthetic enzymes. Besides the aforementioned dehydrated residues and thioether crosslinks, post-translational modifications have also introduced diversified functional groups into lanthipeptides, including but not limited to D-amino acid residues (as in lactocin S and lacticin 3147),\textsuperscript{22} lysinoalanine (as in cinnamycin),\textsuperscript{23} sulfoxide (as in actagardine),\textsuperscript{24} disulfides (as in bovicin HJ50),\textsuperscript{25} N-terminal acetylation (as in paenibacillin),\textsuperscript{26} carbon-carbon crosslink (as in labyrinthopeptin A2),\textsuperscript{27} and hydroxylated and halogenated amino acids (as in microbisporicin) (Figure 1.2).\textsuperscript{28}
Based on the biosynthetic enzymes that introduce the thioether motifs, lanthipeptides are classified into four classes (Figure 1.3). The dehydration of precursor peptides in class I is catalyzed by a dehydratase generically named LanB, while the cyclization is catalyzed by a dedicated LanC cyclase. LanC enzymes possess conserved residues forming a Cys-Cys-His triad by binding to a zinc ion in the active site. The bound zinc activates the Cys thiols in the LanA peptide for nucleophilic attack onto the Dha or Dhb residues formed by LanB catalysis. A transmembrane ATP-binding cassette (ABC) transporter LanT encoded in the biosynthetic gene cluster facilitates the
Figure 1.3 Four classes of lanthipeptide synthetases responsible for the generation of dehydrated amino acid residues and (methyl)lanthionine structures. Dark lines represent the conserved residues in the enzymes that are important for their catalytic activities.

Export of modified precursor peptides from the cytosol. Often the class I biosynthesis pathway also possesses a dedicated subtilisin-like serine protease LanP. The protease either is anchored in the cell membrane or resides in the cytoplasm and is responsible for removing the leader peptide from the modified core peptide to generate the final bioactive compound. The most intensively studied lanthipeptide nisin, an antimicrobial natural product produced by *Lactococcus lactis*, is a typical class I lanthipeptide. The nisin precursor peptide NisA is dehydrated and cyclized by NisB and NisC, respectively (Figure 1.4). Recent pull-down experiments using C-terminally His-tagged NisA
discovered that NisB and NisC form a complex during NisA modification in *L. lactis*. An ABC transporter NisT encoded between *nisB* and *nisC* in the nisin biosynthetic gene cluster transports the modified precursor peptide before the proteolytic removal of the leader peptide by NisP. The epilancin 15X protease ElxP was recently reported as the first *in vitro* reconstituted protein in the LanP family.

Class II lanthipeptides undergo the dehydration and cyclization with the help of a bifunctional enzyme, generically named LanM (Figure 1.5), which has an N-terminal dehydratase domain with no homology with LanB proteins, and a C-terminal LanC-like cyclase domain. ATP and Mg$^{2+}$ are required for the LanM to perform phosphorylation of the Ser/Thr residues and subsequent phosphate ester elimination to generate Dha/Dhb,
as originally demonstrated by *in vitro* study of the lacticin 481 synthetase LctM.\textsuperscript{12,33} Interestingly, no ATP binding site was found in LctM and no ATP is required for the LctM-catalyzed cyclization to occur.\textsuperscript{34} Class II lanthipeptides, such as lacticin 481,\textsuperscript{12} lacticin 3147,\textsuperscript{35} mersacidin,\textsuperscript{36} actagardine,\textsuperscript{24} and haloduracin,\textsuperscript{37} utilize one dedicated LanM synthetase to modify one precursor peptide substrate in nature. However, recent genome-mining studies have identified novel class II lanthipeptide biosynthesis systems where one LanM synthetase possesses much higher substrate tolerance than previously recognized, such as a single synthetase ProcM encoded with 29 potential substrates in the

\textbf{Figure 1.5} Class II lanthipeptide biosynthesis pathway as represented by the maturation of lacticin 481. The bifunctional enzyme LanM catalyzes both the dehydration and cyclization processes. Besides transportation, the ABC-transporter LanT also removes the leader peptide through proteolysis.
cyanobacterium Prochlorococcus MIT9313. More details about the ProcM system is provided in Chapter 2. Leader peptide removal and secretion of class II lanthipeptides is facilitated by ABC-transporter LanT proteins, which typically contain an N-terminal papain-like cysteine protease domain that is not present in class I LanT transporters.

In Class III lanthipeptides, the dehydro amino acids are introduced through phosphorylation of Ser/Thr residues by a kinase domain in a synthetase generically named LanKC, followed by the elimination of the phosphate by a phosphoSer/Thr lyase domain located at the N-terminus of the LanKC. Similar as LanM, the LanKC protein also contains a C-terminal cyclase motif but without the zinc-binding residues conserved among LanC-like cyclase domains. The first discovered LanKC enzyme was RamC which catalyzed the production of the morphogenetic peptide SapB in Streptomyces coelicolor. LabKC was later found as an analog of RamC, which catalyzes the synthesis of labyrinthopeptins in Actinomadura namibiensis, employing GTP instead of ATP to function in vitro. Many LabKCs also catalyze the formation of a carbon-carbon crosslink termed labionin (Lab), a structure exclusively found in class III lanthipeptides, presumably formed by Michael attack onto a dehydroalanine by the enolate generated during the formation of thioether linkages (Figure 1.6). However, the mechanism of LabKC catalysis of this second Michael addition is still elusive.
Figure 1.6 Biosynthetic pathway to labionin.

Class IV lanthipeptide synthetases, generically named LanL, also encompass an N-terminal phosphoSer/Thr lyase domain and a central kinase domain. Unlike LanKC, LanL proteins bare a C-terminal LanC-like domain with conserved zinc-binding residues. The first report on LanL reported in vitro reconstitution of a synthetase from *Streptomyces venezuelae* called VenL. In the presence of ATP and Mg\(^{2+}\), VenL performed phosphorylation, phosphate elimination, and cyclization on the VenA substrate. Remarkably, both class III and class IV lanthipeptide biosynthetic gene clusters contain LanT-like proteins without a protease domain, while no proteases have been found in these clusters so far. Although LanM, LanKC, and LanL all catalyze dehydration through phosphorylation and elimination, no kinase or lyase domain has been found in LanM. In fact, LanB/C, LanM, LanKC and LanL are grouped into different clades in phylogenetic analysis, suggesting independent evolution pathways of these modification enzymes.
1.3 LANTHIPEPTIDE BIOACTIVITY AND MODES OF ACTIONS

A major subclass of lanthipeptides display antimicrobial activity, and are called lantibiotics (abbreviation of lanthionine-containing antibiotics). Thus far, more than 60 lantibiotics have been identified and characterized out of more than 90 known lanthipeptides. Elucidated lantibiotic modes of action fall into four major groups:

(i) Inhibition of bacterial cell wall biosynthesis via binding to lipid II. Lipid II is a membrane-anchored intermediate for bacterial cell wall biosynthesis. Lipid II facilitates cell wall biosynthesis by transferring peptidoglycan subunits from the cytoplasm to the outside of the cell membrane, where the peptidoglycan units are assembled into a peptidoglycan sheet that forms the skeleton of the bacterial cell wall. Lipid II is composed of a polyisoprenoid anchor, attached to two sugar subunits, one N-acetylglucosamine (GlcNAc) and one N-acetylmuramic acid (MurNAc)-pentapeptide, via a pyrophosphate linker. Nisin, mutacin 1140, subtilin, mersacidin, epidermin, and actagardine are known to inhibit peptidoglycan polymerization by binding to lipid II. The N-terminal two rings of nisin bind to the pyrophosphate moiety of lipid II, forming a pyrophosphate cage. Mersacidin and actagardine were suggested to bind to the disaccharide-diphosphate region, unlike the glycopeptide antibiotic vancomycin which binds to the C-terminal dipeptide region (D-Ala-D-Ala) of the pentapeptide part of lipid II.

(ii) Pore formation in the cell membrane. Utilizing bound lipid II as a docking molecule, nisin is able to form pores in the cell membrane with its three C-terminal rings, each pore consisting of eight nisin molecules and four lipid II molecules. Pore formation was also observed with gallidermin against Microoccus and Staphylococcus
strains but not with \textit{Lactococcus lactis}, suggesting the thickness of the target bacterial cell membrane and the length of the lantibiotics determine the mode-of-action of certain lantibiotics, as shorter peptides may not be able to span across the bacterial cell membrane to form pores, while lipid II binding is less affected by the lantibiotic lengths.\textsuperscript{55}

Recently discovered two-component lantibiotics require two peptides (named as the \(\alpha\)- and \(\beta\)-peptides) that synergistically display full activity. Individual \(\alpha\) or \(\beta\) peptides possess little to no antimicrobial activity. Currently known two-component lantibiotics, such as haloduracin,\textsuperscript{37} plantaricin W,\textsuperscript{59} lacticin 3147,\textsuperscript{60} and cytolysin,\textsuperscript{61} are all class II lantipeptides modified by LanM synthetases. As a typical two-component lantibiotic, haloduracin \(\alpha\) (Hal\(\alpha\)) binds to lipid II, facilitating the pore formation by haloduracin \(\beta\) (Hal\(\beta\)) in the cell membrane through a lipid II-Hal\(\alpha\)-Hal\(\beta\) complex in a 1:2:2 stoichiometry.\textsuperscript{62,63} Ring-topology elucidation\textsuperscript{64} and mutagenesis of haloduracin,\textsuperscript{63} together with NMR study of lacticin 3147\textsuperscript{60} and sequence alignments revealed that the \(\alpha\)-peptides bear three C-terminal rings homologous to mersacidin, which are important for lipid II binding activity.\textsuperscript{65}

(iii) Inhibition of the outgrowth of spores. Nisin and subtilin were observed to inhibit spore outgrowth of \textit{Bacillus} and \textit{Clostridium} species.\textsuperscript{66} The activity was abolished when their Dha5 was substituted by alanine, while the bactericidal activity was not affected.\textsuperscript{21,67} However, when the Dha5 was substituted by phenylalanine, neither activity was abolished. Meanwhile, the removal of the C-terminal region (nisin \(\Delta23-34\)) resulted in the decline of both activities.\textsuperscript{68}
(iv) Other modes of actions. Duramycin and its close structural analogue cinnamycin inhibit phospholipase A2 by binding to phosphatidylethanolamine (PE) in the cell membrane.\textsuperscript{69-71} Nisin and Pep5 were also shown to interact with the teichoic acids on the cell membrane to induce autolysis of several staphylococcal strains.\textsuperscript{72}

According to intensive investigations on nisin and duramycin resistant strains, resistance to lantibiotics include the change of membrane components, rigidification of the cell membrane, thickening of the cell wall, alteration of the hydrophobicity of the cell wall, and tuning of cell membrane electronic potential.\textsuperscript{21,73,74} These resistant mechanisms are consistent with known modes of action of lantibiotics, which mainly target the cell membrane components in the course of cell wall biosynthesis. A series of genes, such as the histidine kinase gene \textit{lisK}, the high molecular-weight penicillin binding protein gene \textit{php2229}, the histidine protein kinase gene \textit{hpk1021}, and the D-alanine-poly (phosphoribitol) ligase subunit gene \textit{dltA} were shown to be involved in the alteration of membrane compositions.\textsuperscript{21} However, genetic alterations of target bacteria upon incubation with lantibiotics have not been comprehensively studied, which may unravel more detailed explanations of resistant phenotypes and underexplored modes of actions.

Several lantibiotics such as epilancin K\textsubscript{7}\textsuperscript{75} and epilancin 15X\textsubscript{76} permeabilize the cell membrane through pore formation without binding to lipid II.\textsuperscript{62,75} However, their modest pore formation ability and highly potent antimicrobial activity suggest that unknown targets are yet to be explored. Besides toxicity against bacteria, cytolysin, a two-component lantibiotic, exhibits hemolytic activity by unexplored mechanisms.\textsuperscript{77} Some lanthipeptides exhibit no significant antimicrobial activity but have other functions. These compounds include SapB, a morphogenetic peptide produced by \textit{Streptomyces}
coelicolor; AmfS, a SapB homolog secreted by *S. griseus*; and SapT produced by *S. tendae*, all facilitating aerial hyphae formation during sporulation. In the past, most lanthipeptides were discovered based on their activity mainly through antimicrobial screening. The recently discovered prochlorosins through genome mining have displayed neither growth inhibitory effect against common soil bacteria nor other common activities, suggesting the roles of these lanthipeptides are still to be discovered. With the development of sequencing technologies, bioinformatics and high-throughput analysis, hundreds of lanthipeptides have been identified in the genomes, presumably with more currently unknown biological functions and molecular mechanisms.

### 1.4 SUMMARY AND OUTLOOK

Natural products play important roles in the survival of organisms, and provide a large pool of candidates for drug discovery and pharmaceutical investigations. A major class of peptide-based natural products are ribosomally synthesized and post-translationally modified. Lanthipeptides which possess (methyl)lanthionine structures are a class of intensively studied RiPPs. Based on the biosynthesis enzymes that introduce the thioether motifs, lanthipeptides are classified into four classes. Class I lanthipeptides utilize distinct dehydratase LanB and cyclase LanC, while class II lanthipeptides adopt a bifunctional enzyme LanM. Trifunctional enzymes LanKC and LanL are employed by class III and class IV, respectively. A major subclass of lanthipeptides possess antimicrobial activity, known as lantibiotics with known modes of action including lipid II binding, pore formation, inhibition of spore outgrowth, and phosphatidylethanolamine (PE) binding.
Several lanthipeptides display no antimicrobial activities but other functions important for bacterial life cycles, such as facilitating aerial hyphae formation during sporulation. With the development of genome mining and bioinformatics, a variety of lanthipeptides have been identified in the genome, presumably with more unexplored functions and molecular targets. Prochlorosins are a typical group of lanthipeptides which were discovered through genome mining without known antibiotic activities. Further biological function investigations of prochlorosins require pure products in adequate amounts, which was not achievable through natural product isolation or *in vitro* enzymatic assays. In this thesis, Chapter 2 will describe the development of a novel co-expression system to produce prochlorosins by enzymatic modification in *E. coli*, providing prochlorosins with high purity and sufficient quantity for subsequent bio-function characterizations. Chapter 3 will present use of the co-expression methodology to produce antimicrobial lanthipeptides, with extra modifications besides dehydration and cyclization. The precursor peptide of actagardine, one of the only two lantibiotics that have been evaluated in clinical trials thus far, was fully modified in *E. coli* as presented in Chapter 3. A luciferase-like monooxygenase located in the biosynthetic cluster of actagardine was also reconstituted, both *in vitro* and *in vivo*, to introduce a sulfoxide group onto actagardine and its derivatives. Based on the success of *in vivo* co-expression of lanthipeptides with known or unidentified activities, Chapter 4 will describe efforts towards the incorporation of non-canonical amino acids into lanthipeptides, in order to expand the structural diversity of these natural products, tune their activities, and introduce probes for subsequent functional investigations.
1.5 REFERENCES


"Ribosomally synthesized and post-translationally modified peptide natural products: Overview and recommendations for a universal nomenclature".


(39) Ihnken, L. A. F.; Chatterjee, C.; van der Donk, W. A. Biochemistry 2008, 47, 7352-7363. "In vitro reconstitution and substrate specificity of a lantibiotic protease".


(43) Goto, Y.; Ökesli, A.; van der Donk, W. A. Biochemistry 2011, 50, 891-898. "Mechanistic studies of Ser/Thr dehydration catalyzed by a member of the LanL lanthionine synthetase family".


18


(83) Marsh, A. J.; O'Sullivan, O.; Ross, R. P.; Cotter, P. D.; Hill, C. BMC Genomics 2010, 11, 679. "In silico analysis highlights the frequency and diversity of type 1 lantibiotic gene clusters in genome sequenced bacteria".
CHAPTER 2: *IN VITRO AND IN VIVO RECONSTITUTION OF THE BIOSYNTHESIS OF PROCHLOROSINS*¹

2.1 INTRODUCTION

Based on genome mining using the mersacidin synthetase gene *mrsM* from *Bacillus* sp. strain HIL Y-85 as a query, a single novel *lanM*-like gene was found by previous graduate student Dr. Bo Li in the genomes of two closely related strains of *Prochlorococcus* (MIT9313 and MIT9303) and one distantly related strain of marine *Synechococcus* (RS9961). Interestingly, the genome of *Prochlorococcus* MIT9313 also contains 29 putative *lanA*-like genes as potential substrates of its single LanM-like protein. A smaller number of potential substrates are encoded in the genomes of *Prochlorococcus* MIT9303 and *Synechococcus* RS9961. Thus, *Prochlorococcus* MIT9313 was chosen as a subject of the following investigation.

*Prochlorococcus* is a planktonic marine cyanobacterium that distributes widely throughout the world, from the surface waters to the deep ocean.² This ubiquitous photosynthetic prokaryote has a special significance as the smallest known photosynthetic organism, maintaining its diverse biofunctions with a limited genome size (e.g. 2.4 Mbp for the genome of *Prochlorococcus* MIT9313).³ ⁴ The genomes of

Prochlorococcus strains collected from different oceans of the world and different depths in the ocean are remarkably diverse, which is presumably due to their adaptability to exist in a variety of niches. Studies on the importance of this genetic variation and the stimuli that induced the variation are still under way. Prochlorococcus is found as a ubiquitous organism in the nutrient-poor regions of the ocean, which suggests that it may have special features for surviving under nutrient-limiting conditions that are unfavorable for most other marine organisms.

The lanM-like gene in Prochlorococcus MIT9313 was designated as procM, while the lanA-like genes were designated as procAs. Seven procAs are localized in the same gene cluster as procM (procA1.1 - procA1.7), with 20 other procAs organized in three additional clusters within the same Prochlorococcus genome (procA2.1 - procA2.11, procA3.1 - procA3.5, procA4.1 - procA4.4). Two more isolated procA genes were found elsewhere in the genome (procAs.1 and procAs.2). The translated sequences of procA genes have high homology in their leader peptide regions but their core peptide sequences are strikingly diverse (Figure 2.1). The putative leader and core regions of ProcA peptides are separated by a Gly-Gly or Gly-Ala sequence that is usually present in class II lanthipeptide precursor peptides and is generally a cleavage site of an ABC-transporter with an N-terminal cysteine protease domain, which removes the leader peptide and secretes the mature core peptide. The mature core peptides were designated as prochlorosins (Pcn).
The discovery of 29 genes encoding LanA-like peptides with only one gene encoding a LanM protein in the *Prochlorococcus* genome is of great interest, suggesting substrate tolerance of ProcM. The highly diverse core peptides of the 29 ProcAs have no homology to any known lanthionine-containing compounds, which in turn would significantly enlarge the pool of existing lanthipeptide structures. As the cell density of the plankton in the ocean is relatively low, especially the nutrient depleted areas where

**Figure 2.1** Amino acid sequence alignment of twenty-nine ProcA precursor peptides. Highly conserved residues in the leader peptides are highlighted in blue. Serine and threonine residues in the core peptides are highlighted in orange and pink, respectively. Cysteine residues are highlighted in green. The arrow indicates putative cleavage sites to remove the leader peptides after modification in nature.
Prochlorococci dominate, prochlorosins may have antibiotic activity or may work as signaling molecules.⁶

Among the 29 genes coding for prochlorosins, five (procA1.1, 1.7, 2.5, 3.3 and 4.3) have been shown to be transcribed in exponentially growing Prochlorococcus MIT9313 cultures in ocean water, while the others have not yet been analyzed.⁶ ESI-MS analysis of solid phase extractions of the supernatant from the cultures identified detectable levels of Pcn2.1, 2.11, and 3.2.⁶ In order to perform further structural and biofunctional investigations, production of a relatively large amount of prochlorosins in the laboratory was needed. \textit{In vitro} production of the peptides is one route towards this goal because less than 10 µg of unpurified prochlorosins could be obtained from a 20 L Prochlorococcus culture cultivated in ocean water.⁶ For \textit{in vitro} production, genes encoding ProcA and ProcM were amplified from Prochlorococcus MIT9313 genomic DNA, followed by insertion into the appropriate vectors and utilizing \textit{Escherichia coli} for heterologous expression. A former group member, Dr. Bo Li, accomplished the cloning of 16 of the 29 procA genes into the pET15b vector and the procM gene into the pET28b vector. In addition, some preliminary work, including detailed bioinformatics studies and ring topology investigations, was performed by Dr. Li in collaboration with the group of Professor Sallie Chisholm at the Massachusetts Institute of Technology.⁶ This chapter describes the development of a new production method that involves co-expression of the ProcA precursor peptides with the ProcM enzyme in \textit{E. coli}. ⁶
2.2 RESULTS AND DISCUSSION

2.2.1 ProcM dehydrates and cyclizes ProcAs and their mutants

In addition to the procA/pET15b constructs prepared by Dr. Bo Li, two more procAs (procA2.4 and procA2.8) were cloned in this work from Prochlorococcus MIT9313 genomic DNA into the pET15b vector. ProcA2.4, ProcA2.8 (for sequences, see Figure 2.1), and ProcM were produced in E. coli with an N-terminal hexahistidine tag (His$_6$) to allow convenient purification. At room temperature, peptides ProcA2.4 and ProcA2.8 were incubated with ProcM in 50 mM HEPES buffer (pH 7.5) in the presence of 2.5 mM ATP, 10 mM MgCl$_2$, and 0.5 mM tris(2-carboxyethyl)phosphine (TCEP) for two days. Negative control reactions including ProcA and the other necessary components except ProcM were set up in parallel. After incubation, either endoproteinase Asp-N or Glu-C was added to the reaction mixture to remove a part of the leader peptide together with the His$_6$-tag from the core peptide. The mixtures were incubated at room temperature for 5 h and then subjected to matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-ToF-MS) (for spectra, see Figure 2.2). After incubation with ProcM and Asp-N treatment, the proteolytic fragment containing the ProcA2.4 core peptide exhibited a molecular ion consistent with the loss of one molecule of water (calculated (calcd): 2655.22 [M − H$_2$O + H$^+$], observed: 2654.91), while a small peak corresponding to the loss of two molecules of water was also observed (calcd: 2637.21 [M − 2 H$_2$O + H$^+$], observed: 2637.04). However, the majority of ProcA2.4 core peptides were not dehydrated as shown by MALDI mass spectrometry (Figure 2.2A). After
incubation with ProcM and Glu-C treatment, the loss of two molecules of water from the proteolytic fragment containing the ProcA2.8 core peptide was observed (calcd: 2392.97 $[\text{M} - 2 \text{H}_2\text{O} + \text{H}^+]$, observed: 2393.03, see Figure 2.2B). Without ProcM treatment, no dehydrations were observed in the assay with either peptide. These observations establish

\[ \text{Figure 2.2 In vitro dehydrations of ProcAs by ProcM analyzed by MALDI-ToF-MS.} \]

Representative MALDI-ToF mass spectra of ProcAs modified by ProcM after proteolytic cleavage using A) Asp-N for ProcA2.4, B) Glu-C for ProcA2.8, and C) TEV for ProcA3.2TEV. (D) Amino acids sequences of ProcAs after proteolytic cleavage. Cys residues are in red, and Ser/Thr residues that may be dehydrated are in blue. Underlined amino acids belong to the remaining part of the leader peptide after proteolytic cleavage. Unmodified peptides are denoted as M in the figure.
that ProcM is catalytically competent in dehydrating ProcA2.4 and ProcA2.8, with the latter being a better substrate.

Previously, proteolytic cleavage of wild type ProcAs was not able to remove the leader peptides completely as none of the proteases used could cleave C-terminally at the Gly-Gly motif. In order to completely remove the leader peptide without cleavage of the core, specific protease cleavage sites were introduced via substitution of the C-terminal amino acid residues in the leader peptides. As a trial, the four C-terminal residues of the leader peptides of ProcA2.1, 2.11 and 3.2 (for wild type sequences, see Figure 2.1) were substituted with IEGR, which is the cleavage site for Factor Xa protease. Additionally, the six C-terminal residues of the ProcA2.11 and 3.2 leader peptides were also substituted with ENLYFQ, which is the tobacco etch virus (TEV) protease recognition sequence. The mutant peptides, designated ProcA2.1Xa, ProcA2.11Xa, ProcA2.11TEV, ProcA3.2Xa, and ProcA3.2TEV, were incubated with or without ProcM in 50 mM HEPES buffer (pH 7.5) in the presence of 2.5 mM ATP, 10 mM MgCl₂, and 0.5 mM TCEP for 15 to 20 h. After incubation, the products of the assays with ProcA2.1Xa, ProcA2.11Xa, and ProcA3.2Xa were treated with commercial Factor Xa, while the ProcA2.11TEV and ProcA3.2TEV products were treated with TEV protease. MALDI-ToF MS analysis was then applied to analyze the outcome of the assays. After TEV protease treatment, a peak corresponding to the loss of three molecules of water was observed for the ProcA3.2 core peptide (calcd: 2883.09 [M – 3 H₂O + H⁺], observed: 2881.74, see Figure 2.2C). The dehydration of ProcA3.2TEV was almost complete, with nearly all the core peptides fully
dehydrated (−3 H₂O), after incubation with ProcM at room temperature for 3 days. However, no dehydrations were observed in the spectra of the assay products of ProcA2.1Xa, ProcA2.11Xa, or ProcA3.2Xa, suggesting the mutations interfered with the modification by ProcM. For the reaction of ProcM with ProcA2.11TEV, no peaks corresponding to the ProcA2.11 core peptide were observed, with either sinapinic acid (SA) or α-cyano-4-hydroxycinnamic acid (CHCA) used as the MALDI matrix, although peaks corresponding to the leader peptide were observed. Hence, the in vitro approach was only partially successful as introduction of the protease cleavage sites appeared to hamper in vitro modification of the peptides by ProcM. Even though ProcA3.2TEV was successfully modified by ProcM, longer incubation time (3 days) was required to realize the completion of the modifications, compared with a much shorter time for the wild type ProcA3.2 (12 h). Therefore, multiple-amino-acid substitution at the C-terminus of the leaders of ProcA peptides may hinder the ProcM efficiency in vitro.

2.2.2 Ring topology determination of modified ProcA core peptides

As the masses of cyclized and uncyclized derivatives of dehydrated ProcA peptides are identical, liquid chromatography electrospray ionization tandem mass spectrometry (LC/ESI-MS/MS) was adopted for further characterization of the modified ProcA core peptides (for spectra, see Figure 2.3). For modified ProcA2.4, no fragmentation was observed except in the sequence remaining from the leader peptide (Figure 2.3A). For modified ProcA2.8, two non-overlapping rings can be assigned to the core peptide made
by connecting Cys3 to Dha9 and Cys19 to Dha13, according to the fragmentation pattern observed in tandem mass spectrometry (Figure 2.3).

**Figure 2.3** *In vitro* cyclization patterns of ProcAs modified by ProcM shown by tandem mass spectrometry. Representative tandem mass spectra of ProcAs modified by ProcM after proteolytic cleavage using Asp-N for ProcA2.4 (A & B), and Glu-C for ProcA2.8 (C). (D) Fragmentation patterns of ProcAs after proteolytic cleavage as determined by tandem mass spectrometry. Red arrows indicate the b and y'' ions in MSMS spectra. Core peptide sequences are in blue, the remaining leader peptide sequences in green. Putative dehydroalanines (Dha) and dehydrobutyrines (Dhb) are highlighted in green, cysteine residues highlighted in blue. For the ProcA2.4 fragment with only one dehydration, it is not certain which of the Ser and Thr was dehydrated, thus no Dhb or Dha is indicated in the sequence shown here.
For modified ProcA3.2 core peptide cleaved from ProcA3.2TEV, the fragmentation pattern was similar to that of ProcA3.2 produced in vivo by Prochlorococcus MIT9313 (Figure 2.4), indicating that after incubation with ProcM and treatment with TEV protease, the ProcA3.2 core peptide was dehydrated and cyclized to give the same ring topology as Pcn3.2 produced by the producing Prochlorococcus MIT9313.

Figure 2.4 ESI-MS/MS of ProcA3.2TEV modified by ProcM in vitro and treated by TEV protease. The fragment pattern is similar as Pcn3.2 detected in the culture of Prochlorococcus MIT9313 and wild type ProcA3.2 modified by ProcM in vitro, with the latter two reported in Reference 6.
2.2.3 One-vector co-expression of His\textsubscript{6}-ProcA and ProcM

As shown in the previous section, modified ProcA core peptides with the same ring topology as the corresponding prochlorosins produced \textit{in vivo} by \textit{Prochlorococcus} MIT9313 strains can be obtained via \textit{in vitro} incubation of a linear precursor ProcA with ProcM. However, the \textit{in vitro} method was not able to provide sufficient quantities of the prochlorosins for more detailed structural and biofunctional investigations. One major hurdle was the solubility of ProcA linear precursor peptides, which dissolve in aqueous buffer only in the presence of guanidine hydrochloride. The \textit{in vitro} reactions of linear precursor ProcA and ProcM usually resulted in precipitation after incubation for more than 6 hours. Downstream high-performance liquid chromatography (HPLC) separation of the reaction supernatant was therefore only able to provide trace amount of purified modified core peptides.

It is known that engineering of lanthipeptides can be accomplished by manipulation of the sequence of the precursor peptides in host strains and heterologous expression systems\textsuperscript{8} as well as with \textit{in vitro} reconstituted systems.\textsuperscript{9} The former methods have the advantage of potential scale-up, whereas the \textit{in vitro} approach allows introduction of nonproteinogenic amino acids. Any improved variants identified by the in vitro method could in principle be produced \textit{in vivo} using amber stop-codon suppression technology.\textsuperscript{10} Unfortunately, at the outset of my studies this methodology is not available for most lanthipeptide expression hosts. To be able to take full advantage of \textit{in vivo} introduction of nonproteinogenic amino acids, production of lanthipeptides in \textit{E. coli} would be highly
desirable, but only a single such example has been reported to date for a truncated analog of the class II lantibiotic nukacin ISK-1.\textsuperscript{11} In order to try to improve the production of prochlorosins, a single vector co-expression method was adopted using the \textit{pRSFDuet-1} vector. The genes encoding a series of ProcA peptides fused to an N-terminal His\textsubscript{6}-tag were cloned into multi-cloning site 1 (MCS1) of \textit{pRSFDuet-1}, while the gene encoding ProcM without His\textsubscript{6}-tag was cloned into multi-cloning site 2 (MCS2). Each of the two genes was expressed under the control of the T7 promoter. \textit{E. coli} BL21 (DE3) cells were transformed with the \textit{procA/procM/pRSFDuet-1} plasmid, and the transformants were cultured under conditions previously used for ProcM overexpression. Surprisingly, upon harvest and lysis of the cells, the peptides were mostly present in the soluble portion unlike the unmodified ProcA’s that are typically expressed in inclusion bodies. After purification by immobilized metal affinity chromatography (IMAC), the peptides were isolated in yields ranging from 10 to 35 mg per liter of culture. Remarkably, analysis of the products by MALDI-ToF MS showed that they had been fully modified by ProcM in \textit{E. coli}.

In principle, the proteolytic removal of the leader peptides could be executed \textit{in vivo} by introducing the genes encoding the ABC-transporter. However, in order to avoid the complications often encountered with heterologous expression of proteases and potential toxic effects of the final post-translationally modified products, artificial protease cleavage sites were introduced into the C-termini of ProcA leader peptides to allow for complete removal of the leader peptides \textit{in vitro}. As an example, the glycine residue at
the C-terminus of the leader peptide of ProcA1.7 was substituted with an arginine affording ProcA1.7-G–1R (negative residue numbers are used for amino acids in the leader peptide counting backwards from the leader peptide cleavage site, while positive residue numbers are used for amino acids in the core peptide counting forward from the leader peptide cleavage site), allowing trypsin removal of the leader peptide. Similarly, procA2.11-G–1K, procA3.2TEV, and procA3.3-G–1K were also constructed and cloned into MCS1 of a pRSFDuet-1 vector with procM in MCS2.

After overexpression and purification by IMAC, the modified precursor peptides (for their wild type sequences, see Figure 2.1) were treated with the appropriate commercial proteases (trypsin for ProcA1.7-G–1R, Lys-C for ProcA 2.11-G–1K and 3.3-G–1K, and TEV protease for ProcA3.2TEV). Dehydration of the peptides was confirmed by MALDI-ToF MS (Figure 2.5). Fragmentation patterns shown by tandem mass spectrometry confirmed the same ring topology of the prochlorosins modified by ProcM in E. coli as that of prochlorosins obtained in vitro and produced in vivo from Prochlorococcus MIT9313 (Figure 2.6, the modified peptides are denoted as “Pcn in E. coli”).
To avoid unwanted cleavage within the modified core peptides, the protease cleavage sites described in the previous paragraph were chosen such that they were not present in the core peptide sequences. This consideration limited the pool of cleavage site candidates. However, even when a suitable artificial cleavage site could be identified, complete leader peptide removal was often not successful as will be discussed below. For example, when Arg-C was used in attempts to remove the leader peptide of

**Figure 2.5** MALDI-MS spectra of precursor peptides modified by ProcM in *E. coli* and treated with a protease to remove the leader peptide. A) ProcA1.7-G–1R treated by trypsin, B) ProcA2.11-G–1K treated by Lys-C, C) ProcA3.3-G–1K treated by Lys-C, and D) ProcA3.2TEV_Phe26pBpa (F26B) treated by TEV protease. Unmodified peptides are denoted as M in the figure.
ProcA1.7G−1R it failed to cleave C-terminally at the introduced arginine. Trypsin was able to perform the desired cleavage but also cleaved C-terminally at a lysine residue in the ProcA1.7 core peptide, producing significant amounts of side product. In order to avoid this problem, *procA1.7-G−1E/procM/pRSFDuet-1* was constructed. Glu-C was used for leader peptide removal of this mutant in which Gly−1 of ProcA1.7 was substituted with a glutamic acid. However, Glu−6 was preferred by Glu-C over Glu−1 in ProcA1.7-G−1E and resulted in undesired cleavage. Similar results were observed with ProcA3.3-G−1E and ProcA4.3-G−1E.
Figure 2.6 ESI-MSMS of four ProcA peptides processed by ProcM in *E. coli* and cleaved by commercial proteases. A) Pcn1.7. B) Pcn2.11. C) Pcn3.2. D) Pcn3.3. E) Comparison of fragmentation patterns of Pcn2.11 obtained *in vitro*, *in vivo* from *E. coli* and from *Prochlorococcus* cultures. F) Fragmentation patterns of the four prochlorosins analyzed by ESI-MSMS. The N-terminal threonine of the ProcA1.7-G−1R core peptide is converted to dehydrobutyryl by ProcM. This dehydrobutyryl is hydrolyzed to a 2-oxobutyryl (Obr) group after trypsin cleavage at the cleavage site. Dehydroalanine (Dha) and dehydrobutyryl (Dhb) residues are highlighted in green. Red lines indicate the ring forming after ProcM modifications. Purple lines indicate putative ring topology after ProcM modifications.
Besides the previously discussed *procA* mutants inserted in *procM/pRSFDuet-1*, *procA2.1-G−1R*, *2.1-G−1E*, and *4.3-G−1R* were also inserted into MCS1 of the *procM/pRSFDuet-1* vector. For ProcA2.1 mutants, presumably because Cys1 in the core peptide is involved in forming a thioether ring, neither ProcA2.1-G−1R nor ProcA2.1-G−1E was successfully cleaved by the corresponding proteinase to provide the intact modified core peptide. At present, the ring topology of ProcA2.1 core peptide after ProcM modification still needs to be confirmed, by either tandem MS or NMR spectroscopy. For ProcA4.3-G−1R, Arg−15 was the preferential cleavage position instead of Arg−1 (Figure 2.7), thus only small peaks corresponding to the intact modified core peptide were observed.

![MALDI-MS spectrum](image)

**Figure 2.7** MALDI-MS spectrum of ProcA4.3-G−1R precursor peptides modified by ProcM in *E. coli* and treated with trypsin to remove part of the leader peptide. The unmodified peptide is denoted as M in the figure.
peptide were observed after trypsin cleavage, and the intact modified core peptide could not be purified using analytical HPLC.

In addition to the co-expression constructs described above, *procA1.2−1K*, 1.3-*G−1K*, 1.4-*G−1E*, 1.5-*A−1K*, 1.6, 1.6-*G−1R*, 2.8-*G−1K*, 4.1*TEV*, 4.4-*G−1E*, and *s.1−G−1K* were also cloned into MCS1 of *procM/pRSFDuet-1* vector partially with the help of rotation students Caitlin Deane and Manuel Ortega. Peaks of expected intact fully

![Graph](image)

**Figure 2.8** MALDI-MS spectra of several precursor peptides modified by ProcM in *E. coli* and treated with a protease. A) ProcA1.2−A−1K treated by Lys-C, B) ProcA1.3−G−1K treated by Lys-C, C) ProcA1.5−A−1K treated by Lys-C, D) ProcA1.6 treated by Glu-C, E) ProcA2.8−G−1K treated by Lys-C, and F) ProcA4.1TEV treated by TEV protease. Unmodified peptides are denoted as M in the figure.
modified prochlorosins were detected in MALDI spectra after Lys-C cleavage of ProcA1.2-A–1K, 1.3-G–1K and 1.5-A–1K, Glu-C cleavage of ProcA1.6, Lys-C cleavage of ProcA2.8-G–1K, and TEV protease cleavage of ProcA4.1TEV (Figure 2.8).

In summary, the co-expression methodology worked remarkably well to produce tens of milligrams of modified ProcA peptides. However, the removal of the leader peptides proved challenging and required evaluation of various different strategies. Whereas successful solutions were found for some peptides, no generally applicable method was found.
2.2.4 Large scale prochlorosin preparation for bioactivity investigations

Fully modified prochlorosins for which the leader peptide could be successfully removed were purified by analytical reversed phase HPLC using a C18 column. The correct masses of the purified prochlorosins were confirmed by MALDI-MS as shown in

Figure 2.9 MALDI-MS spectra of four prochlorosins modified by ProcM in *E. coli* after commercial protease cleavage and HPLC purification. A) Pcn1.7, B) Pcn2.11, C) Pcn3.2 and D) Pcn3.3.
Figure 2.9 and the purity of these prochlorosins is shown in Figure 2.10. Purified Pcn1.7, 2.11, 3.2, and 3.3, as well as partially dehydrated ProcA1.7, 3.2 and 3.3 core peptides that were also purified by HPLC, have been sent to our collaborator Professor Sallie Chisholm’s group for bioactivity investigations performed by her graduate student Mr. Andres Cubillos. Partially dehydrated core peptides were also sent because the partially dehydrated peptides are also produced by *Prochlorococcus* MIT9313. Whether they play

![Analytical RP-HPLC of purified prochlorosins. A) Pcn 2.11, B) Pcn3.2, and C) Pcn3.3.](image)

**Figure 2.10** Analytical RP-HPLC of purified prochlorosins. A) Pcn 2.11, B) Pcn3.2, and C) Pcn3.3.
any roles in nature and whether they function independently or in combination with the fully dehydrated and cyclized products, is still to be explored.

Linear ProcA3.2 and ProcA2.11 core peptides mutants, with all the Cys residues substituted by Ser residues, were prepared by solid-phase peptide synthesis (SPPS) and sent to MIT as control compounds. The Cys to Ser substitutions were considered in order to exclude the possibility of non-enzymatic disulfide bond formation in complex experimental environments that would interfere with the desired linear control peptides. Other control compounds included modified ProcA1.7, 2.11, 3.2, and 3.3 precursor peptides (with leader peptides still attached), linear ProcA3.2 core peptide generated by TEV cleavage of unmodified ProcA3.2TEV, and representative lantibiotics nisin and epilancin 15X. Upon the addition of prochlorosins, the transcriptome of prochlorosin-producing and non-producing strains will be analyzed in comparison with the controls, as a preliminary investigation of the biological function of prochlorosins.

Furthermore, fully modified ProcA2.8-G−1K and ProcA3.3-G−1K with a C-terminal His$_6$-tag (ProcA2.8-G−1K-His6 and ProcA3.3-G−1K-His6) instead of a His$_6$-tag at the N-terminus were also prepared by in vivo co-expression. After proteolytic cleavage and purification, C-terminally His$_6$-tagged Pcn2.8 and Pcn3.3 were generated and sent to Mr. Cubillos, for antibody conjugation or pull-down experiments to locate prochlorosins in situ and identify their interacting molecules. The samples that have been prepared and sent to Mr. Cubillos are listed in Table 2.1.
2.2.5 Incorporation of a non-canonical amino acid into modified ProcA

To introduce a photoaffinity labeling group to probe prochlorosin targets in vivo, an amber stop codon (TAG) was introduced into procA3.2TEV replacing the codon encoding Phe26. The mutant gene procA3.2TEV_F26B was co-expressed with ProcM from a pRSFDuet-1 vector, with an orthogonal tRNA/tRNA synthetase pair encoded in a pDule vector (obtained from Professor Peter G. Schultz of the Scripps Research Institute). This tRNA/tRNA synthetase pair was constructed to incorporate the photocrosslinking amino acid into modified ProcA3.2TEV.
acid p-benzoyl-L-Phe (pBpa) at the site of the amber stop codon. The nonproteinogenic amino acid pBpa was added to the medium during the overexpression, and the resulting peptide containing pBpa was fully modified by ProcM. The photo-crosslinking group pBpa has been demonstrated as a useful probe for studying protein-protein and protein-DNA interactions in cells. Upon incubation with producing and non-producing Prochlorococcus strains and UV irradiation at < 365 nm, pBpa-containing prochlorosins are expected to crosslink to biomolecules with which they interact. This Pcn3.2 analog with a photocrosslinking amino acid (Figure 2.5D) may provide a valuable tool for determination of prochlorosin biological activity.
Figure 2.11 MALDI-MS spectrum of ProcA3.2TEV with Phe26 substituted by pBpa, modified by ProcM in *E. coli* and treated with GluC to remove the major part of the leader peptide. The remaining residues of the leader peptides after GluC cleavage are highlighted in yellow in the amino acid sequence shown in the figure. “B” representing the incorporation of pBpa is highlighted in grey. Cysteine residues are highlighted in red. Serine and threonine residues are highlighted in green. After modification by ProcM, these serine and threonine residues were dehydrated, which caused a mass shift in the MALDI-MS compared with the calculated mass of unmodified peptide with the same sequence (denoted as M in the figure). A small amount of peptide resulting from translation termination at the amber stop codon is seen around 3100 Da. The ion with three dehydrations in the core peptide was the major product along with a minor amount of product with two dehydrations, as shown in the inset.
2.3 SUMMARY AND OUTLOOK

Prochlorosins are class II lanthipeptides produced by cyanobacteria of the genus *Prochlorococcus*. *Prochlorococcus* MIT9313 encodes one LanM-like synthetase ProcM and 29 LanA-like substrates ProcAs. ProcM was able to modify 16 wild-type ProcAs, with the 13 remaining ProcAs not investigated yet. Thus far, no biological activities have been observed with the prochlorosins. In the work described in Chapter 2, *in vitro* modification of two wild-type ProcAs and several ProcA mutants for proteolytic removal of the leader peptides were performed. Remarkably, a co-expression technique was successful in producing modified ProcAs in *E. coli*. Combined with *in vitro* proteolysis and reversed-phase purification, the *in vivo* co-expression method was able to produce procholorosins in sufficient quantity for downstream bioactivity assays to decipher the biological role of prochlorosins. The work described in this chapter also facilitated the interpretation of the stereochemistry of the lanthionine and methyllanthionine residues in several prochlorosins.15

2.4 MATERIALS AND METHODS

2.4.1 Materials

All oligonucleotides were purchased from Integrated DNA Technologies. Restriction endonucleases, DNA polymerases, and T4 DNA ligase were purchased from New England Biolabs or Invitrogen. Media components for bacterial cultures were purchased from Difco laboratories. Chemicals were purchased from Fisher Scientific or
from Aldrich unless noted otherwise. Endoproteinases Glu-C, Lys-C, Asp-N, and trypsin were purchased from Roche Biosciences or Worthington Biosciences. Factor Xa protease was purchased from New England Biolabs. Tobacco etch virus (TEV) protease was prepared as described previously. E. coli DH5α was used as host for cloning and plasmid propagation, and E. coli BL21 (DE3) was used as a host for co-expression.

2.4.2 General methods

Negative residue numbers are used for amino acids in the leader peptide counting backwards from the leader peptide cleavage site, i.e. the residue immediately N-terminal to the cleavage site is −1. The plasmids, procA1.7-G−1R/pET15b, procA2.11/pET15b, procA3.2/pET15b, procA3.3/pET15b, and procM/pET28b that were used as templates for PCR were reported previously.6 All polymerase chain reactions (PCR) were carried out on a C1000™ thermal cycler (Bio-Rad). DNA sequencing was performed by the Biotechnology Center at the University of Illinois at Urbana-Champaign, using appropriate primers. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-ToF MS) was carried out on a Voyager-DE-STR (Applied Biosystems). Liquid chromatography electrospray ionization tandem mass spectrometry (LC/ESI-MS/MS) was carried out and processed using a Synapt ESI quadrupole ToF Mass Spectrometry System (Waters) equipped with an Acquity Ultra Performance Liquid Chromatography (UPLC) system (Waters). Solid-phase peptide synthesis was performed on a CEM Liberty microwave peptide synthesizer.
2.4.3 Construction of *procA2.4/pET15b* and *procA2.8/pET15b*

The *procA2.4* and *procA2.8* genes were cloned into a *pET15b* vector between the *NdeI* and *XhoI* restriction sites using nested PCR with genomic DNA of *Prochlorococcus* MIT9313 (provided by Prof. Sallie Chisholm from the Massachusetts Institute of Technology) as the template (primer sequences are listed in Table 2.2). As the sequences at the 5’ end of *procA2.1 - procA2.11* are very similar to each other, a primer (outside FP) annealing specifically to a DNA sequence upstream of the target gene was used in the first round PCR for both genes. For the more diverse 3’ ends, regular reverse primers were used. In the second round, a regular primer complementary to the exact 5’ end of ProcA was employed to amplify the target gene from the product of the first round PCR. The first round PCR amplification was performed by thirty cycles of denaturing (98 °C for 10 s), annealing (55.5 °C for 30 s) and extending (72 °C for 30 s).

*ProcA2.4_outside_FP* and *ProcA2.4_XhoI_RP* were used to generate the larger gene sequence containing the entire sequence of *procA2.4*. In the second round PCR, *ProcA2.4_NdeI_FP* and *ProcA2.4_XhoI_RP* were used for generating the sequence encoding ProcA2.4, using the same PCR program as the first round. The PCR mixture included 1× Phusion HF buffer (New England BioLabs), dNTP mixture (0.2 mM each), primers (stock solution 100 μM each, final concentration 2 μM each), DMSO (4%, v/v), and Phusion High-Fidelity DNA Polymerase (0.02 U/μL). Amplifications were confirmed by 1.5% agarose gel electrophoresis. The PCR products were purified by agarose gel electrophoresis followed by using a QIAquick Gel Extraction Kit (QIAGEN). The
resulting DNA fragments and \textit{pET15b} vector were digested using \textit{NdeI} and \textit{XhoI} restriction enzymes in 1× NEBuffer 4 (New England BioLabs), incubated at 37 °C for 12 h, followed by the addition of Calf Intestinal Alkaline Phosphatase (CIP, 200 U/mL, New England BioLabs) to remove the 5’ phosphate group from the digested vector, preventing the self-ligation of the cleaved vector. The digested products were purified by agarose gel electrophoresis followed by using QIAquick Gel Extraction Kit (QIAGEN). The resulting DNA inserts were ligated with the digested \textit{pET15b} vector using T4 DNA ligase (Invitrogen) in ligase buffer. The mixture was incubated at 24 °C for 30 min, followed by 23 °C for 3 h. The ligation mixture was diluted 5 times with water before \textit{E. coli} DH5α cells were transformed with the ligation product via heat shock and plated onto LB-Ampicillin agar plates, followed by incubating the plates at 37 °C overnight. Several colonies on the plates were picked and individually incubated in 5 mL of LB medium at 37 °C overnight. The plasmids were isolated using QIAprep Spin Miniprep Kit (QIAGEN). The correct sequences of the resulting plasmids were confirmed by DNA sequencing. The plasmid \textit{procA2.8/pET15b} was generated the same way from \textit{Prochlorococcus MIT9313} genomic DNA. (Y.S. Notebook 3, Page 28 – 30).

2.4.4 Mutagenesis of ProcA2.1, ProcA2.11 and ProcA3.2

Mutagenesis of \textit{procA2.1}, \textit{procA2.11} and \textit{procA3.2} to install a cleavage site for a commercial protease to remove the leader sequence was performed by Megaprimer PCR (for primer sequences see Table 2.2). In the first round PCR, ProcA2.1\textsubscript{\textit{NdeI}} FP and
ProcA2.1_Xa_for_FP were used for generating a megaprimer, which covered the 5’ end of procA2.1 and sequences encoding IEGR as the 3’ end of the product, for subsequent rounds. Meanwhile, ProcA2.1_XhoI_RP and ProcA2.1_Xa_for_RP were used for generation of a megaprimer covering the 3’ end of procA2.1 and encoding IEGR as the 5’ end of the megaprimer. The resulting products from the two separate PCR reactions were purified by agarose gel electrophoresis followed by using a QIAquick Gel Extraction Kit (QIAGEN). In this first round, the PCR amplification was performed by thirty cycles of denaturing (98 °C for 10 s), annealing (55 °C for 30 s) and extending (72 °C for 15 s). Wild type procA2.1/pET15b was used as the template for the first round PCR, while the annealed and amplified megaprimer products were used for the second round PCR. The same PCR conditions were repeated for seven cycles in the second round PCR amplifications. After the first seven cycles, ProcA2.1_NdeI_FP and ProcA2.1_XhoI_RP were added to the reaction, followed by twenty-two cycles under the same PCR conditions to generate sequences containing appropriate mutations. The PCR mixture included 1× Phusion HF buffer (New England BioLabs), dNTP mixture (0.2 mM each), primers (1 μM each), DMSO (4%, v/v), and Phusion High-Fidelity DNA Polymerase (0.02 U/μL). Amplifications were confirmed by 1.5% agarose gel electrophoresis. The product was purified by agarose gel electrophoresis followed by using a QIAquick Gel Extraction Kit (QIAGEN), and inserted into a pET15b vector after digestion, ligation and transformation as described before. The desired mutations were confirmed by DNA sequencing. In the case of ProcA2.1, residues AAGG were replaced by IEGR to produce
ProcA2.1Xa, while residues SVAGG were replaced by ENLYFQ to afford ProcA2.11TEV. Plasmids encoding ProcA2.11Xa, ProcA2.11TEV, ProcA3.2Xa, and ProcA3.2TEV were constructed the same way as described here by replacing certain residues using IEGR or ENLYFQ. All mutations were verified by DNA sequencing. (Y.S. Notebook 2, Page 74 – 75, 78 – 81, 82 – 93; Notebook 3, Page 15 – 19).

**Table 2.2** Primer sequences for cloning and mutagenesis of *procA*s in *pET15b*.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>procA2.1 NdeI_FP</td>
<td>GCA ATC CAC ATA TGT CAG AAG AAC AAC TCA AAG C</td>
</tr>
<tr>
<td>procA2.1 XhoI_RP</td>
<td>ATA CAA TCC TCG AGT CAG TAG CAG CCT CTC G</td>
</tr>
<tr>
<td>ProcA2.1_Xa_FP</td>
<td>GAG GAG CTG GAG GGC ATT GAA GGA CGT TGC TGC ATC ACC GGG</td>
</tr>
<tr>
<td>ProcA2.1_Xa_RP</td>
<td>CCC GGT GAT GCA GCA ACG TCC TTC AAT GCC CTC CAG CTC CTC</td>
</tr>
<tr>
<td>procA2.11 NdeI_FP</td>
<td>GCA ATC CAC ATA TGT CAG AAG AAC AAC TCA AAG C</td>
</tr>
<tr>
<td>procA2.11 XhoI_RP</td>
<td>ATA CAA TCC TCG AGC TAG CAG GTA CCG GTC</td>
</tr>
<tr>
<td>ProcA2.11_Xa_FP</td>
<td>GAT GAG CTT GAA AGT ATT GAA GGG CGT GGG AGG ATT GAT ACC</td>
</tr>
<tr>
<td>ProcA2.11_Xa_RP</td>
<td>GGT ATC AAT CCT CCC ACG CCC TTC AAT ACT TTC AAG CTC ATC</td>
</tr>
<tr>
<td>ProcA2.11_TEV_FP</td>
<td>GAG GAT GAG CTT GAA AAT CTG TAT TTC CAA GGG AGG ATT GAT ACC</td>
</tr>
<tr>
<td>ProcA2.11_TEV_RP</td>
<td>GGT ATC AAT CCT CCC TTG GAA ATA CAG ATT TTC AAG CTC ATC</td>
</tr>
<tr>
<td>procA3.2 NdeI_FP</td>
<td>GCA ACC TAC ATA TGT CAG AAG AAC AAC TCA AAG C</td>
</tr>
<tr>
<td>procA3.2 XhoI_RP</td>
<td>ATA CAA TCC TCG AGT CAC TGG TGG AAG CAG</td>
</tr>
<tr>
<td>ProcA3.2_Xa_FP</td>
<td>GAT GAG CTG GAA GGT ATT GAA GGG CGT GGG GGA GGC TGT GAC</td>
</tr>
<tr>
<td>ProcA3.2_Xa_RP</td>
<td>GTC ACA GCC TCC CCC ACG CCC TTC AAT ACC TTC CAG CTC ATC</td>
</tr>
<tr>
<td>ProcA3.2_TEV_FP</td>
<td>GAT GAG CTG GAA AAT TTG TAT TTT CAA GGG GGA GGC TGT GAC</td>
</tr>
<tr>
<td>ProcA3.2_TEV_RP</td>
<td>GTC ACA GCC TCC CCC TTG AAA ATA CAA ATT TTC CAG CTC ATC ATC</td>
</tr>
<tr>
<td>ProcA2.4_outside_FP</td>
<td>GCT GGC AGC TGT AAC TGG ATA TGC</td>
</tr>
<tr>
<td>ProcA2.4_NdeI_FP</td>
<td>GTA AGT ACC ATA TGA GCG AAG AAC AAC TCA AAG C</td>
</tr>
<tr>
<td>ProcA2.4_XhoI_RP</td>
<td>AAT CAT AGC TCG TAA TGG TCT AAC AAG CAC</td>
</tr>
<tr>
<td>ProcA2.8_outside_FP</td>
<td>GTT ATC CGA TCT GTG ACT GGA CGT CG</td>
</tr>
<tr>
<td>ProcA2.8_NdeI_FP</td>
<td>GTA AGT ACC ATA TGT CAG AAG AGC AAC TGA AGG C</td>
</tr>
<tr>
<td>ProcA2.8_XhoI_RP</td>
<td>TTA CAA TCC TCG AGT TAG CAC TCA CCC TCC</td>
</tr>
</tbody>
</table>

### 2.4.5 Overexpression and purification of linear precursor peptides of ProcA and ProcA mutants

*E. coli* BL21 (DE3) cells transformed with a *procA/pET15b* vector constructed as
described in section 2.4.3 and 2.4.4 were grown in 2 L of LB medium containing 100 mg/L ampicillin at 37 °C until the O.D._{600 nm} reached about 0.8. The culture was then induced with 1.0 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and shaken continually at 37 °C for an additional 3 h. The induced cells were harvested by centrifugation (11,900 × g for 20 min at 4 °C, Beckman JLA-10.500 rotor). The cell pellet was resuspended in 20 mL of LanA Start Buffer (20 mM NaH$_2$PO$_4$, pH 7.5 at 25 °C, 500 mM NaCl, 0.5 mM imidazole, 20% glycerol) and lysed by sonication (35% amplitude, 4.0 s pulse, 9.9 s pause, 15 min). The sample was centrifuged (23,700 × g for 40 min at 4 °C) and the pellet was resuspended in 20 mL of LanA Start Buffer again to remove possibly remaining soluble proteins in the pellet. After centrifugation again, the cell pellet was resuspended in 20 mL of LanA Buffer 1 (6 M guanidine hydrochloride, 20 mM NaH$_2$PO$_4$, pH 7.5 at 25 °C, 500 mM NaCl, 0.5 mM imidazole). The soluble portion after centrifugation (23,700 × g for 20 min at 4 °C), which contained His$_6$-tagged ProcAs, was clarified using 0.45 μm syringe filters (Corning), then purified by immobilized metal affinity chromatography (IMAC) using a 5 mL HisTrap HP nickel affinity column (GE Healthcare Life Sciences). The filtered sample was applied to the column and the column was washed with 2 column volumes (CV) of LanA Buffer 1, followed by 2 - 3 CV of LanA Buffer 2 (4 M guanidine hydrochloride, 20 mM NaH$_2$PO$_4$, pH 7.5 at 25 °C, 300 mM NaCl, 30 mM imidazole), and then eluted with 3 CV 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 were immediately used or stored at room temperature overnight and
then desalted and purified by preparative reversed-phase high-performance liquid chromatography (RP-HPLC) using a Waters Delta-pak™ C4; 15 μm; 300 Å; 25 × 100 mm PrepPac® Cartridge. Solvents for the RP-HPLC were solvent A (0.1% TFA in water) and solvent B (0.086% TFA in 80% acetonitrile / 20% water). A gradient of 2-100% of solvent B was executed over 45 min at a flow rate of 8 mL/min, while peptides were detected by absorbance at 220 nm. The fractions were tested by MALDI-ToF-MS for the desired peptides. All the fractions containing the desired pure product were combined and the organic solvents were removed by rotary-evaporation, followed by lyophilization overnight. The product was kept at –20 °C for short-term storage and –80 °C for long-term storage. The yields of the peptides were 25.5 mg, 37.1 mg, 145.3 mg, 38 mg, 11.9 mg, 6.7 mg, and 33.2 mg for His₆-ProcA2.4, His₆-ProcA2.8, His₆-ProcA 2.1Xa, His₆-ProcA2.11Xa, His₆-ProcA2.11TEV, His₆-ProcA3.2Xa, and His₆-ProcA3.2TEV from 4 L cell culture, respectively. (Y.S. Notebook 2, Page 94 – 99; Notebook 3, Page 3 – 33, 36 – 38).

### 2.4.6 Overexpression and purification of His₆-tagged ProcM

*E. coli* BL21 (DE3) cells were transformed with procM/pET28b via electroporation. The cells were grown in 2 L of LB medium containing 50 mg/L kanamycin at 37 °C until the O.D.₆₀₀ nm reached about 0.6. Then the incubation temperature was reduced to 18 °C and the culture was induced with 0.1 mM IPTG and shaken aerobically continuously for an additional 20 h. The cells were harvested by centrifugation (5000 × g for 20 min at
4 °C). The cell pellet was resuspended in 20 mL of ProcM Start Buffer (20 mM Tris, pH 8.0 at 4 °C, 1 M NaCl, 10% glycerol) and was disrupted by multiple passes through an Avestin C5 Emulsiflex Homogenizer. The insoluble aggregates and cellular debris were then discarded after centrifugation (23,700 × g for 30 min at 4 °C). The supernatant was passed through 0.45 μm syringe filters (Corning), followed by nickel affinity fast protein liquid chromatography purification (FPLC) using an Amersham Biosciences/GE Healthcare ÄKTA system. A 5 mL HisTrap HP nickel affinity column (GE Healthcare Life Sciences) was connected to the FPLC system for the purification at 4 °C. Solvents for FPLC were solvent A (ProcM Start Buffer) and solvent B (ProcM Elution Buffer: 20 mM Tris, pH 8.0 at 4 °C, 1 M NaCl, 200 mM imidazole). A gradient of 0 - 100% of solvent B over 45 min was executed with a flow rate of 2 mL/min. Proteins were detected by absorbance at 280 nm and collected in 4 mL fractions. The fractions were analyzed for purity by Tris-Glycine SDS-PAGE (4 - 20% acrylamide gradient). All the fractions containing the pure desired protein were combined and concentrated using an Amicon Ultra-15 Centrifugal Filter Unit (30 kDa MW cut off for His₆-ProcM) to about 2.5 mL. Then the buffer of the concentrated protein was exchanged to ProcM Start Buffer using a PD-10 desalting column (GE Healthcare Life Sciences). The concentration of the protein was determined by absorbance at 280 nm, demonstrating protein yields of 8.5 mg for His₆-ProcM from 2 L overexpression culture. The resulting protein sample was stored at –80 °C with the addition of 10% (v/v, final concentration) glycerol. (Y.S. Notebook 3, Page 16 – 20).
2.4.7 Activity assays for ProcM

Activity assays for ProcM with ProcA peptides and ProcA mutants were carried out in a solution containing 50 mM HEPES (pH 7.5), 2.5 mM ATP, 0.5 mM tris(2-carboxyethyl)phosphine (TCEP), and 10 mM MgCl₂. The final concentrations of ProcA peptides and ProcM were estimated to be 75 μM and 9.8 μM, respectively. The reaction mixture was incubated at room temperature overnight (15 - 20 h) unless noted otherwise. The products were analyzed by MALDI-ToF MS. (Y.S. Notebook 3, Page 10 – 15, 21 – 23, 39, 49).

2.4.8 Construction of pRSFDuet-1 derivatives for co-expression of ProcM and ProcA

The procM gene was cloned by PCR from procM/pET28b (see reference⁶) and inserted into multiple cloning site-2 (MCS2) of the pRSFDuet-1 vector between the Ndel and KpnI restriction sites (for all primer sequences, see Table 2.3). The procA1.4-G−1E, 1.6-G−1R, 1.7-G−1E, 2.1-G−1R/G−1E, 2.8-G−1K, 2.11-G−1K, 3.2TEV, 3.3-G−1K, 4.1TEV, 4.3-G−1R/G−1E, 4.4-G−1E, and s.1-G−1K genes were generated using overlap extension PCR with the originally reported pET15b expression plasmids⁶ as templates, while procA2.8-G−1K-His6 (encoding ProcA2.8 with C-terminal His₆-tag instead of N-terminal) and 3.3-G−1K -His6 were also generated using overlap extension PCR but using the previously generated procA2.8-G−1K/procM/pRSFDuet-1 and procA3.3-G−1K/procM/pRSFDuet-1 as the templates, respectively (Y.S. Notebook 4, Page 52-55,
59, and 93 – 96. Caitlin’s Rotation-Notebook 1 – 14, 16 – 19. Manuel’s Rotation-Notebook 18 – 39). The _procA1.2A−1K_, _1.3G−1K_, _1.5A−1K_, _1.6_, _1.7G−1R_ and _3.3G−1E_ genes were cloned by regular PCR (Y.S. Notebook 4 Page 23, 52 – 55, 59, 93 – 96. Caitlin’s Rotation-Notebook 1 – 14, 16 – 19. Manuel’s Rotation-Notebook 18 – 39), because the mutant constructs already existed in _pET15b_ vectors that were constructed by Dr. Li. In the case of _ProcA3.2_, the residues GVAGG at the C-terminus of the leader peptide were replaced with NLYFQ, generating the ENLYFQ cleavage site recognized by TEV protease. Except _procA2.8-G−1K-His6_ and _procA3.3-G−1K-His6_, which were each cloned into the _pRSFDuet-1_ vector between _NcoI_ and _NotI_ restriction sites (Y.S. Notebook 7 Page 6), the mutant _procA_ genes were individually cloned into MCS1 of the _pRSFDuet-1_ vector between the _EcoRI_ and _NotI_ restriction sites, all with the _procM_ gene inserted in MCS2. The sequences of the resulting plasmids were confirmed by DNA sequencing. The amino acid sequences of _ProcA1.7-G−1R_, _ProcA2.11-G−1K_, _ProcA3.2TEV_ and _ProcA3.3-G−1K_ are shown in Figure 2.12.
Table 2.3 Primer sequences for cloning and mutagenesis of procAs for co-expression.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer Sequence (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>ProcA1.2_EcoR_FP_Duet</td>
<td>GGT GAG TGG AAT TCG ATG TCA GAA GAA CAA CTG</td>
</tr>
<tr>
<td>ProcA1.2_Not_I_RP_Duet</td>
<td>TAA ATT ATG CGG CCG CTC ACC GTA AGT C</td>
</tr>
<tr>
<td>ProcA1.3_EcoR_FP_Duet</td>
<td>GGT TAG TGG AAT TCG ATG TCA GAA GAA CAA CTC</td>
</tr>
<tr>
<td>ProcA1.3_Not_I_RP_Duet</td>
<td>TAA ATT ATG CGG CCG CTC ACT CTG AGG T</td>
</tr>
<tr>
<td>ProcA1.4_EcoR_FP_Duet</td>
<td>CGG TGA GTG GAA TTC GAT TGC AGA AGA ACA ACT GAA GGC</td>
</tr>
<tr>
<td>ProcA1.4_Not_I_RP_Duet</td>
<td>GTA TAA ATA AGC GGC CGC CTA GCT CGG ACA GGC AG</td>
</tr>
<tr>
<td>ProcA1.4_G−1E_FP</td>
<td>GAA GGT GTG GCT GGG GAA GGA AGC TCG TAC AGA AAC</td>
</tr>
<tr>
<td>ProcA1.4_G−1E_RP</td>
<td>GGT TCT GTA CGA GCT TCC TGC TCC ACC AGC CAC ACC TTC</td>
</tr>
<tr>
<td>ProcA1.5_EcoR_FP_Duet</td>
<td>GGT GAG TGG AAT TCG ATG TCA GAA GAA CAA CTC</td>
</tr>
<tr>
<td>ProcA1.5_Not_I_RP_Duet</td>
<td>TAA ATT ATG CGG CCG CTC ATC CCT CAC</td>
</tr>
<tr>
<td>ProcA1.6_EcoR_FP_Duet</td>
<td>GGT GAG TGG AAT TCG ATG TCA GAA GAA CAA CTC AAG</td>
</tr>
<tr>
<td>ProcA1.6_Not_I_RP_Duet</td>
<td>AAT AAA TAT GCG GCC GCT CAC AGC CAA CAC</td>
</tr>
<tr>
<td>ProcA1.6_G−1R_FP</td>
<td>GAA GGT GTA GCT GGC CGC AAA TCT ACT AAT GGA TG</td>
</tr>
<tr>
<td>ProcA1.6_G−1R_RP</td>
<td>CAT CCA TTA GTA GTT TTG CGG CCA GCT ACA CCT TC</td>
</tr>
<tr>
<td>ProcA1.7_EcoR_FP_Duet</td>
<td>GGT GCG AGG AAT TCG ATG AAC CAT AGA CAA CTA AAT CTG</td>
</tr>
<tr>
<td>ProcA1.7_Not_I_RP_Duet</td>
<td>ATA ATA TCG CGG CTC AGC ACA TTT TCC</td>
</tr>
<tr>
<td>ProcA1.7_G−1E_FP</td>
<td>GTG TGG CTG GGG AAA CCA TTG GGG GAA C</td>
</tr>
<tr>
<td>ProcA1.7_G−1E_RP</td>
<td>GTC CCC CCA ATG GTT TCC CCA GCC ACA C</td>
</tr>
<tr>
<td>ProcA2.1_EcoR_FP_Duet</td>
<td>GGT GAG TGG AAT TCG ATG TCA GAA GAA CAA CTC AAA GC</td>
</tr>
<tr>
<td>ProcA2.1_Not_I_RP_Duet</td>
<td>ATA ATT TAG CGG CCG CTC AGT AGC AGC CTC</td>
</tr>
<tr>
<td>ProcA2.1_G−1E_FP</td>
<td>GCA GCC GGA CGC TGC TGC ATC</td>
</tr>
<tr>
<td>ProcA2.1_G−1R_RP</td>
<td>GAT GCA GCA GCC TCC GCC TGC</td>
</tr>
<tr>
<td>ProcA2.1_G−1E_FP</td>
<td>GGG GGG TGG AAT TCG ATG TCA GAA GAA CAA CTC AAG</td>
</tr>
<tr>
<td>ProcA2.1_Not_I_RP_Duet</td>
<td>CAT AGG ATG CGG CCG GCT CAC GTG CAT CAC</td>
</tr>
<tr>
<td>ProcA2.1_G−1E_RP</td>
<td>GTC ATG CAG CAC TCT CCG GCT GCG</td>
</tr>
<tr>
<td>ProcA2.2_EcoR_FP_Duet</td>
<td>GCA ACC TAG AAT TCG ATG TCA GAA GAG CAA CTA AAG G</td>
</tr>
<tr>
<td>ProcA2.2_Not_I_RP_Duet</td>
<td>ATA TAA TCG GGC CGC TTA GCA CTC ACC CTC</td>
</tr>
<tr>
<td>ProcA2.2_G−1K_FP</td>
<td>GTG GCT GGG AAA GGC GGC TGC CAT AAC</td>
</tr>
<tr>
<td>ProcA2.2_G−1K_RP</td>
<td>GGT AGG ACA GGC CGC TTT CCC AGC CAC</td>
</tr>
<tr>
<td>ProcA2.2_G−1E_FP</td>
<td>GGT GGG TGG AAT TCG ATG TCA GAA GAA CAA CTC AAA GC</td>
</tr>
<tr>
<td>ProcA2.2_Not_I_RP_Duet</td>
<td>ATA ATT TAG CGG CCG CTC AGC AAC AGG TAC C</td>
</tr>
<tr>
<td>ProcA2.2_G−1K_FP</td>
<td>GGA AGT GTG GCT GGA AAA GGG AGG ATT GAT ACC</td>
</tr>
<tr>
<td>ProcA2.2_G−1K_RP</td>
<td>GGT ATC AAT CCT CCC TTT GGC AGC CAC ACT TTC</td>
</tr>
<tr>
<td>ProcA3.2_EcoR_FP_Duet</td>
<td>GGT TAG TGG AAT TCG ATG TCA GAA GAA CAA CTC AAG GC</td>
</tr>
<tr>
<td>ProcA3.2_Not_I_RP_Duet</td>
<td>ATA ATT TAG CGG CCG CTC CTC ACT GGT GGA AGC</td>
</tr>
<tr>
<td>ProcA3.2_TEV_FP</td>
<td>GAT GAT GAG CTG GAA AAT TTG TAT TTT CAA GGG GGA GGC TGT GAC</td>
</tr>
<tr>
<td>ProcA3.2_TEV_RP</td>
<td>GTC ACA GCC TCC CCC TTT GAA ATA CAA ATT TTC CAG CTC ATC ATC</td>
</tr>
<tr>
<td>ProcA3.3_EcoR_FP_Duet</td>
<td>GGT GAG TGG AAT TCG ATG TCA GAA GAA CAA CTC AAG GC</td>
</tr>
<tr>
<td>ProcA3.3_Not_I_RP_Duet</td>
<td>ATA ATT TAG CGG CCG CTC ATG CGC GGC</td>
</tr>
<tr>
<td>ProcA3.3_G−1K_FP</td>
<td>CCG CTA CGC CAA AAG GGC ACT AGA</td>
</tr>
<tr>
<td>ProcA3.3_G−1K_RP</td>
<td>CGG TAT CCG CTT TGC CGC TAG CGG</td>
</tr>
<tr>
<td>ProcA4.1_EcoR_FP_Duet</td>
<td>GGT GGG TGG AAT TCG ATG TCA GAA GAG CAA CTC</td>
</tr>
<tr>
<td>ProcA4.1_Not_I_RP_Duet</td>
<td>CAT AGG ATG CGG CCG CTC AGG TTT GGT TTT C</td>
</tr>
</tbody>
</table>
2.4.9 Overexpression and purification of His\textsubscript{6}-tagged modified ProcA and ProcA mutants

*E. coli* BL21 (DE3) cells transformed with a *pRSFDuet-1* vector carrying the genes for one of the ProcA peptides as well as *procM* were grown in 2 L of LB medium containing 50 mg/L kanamycin at 37 °C until the O.D.\textsubscript{600 nm} reached 0.6 - 0.8. The culture was then induced with 0.1 mM isopropyl β-D–1-thiogalactopyranoside (IPTG) and

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ProcA4.1_TEV_FP</td>
<td>AAG TGA GCT GGA AAA TCT GTA TTT TCA AGG GGG GGG AG</td>
</tr>
<tr>
<td>ProcA4.1_TEV_RP</td>
<td>CTC CCC CCC CCT GAA ACA GAT TTT CCA GCT CAC TT</td>
</tr>
<tr>
<td>ProcA4.3_EcoR_I_FP_Duet</td>
<td>GGT GAG TGG AAT TCG ATG TCA GAA GAA CAA CAG AG</td>
</tr>
<tr>
<td>ProcA4.3_G_1R_FP</td>
<td>GTG TGG CTG GGC GCA CTG CAT CTG</td>
</tr>
<tr>
<td>ProcA4.3_G_1R_RP</td>
<td>CAG ATG CAG TGC GCC CAG CCA CAC</td>
</tr>
<tr>
<td>ProcA4.3_G_1E_FP</td>
<td>GTG TGG CTG GGG AAA CTA CAT CTG CTG</td>
</tr>
<tr>
<td>ProcA4.3_G_1E_RP</td>
<td>CAC CAG ATG CAG TTT CCC CAG CCA CAC</td>
</tr>
<tr>
<td>ProcA4.4_EcoR_I_FP_Duet</td>
<td>GCA GAC GTA TGC GGC CGC TCA ACA ATA TTT GCT G</td>
</tr>
<tr>
<td>ProcA4.4_G_1E_FP</td>
<td>GAA AGC GCG GCT GGT GAA AGC CTC AAG AGT GGA TG</td>
</tr>
<tr>
<td>ProcA4.4_G_1E_RP</td>
<td>CAT CCA CTC TTT AGC CTT TCA CCA GCC GCG CTT TC</td>
</tr>
<tr>
<td>ProcAs.1_EcoR_1_FP_Duet</td>
<td>GGT GAG TGG AAT TCG ATG CAT TCA TCA GCC</td>
</tr>
<tr>
<td>ProcAs.1_Not_I_RP_Duet</td>
<td>GCA TAA GTT TGC GCC CGC TTA TCC ATG CCG CTA ATG</td>
</tr>
<tr>
<td>ProcAs.1_G_1K_FP</td>
<td>GAA AGC GTC GCT GGG AAA GCC CAA TCG G</td>
</tr>
<tr>
<td>ProcAs.1_G_1K_RP</td>
<td>CCG ATT GGG CTT TGC CAG CGA CGC TTT C</td>
</tr>
<tr>
<td>ProcA2.8_Nco_I_FP\textsuperscript{a}</td>
<td>GAT AAG TGC CAT GGG CAT GTC AGA AGA GCA ACT GAA GGC</td>
</tr>
<tr>
<td>ProcA2.8_His6_Not_I_RP\textsuperscript{a}</td>
<td>ATA TAA TGG GCC CGC TCA ATG ATG ATG ATG GCA CTC ACC CTC</td>
</tr>
<tr>
<td>ProcA3.3_Nco_I_FP\textsuperscript{b}</td>
<td>GAT AAG TGC CAT GGG CAT GTC AGA AGA ACA ACT CAA GGC</td>
</tr>
<tr>
<td>ProcA3.3_His6_Not_I_RP\textsuperscript{b}</td>
<td>ATA TAA TGG GCC CGC CTA ATG ATG ATG ATG ATG ATG ATG TGG GCC GCA CAT TTT G</td>
</tr>
<tr>
<td>ProcM_Nde_I_FP_Duet</td>
<td>GGT TGG TTC ATA TGG AAA GTC CAT CAT CTG GG</td>
</tr>
<tr>
<td>ProcM_Kpn_I_RP_Duet</td>
<td>GTC TCT GGC CTA CTG AAT AAG GTA CCA ATC ATC T</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Primers used for generating *procA2.8-G\_1K-His6*

\textsuperscript{b} Primers used for generating *procA3.3-G\_1K-His6*
shaken continually at 18 °C for an additional 20 h. The induced cells were harvested by centrifugation (11,900 × g for 20 min at 4 °C). The cell pellet was resuspended in 20 mL of LanA Start Buffer (20 mM NaH$_2$PO$_4$, pH 7.5 at 25 °C, 500 mM NaCl, 0.5 mM imidazole, 20% glycerol) and lysed by sonication (35% amplitude, 4.0 s pulse, 9.9 s pause, 15 min). The sample was centrifuged (23,700 × g for 40 min at 4 °C) and the supernatant was kept for further purification. The pellet was resuspended in 20 mL of LanA Start Buffer to obtain all the soluble proteins remaining in the pellet. Insoluble materials were removed from the combined soluble fractions by centrifugation (23,700 × g for 20 min at 4 °C), and the resulting sample was clarified using a 0.45 μm syringe filter (Corning). The His$_6$-tagged peptides were then purified by immobilized metal affinity chromatography (IMAC) using a 5 mL HisTrap™ HP nickel affinity column (GE Healthcare Life Sciences). The filtered sample was applied to the column and the column was washed with 2 column volumes of LanA Buffer 1 (6 M guanidine hydrochloride, 20 mM NaH$_2$PO$_4$, pH 7.5 at 25 °C, 500 mM NaCl, 0.5 mM imidazole), followed by 2 - 3 column volumes of LanA Buffer 2 (4 M guanidine hydrochloride, 20 mM NaH$_2$PO$_4$, 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 were purified by preparative reverse-phase high-performance liquid chromatography (RP-HPLC) using a Waters Delta-pak™ C4 (15 µm; 300 Å; 25 × 100 mm PrepPac® Cartridge) column. Solvents for RP-HPLC were solvent A (0.1% TFA in water) and solvent B (0.086% TFA in 80% acetonitrile / 20% water). A gradient of 2 -

**Figure 2.12** Precursor peptide sequences of some of the lanthipeptides used in this study. The leader peptides are highlighted in yellow. Black arrows indicate the cleavage sites. Mutated amino acids introduced for commercial protease cleavage are in bold font and underlined. Cysteine residues in the core peptides are highlighted in red. Serine and threonine residues in the core peptides are highlighted in green. Red lines indicate the rings formed after ProcM modification. Purple lines indicate putative ring topology after ProcM modification. Tandem mass spectrometry has confirmed the existence of overlapping rings at the C-terminal core peptide regions of Pen2.11 and Pen3.2 obtained both in vivo (Figure 2.6) and in vitro. However, how the two rings overlap is currently not known and one possible ring topology is shown.

The elution fractions were purified by preparative reversed-phase high-performance liquid chromatography (RP-HPLC) using a Waters Delta-pak™ C4 (15 µm; 300 Å; 25 × 100 mm PrepPac® Cartridge) column. Solvents for RP-HPLC were solvent A (0.1% TFA in water) and solvent B (0.086% TFA in 80% acetonitrile / 20% water). A gradient of 2 -

---

**ProcA1.7G-1R**

**His tag:** MKHRQLNLMSEEQLKAFIAKVQADTSLOEQLKVEGADVVAIAKAGFAITTE DLAHQAANSQKNLSDAELEGVAG TGGIVSITETCDLLVGMCL

**ProcA2.11G-1K**

**His tag:** MSEKFKIKAQVQADTSLOEQLKAEGADVVAIAKAGFAITKEDLNHRQTLSEDELE SVAGK GRIDOC PAGGGETEQGTCC

**ProcA3.2TEV**

**His tag:** MSEQLKANIKAQVQADASLQOEQLRTEGADVVAIAKAGFSITTEDLNHRQN LSDDELENLYQ GGGQGDIQRT DQTVADNTIVPSTSHQ

**ProcA3.3G-1K**

**His tag:** MSEKFKIKAQVQGDSSLQOEQLKAEGADVVAIAKAGFTIKQDLNAAS ELSDELEAGSK LTDGIAVHL TACGYGGKLMRA
100% of solvent B was executed over 45 min at a flow rate of 8 mL/min, and peptides were detected by absorbance at 220 nm. The fractions were analyzed by MALDI-ToF-MS. All the fractions containing the desired product were combined and the organic solvents were removed by rotary evaporation, followed by lyophilization overnight. The product was kept at −20 °C for short-term storage and −80 °C for long-term storage. The yields of the peptides were 22.6 mg, 62.0 mg, 57.9 mg, and 61.7 mg, respectively, for modified His6-ProcA1.7-G−1R, His6-ProcA2.11-G−1K, His6-ProcA3.2TEV, and His6-ProcA3.3-G−1K from 2 L of cell culture. Modified His6-ProcA1.2-A−1K, 1.3-G−1K, 1.4-G−1E, 1.5-A−1K, 1.6, 1.6-G−1R, 1.7-G−1E, 2.1-G−1R/G−1E, 2.8-G−1K, 3.3-G−1E, 4.1TEV, 4.3-G−1R/G−1E, 4.4-G−1E, s.1-G−1K, 2.8-G−1K-His6 and 3.3-G−1K-His6 were also overexpressed and purified in the same way. (Y.S. Notebook 4 Page 24 – 28, 60 – 68, 85, 91 – 92, 97 – 98, Notebook 5 Page 8 – 9, 11 – 12, 27, 34, Notebook 7 Page 16 – 17, Notebook 9 Page 32 – 33. Caitlin’s Rotation-Notebook Page 13 and 15. Manuel’s Rotation-Notebook Page 8 – 16).

2.4.10 Protease cleavage

Modified ProcA mutants from either in vivo co-expression or in vitro enzymatic assay were cleaved by a commercial protease (trypsin for ProcA1.6-G−1R, ProcA1.7-G−1R, and 4.3-G−1R, Lys-C for ProcA1.2-A−1K, 1.3-G−1K, 1.5-A−1K, 2.8-G−1K, 2.11-G−1K, 3.3-G−1K, s.1-G−1K, 2.8-G−1K-His6, and 3.3-G−1K-His6, Glu-C for ProcA1.6, 2.8, and 4.4-G−1E, Asp-N for ProcA2.4 and 2.8, TEV protease for
ProcA2.11TEV, 3.2TEV, and 4.1TEV, and Factor Xa for ProcA2.1Xa, 2.11Xa, and 3.2Xa) to remove the leader peptide. Linear ProcA3.2TEV was also cleaved by TEV protease to generate linear un-modified ProcA3.2 core peptide. MALDI-TOF mass spectra of some of the modified peptides treated by proteases are shown in Figure 2.5. (Y.S. Notebook 4 Page 29, 42, 50 – 51, 62, 68 – 70, and 99 – 100, Notebook 5 Page 18, Notebook 9 Page 28 – 32, 35).

2.4.11 Tandem mass spectrometry analysis of prochlorosins and their derivatives

The ring topology of lanthipeptides was deduced from their MS fragmentation pattern. A sample of 10 μL of in vitro ProcM dehydration assay after protease cleavage, or a sample of 5 μL of each protease cleavage reaction mixture starting from in vivo co-expression products was injected to a BEH C8 column (1.7 μm, 1.0 × 100 mm), and the fully modified product was purified by UPLC using a gradient of 3% mobile phase B (0.1% formic acid in methanol) to 97% mobile phase B in mobile phase A (0.1% formic acid in water) over 12 min. Mass spectra were acquired in ESI positive mode in the range of 50-2000 m/z. The capillary voltage was 3500 V, and the cone voltage was 40 V. The other parameters used were as follows: 120 °C source temperature; 300 °C desolation temperature, 150 L/h cone gas flow, and 600 L/h desolation gas flow. A transfer collision energy of 4 V was used for both MS and tandem MS, while the trap collision energy was set to 6 V for MS and a 20–55 V ramp for MSMS depending on the peptide. Glu-fibrinopeptide B (Sigma) was directly infused as the lock mass. The tandem mass
spectra were processed with MaxEnt3 and analyzed by Protein/Peptide Editor in BioLynx 4.1. The software for analyzing both precursor-ions and fragment-ions was set to report any mass within 0.3 amu of the calculated expected values. Tandem mass spectra of the fully modified peptides treated by proteases (ProcA1.7-G−1R core peptide with five dehydrations, ProcA2.11-G−1K core peptide with five dehydrations, ProcA3.2TEV core peptide with three dehydrations, and ProcA3.3-G−1K core peptide with three dehydrations) are shown in Figure 2.6. Fragmentation patterns of these core peptides were consistent with the ring topology of prochlorosins previously obtained in vivo and in vitro. Therefore, they are denoted as “Pcn in E. coli” (Pcn is short for prochlorosin, i.e. prochlorosin 1.7 is denoted as Pcn1.7). (Y.S. Notebook 4 Page 37 – 38, 42 – 43, 47 – 49, 81 – 82, Notebook 5 Page 15 – 17, 19, Notebook 9 Page 29, 34, 36 – 38).

**2.4.12 HPLC purification of prochlorosins modified in E. coli**

The protease cleavage reactions were quenched with 0.5% TFA, and the desired products purified by analytical RP-HPLC using a Vydac™ C18 column (5 µm; 300 Å; 4.6 i.d. × 250 mm). Solvents for RP-HPLC were solvent A (0.1% TFA in water) and solvent B (0.086% TFA in 80% acetonitrile / 20% water). A gradient of 0.8-100% of solvent B was executed over 50 min at a flow rate of 1 mL/min, and peptides were detected by absorbance at 220 nm. The fractions were tested by MALDI-ToF-MS for the desired peptides. All the fractions containing the desired fully modified product were combined and the solvents were removed by lyophilization overnight. The product was
kept at −20 °C for short-term storage and −80 °C for long-term storage. MALDI spectra of these purified prochlorosins are shown in Figure 2.9. (Y.S. Notebook 4 Page 34 – 36, 56 – 58, 75 – 78, 83 – 85, and 86 – 90).

2.4.13 Solid phase peptide synthesis (SPPS) of linear ProcA2.11 and ProcA3.2 core peptide derivatives

Linear ProcA2.11 and ProcA3.2 core peptides, with all the cysteine residues substituted by serine residues, were synthesized on a 0.1 mmol scale using standard Fmoc solid phase peptide synthesis chemistry. As the starting material, 0.1 mmol Wang-resin preloaded with the C-terminal residue was swollen in 10 mL of 1:1 DCM:DMF. Fmoc-amino acids were used at a scale of 0.2 M in 2.5 mL DMF for each coupling step. The coupling was performed with HCTU (0.5 M in DMF, 1 mL) as the coupling reagent and DIEA (2.0 M in 1-methyl-2-pyrrolidinone, 0.5 mL) as the activating reagent with microwave irradiation (3 min, 20 W, 75 °C). The Fmoc deprotection was performed in DMF containing 20% (v/v) piperidine assisted by microwave irradiation (3 min, 30 W, 75 °C). (Y.S. Notebook 7, Page 85 – 86).

2.4.14 Deprotection of solid phase synthesized products

For a 0.1 mmol scale peptide synthesis, 4.6 mL TFA, 0.125 mL of water, 0.125 mL of TIPS, and 0.125 mL of thioanisole were added to the dried resin-bound peptide. The mixture was blanketeted with nitrogen and stirred vigorously at room temperature for 2 h,
followed by passing through a coarse-grain Büchner funnel. The flow through was collected and the TFA was evaporated with a stream of nitrogen until only about 1 mL of solution was left, followed by the addition of cold diethyl ether. The peptide was precipitated and pelleted from the solution by centrifugation (6,000 rpm in a GA-10 rotor using a Beckman GS-6R centrifuge for 10 min at room temperature). The pellet was then dried under a nitrogen stream and redissolved in 3 mL of 1:1 (v/v) MeCN/H$_2$O. Then, 2 μL of the solution was used for analysis by MALDI-ToF MS. The remaining solution was frozen and lyophilized overnight. (Y.S. Notebook 7, Page 87).

### 2.4.15 Purification of deprotected SPPS products

The crude lyophilized product was redissolved in a minimal amount of 1:1 (v/v) MeCN/H$_2$O and purified by preparative reversed-phase high-performance liquid chromatography (RP-HPLC) using a Waters Delta-pak™ C18 column (2.5 cm × 10.0 cm). Solvents were solvent A (0.1% TFA in water) and solvent B (0.086% TFA in 80% acetonitrile / 20% water). A gradient of 2-100% of solvent B was executed over 45 min at a flow rate of 8 mL/min, and peptides were detected by absorbance at 220 nm. The fractions were analyzed by MALDI-ToF-MS for the desired peptides and all the fractions containing the desired pure product were combined and concentrated by rotary-evaporation to remove acetonitrile and TFA, followed by lyophilization overnight. The product was kept at –20 °C for short-term storage and –80 °C for long-term storage. (Y.S. Notebook 7, Page 88).
2.4.16 Incorporation of para-benzoyl phenylalanine (pBpa or B) in ProcA3.2

The nucleotides encoding phenylalanine at position 26 of the gene for ProcA3.2TEV were substituted by the amber codon TAG. The mutant was generated by nested PCR using primers 5'-GCA GAC AAT ACA ATT GTC CCT TGC AGC TGC TAG CAC CAG TG -3' (3.2_F26B_FP) and 5'-ATA ATT TAG CGG CCG CTC ACT GGT GCT AGC AG -3' (3.2_F26B_RP). The vector procA3.2TEV/procM/pRSFDuet-1 was used as the template. (Y.S. Notebook 5 Page 31 – 33 and 35).

E. coli BL21 (DE3) cells were transformed with a pDule-Tyr vector (obtained from Professor Peter Schultz, the Scripps Research Institute. Basic information about the vector: it encodes pBpaRS and tRNA_{CUA}, tetracycline resistant, has a p15A origin of replication, and uses the constitutive lpp promoter) and a pRSFDuet-1 vector, carrying both procA3.2TEV_F26B and procM. The cells were grown in 400 mL of glycerol minimal medium, supplemented with metals, containing 50 mg/L kanamycin, 12.5 mg/L tetracycline and an 18-amino-acid solution (for components, see Reference 17 and 18). The pDule-Tyr vector carrying an orthogonal aminoacyl tRNA synthetase-tRNA_{CUA} pair can incorporate p-benzoyl-L-phenylalanine (pBpa) at the position encoded by the amber codon TAG. The culture was shaken at 37 °C until the O.D._{600nm} reached 0.5, then pBpa dissolved in 1 M NaOH solution was added to the culture to obtain a final concentration of 1 mM pBpa. After shaking for 0.5 h, the culture was induced with 0.1 mM isopropyl β-D-1 thiogalactopyranoside (IPTG) and shaken continually at 18 °C for an additional 20 h. The induced cells were harvested by centrifugation (11,900 × g for 20 min at 4 °C).
cell pellet was resuspended in 10 mL of LanA Start Buffer and lysed by sonication (35% amplitude, 4.0 s pulse, 9.9 s pause, 15 min). The sample was centrifuged (23,700 × g for 40 min at 4 °C) and the supernatant was kept for further purification. The pellet was resuspended in 10 mL of LanA Start Buffer again. The soluble portions after centrifugation (23,700 × g for 20 min at 4 °C) were combined and clarified using 0.45 μm syringe filters (Corning), then purified by immobilized metal affinity chromatography (IMAC) using a 5 mL HisTrap™ HP nickel affinity column (GE Healthcare Life Sciences). The filtered sample was applied to the column and the column was washed with 2 column volumes of LanA Buffer 1, followed by 2 - 3 column volumes of LanA Buffer 2, and then eluted by 3 column volumes of LanA Elution Buffer. The elution fractions were desalted and purified by solid phase extraction (SPE) using a 3 mL Vydc® BioSelect™ reversed-phase C4 column. A step gradient of 2 column volumes (CV) of 2% solvent B (0.086% TFA in 80% acetonitrile / 20% water) in solvent A (0.1% TFA in water), 2 CV of 50% solvent B in solvent A, and 2 CV of 100% solvent B was applied for the purification. The fractions were analyzed by MALDI-ToF-MS for the desired peptides. The majority of ProcA3.2TEV with Phe26 substituted with pBpa eluted in fractions of 50% solvent B in solvent A. (Y.S. Notebook 5 Page 36 – 47).

An 18 μL sample that eluted at 50% solvent B was incubated with 0.03 U endoproteinase GluC in 50 mM tris buffer at room temperature for 6 h. The mixture was quenched with 0.5% (final concentration) TFA, desalted by Zip-Tip and subjected to MALDI-ToF-MS (for the spectrum, see Figure 2.11). (Y.S. Notebook 5 Page 46 – 47).
**Table 2.4** Plasmid vectors used in Chapter 2.

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>pET15b</td>
<td><em>amp</em> for ampicillin resistance, pBR322 origin</td>
</tr>
<tr>
<td>pRSFDuet-1</td>
<td><em>kan</em> for ampicillin resistance, RSF origin</td>
</tr>
<tr>
<td>pCDFDuet-1</td>
<td><em>aadA</em> for spectinomycin resistance, CDF origin</td>
</tr>
</tbody>
</table>
**Table 2.5** Summary of the constructs and purified peptides mentioned in Chapter 2.

<table>
<thead>
<tr>
<th>ProcA</th>
<th>Ring topology</th>
<th>In producing strain</th>
<th>Transcribed</th>
<th>In pET15b</th>
<th>In procM/pRSFDuet-I</th>
<th>Pcn</th>
<th>purified</th>
</tr>
</thead>
<tbody>
<tr>
<td>ProcA1.1</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y(G−1E)</td>
<td></td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>ProcA1.2</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y(A−1K)</td>
<td></td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>ProcA1.3</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y(G−1K)</td>
<td></td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>ProcA1.4</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y(G−1E)</td>
<td></td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>ProcA1.5</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y(A−1K)</td>
<td></td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>ProcA1.6</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y(WT/G−1R)</td>
<td></td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>ProcA1.7</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y(WT/G−1R)</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>ProcA2.1</td>
<td>Y</td>
<td>Y</td>
<td>Y(WT/Xa)</td>
<td>Y(G−1R/G−1E)</td>
<td></td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>ProcA2.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ProcA2.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ProcA2.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ProcA2.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ProcA2.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ProcA2.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ProcA2.8</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y(WT/G−1K/G−1K-His6)</td>
<td>Y(WT/G−1K/G−1K-His6)</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>ProcA2.11</td>
<td>Y</td>
<td>Y</td>
<td>Y(WT/Xa/TEV)</td>
<td>Y(G−1K)</td>
<td></td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>ProcA3.2</td>
<td>Y</td>
<td>Y</td>
<td>Y(WT/Xa/TEV)</td>
<td>Y(TEV)</td>
<td></td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>ProcA3.3</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y(G−1K/G−1E/G−1K-His6)</td>
<td>Y(G−1K/G−1E/G−1K-His6)</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>ProcA4.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ProcA4.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ProcA4.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ProcAs.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a: Ring topology confirmed by tandem mass spectrometry and NMR spectroscopy. Preparation and ring topology elucidation by tandem mass spectrometry for ProcA2.8 was described in this chapter.

b: Products detected in the supernatant of *Prochlorococcus* MIT9313 by tandem mass spectrometry as reported.

c: mRNA detected in *Prochlorococcus* MIT9313 by transcription analysis as reported.

d: WT represents “wild type”. Y without further explanations in brackets indicates that the “wild type” peptide was used.

e. The genes encoding ProcAs inserted in the MCS-1 of procM/pRSFDuet-1

f. Prochlorosins prepared by co-expression in *E. coli*.

g. Purified ProcM-modified ProcA peptides generated by co-expression in *E. coli.*
2.5 REFERENCES


NukM in *Escherichia coli*.


CHAPTER 3: HETEROLOGOUS PRODUCTION OF THE LANTIBIOTIC ALA(0)ACTAGARDINE IN ESCHERICHIA COLI

3.1 INTRODUCTION

The majority of lanthipeptides possess antimicrobial activities, and they are called lantibiotics. Nisin has been used as a food preservative for more than fifty years without significant antibiotic resistance. Therefore, lanthipeptides are a class of promising compounds as potential medicinal antibiotics, given that the rising antibiotics resistance problem requires the adoption of new scaffolds for new antibiotics. However, most elongated lanthipeptides, such as nisin, that have well-known antimicrobial activities ex vivo, degrade too fast in the intestine or blood to work as potential drugs. In contrast, globular lanthipeptides with more rigid conformations have higher resistance to protease degradation and thus are expected to have longer life time in the serum, making them more promising as drug candidates. Thus far, two lantibiotic derivatives, NAI-107 and NVB302 have been evaluated in preclinical studies. NAI-107 is also known as microbisporicin, while NVB302 is derived from actagardine (Figure 3.1). In fact, NVB302 has successfully completed a Phase II clinical trial recently.

1Parts of this chapter have been reproduced with permission from: "Heterologous production of the lantibiotic Ala(0)actagardine in Escherichia coli." Chem. Commun. 2012, 48, 10966-10968, published by The Royal Society of Chemistry.
Actagardine is a 19-amino-acid lahtipeptide produced by an actinomycete, *Actinoplanes garbadinensis*. It possesses a globular conformation with an N-terminal lanthionine ring and three C-terminal intertwined methyllanthionine rings (Figure 3.1). Actagardine possesses potent antimicrobial activity against gram-positive pathogens, and is most effective against streptococci and obligate anaerobes. The discovery of the compound and preliminary investigation of its antimicrobial activity was reported in 1976 and 1977, when it was named gardimycin. Twenty years later, the two- and
three-dimensional structure of actagardine in solution was determined in a water-acetonitrile environment.\textsuperscript{9,10} It has been suggested that actagardine, similarly to its structural relative mersacidin, binds to lipid II and blocks transglycosylation thus inhibiting cell wall biosynthesis. Its activity is not antagonized by vancomycin, which is known to bind to the C-terminal D-Ala-D-Ala of the pentapeptide of lipid II during cell wall synthesis, suggesting that actagardine interacts with the disaccharide-pyrophosphate moiety of lipid II.\textsuperscript{11}

Actagardine and its derivatives recently entered clinical trials at Novacta Therapeutics due to its selectivity against \textit{Clostridium difficile}, with reduced activity against other bacteria in the gut flora, especially anaerobic gram-positive cocci.\textsuperscript{4} Actagardine was also shown to be stable in simulated gastric fluid. Therefore, it is evaluated as a drug candidate treating \textit{C. difficile} associated diarrhea.\textsuperscript{4}

The biosynthetic gene cluster of actagardine in \textit{Actinoplanes garbadinensis} ATCC31049 was reported in 2009 (Figure 3.2).\textsuperscript{12} In the cluster, \textit{garA} encodes the precursor peptide, and \textit{garM} encodes a LanM protein that performs dehydration and cyclization on the core peptide. Also present in the cluster is a gene encoding a novel luciferase-like monooxygenase \textit{GarO} that is believed to catalyze the formation of the sulfoxide group in actagardine, as postulated based on gene knockout studies.\textsuperscript{12} Heterologous expression of all genes in the biosynthetic gene cluster of actagardine in \textit{Streptomyces lividans} generated Ala(0)-actagardine, which was also reported as an antimicrobial compound,\textsuperscript{13} suggesting the protease responsible for the removal of the alanine residue N-terminally to actagardine is located outside the biosynthetic gene cluster of actagardine.\textsuperscript{12}
Deoxyactagardine B (DAB) is a sequence and structural analogue of actagardine, differing only by two residues (V15L and I16V in DAB compared to actagardine, Figure 3.1). The biosynthetic gene cluster of DAB in *Actinoplanes liguriae* NCIMB41362 was reported in 2010 (Figure 3.2). Although a luciferase-like monooxygenase (LigO) was also present in the DAB biosynthetic gene cluster, actagardine B with a sulfoxide group was not detected from *A. liguriae* NCIMB41362 culture. NVB302, a derivative of DAB with a diaminoheptane tail attached semi-synthetically to the C-terminus of the lanthipeptide structure, has entered clinical trials for treating *C. difficile* associated diarrhea.

To the best of my knowledge, current production of actagardine derivatives follows semisynthetic routes starting from the purification of actagardine or DAB from *A. garbadinensis* or *A. liguriae*, followed by the chemical addition of the C-terminal hydrophilic tail. In contrast, heterologous biosynthesis of actagardine in *Escherichia coli*...
has not been reported. Sequence-specific engineering is often more facile for heterologously produced compounds compared with analogue generation from the native producing strain. As my previous work described in Chapter 2 had shown the success of modifying lanthipeptides directly in *E. coli* by co-expression of the substrate and corresponding modification enzymes,\(^\text{15}\) I applied co-expression to the heterologous production of actagardine and its derivatives, followed by the affinity purification of modified precursor peptides and proteolytic removal of the leader peptides. In this way, I aimed to develop an alternative and more convenient route to produce actagardine and its derivatives, with advantages in the ease of purification and sequence engineering of substrates, while eliminating the need for organic synthesis.

Actagardine is the only known lanthipeptide that possesses a sulfoxide group in its structure.\(^\text{3}\) It was suggested by gene knock-out studies that the sulfoxide group was introduced enzymatically by GarO, the only known luciferase-like monooxygenase in all published lanthipeptide biosynthetic gene clusters.\(^\text{12}\) However, GarO had never been reconstituted *in vitro* before the work described in this Chapter. Thus, investigation on the formation of the sulfoxide group was envisioned to expand current knowledge on lanthipeptide structures and lanthipeptide biosynthetic enzymes.

### 3.2 RESULTS AND DISCUSSION

#### 3.2.1 Expression vector construction for GarA, GarM and GarO

*Actinoplanes garbadinensis* ATCC 31049 was grown in GYM and LB liquid media and on GYM or oatmeal agar plates. Genomic DNA of *Actinoplanes garbadinensis* ATCC 31049 was extracted from GYM and LB liquid cultures. The genes
encoding GarA and GarO were cloned using a pair of primers (Table 4.1). In contrast, 
garM was cloned using stepwise overlap extension PCR using Phusion polymerase.

The garA gene was inserted into multiple cloning site-1 (MCS1) of pRSFDuet-1 such that the GarA product contained an N-terminal His\textsubscript{6}-tag, while garM was inserted into MCS2 of pRSFDuet-1. Several garA derivatives substituting the garA in MCS1 in 
garM/pRSFDuet-1 were also constructed, including garA-A−1K to introduce a Lys-C cleavage site C-terminally to the leader peptide, garA-A−1L/A−2L to introduce two leucine residues for proteolytic removal of the leader peptide with Leu aminopeptidase, 
garA-IEGR (in which the nucleotides encoding the four C-terminal residues of GarA leader peptide were substituted with codons encoding IEGR residues) to allow Factor Xa cleavage, garA-19insK to add a lysine residue C-terminally to the core peptide in order to increase the solubility of actagardine, garA-V15L/I16V to mimic the sequence of DAB, and garA-V15L/I16V/19insK to mimic NVB302. Negative residue numbers are used for amino acids in the leader peptide counting backwards from the leader peptide cleavage site, while positive residue numbers are used for amino acids in the core peptide counting forward from the leader peptide cleavage site.

Two copies of the garO gene were inserted into MCS1 and MCS2 of pCDFDuet-1, generating GarO without a His\textsubscript{6}-tag. Vectors with one garO only in MCS1 of pCDFDuet-1 that did not encode a His\textsubscript{6}-tag, were also generated. A pRSFDuet-1 was also prepared with garA in MCS1, garM in MCS2, and garO (no attached His\textsubscript{6}-tag gene) with its own T7 promoter, ribosomal binding site, and T7 terminator. The garO gene was also inserted in pET28b producing GarO with an N-terminal His\textsubscript{6}-tag for \textit{in vitro} protein reconstitution. In order to perform protein purification for crystallization attempts, the garM gene was
also inserted in pET28b and MCS1 of pRSFDuet-1 to produce GarM with an N-terminal His$_6$-tag.

3.2.2 Modification of GarA and GarA derivatives by GarM in *E. coli*

Co-expression of GarM and His$_6$-tagged GarA using the pRSFDuet-1 vector successfully generated four-fold dehydrated GarA. The peptide was purified by immobilized metal affinity chromatography (IMAC; Ni$^{2+}$) and reversed-phase high-performance liquid chromatography (RP-HPLC). After lyophilization, 4.2 mg modified GarA was obtained from 1 L of culture. Proteolytic removal of the majority of the leader peptide with trypsin (cleavage after Arg$^{-5}$ of the leader peptide) followed by treatment with aminopeptidase A8200 to remove residues $-4$ to $-2$, generated a product consistent with Ala(0)-deoxyactagardine (Figure 3.3A). After incubation with 2-iodoacetamide (IAA), no IAA adducts were detected by matrix-assisted laser desorption ionization (MALDI) mass spectrometry (MS) (Figure 3.4), suggesting that free cysteines were not present in the core peptide.
Ala(0)-deoxyactagardine (Ala(0)-DA) thus produced displayed low solubility. NVB302 that is in clinical development contains a C-terminal 1,7-diaminoheptane tail, which increases solubility and bioactivity compared with deoxyactagardine. In order to produce a structure similar to NVB302, we substituted Val15 and Ile16 in GarA with Leu and Val, respectively, and introduced a lysine residue at its C-terminus, forming GarA-V15L/I16V/19insK (Figure 3.1) to similarly increase the solubility of Ala(0)-DA produced in E. coli. The dehydration and cyclization of this mutant GarA by GarM was not affected by these mutations or by the addition of the C-terminal lysine as confirmed.

![MALDI-MS spectra](image)

**Figure 3.3** MALDI-MS spectra of GarA and GarA-V15L/I16V/19insK modified by GarM in *E. coli* and treated with trypsin and aminopeptidase A8200 to remove the leader peptide. (A) Spectrum of Ala(0)-deoxyactagardine, [M + H]^+ calculated m/z: 1945.34, observed m/z: 1945.68; (B) Spectrum of Ala(0)-DAB-Lys, [M + H]^+ calc. m/z: 2073.51, obs. m/z: 2074.01.
by MALDI MS after proteolytic cleavage with trypsin and aminopeptidase (Figure 3.3B). The compound produced was termed Ala(0)-DAB-Lys (Figure 3.1).

In order to completely remove the leader peptide, such that Ala(0) was no longer attached to the deoxyactagardine core peptide, the peptides GarA-A−1K, GarA-A−1L/A−2L, and GarA-IEGR were co-expressed with GarM using pRSF-Duet-1. Peptides were purified by Ni-NTA and C4 RP-HPLC. All of these precursor peptides were successfully generated with four dehydrations, although GarA-A−1K was obtained in very low yield. After treatment with trypsin followed by the addition of aminopeptidase A8200, which worked better than leucine aminopeptidase in removing the leader peptide in the case of GarA-A−1L/A−2L, deoxyactagardine was generated, but

**Figure 3.4** MALDI-MS spectra of GarM-modified GarA incubated with IAA (blue) and without IAA (red) and treated with trypsin. M indicates the fragment of GarM-modified GarA after trypsin cleavage (sequence: TIYA-DAB, [M − 4 H2O + H]+ calc. m/z: 2393.08, obs. m/z: 2393.00)
the product with Leu(0) attached was still the major product. Factor Xa was not able to cleave C-terminally at the arginine residue of IEGR in GarA-IEGR. Instead, it preferentially cleaved C-terminally to Arg(−6), generating TIEGR-deoxyactagardine. The proteolysis reactions described here were all complete with no starting material peak observed in the final reaction mixtures. Given the difficulties to produce DAB I focused my efforts on Ala(0)-DAB and its derivatives.

Figure 3.5 ESI-MS/MS of GarA-V15L/I16V/19insK modified in E. coli by GarM and GarO and treated in vitro with trypsin and aminopeptidase A8200.

### 3.2.3 Ring topology confirmation of modifications on GarA derivatives

To confirm that the Ala(0)-DAB-Lys product had the correct ring topology, it was analyzed by tandem MS indicating a globular structure as no fragments were generated from the C-terminal fourteen amino acid region (Figure 3.5). Furthermore, Ala(0)-DAB-Lys displayed antimicrobial activity against *Bacillus subtilis* W23, *Microccocus luteus*
4698 (Figure 3.6), and Enterococcus faecalis 29212. The MIC of Ala(0)-DAB-Lys against E. faecalis was 7.3 μM compared to 4.2 μM reported for deoxyactagardine and actagardine. Thus, the addition of a C-terminal lysine largely maintained the antimicrobial activity of Ala(0)-deoxyactagardine, while significantly increasing solubility.

**Figure 3.6** Antimicrobial assays of actagardine derivatives. (A) Antimicrobial assays of Ala(0)-DAB-Lys modified in E. coli. (1) Nisin, (2) Ala(0)-DAB-Lys from GarA/GarM coexpression in E. coli, (3) negative control containing buffer and proteinases. The applied samples in (1) and (2) consisted of 20 μL of 100 μM stock solutions. (B) Antimicrobial activity of Ala(0)-deoxyactagardine and Ala(0)-deoxyactagardine-Lys generated in E. coli against M. luteus ATCC 4698. Compounds were produced by incubating His6-GarA and His6-GarA-19insK with trypsin and aminopeptidase A8200 to remove the leader peptide. Compounds spotted in the wells were: (1) Ala(0)-deoxyactagardine from GarA, final concentration (f.c.) 1.87 mg/mL, (2) Ala(0)-deoxyactagardine-Lys from GarA-19insK, f.c. 1.96 mg/mL (40 μL was spotted in each well).

### 3.2.4 Modification of GarA and GarA derivatives by GarM and GarO in E. coli

Deoxyactagardine is less active than actagardine against some bacterial strains with a two-fold increased MIC for inhibition of the growth of M. luteus 4698. In an attempt to produce sulfoxide-containing analogs, co-expression with the monooxygenase GarO was pursued. Co-expression of garA/garM/pRSFDuet-1 and garO/pCDFDuet-1 only
produced GarM-modified GarA without oxidation. Therefore, two systems were further evaluated in which either the *garA/garM/pRSFDuet-1* plasmid was used in combination with a *pCDFDuet-1* vector containing two copies of *garO* (*2garO/pCDFDuet-1*), or the *garO* gene was inserted into the *garA/garM/pRSFDuet-1* plasmid (see Experimental section for details). His-tagged GarA was purified by IMAC followed by C8 solid phase extraction (SPE). The purified product was incubated with trypsin and aminopeptidase to remove the leader peptide and the product was analyzed by MALDI-MS. For both approaches, the mass of the resulting four-fold dehydrated core peptide was increased by

![Figure 3.7 MALDI-MS spectra of GarA modified by GarM and GarO.](image_url)

(A) Spectrum of GarA from co-expression of *garA/garM/pRSFDuet-1* with *2garO/pCDFDuet-1* in *E. coli*, followed by trypsin digestion. M indicates GarM-modified GarA treated with trypsin, [M + O + Na]^+ calc. m/z: 2431.06, obs. m/z: 2431.45; (B) product of incubation of GarA-V15L/I16V/I19insK with GarO in vitro and subsequent trypsin digest; [M + O + Na]^+ calc. m/z: 2559.23, obs. m/z: 2559.55; (C) product of incubation of purified Ala(0)-DAB-Lys with GarO in vitro, M indicates Ala(0)-DAB-Lys, [M + O + Na]^+ calc. m/z: 2111.49, obs. m/z: 2111.77.
16 Da (Figure 3.7A and Figure 3.8), consistent with the formation of one sulfoxide group. Unfortunately, when the same experiments were performed with GarA-V15L/I16V/19insK, the GarO-mediated oxidation was incomplete (Figure 3.9), suggesting that the additional C-terminal Lys interferes with efficient oxidation in E. coli.

![MALDI-MS spectrum](image)

**Figure 3.8** MALDI-MS spectrum of GarA modified by GarO from co-expression of *garA/garM/garO/pRSFDuet-1*. M indicates GarM-modified GarA with most of the leader peptide removed by trypsin (sequence: TIYA-deoxyactagardine, [M + O + Na]⁺
calc. *m/z*: 2431.06, obs. *m/z*: 2431.34; [M + O + K]⁺ calc. *m/z*: 2447.17, obs. *m/z*: 2447.38)

### 3.2.5 In vitro reconstitution of GarO

Actagardine is the only lantibiotic containing a sulfoxide group. To investigate the GarO enzymatic activity *in vitro*, *garO* was cloned into a pET28b vector. The protein was expressed in *E. coli* with an N-terminal His₆-tag, and was purified by IMAC. His₆-GarO behaved as a monomer in solution as determined by analytical gel filtration chromatography (Figure 3.10). Sequence analysis suggests that GarO is a luciferase-like
monooxygenase that possesses a typical binding domain for flavin mononucleotide (FMN). Previous studies on luciferase-like enzymes\textsuperscript{17,18} have shown that reduced FMN reacts with molecular oxygen, forming a flavin-hydroperoxide, which is followed by transfer of an oxygen atom to the substrate. NAD(P)H then commonly reduces the oxidized flavin back to its reduced form. We incubated GarM-modified GarA and GarA-V15L/I16V/19insK with purified GarO, FMN, and NADH, followed by proteolytic removal of the leader peptide as described above. MALDI analysis showed the formation of a molecular ion consistent with a single oxidation in the core peptide (Figure 3.7B). Substitution of NADH with NADPH resulted in abolishment of the GarO activity.

![MALDI-MS spectrum of modified GarA-V15L/I16V/19insK from co-expression of garA-V15L/I16V/19insK/garM/pRSFDuet-1 and 2garO/pCDFDuet-1. M indicates GarM-modified GarA-V15L/I16V/19insK with the leader peptide removed by trypsin and aminopeptidase A8200 (sequence: Ala(0)-DAB-Lys, [M + O + Na]\textsuperscript{+} calc. m/z: 2095.49, obs. m/z: 2095.64; [M + O + Na]\textsuperscript{+} calc. m/z: 2111.49, obs. m/z: 2111.68).](image)

Figure 3.9 MALDI-MS spectrum of modified GarA-V15L/I16V/19insK from co-expression of garA-V15L/I16V/19insK/garM/pRSFDuet-1 and 2garO/pCDFDuet-1. M indicates GarM-modified GarA-V15L/I16V/19insK with the leader peptide removed by trypsin and aminopeptidase A8200 (sequence: Ala(0)-DAB-Lys, [M + O + Na]\textsuperscript{+} calc. m/z: 2095.49, obs. m/z: 2095.64; [M + O + Na]\textsuperscript{+} calc. m/z: 2111.49, obs. m/z: 2111.68).
The dehydration and cyclization reactions involved in lantibiotic biosynthesis require the presence of the leader peptide, but other tailoring-type transformations such as hydroxylation\(^\text{19}\) and oxidative decarboxylation\(^\text{20}\) do not require the leader peptide to be attached. To determine whether oxidation by GarO required the leader peptide, the enzyme was incubated for 8 h with purified Ala(0)-DAB-Lys, FMN, and NADH. The major product was a peptide with m/z of +16 Da (Figure 3.7C). Thus, the leader peptide is not required for GarO catalysis, and the C-terminal Lys does not prevent GarO activity. The lack of oxidation in *E. coli* when co-expressing GarA-V15L/I16V/19insK with GarM and GarO (Figure 3.9) may therefore be the result of relatively inefficient production of GarO. A peak corresponding to the addition of three oxygen atoms was also observed in the product of *in vitro* GarO modification. This product is likely the result of uncoupling.

**Figure 3.10** Analysis of His\(_6\)-GarO by gel filtration chromatography. Peak A corresponds to the monomer of His\(_6\)-tagged GarO.
of the formation of the flavin peroxide and the oxygen transfer to the substrate, resulting in the formation of hydrogen peroxide or superoxide, which are known general side products of in vitro reactions with flavin-dependent monooxygenases.\textsuperscript{21} Hydrogen peroxide formation in turn results in non-enzymatic oxidation of the thioether rings.

3.2.6 Overexpression and purification of GarM

In order to perform crystallization trials on GarM, the gene encoding GarM was cloned into MCS1 of pRSFDuet-1 and pET28b. The protein was expressed in E. coli with an N-terminal His\textsubscript{6}-tag, and was purified by IMAC and gel filtration fast protein liquid chromatography purification (FPLC). However, GarM always eluted with an impurity, displaying a strong band with lower molecular weight (between 70 kDa to 100 kDa) than His\textsubscript{6}-GarM (114 kDa) on the SDS-PAGE gel of the FPLC fractions. The impurity could not be resolved well from His\textsubscript{6}-GarM either by IMAC or further gel filtration FPLC purification. Therefore, the impurity was assumed to be a truncation product. However, addition of protease inhibitor before lysis of the cells and to all the buffers throughout the whole purification process did not result in improvements.

3.3 SUMMARY AND OUTLOOK

Actagardine is a globular lantibiotic which is stable in simulated gastric fluid and highly potent against Clostridium difficile, with reduced activity against other bacteria in the gut flora. Its derivative NVB302 has already been evaluated in trials for treating C. difficile associated diarrhea. In the work described in Chapter 3, Ala(0)actagardine and derivatives were heterologously produced in E. coli by co-expression of the substrate
peptide, wild-type and mutated GarA, with GarA modification enzymes GarM and GarO. The activity of GarO, a luciferase-like monooxygenase that introduces the unique sulfoxide group of actagardine, was also investigated in vitro. The work described in Chapter 3 also confirmed that the co-expression system can be utilized to introduce multiple post-translational modifications, such as the formation of a sulfoxide structure, onto lanthipeptides besides dehydration and cyclization. In fact, based on the success of co-expression of prochlorosins as described in Chapter 2, the co-expression methodology has also been applied to produce multiple other lanthipeptides, including nisin, haloduracin, cinnamycin, geobacillin, and cytolysin.

3.4 MATERIALS AND METHODS

3.4.1 Materials

All oligonucleotides were purchased from Integrated DNA Technologies. Restriction endonucleases, DNA polymerases, and T4 DNA ligase were purchased from New England Biolabs. Media components for culturing bacteria were purchased from Difco laboratories. Chemicals were purchased from Fisher Scientific or from Aldrich unless noted otherwise. Trypsin was purchased from Worthington Biosciences. Aminopeptidase A8200 was purchased from Sigma. E. coli DH5α was used as host for cloning and plasmid propagation, and E. coli BL21 (DE3) was used as a host for co-expression.

3.4.2 General methods

Positive residue numbers are used for amino acids in the core peptide counting forward starting from the N-terminal residue in the core peptide. Actagardine can be
produced with an additional N-terminal alanine in *Actinoplanes garbadinensis*, and this N-terminal alanine residue is numbered Ala(0).\textsuperscript{12,13} We also constructed actagardine derivatives with an extra lysine residue located at the C-terminus of GarA generating GarA-19insK. Another derivative with Val15 and Ile16 substituted with Leu and Val residues and an extra lysine residue attached to the C-terminus was also constructed, and this mutant was named GarA-V15L/I16V/19insK.

All polymerase chain reactions (PCR) were carried out on a C1000™ thermal cycler (Bio-Rad). DNA sequencing was performed by ACGT Inc. and the Biotechnology Center at the University of Illinois at Urbana-Champaign, using appropriate primers. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) was carried out on an ultrafleXtreme™ MALDI TOF/TOF system (Bruker Daltonics). Liquid chromatography electrospray ionization tandem mass spectrometry (LC/ESI-MS/MS) was carried out and processed using a Synapt ESI quadrupole TOF Mass Spectrometry System (Waters) equipped with an Acquity Ultra Performance Liquid Chromatography (UPLC) system (Waters). Fast protein liquid chromatography (FPLC) purification was carried out on an Amersham Biosciences/GE Healthcare ÄKTA system.

3.4.3 Genomic DNA extraction of *Actinoplanes garbadinensis* ATCC 31049

*Actinoplanes garbadinensis* ATCC 31049 bacteria were purchased from the American Type Culture Collection (ATCC). Lyophilized bacteria were inoculated in 50 mL of GYM or LB media. The mixtures were shaken at 30 °C and 200 r.p.m. for a week with autoclaved glass beads present in the flask and then centrifuged. The *Actinoplanes garbadinensis* ATCC 31049 genomic DNA was extracted from the cell pellet, using Ultra
Clean™ Microbial DNA Isolation Kit (MO Bio Laboratories Inc.) or Wizard® Plus SV Minipreps kit (Promega), with the Ultra Clean™ kit giving higher yield. (See Y.S. Notebook 5, Page 084).

### 3.4.4 Construction of pRSFDuet-1 derivatives for garM, garA, and garO

Attempts to clone *garM* using terminal forward and reverse primers all failed, resulting in either very weak bands or no bands under various conditions using a variety of polymerases. The *garM* gene was therefore cloned by PCR from *A. garbadinensis* ATCC 31049 genomic DNA using stepwise overlap extension PCR (for all primer sequences, see Table 3.1): the first round PCR generated fragment GarM_NdeI_1456 using GarM_NdeI_FLP_Duet-1 and GarM_1456seq_RP primer pair, fragment GarM_1117_2074 using GarM_1117seq_FP and GarM_2074seq_RP primer pair, fragment GarM_576_2074 using GarM_576seq_FP and GarM_2074seq_RP primer pair, and fragment GarM_1689_KpnI using GarM_1689seq_FP and GarM_KpnI_RLP_Duet-1 primer pair. The second round PCR generated fragment GarM_NdeI_2074 using GarM_NdeI_1456 and GarM_1117_2074 megaprimer pair, and fragment GarM_576_KpnI using GarM_576_2074 and GarM_1689_KpnI megaprimer pair. The third round PCR generated the GarM_NdeI_KpnI full-length product using GarM_NdeI_2074 and GarM_576_KpnI megaprimer.

AccuPrime™ GC rich polymerase (Invitrogen) was able to amplify full length *garM* by PCR providing sufficient amount of product to conduct digestion and ligation, yielding more than Phusion polymerase, Platinum Taq, Platinum Pfu or OneTaq polymerase. However, the resulting ligation products showed a large number of point
mutations in the amplified region, indicating the low fidelity of AccuPrime™ GC rich polymerase, which was considered too low to be used to correctly amplify an insert as long as 3 kb like garM. The gene garM, obtained by stepwise overlap extension PCR as described above, was inserted into multiple cloning site 2 (MCS2) of the pRSFDuet-1 vector between the NdeI and KpnI restriction sites for co-expression, and into pET28b and pRSFDuet-1 between the EcoRI and HindIII restriction sites for GarM protein purification.

The garA gene was cloned by PCR from A. garbadinensis ATCC 31049 genomic DNA with the GarA_EcoRI_FP_Duet-1 and GarA_HindII_RP_Duet-1 primers, and inserted between the EcoRI and HindIII restriction sites of MCS1 in pRSF-1. The garA mutant gene GarA-19insK was generated using overlap extension PCR with the GarA_EcoRI_FP_Duet-1 and GarA_C+K_RP primers, and the garA/pRSFDuet-1 vector as the template. The garA mutant gene GarA-V15L/I16V/I19insK was generated using overlap extension PCR with the GarA_EcoRI_FP_Duet-1 and GarA_LV_C+K_RP primers, and the garA/pRSFDuet-1 vector as the template. The wild type and mutant garA genes were individually cloned into MCS1 of the pRSFDuet-1 vector between the EcoRI and HindIII restriction sites, with the garM gene already inserted in MCS2. Plasmids garA-A−1L/A−2L/garM/pRSFDuet-1 and garA-A−1K/garM/pRSFDuet-1 were generated similarly using primers listed in Table 3.1. The sequences of the resulting plasmids were confirmed by DNA sequencing.

The garO/pCDFDuet-1 plasmid was generated by inserting a garO gene into MCS2 of pCDFDuet-1 between the NdeI and XhoI restriction sites, resulting in garO/pCDFDuet-1. In order to make 2garO/pCDFDuet-1, one copy of garO was
inserted into MCS1 of \textit{garO/pCDFDuet-1} between the \textit{NcoI} and \textit{HindIII} restriction sites resulting in two proteins encoded without His\textsubscript{6}-tags. In order to purify the \textit{garO} protein to perform \textit{in vitro} assays, the \textit{garO} gene was inserted into a \textit{pET28b} vector between \textit{NdeI} and \textit{XhoI} sites. In order to construct a vector for three-gene co-expression (\textit{garA/garM/garO}), the \textit{garO} gene was amplified from \textit{garO/pCDFDuet-1} including the T7 promoter and lac operator and inserted into \textit{garA/garM/pRSFDue-1} or \textit{garA-V15L/I16V/19insK/garM/pRSFDuet-1} vector between the \textit{PacI} and \textit{AvrII} sites.


\begin{table}
\centering
\begin{tabular}{|l|l|}
\hline
\textbf{Primer Name} & \textbf{Primer Sequence (5'-3')} \\
\hline
GarA\_EcoRI\_FP\_Duet-1 & GTA AGT ACG AAT TCG ATG ATG TCT TCT GCC ATC GAC AGG AAG \\
GarA\_HindIII\_RP\_Duet-1 & ATA ATA TCA AGC TTT CAT TTA CGG GGC GAC ATC ACG GTG \\
GarA\_C+K\_RP & ATA ATA TCA AGC TTT CAT TTA CGG GGC GAC ATC ACG GTG \\
GarA\_LV\_C+K\_RP & ATA ATA TCA AGC TTT CAT TTA CGG GGC GAC ATC ACG GTG \\
GarM\_NdeI\_FP\_Duet-1 & GTA AGT ACC ATA TGT CAC CGG TTC CTT CAC TCA ATT CCC CCT CGG TAC GCG \\
GarM\_KpnI\_RP\_Duet-1 & ATA ATA TGG GTA CTA CCT GTC GTC AGC GGC GGC TGC AGC CAG \\
GarM\_576seq\_FP & GCT GCT GAC CGA CCC CGC CTA C \\
GarM\_1117seq\_FP & CCG ATC GAC CTA CAG ACC GTG CTG \\
GarM\_1689seq\_FP & CGA GTC GGA AAA GCC ACA ACT GCT CG \\
GarM\_2249seq\_FP & CAT CGT CGA GAC CTG CGG CGC CTA CAG \\
GarM\_303seq\_RP & GTG ACC ACA GTT TGG TGG ACA ACT CG \\
GarM\_867seq\_RP & CAT CGT GTC GCC GTA GCC GCC GCG TTC \\
GarM\_1456seq\_RP & GCA TGA TGT CTT AGG CAG CGG TGA ACC C \\
GarM\_2074seq\_RP & GCT TGT AGC TGT CCT GGG C \\
GarM\_2616seq\_RP & GAG TGC CCT GGC GAG CCG GTG CAG \\
GarM\_EcoRI\_FP\_Duet-1 & GTA AGT ACG AAT TCG ATG ATG TCT TCT GCC ATC GAC AGG AAG \\
GarM\_HindIII\_RP\_Duet-1 & ATA ATA TCA AGC TTT CAT TTA CGG GGC GAC ATC ACG GTG \\
GarM\_NcoI\_FP & GTA AGT ACG AAT TCG ATG CTC AGC GTG CTG GAC CAG \\
GarM\_T7\_PacI\_FP & TCC AGG CA TTAATT TCAAATTAATACGACTC ACT ATAGGGG \\
GarM\_AvrII\_RP & ATA ATA TGA CCT GTG CAG GCC GCC GCG GGG GTG GAG \\
\hline
\end{tabular}
\caption{Primer sequences for cloning and mutagenesis of \textit{garA}, \textit{garM} and \textit{garO}.}
\end{table}
3.4.5 Overexpression and purification of GarM-modified His$_6$-tagged GarA and GarA-V15L/I16V/19insK

*E. coli* BL21 (DE3) cells transformed with a *pRSFDuet-1* vector carrying the genes encoding GarM and one of the GarA peptides were shaken at 225 r.p.m. in 2 L of LB medium containing 50 mg/L kanamycin at 37 °C until the O.D.$_{600\text{ nm}}$ reached 0.6 − 0.8. The culture was then induced with 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and shaken continually at 18 °C for an additional 20 h. (Overexpression and purification see Y.S. Notebook 6, Page 50, 53, 54, 64, 66, Notebook 7 Page 32).

For the co-expression of GarA, GarM and GarO, *E. coli* BL21 (DE3) cells transformed with a *pRSFDuet-1* vector carrying *garA/garM/garO*, or transformed with two vectors (one was *garA/garM/pRSFDuet-1* or *garA-V15L/I16V/19insK/garM/pRSFDuet-1*, while the other was *garO/pCDFDuet-1* or 2*garO/pCDFDuet-1*) were grown in 2 L of LB medium containing 50 mg/L kanamycin and 25 mg/L spectinomycin (spectinomycin was not needed for *garA/garM/garO/pRSFDuet-1* overexpression) at 37 °C until the O.D.$_{600\text{ nm}}$ reached 0.6 − 0.8. The culture was then induced with 0.2 mM IPTG and shaken continually at 18 °C for an additional 30 h with *garO/pCDFDuet-1* or 20 h with 2*garO/pCDFDuet-1*. (Overexpression and purification see Y.S. Notebook 6, Page 79, 93 − 94; Y.S. Notebook 7, Page 4, 33, 35, 63 − 65).

The induced cells were harvested by centrifugation (11,900 × g for 20 min at 4 °C). The cell pellet was resuspended in 20 mL of LanA Start Buffer (20 mM NaH$_2$PO$_4$, pH 7.5 at 25 °C, 500 mM NaCl, 0.5 mM imidazole, 20% glycerol) and lysed by sonication (35% amplitude, 4.0 s pulse, 9.9 pause, 15 min). The sample was centrifuged (23,700 × g
for 40 min at 4 °C) and the supernatant was kept for further purification. The pellet was resuspended in another 20 mL of LanA Start Buffer to obtain any soluble proteins remaining in the pellet. Insoluble materials were removed from the combined soluble fractions by centrifugation (11,900 × g for 20 min at 4 °C), and the resulting sample was clarified using a 0.45 μm syringe filter (Corning). The His$_6$-tagged peptides were then purified by immobilized metal affinity chromatography (IMAC) using a 5 mL HisTrap™ HP nickel affinity column (GE Healthcare Life Sciences). The filtered sample was applied to the column and the column was washed with 2 column volumes of LanA Buffer 1 (6 M guanidine hydrochloride, 20 mM NaH$_2$PO$_4$, pH 7.5 at 25 °C, 500 mM NaCl, 0.5 mM imidazole), followed by 2 ~ 3 column volumes of LanA Buffer 2 (4 M guanidine hydrochloride, 20 mM NaH$_2$PO$_4$, 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 were purified by preparative reversed-phase high-performance liquid chromatography (RP-HPLC) using a Waters Delta-pak™ C4; 15 μm; 300 Å; 25 × 100 mm PrepPac® Cartridge. Solvents for the RP-HPLC were solvent A (0.1% TFA in water) and solvent B (0.086% TFA in 80% acetonitrile / 20% water). A gradient of 2 - 100% of solvent B was executed over 45 min at a flow rate of 8 mL/min, and peptides were detected by absorbance at 220 nm. The fractions were analyzed by MALDI-TOF-MS. All the fractions containing the desired product were combined and the organic solvents were removed by rotary evaporation, followed by lyophilization overnight. The yields of modified GarA and GarA-V15L/I16V/19insK from *garA/garM/pRSFDuet-1* were 4.2 mg and 2.1 mg per liter, respectively, while the yields of modified GarA and
GarA-V15L/I16V/19insK from garA/garM/pRSFDuet-1 and garA-V15L/I16V/19insK/garM/pRSFDuet-1 with 2garO/pCDFDuet-1 were 0.17 mg and 0.74 mg per liter, respectively. The products were kept at −20 °C for short-term storage and −80 °C for long-term storage.

3.4.6 Overexpression and purification of His$_6$-tagged GarO

*E. coli* BL21 (DE3) cells transformed with garO/pET28b were shaken at 225 r.p.m. in 2 L of LB medium containing 50 mg/L kanamycin at 37 °C until the O.D.$_{600}$ nm reached 0.6 − 0.8. The culture was then induced with 0.33 mM IPTG and shaken at 18 °C for an additional 20 h. (See Y.S. Notebook 6, Page 70).

The purification steps of GarO were all performed at 4 °C. Induced cells were first harvested by centrifugation (11,900 × g for 20 min at 4 °C). The cell pellet was resuspended in 20 mL of Start Buffer (20 mM Tris, pH 8.0 at 4 °C, 500 mM NaCl, 10% glycerol) and lysed by a high pressure homogenizer (Avestin, Inc.). After centrifugation (23,700 × g for 40 min at 4 °C), the supernatant was clarified using a 0.45 μm syringe filter (Corning), followed by injection onto a fast protein liquid chromatography (FPLC) system (ÄKTA, GE Heathcare Life Sciences), equipped with a 5 mL HisTrap™ HP nickel affinity column (GE Healthcare Life Sciences) pre-charged with Ni$^{2+}$ and pre-equilibrated with Start Buffer. After loading, the column was washed with Buffer A (20 mM Tris, pH 8.0 at 4 °C, 1 M NaCl, 30 mM imidazole). His$_6$-GarO was eluted using a linear gradient of 0-100% of Buffer B (20 mM Tris, pH 8.0 at 4 °C, 1 M NaCl, 200 mM imidazole), over 45 min at a flow rate of 2 mL/min and detection by UV absorbance at 280 nm. The fractions were analyzed by SDS-PAGE (4 – 20% tris-glycine gel, Bio-Rad).
All the fractions containing the desired product were combined and concentrated to 2.5 mL using an Amicon Ultra-15 Centrifuge Filter Unit (10 kDa MWCO, Millipore). Then the buffer of the concentrated protein was exchanged to Start Buffer using a PD-10 desalting column (GE Healthcare Life Sciences). The concentration of the protein was determined by absorbance at 280 nm, and the resulting protein sample was stored at −80 °C. (See Y.S. Notebook 6, Page 71 – 75, 81).

To evaluate the oligomerization state of GarO, purified GarO (Calc. M.W.: 38.3 kDa) was injected into the FPLC system equipped with a Superdex 200 (GE Healthcare) column pre-equilibrated with Column Buffer (20 mM NaH₂PO₄, 500 mM NaCl, 10% glycerol, pH 8.0 at 4 °C). The protein was eluted at a flow rate of 1 mL/min and proteins were detected by UV absorbance at 280 nm. The major peak was eluted with a retention time of 78 min, indicating a monomer oligomerization state when compared with molecular weight standards (Figure 3.10). (See Y.S. Notebook 6 Page 76 – 78).

3.4.7 Proteolytic removal of GarA leader peptide

Modified GarA and its mutants were dissolved in water generating 6 mg/mL peptide solutions. An aliquot of 10 μL of GarA stock solution was incubated with 1 μL of trypsin solution (0.6 mg/mL) in 50 mM Tris buffer (pH = 8.3) for at least 1 h to remove most of the N-terminal portion of the leader peptide. An aliquot of 4 μL of aminopeptidase A8200 (0.1 Unit/ mL) was then added to the mixture and incubated for at least three hours to remove the residual C-terminal portion of the leader peptide. (See Y.S. Notebook 6 Page 59, 60, 66 – 68, 85 – 86, 89, 91; Notebook 7 Page 27 – 28, 64).
3.4.8 Product evaluation for *in vivo* coexpression of GarA, GarM and GarO

For the co-expression of GarA, GarM and GarO in *E. coli*, the same expression procedure was used as described above under “Overexpression and purification of His6-tagged GarM-modified GarA and GarA-V15L/I16V/19insK”. GarA with four dehydrations and one oxygen addition was observed after proteolytic cleavage, regardless of whether one-vector (*garA/garM/garO/pRSFDuet-1*) or two vectors (*garA/garM/pRSFDuet-1* together with *2garO/pCDFDuet-1*) were used (Figure 3.8 and Figure 3.7A, respectively). Two-vector co-expression using *garA-V15L/I16V/19insK/garM/pRSFDuet-1* together with *2garO/pCDFDuet-1* generated GarA-V15L/I16V/19insK with four dehydrations but incomplete addition of one oxygen (Figure 3.9). (See Y.S. Notebook 7 Page 63).

3.4.9 HPLC purification of Ala(0)-DAB-Lys

The Ala(0)-DAB-Lys generated in the trypsin and aminopeptidase A8200 cleavage reaction was purified by analytical RP-HPLC using a Vydac™ C18; 5 µm; 300 Å; 4.6 i.d. × 250 mm column. Solvents for RP-HPLC were solvent A (0.1% TFA in water) and solvent B (0.086% TFA in 80% acetonitrile / 20% water). A gradient of 0.8-100% of solvent B was executed over 50 min at a flow rate of 1 mL/min, and peptides were detected by absorbance at 220 nm. The fractions were tested by MALDI-ToF-MS for the desired peptides. All the fractions containing the desired fully modified product were combined and the solvents were removed by lyophilization overnight. The product was kept at −20 °C for short-term storage and −80 °C for long-term storage. (See Y.S. Notebook 7 Page 64 – 65).
### 3.4.10 Iodoacetamide treatment of modified His₆-GarA and its derivatives to detect free cysteines

An aliquot of 20 μL of the trypsin cleavage reaction of GarM-modified GarA or its derivatives was mixed with 50 mM Tris-HCl (pH 8.3), 50 mM IAA, and 0.5 mM TCEP. The resulting solution was then incubated at room temperature overnight, followed by desalting using ZipTipC₁₈ (Millipore) and MALDI-TOF-MS analysis. No carboxyamidomethyl (CAM) thiol modification was observed after IAA treatment (Figure 3.4), suggesting the cyclization was complete and free cysteine residues were not present. (See Y.S. Notebook 6 Page 56).

### 3.4.11 GarO enzymatic assays

Activity assays for GarO with dehydrated and cyclized GarA, GarA-V15L/I16V/19insK, and purified Ala(0)-DAB-Lys were carried out in an assay mix containing 50 mM Na₂HPO₄ (pH 8.0), 0.2 mM FMN, and 0.5 mM NADH. The final concentrations of GarA (or its derivatives) and GarO were 75 μM and 10 μM, respectively. The reaction mixtures were incubated at room temperature overnight. An aliquot of 20 μL of the reaction was treated with 1 μL of trypsin stock solution (0.6 mg/mL) before analysis by MALDI-ToF mass spectrometry. When NADH was substituted with NADPH, no product was observed after overnight incubation. (See Y.S. Notebook 6 Page 76 – 78; Y.S. Notebook 7 Page 15, 25 – 28, 70 – 72).
3.4.12 Tandem mass spectrometry analysis of GarO modified Ala(0)-DAB-Lys

To obtain insights into the site of oxidation, an aliquot of 5 μL of GarO-modified Ala(0)-DAB-Lys was analyzed by LC/ESI-MS/MS. The trypsin cleavage reaction mixture was injected to a BEH C8 column (1.7 μm, 1.0 × 100 mm), and the fully modified product was purified by UPLC using a gradient of 3% mobile phase B (0.1% formic acid in acetonitrile) to 97% mobile phase B in mobile phase A (0.1% formic acid in water) over 12 min. Mass spectra were acquired in ESI positive mode in the range of 50-2000 m/z. The capillary voltage was 3200 V, and the cone voltage was 40 V. The other parameters used were as follows: 120 °C source temperature, 300 °C desolvation temperature, 150 L/h cone gas flow, and 600 L/h desolvation gas flow. A transfer collision energy of 4 V was used for both MS and tandem MS, while the trap collision energy was set to 6 V for MS and 35 V for MSMS. Glu-fibrinopeptide B (Sigma) was directly infused as the lock mass. The tandem mass spectra were processed with MaxEnt3 and analyzed by Protein/Peptide Editor in BioLynx 4.1. The software for analyzing both precursor ions and fragment ions was set to report any mass within 0.3 amu of the calculated values. Fragments corresponding to b7 and y”14 were observed (Figure 3.5), confirming the sulfoxide group was added to the C-terminal intertwined rings. The loss of one molecule of H₂O was observed during MSMS when collision energy was applied on the parent ion. (See Y.S. Notebook 7 Page 74 – 76).
3.4.13 Bioactivity assay for deoxyactagardine derivatives

*Bacillus subtilis* W23 was inoculated from frozen stock into 5 mL of DSMZ Medium 1 Nutrient Broth and grown overnight at 37 °C and 250 r.p.m. An aliquot of 100 μL of overnight culture was inoculated into 20 mL of molten DSMZ Medium 1 Nutrient Agar kept at 55 °C, followed by dispensing into a round petri dish. While the bacteria-containing agar was cooled at room temperature for 0.5 h, 20 μL of nisin (100 μM), GarM-modified GarA-V15L/I16V/I19insK treated with trypsin and aminopeptidase A8200 (peptide final concentrations: 100 μM; other components included 50 mM Tris buffer, 0.03 mg/mL trypsin, and 0.1 U/mL aminopeptidase A8200, pH = 8.3), or negative control with buffer and proteinases only, was spotted onto the agar (Figure 3.6A). (See Y.S. Notebook 7 Page 60 – 61, 66 – 69).

*M. luteus* ATCC 4698 was inoculated from frozen stock into 5 mL of Mueller-Hinton Broth (MHB) and grown overnight at 37 °C and 250 r.p.m. An aliquot of 100 μL of overnight culture was inoculated into 20 mL of molten Mueller-Hinton Agar (MHA) kept at 42 °C, followed by dispensing into a Nunc OmniTray. While the bacteria-containing MHA agar was cooled at room temperature for 1 h, another aliquot of 200 μL of overnight culture was inoculated into 40 mL of molten MHA kept at 42 °C. The mixture was then dispensed onto the solidified layer of MHA. Wells were made by placing a sterile 96-well PCR plate in the molten seeded agar. Once the agar was solidified, the 96-well PCR plate was removed and aliquots of 60 μL samples were loaded in wells independent from each other by spatial separation. Then the plate was transferred to a 30 °C incubator and kept in the incubator for 36 h. Both Ala(0)-deoxyactagardine and Ala(0)-DAB-Lys displayed antimicrobial activity against *M. luteus* ATCC4698. It was
noted that the addition of one lysine residue to the C-terminus of deoxyactagardine increased the core peptide solubility, as precipitate was observed in the well bottom for Ala(0)-deoxyactagardine, but not for the Ala(0)-deoxyactagardine-Lys which was generated for comparison (Figure 3.6). (See Y.S. Notebook 6 Page 95).

To further determine the minimum inhibitory concentration (MIC) of Ala(0)-DAB-Lys, overnight cultures of *Bacillus subtilis* W23 in DSMZ Medium 1 Nutrient Broth or *Enterococcus faecalis* 29212 in Brain Heart Infusion Broth (BHI) were inoculated into fresh DSMZ Medium 1 Nutrient Broth or BHI, respectively, to generate cell cultures with an O.D.\textsubscript{600 nm} of 0.01. An aliquot of 30 µL of purified Ala(0)-DAB-Lys aqueous solution was added into 250 µL of the cell cultures in a 48-well plate (at final peptide concentrations of 60 µM, 20 µM, 10 µM, 5 µM, 2.5 µM, 1.25 µM, 625 nM, 312 nM, 156 nM, 78 nM, 39 nM, 20 nM, 10 nM, and 0 nM). Every concentration was tested in triplicate. The plate was shaken at 30 °C at 250 r.p.m. for 6 h, and the cell density was monitored by the plate reader at 600 nm. The MIC of Ala(0)-DAB-Lys against *Bacillus subtilis* W23 in DSMZ Medium 1 Nutrient Broth was 1.20 µM. The MIC of actagardine and deoxyactagardine against *Enterococcus faecalis* 29212 were reported to be 4.24 µM and 4.27 µM, respectively. In comparison, the MIC of Ala(0)-DAB-Lys generated in this work against *Enterococcus faecalis* 29212 in BHI was 7.28 µM. (See Y.S. Notebook 7, Page 60 – 61, 66 – 69. Bioactivity primary test see Y.S. Notebook 7 Page 42 – 45).

3.4.14 Overexpression and purification of His\textsubscript{6}-tagged GarM

*E. coli* BL21 (DE3) cells were transformed with garM/pET28b or garM/pRSFDuet-1 via electroporation. The cells were grown in 2 L of LB medium containing 50 mg/L
kanamycin at 37 °C until the O.D._{600nm} reached about 0.6. Then the incubation temperature was reduced to 18 °C and the culture was induced with 0.1 mM IPTG and shaken aerobically continuously for an additional 20 h. The cells were harvested by centrifugation (5000 × g for 20 min at 4 °C). The cell pellet was resuspended in 20 mL of GarM Start Buffer (20 mM Tris, pH 8.0 at 4 °C, 1 M NaCl, 10% glycerol) and was disrupted by multiple passes through a high pressure homogenizer (Avestin, Inc.). The insoluble aggregates and cellular debris were then discarded after centrifugation (15,000 × g for 30 min at 4 °C). The supernatant was passed through 0.45 μm syringe filters (Corning), followed by nickel affinity fast protein liquid chromatography (FPLC) purification using an Amersham Biosciences/GE Healthcare ÄKTA system. A 5 mL HisTrap HP nickel affinity column (GE Healthcare Life Sciences) was connected to the FPLC system for the purification at 4 °C. Solvents for FPLC were solvent A (GarM Start Buffer) and solvent B (GarM Elution Buffer: 20 mM Tris, pH 8.0 at 4 °C, 1 M NaCl, 200 mM imidazole). A gradient of 0-100% of solvent B over 45 min was executed with a flow rate of 2 mL/min. Proteins were detected by absorbance at 280 nm and collected in 4 mL fractions. The fractions were analyzed for purity by Tris-Glycine SDS-PAGE (4-20% acrylamide gradient). All the fractions containing the pure desired protein were combined and concentrated using an Amicon Ultra-15 Centrifugal Filter Unit (30 kDa MW cut off for His_6-GarM) to about 2.5 mL. Then the buffer of the concentrated protein was exchanged to GarM Final Buffer (20 mM HEPES, pH 7.5 at 4 °C, 500 mM NaCl) using the FPLC system equipped with a SuperDex 75 self-packed gel filtration column. The protein was eluted with a flow rate of 0.6 mL/min. Proteins were detected by absorbance at 280 nm and collected in 4 mL fractions. A major impurity with a molecular weight
between 70 kDa and 100 kDa could not be separated from GarM after gel filtration, while GarM eluted out mainly as the monomer. The concentration of the protein was determined by absorbance at 280 nm, demonstrating protein yields of 9.1 mg/L for His$_6$-GarM (impurity included). The resulting protein sample was stored at -80 °C. (See Y.S. Notebook 7 Page 90).

### Table 3.2 Plasmids mentioned in Chapter 3.

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>pRSFDuet-1</td>
<td>kan for kanamycin resistance, RSF origin</td>
</tr>
<tr>
<td>garA/garM/pRSFDuet-1</td>
<td></td>
</tr>
<tr>
<td>garA-A$^{-1}$K/garM/pRSFDuet-1</td>
<td></td>
</tr>
<tr>
<td>garA-A$^{-1}$L/garM/pRSFDuet-1</td>
<td></td>
</tr>
<tr>
<td>garA-19insK/garM/pRSFDuet-1</td>
<td></td>
</tr>
<tr>
<td>garA-V15L/I16V/19insK/garM/pRSFDuet-1</td>
<td></td>
</tr>
<tr>
<td>garM/pRSFDuet-1</td>
<td></td>
</tr>
<tr>
<td>garA/garM/garO/pRSFDuet-1</td>
<td></td>
</tr>
<tr>
<td>pCDFDuet-1</td>
<td>aadA for spectinomycin resistance, CDF origin</td>
</tr>
<tr>
<td>garO/pCDFDuet-1</td>
<td></td>
</tr>
<tr>
<td>2garO/pCDFDuet-1</td>
<td></td>
</tr>
<tr>
<td>pET28b</td>
<td>kan for kanamycin resistance, pBR322 origin</td>
</tr>
<tr>
<td>garM/pET28b</td>
<td></td>
</tr>
</tbody>
</table>

### 3.5 REFERENCES


CHAPTER 4: EFFORTS TOWARDS THE INCORPORATION OF NON-CANONICAL AMINO ACID RESIDUES INTO LANTHIEPTIDES

4.1 INTRODUCTION

Lantibiotics, as a major subclass of lanthipeptides, have antimicrobial activity against a variety of bacterial pathogens. With the rapid development of bioinformatics, an increasing number of lanthipeptides has been discovered through genome mining, but the potential application of these newly discovered compounds is hindered by their unknown functions. Meanwhile, no lantibiotics are currently applied in human therapy mainly due to their low metabolic stability and unoptimized physical properties. Therefore, expansion of the structural diversity of lanthipeptides is highly desired, in order to probe their unknown functions, as well as to increase their stability, tune their solubility and increase their activity. However, these ribosomally synthesized peptides generated from precursor peptides made with 20 canonical amino acids possess limited functional groups for such exploration.

Chemical synthesis mainly through solid phase peptide synthesis (SPPS) techniques is currently a major manufacture technology for peptides containing diversified non-canonical amino acids (NAAs). SPPS has been utilized to synthesize up to 100-amino-acid-long peptides. Combined with solution-phase chemical synthesis, SPPS has been successfully applied in large-scale linear peptide drug synthesis such as Fuzeon®, a potent HIV-1 membrane fusion inhibitor. However, chemical synthesis is most efficient
for the synthesis of peptides with less than 20 amino acid residues, as the purification becomes more difficult as the length increases.\textsuperscript{11} Cyclic peptide structures such as lanthipeptides add metabolic stability but also increase the complexity of the chemical synthesis, although successful syntheses have been reported.\textsuperscript{12-15} As an alternative to organic synthesis of peptides containing non-canonical amino acids, semisynthesis by expressed protein ligation (EPL) has been shown to effectively incorporate NAAs.\textsuperscript{16,17} Intensive research has also been performed to incorporate NAAs during recombinant biosynthesis with the help of natural or evolved tRNA synthetases.\textsuperscript{18,19} Schultz and coworkers have engineered the \textit{Methanococcus jannaschii} tyrosyl-tRNA synthetase (MjTyrRS) to conjugate various NAAs to a tyrosyl-tRNA\textsubscript{CUA}\textsuperscript{Tyr} from \textit{M. jannaschii} that recognizes the amber codon, in order to position-specifically incorporate NAAs \textit{in vivo}.\textsuperscript{20-22} Tirrell and coworkers have utilized the methionyl-tRNA (Met-tRNA) synthetase to incorporate methionine analogues \textit{in vivo}. In the absence of methionine, methionyl-tRNA synthetase was able to incorporate certain methionine analogues recognizing the methionine natural codon AUG (Figure 4.1A) in a residue-specific manner \textit{in vivo}.\textsuperscript{23} Among all the methionine analogues that the authors tried, L-homopropargylglycine (H-Hpg-OH) and norleucine presented the highest incorporation efficiency (Figure 4.1A).\textsuperscript{19,24} Furthermore, engineered pyrrolysyl-tRNA synthetase (PylRS)/tRNA\textsubscript{CUA}\textsuperscript{Pyl} (pylT) pairs for orthogonal NAA incorporation were recently developed, which were reported to function in rich media.\textsuperscript{25-29} Wild type pyrrolysyl-tRNA synthetase recognizes the TAG amber codon and was found in \textit{Methanosarcina barkeri} but not in \textit{E. coli}, yeast, or mammalian cells. An evolved pyrrolysyl-tRNA synthetase-pylT pair has subsequently
been utilized to incorporate diversified L-lysine and L-phenylalanine derivatives with aliphatic or aromatic side chains (Figure 4.2) into proteins for various biochemical applications.\textsuperscript{26,30-33}

As shown in Chapters 2 and 3, lanthipeptides can be produced by co-expression of the precursor peptide and the corresponding modification enzymes in \textit{E. coli}, followed by \textit{in vitro} proteolytic removal of the leader peptides and HPLC purification of the modified core peptides. The co-expression system was able to produce pure prochlorosins at the
level of tens of milligrams per liter, enough for the initiation of prochlorosin biological activity investigation. Further evaluation of the biological functions of prochlorosins in nature, including in situ monitoring of the localization of prochlorosins and identification of molecules interacting with prochlorosins in vivo, requires efficient identification of their binding partners in complex environments. Therefore, methods are desired to introduce probes, which should be inert and small, and not interfere with the interaction of prochlorosins with their targets in nature. One example in this category is the photocrosslinking amino acid $p$-benzoyl-L-Phe ($p$Bpa) that can covalently link to adjacent peptide chains under exposure of UV light. Thus, in the work described in this chapter

**Figure 4.2** Examples of reported NAAs incorporated into proteins by the evolved pyrrolysyl-tRNA synthetase-pylT pair.
we investigated the incorporation of \( p \)Bpa at various positions of four prochlorosins (Figure 4.1B).

![Figure 4.3 Nisin A structure. The residues substituted by amber codons are highlighted in green.](image)

Nisin A (Figure 4.3) has been used as a food preservative for over fifty years throughout the world without development of significant antimicrobial resistance. However, nisin A has never entered clinical trials for application in human therapy, mainly because of the low stability of this compound. Nisin A degrades by cleavage at dehydrobutyryne 2, dehydroalanine 5 and dehydroalanine 33 residues in a wide pH range at room temperature.\(^3^4\) The activity of nisin A was also reduced by proteolytic cleavage N-terminally at serine 29.\(^3^5\) The substitution of dehydroalanine 5 with dehydrobutyryne increases nisin’s stability against acid-catalyzed chemical degradation.\(^3^6\) Previous research has also shown that the nisin A derivative T2S displayed increased antimicrobial activity against certain gram-positive bacteria, and that the M21K mutant possesses enhanced activity against several gram-negative bacteria such as *Pseudomonas* and *Salmonella* spp. Meanwhile M21V and K22T analogs of nisin A displayed higher potency against notorious pathogens including *Streptococcus agalactiae*, *Clostridium difficile*, and *Listeria monocytogenes*,\(^3^7\) and the mutant NisA-I4V/S5F/L6G was reported
to inhibit the growth of the nisin-producing strain *Lactococcus lactis*. Based on these previous studies, several residues, including Thr2, Met21, Lys22, Ser29, Ser33, as well as the Ile4/Ser5/Leu6 region, were chosen to expand the structural diversity of nisin A with the aim of increased stability and increased bioactivity. Based on the previous success, we adopted the evolved tRNA synthetase systems as well as Met-tRNA synthetase to expand lanthipeptide structural diversity.

**4.2 RESULTS AND DISCUSSION**

**4.2.1 Overexpression test with an engineered tyrosyl-tRNA synthetase**

We obtained the engineered *pDULE-Tyr* and *pEVOL-Tyr* vectors from Professor Schultz (the Scripps Research Institute) as a generous gift. The latter was reported to perform *p*-benzoyl-L-phenylalanine (*p*Bpa) incorporation into proteins in rich medium, whereas minimal medium is needed for *pDULE-Tyr*. Both vectors carry an orthogonal tyrosyl-tRNA synthetase (*MjTyrRS*) / tRNA$_{CUA}^{Tyr}$ pair to incorporate *p*Bpa at the position encoded by the amber codon TAG. We also constructed a series of *procA/procM/pRSFDuet-1* plasmids with amber codon mutations in the *procA* genes, including *procA1.1-G\_1K/F1B*, *procA1.1-G\_1K/F2B*, *procA1.1-G\_1K/F11B*, *procA1.7-G\_1R/M21B*, *procA3.3-G\_1K/Y15B*, and *procA3.3-G\_1K/A23B* (B represents the amber codon. Negative residue numbers are used for amino acids in the leader peptide counting backwards from the leader peptide cleavage site, i.e. the glycine residue immediately N-terminal to the cleavage site is −1 and when substituted by a lysine residue the mutant is denoted as G\_1K). *E. coli* BL21 (DE3) cells were co-transformed with the *pDule-Tyr* or *pEVOL-Tyr* vector, and a *procA/procM/pRSFDuet-1* vector carrying a TAG codon in the
ProcA gene for co-expression. However, no reproducible products containing pBpa were obtained from the overexpression tests utilizing the aforementioned vectors.

ProcA1.1G–1K_F1B and ProcA1.1G–1K_F11B appeared to show the desired products by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI MS) after proteolytic cleavage, but further investigation of these fragments by tandem MS did not generate the expected fragmentation for modified Pcn1.1. Other constructs did not generate fragments with m/z values that were consistent with calculated values. The pBpa was successfully incorporated into ProcA3.2TEV after mutating the codon of Phe26 to an amber stop codon, as shown in Chapter 2, but trials with other ProcAs were not successful thus far. Prof. Mehl from Franklin and Marshall College also tried to use their tRNA/tRNA synthetase pair encoded in an engineered pDule vector to incorporate pBpa into Pcn3.2 by using the procA3.2TEV-F26B/procM/pRSFDuet-1 vector, but has obtained no success thus far, with only truncation products observed in MALDI MS, suggesting the introduced amber codon was read as a stop codon.

4.2.2 Overexpression test with an engineered methionyl-tRNA synthetase

Endogenous methionyl-tRNA synthetase was utilized to incorporate H-Hpg-OH to introduce an alkyne group into prochlorosins. Peptides ProcA1.7-G–1E, ProcA2.8 and ProcA3.3-G–1K were chosen as the templates as the ring topology of prochlorosin (Pcn) 1.7, 2.8 and 3.3 has been determined (Figure 4.1B), and they all possess methionine (Met) residues. Pcn1.7 has a Met within a methyllanthionine ring, and Pcn2.8 contains a
Met between two lanthionine rings. Pcn3.3 has a large methyllanthionine ring wrapping a smaller methyllanthionine ring, with the larger ring containing one Met (Figure 4.1B).

A methionine auxotrophic *E. coli* strain and M9 minimal media supplemented with an amino acid mixture were utilized for overexpression. The full procedure for the incorporation is described in the Materials and Methods section. For proof of concept, the less costly NAA norleucine (Nle) was chosen, which was successfully incorporated into ProcM-modified ProcA1.7-G–1E, ProcA2.8, and ProcA3.3-G–1K. Subsequently, H-Hpg-OH was also successfully incorporated into the fully modified core peptides of ProcA2.8 and ProcA3.3-G–1K as confirmed by MALDI MS (Figure 4.4). No ions corresponding to methionine incorporation into the ProcA1.7-G–1E core peptide were observed in the MALDI mass spectra. One experimental difficulty in these experiments is

![Figure 4.4 MALDI-MS spectra of incorporation of Hpg into (A) ProcA2.8, and (B) Proc3.3-G–1K. M indicates the linear ProcA fragment containing intact core peptide, generated by GluC cleavage.](image-url)
that the calculated m/z value of a peptide containing Hpg and one sodium adduct is the
same m/z as that of a peptide with the incorporation of methionine instead of Hpg. As
wild type methionyl-tRNA synthetase was utilized, it is possible to have methione
incorporated if a small amount of Met was present in the final media supplemented with
H-Hpg-OH, or if the Met pool in the cells had not been completely depleted prior to
adding H-Hpg-OH. However, various pre-incubation periods between the medium
change and induction (see Materials and Methods for a detailed description of the
incorporation procedure) ranging from 15 min to 30 min did not result in any difference
regarding the peak intensity ratio of the fragment with only Hpg incorporated and its +22
Da peak. As the pre-incubation step was desired for the bacteria to consume the residual
methionine, elongated pre-incubation time was expected to reduce any methionine-
incorporated product. Thus, our results suggest that the +22 Da peak shown in Figure 4.4
was more probably arising from a sodium adduct of the Hpg incorporated peptide instead
of methionine incorporation.

The NAAs that can be utilized efficiently by wild-type methionyl-tRNA synthetase
are limited to norleucine, Hpg, and azidohomoalanine (Aha) (Figure 4.1A),40 with the
latter two expensive (L-azidohomoalanine: $1600/g; L-homopropargylglycine: $1200/g,
as listed by their major suppliers), considering at least 200 mg of the starting material is
required for 1 L overexpression to obtain at most 3 mg of modified precursor peptide.
Therefore, methionyl-tRNA synthetase proved not to be ideal for NAA incorporation in
large scale lanthipeptide production.
4.2.3 Overexpression test with an engineered pyrrolysyl-tRNA synthetase

We constructed six nisA/nisB/pRSFDuet-1 mutants with the codon for one amino acid residue in the NisA core peptide substituted by an amber codon, including nisA-T2B, nisA-I4V/S5B/L6G, nisA-M21B, nisA-K22B, nisA-S29B, and nisA-S33B. A nisC/tRNA/pCDFDuet-1 was constructed per Professor Liu’s (Texas A&M University) suggestions, with a higher copy number of pyrrolysyl-tRNA\(^{\text{pyl}}\)\(_{\text{CUA}}\) expected to lead to higher levels of incorporation. A pEVOL-pylT-mmPylRS(2X)-N346A/C348A vector was generously provided by Professor Liu, which contained one copy of pyrrolysyl-tRNA (pylT) under the control of a constitutive proK promoter, and two copies of the engineered pyrrolysyl RNA synthetase (mmPylRS-N346A/C348A), one of which is under the control of a constitutive glnS promoter while the other is controlled by an arabinose inducible araBAD promoter. E. coli BL21 (DE3) cells containing the pEVOL-pylT-mmPylRS(2X)-N346A/C348A vector were co-transformed with the nisA/nisB/pRSFDuet-1 mutant plasmids and nisC/tRNA/pCDFDuet-1 for co-expression.

Incorporation of 3-trifluoromethyl-L-phenylalanine (L-Phe(3-CF\(_3\))-OH) was attempted with NisA-T2B, I4V/S5B/L6G, K22B and S29B, while incorporation of 3-methyl-L-phenylalanine (L-Phe(3-Me)-OH) was tried with NisA-S29B and S33B. Co-expression of NisA amber mutants, NisB and NisC with L-Phe(3-Me)-OH present in the medium as the NAA source produced only truncation products where the introduced amber codon was recognized only as a stop codon. Co-expression in the presence of L-Phe(3-CF\(_3\))-OH generated products containing the NAA, with the desired peptide detected as the major peak in MALDI-MS analysis of NisA-I4V/S5B/L6G after Ni-NTA purification (Figure 4.5). Similarly ions were observed corresponding to successful
incorporation into NisA-K22B and S29B in MALDI spectra after Ni-NTA purification (Figure 4.6). Noteworthy, NisA-I4V/S5B/L6G containing L-Phe(3-CF₃)-OH possessed 6 dehydrations out of 7 expected dehydrated residues (Figure 4.5). NisA-K22B displayed 6 dehydrations out of 8 anticipated dehydrated residues (Figure 4.6A), while NisA-S29B was dehydrated 8 times out of 8 proposed dehydrated residues (Figure 4.6B), as indicated by MALDI MS. Truncation products with the introduced amber codon functioning as a stop codon dominated in the Ni-NTA elution fractions of NisA-T2B, K22B and S29B.

Use of minimal medium and lysogeny broth (LB) medium resulted in no significant

Figure 4.5 MALDI-MS spectra of the incorporation of L-Phe(3-CF₃)-OH into NisA-I4V/S5B/L6G. (A) M9 medium. (B) LB medium. M indicates the linear NisA-I4V/S5B/L6G fragment containing intact core peptide, generated by trypsin cleavage. NAA represents L-Phe(3-CF₃)-OH residue.
differences in relative ratios of truncation products and incorporated modified products (Figure 4.5), although cells in LB medium generally grew faster than in the minimal medium.

![MALDI-MS spectra](image)

**Figure 4.6** MALDI-MS spectra of the incorporation of L-Phe(3-CF₃)-OH into (A) NisA-K22B and (B) NisA-S29B. M indicates the linear fragment containing intact core peptide, generated by trypsin cleavage. NAA represents L-Phe(3-CF₃)-OH residue.

Among the L-Phe(3-CF₃)-OH incorporation trials with NisA mutants, truncation products in which the amber codon was recognized as a stop codon were observed for most of the constructs together with L-Phe(3-CF₃)-OH incorporated products. The NisA-S29B truncation product showed up to 7 dehydrations out of 7 proposed dehydrated residues (Figure 4.7), as indicated by MALDI MS. Per suggestion of Prof. Liu, minimizing the number of plasmids is beneficial to maintain the health of the *E. coli*.  

119
However, co-transformation or co-expression of wild type \textit{nisA/nisB/nisC/pRSFDuet-1} and \textit{pEVOL-pylT-mmPylRS(2X)-N346A/C348A} always resulted in limited cell growth. Thus the two-plasmid strategy was not further pursued.

![MALDI-MS spectrum](image)

**Figure 4.7** MALDI-MS spectrum of the truncation product from NisA-S29B during co-expression trial to incorporate L-Phe(3-CF$_3$)-OH. M* indicates the linear fragment containing the truncated core peptide with the introduced amber codon serving as the stop codon, generated by trypsin cleavage.

### 4.3 SUMMARY AND OUTLOOK

Non-canonical amino acids (NAAs) can expand structural diversity of peptide-based natural products, introduce probes for investigation of peptide activity, and modify biophysical properties of peptides for drug development. In Chapter 4, endogenous methionyl-tRNA synthetase and an evolved \textit{M. jannaschii} tyrosyl-tRNA synthetase
(MjTyrRS)/tyrosyl-tRNA\textsubscript{Tyr\text{\textsuperscript{CUA}}} pair were utilized in trials to incorporate Hpg and pBpa into modified ProcAs, respectively. Similar to a recent report on the production of lichenicidin in \textit{E. coli} containing NAAs,\textsuperscript{41} general success was observed using methionyl-tRNA synthetase to incorporate Hpg. An engineered pyrrolysyl-tRNA synthetase/tRNA\textsubscript{CUA} (pylT) pair was adopted for trials to incorporate L-Phe(3-CF\textsubscript{3})-OH into dehydrated and cyclized NisA, with the NAA incorporation observed in certain NisA mutants. Further optimization is needed and will be pursued by graduate student Ms. Nidhi Kakkar in our laboratory.

\textbf{4.4 MATERIALS AND METHODS}

\textbf{4.4.1 Materials}

L-Homopropargylglycine (H-Hpg-OH) was purchased from Chiralix in the Netherlands. 3-Trifluoromethyl-L-phenylalanine (L-Phe(3-CF\textsubscript{3})-OH) and 3-methyl-L-phenylalanine (L-Phe(3-Me)-OH) were purchased from Chem-Impex International, Inc. All oligonucleotides were purchased from Integrated DNA Technologies. Restriction endonucleases, DNA polymerases and T4 DNA ligase were purchased from New England Biolabs or Invitrogen. Media components for bacterial cultures were purchased from Difco laboratories. All other chemicals were purchased from Fisher Scientific or from Aldrich unless noted otherwise. Endoproteinase GluC, LysC, and trypsin were purchased from Roche Biosciences or Worthington Biosciences. Factor Xa protease was purchased from New England Biolabs. \textit{E. coli} DH5\textalpha{} was used as the host for cloning and plasmid propagation. \textit{E. coli} BL21 (DE3) and methionine auxotrophic T7 express crystal
competent *E. coli* (NEB) were used as the hosts for co-expression. Vectors *pDULE-Tyr* and *pEVL-Tyr* were obtained from Prof. Peter Schultz at the Scripps Research Institute. *pEVOL-pylT-mmPylRS(2X)-N346A/C348A* was obtained from Prof. Wenshe Liu at Texas A&M University. Vectors *procA/procM/pRSFDuet-1*, *nisA/nisB/pRSFDuet-1* and *nisC/pACYCDuet-1* were constructed previously, and *nisA_I4V_S5B_L6G/nisB/pRSFDuet-1* was provided by Neha Garg of our laboratory.

### 4.4.2 General methods

All polymerase chain reactions (PCR) were carried out on a C1000™ thermal cycler (Bio-Rad). DNA sequencing was performed by ACGT Inc. and the Biotechnology Center at the University of Illinois at Urbana-Champaign, using appropriate primers. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-ToF MS) was carried out on a Voyager-DE-STR (Applied Biosystems). Liquid chromatography electrospray ionization tandem mass spectrometry (LC/ESI-MS/MS) was carried out and processed using a Synapt ESI quadrupole ToF Mass Spectrometry System (Waters) equipped with an Acquity Ultra Performance Liquid Chromatography (UPLC) system (Waters).

### 4.4.3 Construction of *procA/procM/pRSFDuet-1* mutants containing an amber codon

overlap extension PCR with the procA/procM/pRSFduet-1 vector as the template. Mutated procA genes were introduced into multi cloning site 1 (MCS1) of the procM/pRSFduet-1 vector through digestion and ligation. The general PCR, digestion and ligation procedures were the same as described in Chapter 2. The primer sequences are listed in Table 4.1. (Y.S. Notebook 5 Page 48 – 51).

### Table 4.1 Primer sequences for cloning and mutagenesis of procAs.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ProcA1.1_EcoRI_FP</td>
<td>GGT GCG AGG AAT TCG ATGAAAAAGCGACTCAACAAT</td>
</tr>
<tr>
<td>ProcA1.1_NotI_RP</td>
<td>AAG CGG CCG CTC AGC ACA CAT TGA TAG</td>
</tr>
<tr>
<td>ProcA1.1G-1K_F1B_FP</td>
<td>GGT GTG GCT GGG AAA TAG TTC TGT GTG CAG GG</td>
</tr>
<tr>
<td>ProcA1.1G-1K_F1B_RP</td>
<td>CCC TGC ACA CAC AAT TTT CCA GCC ACA CC</td>
</tr>
<tr>
<td>ProcA1.1G-1K_F2B_FP</td>
<td>GGT GTG GCT GGG AAA TTT TAG TGT GTG CAG GG</td>
</tr>
<tr>
<td>ProcA1.1G-1K_F2B_RP</td>
<td>CCC TGC ACA CAC TAA AAT TTC CCA GCC ACA CC</td>
</tr>
<tr>
<td>ProcA1.1G-1K_F11B_FP</td>
<td>GCA GGG TAC TGC TAA CCG TTA GAC TAT CCA TGT GTG C</td>
</tr>
<tr>
<td>ProcA1.1G-1K_F11B_RP</td>
<td>GCA CAC ATT GAT AGT CTA ACG GTT AGC AGT ACC CTG C</td>
</tr>
<tr>
<td>ProcA1.7_EcoRI_FP</td>
<td>GGT GCC AGG AAT TCG ATG AAG CAT AGA CAA CTG AAT CTG</td>
</tr>
<tr>
<td>ProcA1.7M21B_NotI_RP</td>
<td>ATA ATA TCG CCG CTC AGC ACT TCC CCA C</td>
</tr>
<tr>
<td>ProcA3.3_EcoRI_FP</td>
<td>GGT GAG TGG AAT TCG ATG AAG CAT AGA CAA CTG AAT CTG</td>
</tr>
<tr>
<td>ProcA3.3_NotI_RP</td>
<td>ATA ATT TAG CCG CCT ATG CTC GCG</td>
</tr>
<tr>
<td>ProcA3.3_Y15B_FP</td>
<td>CAC GGC TGG ATG TTA GGG CGG GAC CAA AAT G</td>
</tr>
<tr>
<td>ProcA3.3_Y15B_RP</td>
<td>CAT TTT GGT CCC GCC CTA ACA TCC AGC CGT G</td>
</tr>
<tr>
<td>ProcA3.3_A23B_NotI_RP</td>
<td>ATA ATT TAG CCG CTC ACT AGC GGC ACA TTT TG</td>
</tr>
</tbody>
</table>

4.4.4 Incorporation of para-benzoyl phenylalanine (pBpa) in ProcA

*E. coli* BL21 (DE3) cells transformed with a *pDule-Tyr* vector or *pEVOL-Tyr*, and a mutated procA/procM/pRSFduet-1 vector were grown overnight at 37 °C in 5 mL of supplemented GMML medium (glycerol minimal medium, supplemented with metals and 18-amino-acid solution; for components, see ref. 21,43) containing 50 mg/L kanamycin and 12.5 mg/L tetracycline. An aliquot of 1 mL overnight culture was inoculated into 50 mL of fresh supplemented GMML medium. The culture was shaken at
37 °C until the O.D._{600nm} reached 0.5, then pBpa dissolved in 1 M NaOH was added to the culture to obtain a final concentration of 1 mM pBpa. After shaking for 0.5 h, the culture was induced with 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and shaken continually at 18 °C for an additional 20 h. The induced cells were harvested by centrifugation (11,900 × g for 20 min at 4 °C). The cell pellet was resuspended in 10 mL of LanA Start Buffer and lysed by sonication (35% amplitude, 4.0 s pulse, 9.9 s pause, 15 min). The cell lysate was centrifuged (23,700 × g for 40 min at 4 °C) and the supernatant was kept for further purification. The pellet was resuspended in 10 mL of LanA Start Buffer again. The soluble portions after centrifugation (23,700 × g for 20 min at 4 °C) were combined and clarified using 0.45 μm syringe filters (Corning), then purified by immobilized metal affinity chromatography (IMAC) using a 1 mL HisTrap™ HP nickel affinity column (GE Healthcare Life Sciences). The filtered sample was applied to the column and the column was washed with 2 column volumes of LanA Buffer 1, followed by 2 - 3 column volumes of LanA Buffer 2, and then eluted by 3 column volumes of LanA Elution Buffer. The elution fractions were desalted and purified by solid phase extraction (SPE) using a 3 mL Vydac® BioSelect™ reversed-phase C4 column. A step gradient wash and elution using 2 column volumes (CV) of 2% solvent B (0.086% TFA in 80% acetonitrile / 20% water) in solvent A (0.1% TFA in water), 1 CV of 50% solvent B in solvent A, 1 CV of 100% solvent B and 1 CV of acetonitrile was applied for the purification. The fractions were tested by MALDI-ToF MS for the desired peptides. The majority of the desired peptides eluted in fractions of 50% solvent B in solvent A. (Y.S. Notebook 5 Page 52 – 59, 67 – 68).

An 10 µL aliquot of cell lysate in LanA Start Buffer after sonication was incubated
with 0.03 U endoproteinase GluC in 50 mM HEPES buffer (pH 7.5) at room temperature for 6 h. The mixture was quenched with 0.5% (final concentration) TFA, zip-tipped and subjected to MALDI-ToF MS. (Y.S. Notebook 5 Page 68 – 69).

4.4.5 Incorporation of H-Hpg-OH using methionyl-tRNA synthetase

Methionine auxotrophic E. coli cells were transformed with a mutated procA/procM/pRSFDuet-1 vector. Modified M9aa starter culture\(^{44}\) inoculated with methionine auxotrophic E. coli was incubated at 37 °C overnight. An aliquot of 50 mL M9aa medium inoculated with 1 mL of the starter culture was grown at 37 °C to O.D.\(_{600}\) 0.55. The culture was centrifuged at 6,500 × g for 10 min and washed with 0.9% aqueous NaCl 3 times. The pellet was divided into two halves and each pellet was resuspended in 1 L M9aa medium without methionine, and shaken at 37 °C for 20 min. Then the two cultures were cooled on ice for 10 min, supplemented with 300 µg/mL H-Hpg-OH (O.D.\(_{600}\) = 0.7) or 200 µg/mL methionine (O.D.\(_{600}\) = 0.5) as a control experiment, and induced with 0.2 mM IPTG. The cultures were shaken at 18 °C for 20 h before harvesting. (Y.S. Notebook 6 Page 40 – 42, 55).

4.4.6 Construction of nisA/nisB/pRSFDuet-1 mutants and nisC/tRNA/pCDFDuet-1 containing an amber codon

The amber codon TAG was introduced into nisA/nisB/pRSFDuet-1 vector at different positions of nisA, generating co-expression vectors encoding NisA-T2B, NisA-M21B, NisA-K22B, NisA-S29B, and NisA-S33B. The mutants were generated using QuickChange PCR with the nisA/nisB/pRSFDuet-1 vector as the template. The primer
sequences are listed in Table 4.2. The PCR mixture included 1× Phusion HF buffer (New England BioLabs), dNTP mixture (2 mM each), primers (stock solution 100 μM each, final concentration 2 μM each), DMSO (4%, v/v), template plasmid nisA/nisB/pRSFDuet-1 (~ 20 ng) and Phusion High-Fidelity DNA Polymerase (0.02 U/μL). The PCR amplification was performed by thirty cycles of denaturing (98 °C for 10 s), annealing (57.5 °C for 30 s), and extending (72 °C for 150 s). The product was purified by QIAquick PCR Purification Kit (QIAGEN). An aliquot of the purified PCR product was mixed with 2 μL of 10x NEBuffer 4 and 1 μL of DpnI (stock solution 20,000 U/mL, New England Biolabs), followed by incubation at 37 °C for 1 h. DpnI was heat inactivated by incubation at 80 °C for 20 min. An aliquot of 5 μL of the resulting solution was used for E. coli DH5α heat shock transformation. (Y.S. Notebook 7 Page 78 – 80).

One copy of nisC was inserted into MCS2 of pCDFDuet-1 vector between BglII and XhoI restriction sites through PCR amplification, digestion, and ligation using nisC/pACYCDuet-1 as the PCR template. In order to increase the copy number presented in the co-expression host cells, one copy of pyrrolysyl-tRNA (pylT) was introduced into MCS1 of the constructed nisC/pCDFDuet-1 between BamHI and NotI restriction sites, using the pEVOL-pylT-mmPylRS(2X)-N346A/C348A vector as the PCR template. The general PCR, digestion and ligation procedures were the same as described in Chapter 2. The primer sequences are listed in Table 4.2. (Y.S. Notebook 9 Page 3 – 4).
Table 4.2 Primer sequences for cloning and mutagenesis of nisA, nisC and pylT.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer Sequence (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>proK_BamHI_FP</td>
<td>ATA ATA TGG GAT CCG TGT TCT CAA ATG CCT GAG GCC</td>
</tr>
<tr>
<td>tRNA_NotI_RP</td>
<td>ATA ATA TGG CCG CCG CCA TGC AAA AAA GCC TGC TCG TTG</td>
</tr>
<tr>
<td>NisA_T2B_FP</td>
<td>CAG GTG CAT CAC CAC GCA TTT AGA GTA TTT CGC TAT G</td>
</tr>
<tr>
<td>NisA_T2B_QRP</td>
<td>GCG TGG TGA TGC ACC TGA ATC TTT CTT CG</td>
</tr>
<tr>
<td>NisA_M21B_FP</td>
<td>GAG CTC TGA TGG GTT GTA ACT AGA AAA CAGCAA CTT G</td>
</tr>
<tr>
<td>NisA_M21B_QRP</td>
<td>GTT ACA ACC CAT CAG AGC TCC TGT TTT ACA ACC</td>
</tr>
<tr>
<td>NisA_K22B_FP</td>
<td>CTC TGA TGG GTT GTA ACA TGT AGA CAG CAA CTT GTC</td>
</tr>
<tr>
<td>NisA_K22B_QRP</td>
<td>CAT GTT ACA ACC CAT CAG AGC TCC TGT TTT AC</td>
</tr>
<tr>
<td>NisA_S29B_FP</td>
<td>GAA AAC AGC AAC TTG TCA ACG GAG G</td>
</tr>
<tr>
<td>NisA_S29B_QRP</td>
<td>ACA ATG ACA AGT TGC TGT TTT CAT GTT ACA ACC</td>
</tr>
<tr>
<td>NisA_S33B_FP</td>
<td>GTT ACC TAA TTA ACG AAA TTA ATA CTA CTC ACT ATA GGA G</td>
</tr>
<tr>
<td>NisA_S33B_QRP</td>
<td>CGT GAA TAC TAC AAT GAC AAG TTG CTG TTT TC</td>
</tr>
<tr>
<td>NisC_T7_PacI_FP</td>
<td>TCC AGG CAT TAA TTA ACG AAA TTA ATA CGA CTC ACT ATA GGA G</td>
</tr>
<tr>
<td>NisC_AvrII_RP</td>
<td>ATA ATA TGC CTA GGT CAT TTC CTC TTC CCT CTT TTC</td>
</tr>
</tbody>
</table>

4.4.7 Incorporation of non-canonical amino acids in NisA

*E. coli* BL21 (DE3) electro competent cells baring *pEVOL-pylT-mmPylRS(2X)-N346A/C348A* were co-transformed with mutated plasmids *nisA/nisB/pRSFDuet-1* and *nisC/tRNA/pCDFDuet-1*. Single colonies were picked and inoculated into 5 mL of modified M9\(^{43}\) or LB culture containing 50 mg/L kanamycin, 8.3 mg/L spectinomycin, and 8.3 mg/L chloramphenicol. An aliquot of 2 mL of the overnight culture was inoculated into 100 mL of LB or supplemented M9 medium. The culture was shaken at 37 °C until the O.D.\(_{600nm}\) reached 0.5, then L-Phe(3-CF\(_3\))-OH or L-Phe(3-Me)-OH dissolved in 1 M NaOH solution (for solubility) was added to the culture to obtain a final concentration of 2 mM. The culture was immediately neutralized with 5 M HCl to obtain pH ~ 7.5. Arabinose (stock solution concentration: 20%) was also added to obtain a final concentration of 0.2%. After shaking for 0.5 h, the culture was induced with 0.3 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and shaken continually at 18 °C for an
additional 20 h. The induced cells were harvested by centrifugation (11,900 × g for 20 min at 4 °C). The cell pellet was resuspended in 10 mL of LanA Buffer 1 and lysed by sonication (35% amplitude, 4.0 s pulse on, 9.9 s pause, 15 min). The soluble portion after centrifugation (23,700 × g, 20 min, 4 °C) was clarified using 0.45 µm syringe filters (Corning), then purified by immobilized metal affinity chromatography (IMAC) using a 1 mL HisTrap™ HP nickel affinity column (GE Healthcare Life Sciences). The filtered sample was applied to the column and the column was washed with 2 column volumes (CV) of LanA Buffer 1, followed by 2 - 3 CV of LanA Buffer 2, and then eluted with 3 CV of LanA Elution Buffer. The elution fractions were desalted and purified by solid phase extraction (SPE) using a 3 mL Vydac® BioSelect™ reversed-phase C4 column. A step gradient of 2 CV of 2% solvent B (0.086% TFA in 80% acetonitrile / 20% water) in solvent A (0.1% TFA in water), 1 CV of 50% solvent B in solvent A, 1 CV of 100% solvent B, and 1 CV of acetonitrile was applied for the purification. The fractions were tested by MALDI-ToF MS for the desired peptides. The majority of peptides were eluted in fractions of 50% solvent B in solvent A (Y.S. Notebook 9 Page 7 – 9).

An 18 µL sample of lyophilized 50%-solvent B-eluted fraction was incubated with 0.6 mg/mL trypsin in 50 mM tris buffer (pH 8.0) at room temperature for 0.5 h. The mixture was zip-tipped and subjected to MALDI-ToF MS. (Y.S. Notebook 9 Page 10).
Table 4.3 Plasmids mentioned in Chapter 4.

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>pRSFDuet-1</td>
<td>kan for kanamycin resistance, RSF origin</td>
</tr>
<tr>
<td>procA1.1-G−1K/F1B/procM/pRSFDuet-1</td>
<td></td>
</tr>
<tr>
<td>procA1.1-G−1K/F2B/procM/pRSFDuet-1</td>
<td></td>
</tr>
<tr>
<td>procA1.1-G−1K/F11B/procM/pRSFDuet-1</td>
<td></td>
</tr>
<tr>
<td>procA1.7-G−1R/M21B/procM/pRSFDuet-1</td>
<td></td>
</tr>
<tr>
<td>procA3.3-G−1K/Y15B/procM/pRSFDuet-1</td>
<td></td>
</tr>
<tr>
<td>procA3.3-G−1K/A23B/procM/pRSFDuet-1</td>
<td></td>
</tr>
<tr>
<td>nisA-T2B/nisB/pRSFDuet-1</td>
<td></td>
</tr>
<tr>
<td>nisA-J4V/S5B/L6G/nisB/pRSFDuet-1</td>
<td></td>
</tr>
<tr>
<td>nisA-M21B/nisB/pRSFDuet-1</td>
<td></td>
</tr>
<tr>
<td>nisA-K22B/nisB/pRSFDuet-1</td>
<td></td>
</tr>
<tr>
<td>nisA-S29B/nisB/pRSFDuet-1</td>
<td></td>
</tr>
<tr>
<td>nisA-S33B/nisB/pRSFDuet-1</td>
<td></td>
</tr>
<tr>
<td>pCDFDuet-1</td>
<td>aadA for spectinomycin resistance, CDF origin</td>
</tr>
<tr>
<td>nisClpylT/pCDFDuet-1</td>
<td></td>
</tr>
<tr>
<td>pACYCDuet-1</td>
<td>cat for chloramphenicol resistance, p15A origin</td>
</tr>
<tr>
<td>nisClpACYCDuet-1</td>
<td></td>
</tr>
<tr>
<td>pDule-Tyr</td>
<td>tet for tetracycline resistance, p15A origin, one copy of pBpaRS , one copy of tRNA_CUA</td>
</tr>
<tr>
<td>pEVL-Tyr</td>
<td>cat for chloramphenicol resistance, p15A origin, one copy of pBpaRS, one copy of tRNA_CUA</td>
</tr>
<tr>
<td>pEVL-pylT-mmPylRS(2X)-N346A-C348A</td>
<td>cat for chloramphenicol resistance, p15A origin, one copy of pyrolysyl-tRNA , two copies of engineered pyrolysyl RNA synthetase gene (mmPylRS )</td>
</tr>
</tbody>
</table>

4.5 REFERENCES


"A mass spectrometry-guided genome mining approach for natural product peptidogenomics".


(6) Wang, H.; Fewer, D. P.; Sivonen, K. *PLoS ONE* 2011, 6, e22384. "Genome mining demonstrates the widespread occurrence of gene clusters encoding bacteriocins in cyanobacteria".


(11) Bray, B. L. *Nat. Rev. Drug Discov.* 2003, 2, 587-593. "Large-scale manufacture of peptide therapeutics by chemical synthesis".


131


