DISCOVERY AND ENGINEERING OF LASSO PEPTIDES:
MOLECULAR SWITCHES, POST-TRANSLATIONAL
MODIFICATIONS, AND FUSION PROTEINS

CHUHAN ZONG

A DISSERTATION PRESENTED TO
THE FACULTY OF PRINCETON UNIVERSITY
IN CANDIDACY FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF CHEMISTRY
ADVISOR: PROFESSOR A. JAMES LINK

APRIL 2018
Abstract

Lasso peptides are a unique class of natural products, characterized by a threaded topology and remarkable stability. In addition, lasso peptides exhibit a wide range of bioactivities. These properties make them strong candidates for peptide therapeutics and as peptide scaffolds. Despite progress made in the discovery of novel lasso peptides and engineering of lasso peptide based scaffolds, there is still much remaining for further exploration. Thus, this thesis aims 1) to discover and characterize novel lasso peptide gene clusters, 2) to develop an engineering platform for expanding the lasso peptide based scaffolds for potential therapeutic applications.

Chapter 2 presents the optimization of a previous developed precursor-centric genome mining algorithm using new lasso peptide precursor sequence features and new motifs of lasso peptide maturation enzymes. Applying the updated algorithm on DNA contig sets, complete and draft genomes from bacteria and archaea, I was able to identify 948 high confidence putative lasso peptide gene clusters. In addition, I discovered several new gene clusters with novel precursor sequence features and post-translational elements.

I found two lasso peptide gene clusters associated with isopeptidases from Gram-negative organism Asticcacaulis benevestitus through the genome mining. In chapter 3, along with Michelle Wu and Jason Qin, I successfully expressed both peptides, named beneondin-1 and benendin-2, in E. coli. The molecular structure of benenodin-1 was solved using solution NMR. I found
benenodin-1 exhibits conformational switching between two distinct threaded conformers upon heat stimulation. This makes benenodin-1 the first natural peptide-based molecular switch. I also showed only the native conformer of benenodin-1 is cleaved by its associated isopeptidase, which may also be relevant to the biological function of these molecules.

Chapter 4 focuses on characterization of a lasso peptide, albusnodin, from Streptomyces albus DSM 41398. I found albusnodin is acetylated, the first example of a lasso peptide with this modification. Performing genetic studies with Wai Ling Cheung-Lee, I further showed that the acetyltransferse colocalized with the albusnodin gene cluster is required for the biosynthesis of this lasso peptide. In addition, bioinformatics analysis showed this type of lasso peptide is widespread in Actinobacteria, with 44 examples found in currently sequenced genomes.

Finally, in chapter 5, I developed a new engineering platform on the lasso peptide, astexin-1. I demonstrated a fusion of two model proteins, the artificial leucine zipper A1 and the superfolder variant of GFP, to the C-terminus of astexin-1. Moreover, along with Kenneth Hubbell and Shubham Chatterjee, I combined this fusion approach with cysteine bio-conjugation chemistry to graft two anticancer epitopes on astexin-1.

Overall, these exciting projects carried out in this thesis have enriched and opened new research areas in the field of lasso peptides.
Acknowledgement

My time at Princeton has been the most rewarding learning experience in my life so far. I would like to reflect on the people who have supported and helped me so much along the road to completing this thesis.

Firstly, I would like to express my special appreciation and thanks to my advisor Professor A. James Link: you have been a tremendous mentor for me. I would like to thank you for the continuous support of my research and for allowing me to grow as a research scientist.

I would like to thank the rest of my committee members, Professor Michael Hecht and Professor Mohammad Seyedsayamdost, not only for their insightful comments and encouragement but also for the hard questions.

I would like to thank current and former members in the Link lab for the stimulating discussions, for the sleepless nights we were working together, and for all the fun and friendships we have had in the past five years: Dr. Caitlin Allen, Wai Ling Cheung, Dr. Alan Futran, Joe Koos, Dr. Jack Lu, Dr. Mikhail Maksimov and Frank Piscotta.

Thank you all my dear undergraduates: Shubham Chatterjee, Cathy Chen, Maria Chen, Kenneth Hubbell, Caroline Kim, Nini Le, Jason Qin, Danielle Taylor, Catlin Wood, Michelle Wu, Aded Yako, and Lawrence Yu. It was such a great honor to mentor and know each of you. Your excellent problem-solving skills and creativity inspired me to resolve issues that might have easily escalated into roadblocks. More importantly, you gave me great opportunities to learn the things
that make a good mentor, a well-rounded manager, and a caring and kind person.

I would like to thank my advisors and mentors at the University of Minnesota Duluth: Professor Dr. Robert Carlson, Dr. Steven Hedman, Dr. Anne Hinderliter and Dr. Elizabeth Minor, for your support and encouragement through all these years. I would like to thank Mr. Chris Haidos who brought me to the United States.

My Ph.D. was made enjoyable in large part due to the many friends who became a part of my life. I am especially indebted to Li Chen, Cheng Cheng, Mircea Davidescu, Xinyi Fan, Alta Fang, Yaosheng Fu, Bo Guo, Siyang Han, Xiaogang He, Mengyi Jiang, Granthor Jindal, Hao Li, Sophia Li, Xiaoyuan Li, Keira Liu, Linglu Liu, Yajun Peng, Yue Qin, Xiao Shi, He Sun, Xin Teng, Rurun Wang, Tong Wu, Xue Wu, Yihan Wu, Qian Xu, Xiangtao Xu, Zhuozhi Yao, Aishan Zhao, Wang Zhan, Hao Zhang, Ran Zhang, and Boxiao Zheng.

The deepest thanks goes to my family. Words cannot express how grateful I am to my parents Xiaoping Wang and Jun Zong, and to my American mom and dad, Patricia and William Litchy, for your unconditional love. Your encouragement and belief in me were what sustained me thus far. I love you and I will always love you. Finally, I would like to express heartfelt appreciation to my beloved significant other: Tian Xie. You were always my stone-hard support in the tough moments when there was no one to answer my queries. The strong partnership, trust, and commitment that we built together along the way is a priceless treasure for our future. Let’s carry on, the best is yet to come.
# Contents

Abstract iii  
Acknowledgement v  
Table of Contents vi  
List of Figures x  
List of Tables xiii  

## Chapter 1  
1.1 Introduction 1  
1.2 Natural products 2  
1.3 Ribosomally synthesized and post-translationally modified peptides 3  
1.4 Lasso peptides 5  
1.4.1 Overview of lasso peptides 5  
1.4.2 Biosynthesis of lasso peptides 9  
1.4.3 Gene cluster architectures of lasso peptides 10  
1.4.4 Genome mining of lasso peptides 12  
1.4.5 Lasso peptide structure and stability 13  
1.4.6 Lasso peptide engineering 15  
1.5 Research overview 16  
1.6 References 19  

## Chapter 2  
2.1 Introduction 26  
2.2 Results 30  
2.2.1 Updating precursor peptide pattern 30  
2.2.2 Updating maturation enzyme motifs 31  
2.2.3 Overview of the precursor-centric algorithm 34  
2.2.4 Analysis of lasso peptide cluster hits 35  
2.2.5 Lasso peptide gene clusters from Asticcacaulis benevostitus 36  
2.2.6 Lasso peptide gene clusters from Sphingopyxis fribergensis 37  
2.2.7 Lasso peptide gene cluster from Streptomyces albus 39  
2.2.8 Lasso peptide cluster from Methanobacterium paludis 40  
2.3 Conclusion and Future Works 41  
2.4 References 44  

## Chapter 3  
3.1 Introduction 50
3.2 Results

3.2.1 Lasso peptide gene clusters from Asticcacaulis benevestitus

3.2.2 Heterologous expression of benenodin-1 lasso peptide

3.2.3 NMR structure of benenodin-1 ΔC5

3.2.4 Thermally induced conformational switching of benenodin-1

3.2.5 Kinetics and energetics of conformational switching

3.2.6 Conversion of benenodin-1 conformers into [2]rotaxanes

3.2.7 Mutagenesis of steric lock residues in benenodin-1

3.2.8 Substrate recognition by benenodin-1 isopeptidase

3.3 Conclusion

3.4 Future works

3.5 Materials and methods

3.5.1 Strains and reagents

3.5.2 Plasmid construction

3.5.3 Lasso peptide expression and purification

3.5.4 Benenodin-1 ΔC5, conformer 2 production

3.5.5 Heating assay on lasso peptides

3.5.6 NMR structure analysis of benenodin

3.5.7 BenE isopeptidase expression and purification

3.5.8 BenE and AtxE2 isopeptidase assay

3.5.9 Carboxypeptidase assay

3.5.10 Trypsin assay

3.6 Supplementary materials

3.6.1 Supplementary figures

3.6.2 Supplementary tables

3.7 References

Chapter 4

4.1 Introduction

4.2 Results

4.2.1 Discovery of the lasso peptide gene cluster of albusnodin

4.2.2 Heterologous expression of albusnodin

4.2.3 Characterization of albusnodin

4.2.4 Locating the acetylated residue

4.2.5 The cyclization residue of albsnodin

4.2.6 Conclusion and future direction

4.3 Materials and methods

4.3.1 Bioinformatics

4.3.2 Growth of S. albus and genomic DNA extraction

4.3.3 Plasmid construction and conjugation

4.3.4 Albusnodin expression and purification

4.3.5 HPLC analysis

4.3.6 MALDI analysis

4.3.7 LC-MS analysis

4.3.8 Carboxypeptidase assay

4.3.9 Trypsin assay

4.3.10 Chemical cleavage
Chapter 5

5.1 Introduction

5.2 Results

5.2.1 Astexin-1 as the model fusion peptide

5.2.2 Leucine zipper protein A1 as the model fusion protein

5.2.3 Direct fusion to astexin-1

5.2.4 Introduction of Flexible Linkers

5.2.5 Introduction of thrombin digestion site

5.2.6 Fusing Astexin-1 to GFP

5.2.7 Tail extension and epitope grafting

5.3 Conclusion

5.4 Materials and methods

5.4.1 Strains and reagents

5.4.2 Plasmid construction

5.4.3 Purification of astexin-1

5.4.4 Fusion protein expression

5.4.5 Fusion protein purification

5.4.6 ESI-MS analysis of intact fusion protein

5.4.7 Thrombin digestion of fusion proteins and HPLC analysis

5.4.8 Carboxypeptidase digestion assays

5.4.9 sfGFP fluorescence intensity measurement

5.5 References

Chapter 6
List of Figures

Chapter 1
Figure 1.1 General scheme for RiPP biosynthesis 4
Figure 1.2 General topology and nomenclature of lasso peptides. 5
Figure 1.3 Representatives of lasso peptides in each class 6
Figure 1.4 Schematic of the biosynthesis of lasso peptides 9
Figure 1.5 Examples of lasso peptide gene clusters with and without posttranslational modification enzymes 11
Figure 1.6 Schematic of lasso peptide unthreading via loop pulling and tail pulling 14

Chapter 2
Figure 2.1 The steps of the biosynthesis of lasso peptides 26
Figure 2.2 The updated lasso peptide precursor pattern 30
Figure 2.3 The conserved motifs in lasso peptide B homologs 32
Figure 2.4 The conserved motifs in lasso peptide C homologs 33
Figure 2.5 Outline of the precursor-centric genome mining method 34
Figure 2.6 Overview of the lasso peptide clusters from Asticcacaulis benevestitus 36
Figure 2.7 Overview of the lasso peptide clusters from Sphingopyxis fribergensis 37
Figure 2.8 The putative lasso peptide gene cluster from Streptomyces albus 38
Figure 2.9 Comparison of the lasso peptide clusters of paeninodin and paludinodin 39
Figure 2.10 The peptide sequences of fuscanodin and pandonodin 42

Chapter 3
Figure 3.1 The gene clusters and peptide sequences from Asticcacaulis benevestitus 51
Figure 3.2 HPLC and MALDI-MS of cell extracts of benenodin-1 52
Figure 3.3 HPLC and MALDI-MS of cell extracts of benenodin-2 53
Figure 3.4 NMR structure of benenodin-1 ΔC5 54
Figure 3.5 Benenodin-1 establishes an equilibrium between two conformers upon heating 55
Figure 3.6 Carboxypeptidase assays on benenodin-1, conformers1 56
Figure 3.7 Conformer 2 of benenodin-1 ΔC5 is partially unthreaded 57
Figure 3.8 Kinetics and energetics of conformational switching in benenodin-1 58
Figure 3.9 Arrhenius plots and van't Hoff plots 60
Figure 3.10 Generating a protease-cleavable variant of benenodin-1 63
Figure 3.11 The K17A variant of benenodin-1 ΔC5 unthreads upon heating at 95 °C 65
Figure 3.12 The Q15A variant of benenodin-1 ΔC5 does not undergo a conformation change upon heating at 95 °C 66
Figure 3.13 The E14A benenodin-1 \( \Delta C5 \) variant switches to conformer 2 upon heating
Figure 3.14 Homology modeling of the benenodin-1 isopeptidase BenE
Figure 3.15 Substrate specificity of benenodin-1 isopeptidase BenE
Figure 3.16 MS of isopeptidase cleavage reactions on benenodin-1 \( \Delta C5 \)
Figure 3.17 A chimeric substrate with the loop residues of astexin-3
Figure 3.18 Hydrolysis of a benenodin-1/astexin-3 chimera by BenE and AtxE2
Figure 3.19 Control isopeptidase hydrolysis experiments
Figure 3.20 Comparison of the backbone of the isopeptide-bonded rings of benenodin-1 and astexin-3 (PDB: 2M8F)
Figure 3.21 Isopeptidase reactions on benenodin-1 alanine variants
Figure 3.22 Isopeptidase reaction on alanine variants of benenodin-1 after heating
Figure 3.23 Potential benenodin based MIMs
Figure 3.24 Potential benenodin based polymers
Figure 3.25 Synthetic scheme of a daisy chain using lasso peptides as the starting materials
Figure 3.26 Schematic representation for the synthesis lassos peptide based polymers
Figure 3.27 NOESY spectrum of benenodin-1 \( \Delta C5 \), conformer 1
Figure 3.28 TOCSY spectrum of benenodin-1 \( \Delta C5 \), conformer 1
Figure 3.29 NOESY spectrum of benenodin-1 \( \Delta C5 \), conformer 2
Figure 3.30 TOCSY spectrum of benenodin-1 \( \Delta C5 \), conformer 2

Chapter 4
Figure 4.1 Acetylation reaction catalyzed by GCN5-related N-acetyltransferase (GNATs)
Figure 4.2 The gene cluster of albusnodin
Figure 4.3 Complete list of distinct lasso peptide precursors found in actinobacteria with an ACBT gene cluster architecture
Figure 4.4 Structure prediction of the acetyltransferase found in the albusnodin gene cluster
Figure 4.5 MALDI-MS spectra of partially purified supernatants from \textit{S. lividans} 66 cultures heterologously expressing albusnodin
Figure 4.6 MALDI-MS spectra of partially purified supernatants from cultures expressing albusnodin in \textit{S. coelicolor} M1146
Figure 4.7 MALDI-MS spectra of cell lysates of \textit{S. coelicolor} M1146 or \textit{S. lividans} 66
Figure 4.8 HPLC traces of albusnodin fractionation
Figure 4.9 LC trace (total ion chromatogram) of partially purified albusnodin using qTOF mass detector
Figure 4.10 MS/MS analysis of albusnodin
Figure 4.11 Additional MS/MS fragmentation of albusnodin
Figure 4.12 LC traces (mass detector) of partially purified albusnodin
Figure 4.13 Mass spectra of partially purified albusnodin
Figure 4.14  Chemical cleavage of albusnodin  116
Figure 4.15  Cartoon model of the likely topology of full length albusnodin  117

Chapter 5
Figure 5.1  Lasso peptide structure and gene cluster architecture  132
Figure 5.2  Electrospray mass spectra of direct fusions of astexin-1 to A1 leucine zipper  133
Figure 5.3  Maturation of astexin-1 precursor fusion proteins  135
Figure 5.4  Mass spectrometry of astexin-1-A1 fusion proteins with flexible linkers  136
Figure 5.5  Deconvoluted mass spectra of astexin-1-(GSSG)x-A1 fusion proteins  137
Figure 5.6  Quantification of the extent of lasso formation in astexin-1 fusion proteins  139
Figure 5.7  MALDI-MS of thrombin-cleaved lasso and linear peptides  140
Figure 5.8  MALDI-MS of carboxypeptidase-digested lasso peptides  142
Figure 5.9  Fusion of astexin-1 to superfolder GFP (sfGFP)  143
Figure 5.10  Comparison of the fluorescence of astexin-1-(GSSG)5-Thrombin superfolder GFP (red) to native superfolder GFP (blue)  144
Figure 5.11  Production schematic of tail extended lasso epitope  145
Figure 5.12  Schematic of VEGF lasso epitope targeting VEGF receptor  146
Figure 5.13  Schematic of TEAD4 lasso epitope interfering YAP to bind TEAD4  147
## List of Tables

### Chapter 1
Table 1.1 The list of all the characterized lasso peptides 8

### Chapter 2
Table 2.1 Lasso peptide gene clusters in the MEME training set 31
Table 2.2 The analysis of cluster hits in this study 34

### Chapter 3
Table 3.1 Kinetic data of benenodin-1 conformational switching 62
Table 3.S1 Proton assignments for Conformers 1 and 2 91
Table 3.S2 Lists of NOEs observed between ring residues of benenodin-1 ΔC5 92
Table 3.S3 Sequences of primers used in Chapter 3 93
Table 3.S4 Plasmids constructed in Chapter 3 94

### Chapter 4
Table 4.1 Plasmids used in Chapter 4 119
Table 4.2 Primers used in Chapter 4 120
Table 4.3 Solvent gradient 1 122
Table 4.4 Solvent gradient 2 122
Table 4.5 Solvent gradient 3 123

### Chapter 5
Table 5.1 Sequences of bioactive epitopes that are grafted on lasso peptide astexin-1(V6C) 146
Table 5.S1 Plasmids used in Chapter 5 155
Table 5.S2 Primers used in Chapter 5 156
Chapter 1

Introduction

1.1 Introduction

Lasso peptides are an emerging class of ribosomally synthesized and post-translationally modified natural products, characterized by a tremendously stable knot structure. They carry out diverse bioactivities ranging from antimicrobial activity to HIV inhibition. In addition, lasso peptides exhibit outstanding stability against enzymatic, chemical and thermal degradation. These interesting features make lasso peptides promising candidates for peptide drug discovery and as scaffolds for peptide epitopes. In recent years, genome mining and bioinformatics tools have helped scientists and researchers to uncover hundreds of putative lasso peptide gene clusters. In addition, methods of heterologous expression and new chemical biology techniques have been widely applied in the production and characterization of lasso peptides, especially for lasso peptides from gram-negative organisms. In the area of engineering, saturation mutagenesis and epitope grafting have been used to redesign and engineer the well-studied lasso peptide microcin J25. Despite the progress made in lasso peptide fields, there is still much that remains to be done in discovery, characterization and engineering. First, we need to improve existing gene mining methods to discover novel gene clusters that diverge from known lasso peptide examples. Second, we need to develop a robust heterologous expression and characterization system for the understudied lasso peptides from gram-positive
organisms. Third, we need to build a lasso peptide display system to screen lasso peptides for new functions in a high-throughput fashion.

To attack these problems, this thesis focuses on 1) optimizing a genome mining method for novel lasso peptide discovery, 2) heterologous expression and structural characterization of novel lasso peptides from both gram positive and negative organisms, and 3) developing an engineering platform for constructing lasso peptide fusion protein for future peptide display application and lasso-peptide-based epitopes grafting. This introductory chapter provides background information on these mentioned topics and presents a research outlook for the entire thesis.

1.2 Natural products

Natural products are primary and secondary metabolites produced by living organisms. The history of modern medicine is full of remarkable stories of the discovery of natural products that led to the birth of pharmaceutical industry and profoundly impacted human health.\(^1\) Natural-product-based drugs, with a market size of $250 billion, account for about sixty percent of the approved drugs in today’s pharmaceutical market.\(^2\) Natural-product-based drugs cover a broad spectrum of therapeutic areas such as infectious, immunological, cardiovascular, logical and oncology applications.\(^3,4,5\) For instance, penicillin is a well-known natural product that is used as an antibiotic drug to treat bacterial infections caused by *Staphylococci* and *Streptococci*.\(^6\) Another example, Taxol, an anticancer drug isolated from Pacific yew, is used to treat a number of types of
cancers including ovarian cancer, breast cancer, and lung cancer. Taxol prevents the progression of mitosis by stabilizing microtubule polymers from disassembly and thus inhibits cell division. The enormous structural and chemical diversity of natural products continue to inspire generations of medicinal chemists to discover and optimize potential drug leads and develop novel therapeutics.

Traditionally natural product discovery focused on microbial and plant extraction, fermentation, and purification for bioactivity screening and structure elucidation. The advancement in genome sequencing and high-throughput characterization technologies of natural products has fundamentally changed the landscape of natural product research. Bioinformatics tools have allowed us to bridge biosynthetic gene clusters to known natural products and predict yet-to-be isolated novel compounds on the basis of gene sequences. Mass spectrometry tools have enabled us to detect low abundance natural products. Synthetic biology and heterologous host engineering methods have empowered us to generate novel natural products by engineering genes governing their biosynthesis. Despite the developments in this field, there are still far more novel natural products awaiting discovery.

1.3 Ribosomally Synthesized and Post-translationally Modified Peptides

Ribosomally synthesized and post-translationally modified peptides (RiPPs) are an emerging class of natural products. Their wide-range of biological activities and great chemical diversity have excited new research interest in RiPPs. Comparing to non-ribosomal peptides and polyketides, the chemical
structures of RiPPs are closely linked with their genetic-coded ribosomal peptide precursors.\textsuperscript{15} Thus RiPPs can be predictable from their genomic data, and they have gained increasing interest in the research community for bioinformatics-guided natural product discovery.\textsuperscript{16}

**Figure 1.1** General scheme for RiPP biosynthesis. First, the precursor peptide is produced by ribosomes. Second, the precursor peptide is post-translationally modified by the maturation enzymes, where the leader peptide is removed and the core peptide is transformed into the mature product.

The biosynthetic gene cluster of a typical RiPP includes a peptide precursor gene and a set of genes encoding for post-translational tailoring enzymes which later on assemble together as biosynthetic machinery. The peptide precursor consists of an N-terminal leader peptide and a C-terminal core peptide containing post-translational modification sites. The assembled biosynthetic machinery recognizes and removes the leader peptide and the
modified core peptide becomes the final RiPP, Figure 1.1.

Based on their biosynthetic machinery and structure features, RiPPs are grouped in different subclasses including lanthipeptide, cyanobactins, and amatoxins. Up to date, there are more than 20 subclasses of RiPPs that have been characterized involving many types of post-translational modifications such as alkylation, acetylation, cyclization, glycosylation, and sulfonation. The extensive post-translational modifications on RiPPs allow RiPPs to carry tremendous structural diversity and empower them with a wide range of bioactivities such as antifungal, antibacterial, and antiviral activities.

1.4 Lasso peptides

1.4.1 Overview of lasso peptides

Lasso peptides are a type of RiPPs that are characterized by a unique threaded slipknot shape, Figure 1.2. A typical lasso peptide contains a macrolactam ring forming via an isopeptide bond between its N-terminus and an aspartate or glutamate side chain at the 7 to 9th position. The C-terminal portion of the peptide gets threaded through the ring to create the lasso fold. The region after the isopeptide bond and before passing through the ring is the loop, and the portion with the C-terminus after the ring is the tail of the lasso peptide, Figure 1.2.
Figure 1.2 General topology and nomenclature of lasso peptides. This figure is adopted from the reference 25.

Lasso peptides are subdivided into class I, class II, and class III based on the number of disulfide bonds they exhibit. Some lasso peptide representatives are shown in Figure 1.3.

Class I lasso peptides carry two disulfide bonds, and the N-terminal residue of the core peptide is a cysteine. In addition to the thread, the structures of Class I lasso peptides are stabilized by two disulfide bridges that connect the ring to the loop and tail. BI-32169 is the only representative of the third class of lasso peptides and has a single disulfide bond linking the ring and the tail.26 So far, Class II lasso peptides are the largest subclass, containing more than 40 characterized members, Table 1.1.24 Class II lasso peptides have no disulfide bond and use steric interactions to help maintain the threaded shape. Usually, there are bulky amino acids that serve as steric locks sitting right above and below the ring, preventing the tail from sliding and unthreading out of the ring.24,26 This thesis is mainly focused on Class II lasso peptides.
### Table 1.1

List of all characterized lasso peptides.

<table>
<thead>
<tr>
<th>Class I</th>
<th>Lasso peptide from Gram-negative organisms</th>
<th>Lasso peptide from Gram-positive organisms</th>
</tr>
</thead>
<tbody>
<tr>
<td>aborycin</td>
<td>LGIGSGNFAOGYAVVFW</td>
<td>Actinokineospora sphericospiaegae</td>
</tr>
<tr>
<td>humidimycin</td>
<td>LGIGSFGQNYAWA</td>
<td>Acinetobacter gyllebernii</td>
</tr>
<tr>
<td>siamycin I</td>
<td>LGVGSGNFAOGYAVVFW</td>
<td>Asticacaulia excentrica</td>
</tr>
<tr>
<td>siamycin II</td>
<td>LGIGSGNFAOGYAVVFW</td>
<td>Asticacaulia excentrica</td>
</tr>
<tr>
<td>stievecin</td>
<td>VVGGTDNPFGKTTAW</td>
<td>Burkholderia rhizoxica</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Class II</th>
<th>Lasso peptide from Gram-negative organisms</th>
<th>Lasso peptide from Gram-positive organisms</th>
</tr>
</thead>
<tbody>
<tr>
<td>acinetolin</td>
<td>GSKGFPF-EFWTEGNYG</td>
<td>Acinetobacter gyllebernii</td>
</tr>
<tr>
<td>astexin-1</td>
<td>LGSGVFWPSQVQQQFFSESRNQD</td>
<td>Asticacaulia excentrica</td>
</tr>
<tr>
<td>astexin-2</td>
<td>LGVQGALQVSVGQFWQGLAD</td>
<td>Asticacaulia excentrica</td>
</tr>
<tr>
<td>astexin-3</td>
<td>GTPMVGGLDSVSGQYQAPLAD</td>
<td>Asticacaulia excentrica</td>
</tr>
<tr>
<td>burhizin</td>
<td>GAGQYK-EVAPONGWDD</td>
<td>Burkholderia rhizoxica</td>
</tr>
<tr>
<td>capistrin</td>
<td>GQFQFTPAARQSRGFPN</td>
<td>Burkholderia thailandensis</td>
</tr>
<tr>
<td>caulocrinol I</td>
<td>GDLNAP-EPEGREPTGLSRE</td>
<td>Caulobacter sp.</td>
</tr>
<tr>
<td>caulocrinol II</td>
<td>GDLFAP-APEVGRPRNQSL</td>
<td>Caulobacter sp.</td>
</tr>
<tr>
<td>caulocrinol III</td>
<td>GQITRNP-SVIGACIEGQRQ</td>
<td>Caulobacter sp.</td>
</tr>
<tr>
<td>caulocrinol IV</td>
<td>SFVNVGTVQLVGSQF</td>
<td>Caulobacter sp.</td>
</tr>
<tr>
<td>caulocrinol V</td>
<td>SIEGDGLRBNBSQTYK</td>
<td>Caulobacter sp.</td>
</tr>
<tr>
<td>caulocrinol VI</td>
<td>AGTVLPLPETNQIKFYDAP</td>
<td>Caulobacter sp.</td>
</tr>
<tr>
<td>caulocrinol VII</td>
<td>SGIQGFPFTEPNRRW</td>
<td>Caulobacter sp.</td>
</tr>
<tr>
<td>caulosegin</td>
<td>GAVQGQP-2AVNPGLREIQG</td>
<td>Caulobacter segnis</td>
</tr>
<tr>
<td>caulosegin II</td>
<td>GATLGPLEDFLPGNYMP</td>
<td>Caulobacter segnis</td>
</tr>
<tr>
<td>caulosegin III</td>
<td>GAVYGLLEDITARYDM</td>
<td>Caulobacter segnis</td>
</tr>
<tr>
<td>klebaldin</td>
<td>GSGIPGI-EPFNVNQH</td>
<td>Klebsiella pneumoniae</td>
</tr>
<tr>
<td>microcin J25</td>
<td>GQAGSHP-2EFGCSPISFYG</td>
<td>Escherichia coli</td>
</tr>
<tr>
<td>rhodanoin</td>
<td>CVLPGFQ-FEPNGHAPPQRT</td>
<td>Rhodobacter thioxydans</td>
</tr>
<tr>
<td>rubrivin</td>
<td>GSPGLQNSIENPAPQORQV</td>
<td>Rubrivivans gelatinosus</td>
</tr>
<tr>
<td>sphingonodin I</td>
<td>GFGITG-GVQFLNSEQ</td>
<td>Sphingobium japonicum</td>
</tr>
<tr>
<td>sphingonodin II</td>
<td>MGSSGQZ-DONGQKPNSLGSDD</td>
<td>Sphingobium japonicum</td>
</tr>
<tr>
<td>sphingopyxin I</td>
<td>GIEGFGVQPEDEQGEHFLAQGITTADD</td>
<td>Sphingopyxis alaskensis</td>
</tr>
<tr>
<td>sphingopyxin II</td>
<td>GEAQQQGQFTZQAGIQ</td>
<td>Sphingopyxis alaskensis</td>
</tr>
<tr>
<td>subterminalis</td>
<td>GSPFQGRL-EPFVLAQ</td>
<td>Sphingomonas subterranea</td>
</tr>
<tr>
<td>syanadin</td>
<td>GSGGTV-2APAGGQLAGG</td>
<td>Sphingobium yanoikuyae</td>
</tr>
<tr>
<td>xanthomonin I</td>
<td>GEPFLA--EJEGGPNQIS</td>
<td>Xanthomonas gardneri</td>
</tr>
<tr>
<td>xanthomonin III</td>
<td>GGAGAG--EYNGPASGIAE</td>
<td>Xanthomonas gardneri</td>
</tr>
<tr>
<td>xanthomonin</td>
<td>GIGGDG-DFGQKPDV</td>
<td>Phenylbacterium zucineum</td>
</tr>
</tbody>
</table>

### Class III | Lasso peptide from Gram-positive organisms | Lasso peptide from Gram-positive organisms |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>actinoin</td>
<td>GYPMSNROSPGFTYFG</td>
<td>Actinokineospora sphericospiaegae</td>
</tr>
<tr>
<td>lassomycin</td>
<td>LRRFA-QLQVRNRI</td>
<td>Lentzea kentuckyensis</td>
</tr>
<tr>
<td>propeptin</td>
<td>GFYMWTDLFLGVTTFSP</td>
<td>Microbiosa sp.</td>
</tr>
<tr>
<td>peenin</td>
<td>AGQGTSFAPAQDPEDVEYD</td>
<td>Paeonbacillus dendiriformis</td>
</tr>
<tr>
<td>sphaerican</td>
<td>LGFNVWPEPQGNYFPI</td>
<td>Planomonospora sphaerica</td>
</tr>
<tr>
<td>lariatin A</td>
<td>QSVLVR-2WGHzVXIKP</td>
<td>Rhodococcus jostii K01-0171</td>
</tr>
<tr>
<td>lariatin B</td>
<td>QSVLVR-2WGHzVXIKP</td>
<td>Rhodococcus jostii K01-0171</td>
</tr>
<tr>
<td>streptomycosemin</td>
<td>LSIGSPYNOILGYPALIVYYF</td>
<td>Streptomycospora alba YIM 00003</td>
</tr>
<tr>
<td>anatin</td>
<td>GIQGWGN-DFGQKPDV</td>
<td>Streptomyces coerulescens</td>
</tr>
<tr>
<td>cattilein</td>
<td>SYNHGWDEHWNHWQGW</td>
<td>Streptomyces sivice ATCC 20983</td>
</tr>
<tr>
<td>chaxapeptin</td>
<td>GFSGKFL-ISFGLNNF</td>
<td>Streptomyces leewenhoekii</td>
</tr>
<tr>
<td>citralassin A</td>
<td>LGQRAG-LRQLVSKN</td>
<td>Streptomyces albus</td>
</tr>
<tr>
<td>lagmycin</td>
<td>LGASQSF-PLGQSLL</td>
<td>Streptomyces sp.</td>
</tr>
<tr>
<td>moomiycin</td>
<td>SYNHGWDEHWNHWQGW</td>
<td>Streptomyces sivice ATCC 20983</td>
</tr>
<tr>
<td>RES-701</td>
<td>GNMQZAPWPNFYVY</td>
<td>Streptomyces sivice ATCC 20983</td>
</tr>
<tr>
<td>so15-2005</td>
<td>GVFYGKQVYWSRII</td>
<td>Streptomyces roesporous</td>
</tr>
<tr>
<td>sungsanpin</td>
<td>GFSGKFL-ISFGLNW</td>
<td>Streptomyces sp.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Class III</th>
<th>Lasso peptide from Gram-negative organisms</th>
<th>Lasso peptide from Gram-positive organisms</th>
</tr>
</thead>
<tbody>
<tr>
<td>BI-32169</td>
<td>GLPGW-PSDIGPGWNTFPAC</td>
<td>Streptomyces sp.</td>
</tr>
</tbody>
</table>
Figure 1.3  Representatives of lasso peptides in each class. The structures were resolved either using solution NMR or X-Ray crystallography.

Some of the characterized lasso peptides have specific bioactivities including antimicrobial and antiviral activities. For instance, microcin J25, the most investigated lasso peptide, is known to inhibit RNA polymerase and serve as an antimicrobial molecule towards *E. coli, Salmonella, and Shigella* species.\textsuperscript{27} Lassomycin is reported to poison *Mycobacterium tuberculosis* by stimulating the
activity of an ATP-dependent protease. RP 71955 has been found to inhibit the HIV-1 aspartic proteinase and the production of HIV reverse transcriptase.

1.4.2 Biosynthesis of lasso peptides

In general, the biosynthesis of lasso peptides involves three essential genes: A encoding for the precursor peptide, B encoding for a cysteine protease homolog and C encoding for an ATP dependent asparagine synthase homolog. In brief, once the precursor peptide A gets produced from the ribosome, the maturation machinery consisting of B and C proteins then processes the precursor to a final lasso, Figure 1.4.

The precursor peptide is usually around 50 amino acids long with a leader peptide (~30 aa) and a core peptide (~20 aa). In the leader peptide, a highly-conserved threonine is located at the penultimate position. Previous mutagenesis studies of lasso peptides such as microcin J25 and capistruin suggested that the threonine residue may be a recognition element for docking the precursor on the maturation machinery. In the core peptide, the residue at position 1 and ring formation residue aspartate or glutamate at position 7, 8, or 9 are crucial for lasso peptide synthesis. Among the characterized Class II lasso peptides, the residue at position 1 is usually a glycine, with a few exceptions where lasso peptide caulonodin IV-VI start with alanine, and streptomonomicin starts with serine. The B protein, which contains a Cys-His-Asp catalytic triad similar to cysteine proteases and transglutaminases, catalyzes the cleavage of the leader peptide. The C protein catalyzes the backbone-sidechain isopeptide
bond formation and matures the peptide into its final threaded shape.\textsuperscript{26}

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{figure14.png}
\caption{Schematic of the biosynthesis of lasso peptides.}
\end{figure}

1.4.3 Gene cluster architectures of lasso peptides

Besides the three crucial genes of ABC for lasso peptide biosynthesis, additional conserved features and gene variations have been identified in lasso peptide gene clusters. The gene clusters of lasso peptides such as microcin J25 (\textit{mcjABCD}) and capistruin (\textit{capABCD}) include a fourth gene encoding an ABC transporter, McjD or CapD. The transporter exports lasso peptides out of cells for self-defense related activities. In contrast to microcin J25 and capistruin, the B gene of lasso peptide lariatin is split into two genes, resulting in a gene cluster of \textit{larAC B1B2D}. The B1 protein is a homolog of the PqqD enzyme superfamily that binds to the precursor peptide, while the B2 protein carrying a cysteine catalytic
The presence of an ABC transporter correlates with antimicrobial activities of lasso peptides. Many characterized ABC-transporter associated lasso peptides obtain antimicrobial activities. 24

The discovery of the astexin family of lasso peptides has revealed a new type of cluster architecture atxABCEFIR where the clusters lack the transporter gene D. 37,38,39 Instead, the clusters include genes encoding for an isopeptidase (E), a putative TonB-dependent receptor (F) and a σ-/anti-σ-factor pair (I/R) as their conserved features. The isopeptidase, homologous to serine proteases, highly specifically linearizes its associated lasso peptide. The TonB-dependent receptor and a σ-/anti-σ-factor pair are used for siderophore uptake and regulation. Thus lasso peptides connected to isopeptidases are proposed to have metal uptake functions. 25

**Figure 1.5** Examples of lasso peptide gene clusters with and without posttranslational modification enzymes.

Recent genome mining studies and natural product isolation studies have also revealed lasso peptide gene clusters contain additional PTM elements.
involving phosphorylation, C-terminal methylation. A summary of characterized lasso gene clusters is shown in Figure 1.5.

1.4.4 Genome mining of lasso peptides

With the explosion of genome data, the discovery of lasso peptides has shifted from using screening methods to genome-guided approaches. Currently, there are two popular genome-guided methods for the discovery of lasso peptides, namely homology-based genome mining and precursor-centric genome mining. Homology-based genome mining uses the characterized maturation enzymes B and C as the templates for BLAST searches. Once the hits of B homologs and C homologs are identified, the precursor A genes then get manually searched and annotated. This method allowed the discovery of lasso peptides such as capistruin and caulosegnins. The Magarvey group and the Mitchell group recently updated this method using a combination of profile hidden Markov models of B and C enzymes training instead of using BLAST.

The precursor-centric genome mining method developed in our group is an automated method for surveying the gene clusters of lasso peptides through all genomes. This method first searches small open reading frames that match the lasso peptide precursor pattern with conserved chemical features and sequence length restrictions. For each precursor hit, open reading frames within 10,000 bases of the precursor were searched for the conserved motifs identified in the B and C homologs of lasso peptides using the MEME software suite. The number of motifs matched was termed into the rank, a score which was used
to determine the likelihood of putative lasso peptide clusters. This method led to the identification of the astexin family of lasso peptides. Since 2012, there are lasso peptides with new precursor sequence features and maturation enzyme motifs have been discovered. It is necessary to improve this precursor-centric genome mining algorithm by including the new findings. In addition, the optimized algorithm should be applied to survey the updated genome database and to discover many new families of lasso peptides.

1.4.5 Lasso peptide structure and stability

While all the lasso peptides share a slipknot-like 3D topology, lasso peptides have diverse structural features and sequences. For Class II lasso peptides, the size of the ring varies from 7 to 9 amino acids. In addition, there is no restriction on amino acid compositions of the ring sequence except for isopeptide-bonded residues. The loops and tails also exhibit a high degree of variability in sizes and amino acid compositions. The steric locks holding the thread are usually large and hydrophobic residues such as phenylalanine, tyrosine, and tryptophan. For smaller ring size lasso peptides, smaller residues like arginine, lysine, and isoleucine can also be used, including caulonodins and caulosegnins.

The slipknot structure of lasso peptides empowers many of these molecules with hyper-stability against enzymatic, chemical and thermal degradation. For example, lasso peptides like microcin J25 and capistruin exhibit extraordinary thermostability. However, in the recent studies, more examples
of lasso peptides are found to be thermolabile, undergoing unthreading when subjected to temperatures above physiological conditions, including astexins and caulosegnins. The unthreading process of lasso peptides can be monitored using HPLC, mass spectrometry, and carboxypeptidase assays.

**Figure 1.6** Schematic of lasso peptide unthreading via loop pulling and tail pulling

Lasso peptides are classified as [1]rotaxane molecules. [1]Rotaxane is assembled from a single chemical backbone but still threads a ring and has capped ends on the thread, **Figure 1.6**. During the unthreading process, the ring of a lasso peptide slides between the loop and tail, **Figure 1.6**. Because of this, lasso peptides are interesting starting materials for molecular switches where the
position of the ring on the thread can be controlled with certain stimuli including heat, pH, and current changes. It is one of my research interests to characterize new lasso peptides with molecular switching behaviors and to construct lasso peptide-based molecular motors.

1.4.6 Lasso peptide engineering

Some lasso peptides exhibit tremendous stability against proteolytic and thermal stress. This makes lasso peptides promising scaffolds for peptide-based drug discovery and engineering. Since lasso peptides are coded by the precursor genes and synthesized by the ribosome, the engineering of lasso peptides is more straightforward. Several engineering studies were conducted on the well-known lasso peptide microcin J25. Pan et al. screened a library of microcin J25 mutants and examined the ability to take multiple mutations and still produce lasso folding to increase antimicrobial activity. 49 Piscotta et al. successfully incorporated unnatural amino acids at four positions in lasso peptide microcin J25. 50 This effort opened up opportunities to further expand the chemical diversity of lasso peptides. In addition, the Marahiel group inserted a tripeptide, Arg-Gly-Asp(RGD), into the loop area of microcin J25. The resulting peptide microcin J25-RGD was successfully produced and acted as an integrin antagonist with nanomolar affinity. 51 We believe that we can carry out similar engineering approaches and implement new engineering platforms on other lasso peptides. In addition, high-throughput methods for directed evolution of lasso peptides screening for new functions and bio-activities are still waiting to be developed. Thus, I am interested in constructing a platform to fuse lasso peptides
to carrier proteins, which is the first step for developing a lasso peptide display system.

1.5 Research overview

This thesis mainly focuses on the discovery, characterization and engineering of lasso peptides. In Chapter 2, I optimized the precursor-centric genome mining algorithm with new lasso peptide precursor patterns and maturation enzyme motifs. By applying the updated algorithm on full/draft genomes and DNA contig sets from bacteria and archaea, I was able to identify many new putative lasso peptide gene clusters with novel peptide sequence features and cluster architectures that have not been revealed before. Among these putative clusters, I chose the clusters from *Asticcacaulis benevestitus* and *Streptomyces albus* for further characterization in Chapter 3 and Chapter 4 accordingly.

Chapter 3 describes our in-depth investigation of two gene clusters from *Asticcacaulis benevestitus*. In collaboration with Michelle Wu and Jason Qin, I was able to heterologously express two lasso peptides from *Asticcacaulis benevestitus*, named benenodin-1 and benenodin-2, in *E. coli*. The low solubility of benenodin-2 prevented it from being further studied. I solved the solution NMR structure and conducted a series of thermal stability tests of benenodin-1. Surprisingly, benenodin-1 exhibits conformational switching between two distinct threaded conformers upon actuation by heat. I also characterized the kinetics and energetics of the conformational switch. In addition, I found that only the
native benenodin-1 can be recognized and cleaved by its associated isopeptidase. This thermally induced conformational switching of benenodin-1 may also be relevant to the biological function of this molecule.

In Chapter 4, I characterized a lasso peptide, albusnodin, from *Streptomyces albus*. Along with Wai Ling Cheung-Lee, I used heterologous host engineering techniques to express albusnodin in two *Streptomyces* stains. Unlike other known lasso peptides, albusnodin carries a new post-translational modification: acetylation of lysine-10. I showed through genetic and biochemical studies that the gene *albT*, encoding an acetyltransferase within the lasso peptide gene cluster, is essential for the biosynthesis of albusnodin. This study provided new insights into diverse post-translational modifications on lasso peptides and illustrates a heterologous expression system for lasso peptides from Gram-positive bacteria.

The work in Chapter 5 laid out the foundation for a high-throughput screening system on lasso peptide libraries and a platform for engineering lasso peptides as a robust reprogrammable molecular scaffold. I demonstrated a fusion of two model proteins, the artificial leucine zipper A1 and the superfolder variant of GFP, to the C-terminus of the lasso peptide astexin-1. The ability to fuse lasso peptides to a protein of interest opens the doors for the development of lasso peptide display and screening systems. Moreover, in collaboration with Kenneth Hubbell and Shubham Chatterjee, I combined this fusion approach with cysteine bio-conjugation chemistry to graft two anticancer epitopes on astexin-1. This affirmed the possibility of using lasso peptides as a stable scaffold for small
peptide sequences with promising pharmaceutical applications.

Finally, in Chapter 6, I reviewed our findings and discussed the impact and future directions in the discovery and engineering of lasso peptides.
1.6 References


Ling, L.; Fettermann, A.; Hughes, D.; Bissell, A.; et al. Lassomycin, a
ribosomally synthesized cyclic peptide, kills mycobacterium tuberculosis by

(29) Helynck, Gerard; Dubertret, Catherine; Mayaux, J. F. L. J. Isolation of RP
71955, a new anti-HIV-1 peptide secondary metabolite. *J. Antibiot. (Tokyo)*.
1993, 46 (11), 1756–1757.

(30) Maksimov, M. O.; Pan, S. J.; Link, A. J. Lasso peptides: structure, function,

(31) Cheung, W. L.; Pan, S. J.; Link, A. J. Much of the microcin J25 leader

(32) Knappe, T. A.; Linne, U.; Robbel, L.; Marahiel, M. A. Insights into the
2009, 16 (12), 1290–1298.

(33) Zimmermann, M.; Hegemann, J. D.; Xie, X.; Marahiel, M. A.
Characterization of caulonodin lasso peptides revealed unprecedented N-
terminus residues and a precursor motif essential for peptide maturation.

(34) Metelev, M.; Tietz, J. I.; Melby, J. O.; Blair, P. M.; Zhu, L.; Livnat, I.;
Severinov, K.; Mitchell, D. A. Structure, bioactivity, and resistance
mechanism of streptomonomicin, an unusual lasso peptide from an

Marahiel, M. A. The B1 Protein Guides the Biosynthesis of a Lasso

(36) Cheung, W. L.; Chen, M. Y.; Maksimov, M. O.; Link, A. J. Lasso peptide


(44) Tietz, J. I.; Schwalen, C. J.; Patel, P. S.; Maxson, T.; Blair, P. M.; Tai, H.-C.; Zakai, U. I.; Mitchell, D. A. A new genome-mining tool redefines the


Chapter 2

Genome Mining of Lasso Peptides

Recently genome-mining tools have been widely used in the discovery of new biosynthetic gene clusters (BGCs) and predict novel natural products. Lasso peptides are a family of natural products with a unique 3D structure resembling a threaded slipknot. Previous work in our lab, we developed a precursor-centric genome mining method that allowed us to reveal putative lasso peptide gene clusters across bacteria strains. In this chapter, I optimized this precursor-centric method using a new lasso peptide precursor pattern and updated training sets of lasso peptide maturation enzymes. Our approach yielded a comprehensive survey of genomes from bacteria and archaea, identifying 948 putative lasso peptide gene clusters. I also revealed novel precursor sequence features and post-translational elements among several newly discovered gene clusters. This work lays out the genetic and bioinformatics foundation to conduct further investigations on the characterization of these novel clusters in later chapters.
2.1 Introduction

Genome sequencing is becoming increasingly important and has opened up new research areas in microbiology and natural product discovery.\textsuperscript{1} Up to date, thousands of microbes have been fully sequenced and are publicly available across many research communities.\textsuperscript{2} In the meantime, new sets of genomes are assembled and uploaded into the public domain at an accelerating rate.\textsuperscript{3} Among the wealth of microbial genome data, a large number and diversity of biosynthetic gene clusters (BGCs) await to be annotated and linked with their encoded natural products.\textsuperscript{4} To discover novel BGCs and predict their associated natural products, the development of effective bioinformatics mining algorithms bridging BGCs to molecules becomes the major task in natural product research community.

Along with the well-studied natural products such as polyketide and non-ribosomal peptide metabolites, another family of natural products, ribosomally synthesized and post-translationally modified peptides (RiPPs), have gathered increasing attention. RiPPs usually consist of a leader peptide and a core peptide. The leader peptide guides post-translational modification enzymes tailoring on the precursor peptide, and the core peptide becomes the final product after removal of the leader peptide by proteases.\textsuperscript{5} Based on the high sequence similarity between precursor peptides and final RiPPs products, and the conserved arrangement of post-translational machinery genes, many RiPPs products can be predicted from genome sequence data.\textsuperscript{6} The conventional RiPPs genome mining methods focus on searching for the genes encoding
maturation enzymes involved in the biosynthesis of RiPPs products. This can be done manually with tools such as BLAST. Recently, several automated RiPPs mining software packages including antiSMASH and BAGEL have become publicly available and gained great popularity. These tools survey genomes using profiles of highly conserved biosynthetic enzymes and evaluate BGC hits with pre-defined cutoffs. Applying these tools, researchers were able to discover many new natural products belonging RiPPs families such as lanthipeptides and cyanobactins.

**Figure 2.1** The steps of the biosynthesis of lasso peptides. The protease B cleaves the lasso peptide precursor A removing the leader peptide (shown in green). The core peptide then gets folded into a lasso shape by an asparagine synthetase-like protein.
Lasso peptides are an interesting family of RiPPs that they hold a unique 3D slipknot shape. This shape is made by an isopeptide bond between the N-terminus of the peptide and an aspartate or glutamate side chain forms eight to nine amino acid macrocycle through which the C-terminal portion is threaded forming a loop and a tail.\textsuperscript{13} The biosynthesis of lasso peptides involves three essential genes, A, B and C. The A gene encodes for a precursor peptide which is synthesized by the ribosome. This precursor is then post-translationally modified by an ATP-dependent cysteine protease encoded by the B gene, which cleaves off the leader peptide. The C gene, encoding an asparagine synthetase-like protein which catalyzes the lactam formation and folds the peptide into a lasso shape (Figure 2.1).\textsuperscript{14}

Recently genome mining has reshaped the way of lasso peptide discovery. So far, homology-based and precursor-centric approaches are two commonly used bioinformatics methods for lasso peptide mining.\textsuperscript{15} The homology-based method focuses on surveying proteins homologous to the maturation enzymes for lasso peptide. Capistruin was the first example discovered by the Marahiel group using this method.\textsuperscript{16} Specifically, the authors BLAST searched proteins homologous to the maturation enzymes McjB, McjC and McjD from the well-characterized microcin J25 biosynthetic gene clusters. They were able to find a \textit{capABCD} gene cluster for the biosynthesis of capistruin from \textit{Burkholderia thailandensis}.\textsuperscript{16} This maturation enzyme homology-based approach has recently advanced by Magarvey group and Mitchell group, where they combined Hidden-Markov-model-based analysis and machine learning to localize biosynthetic gene
clusters and predict lasso peptides$^{7,17}$.

The precursor-centric method developed from our group first identifies precursor genes and then surveys the neighborhood genes for likely maturation enzymes.$^{15}$ This approach incorporates the conserved amino acid sequence features and length restrictions of precursor peptides to generate a precursor sequence pattern. First, the glycine is highly conserved at the first position of core peptides. Second, there is a requirement for a glutamic or aspartic acid to be present in position 8 or 9 for isopeptide cyclization. Third, the threonine is conserved at the penultimate position of leader peptides. In addition, the sequence length ranges of both leader and core peptides are additional restrictions for lasso peptide precursors. Once a short ORF meeting the four requirements is identified, a 10,000 base region centered on the ORF is searched for genes encoding proteins that contain conserved motifs from the lasso peptide B and C maturation enzymes using the MAST algorithm.$^{18}$ Together there are seven motifs in maturation enzymes including four motifs from B homologs and three motifs for C homologs. Finally, the algorithm ranks the putative lasso peptide gene clusters based on the number of motifs founded, with the ranking score from 1 to 7. For instance, if five motifs are found in a putative cluster, the ranking score of the gene cluster would be 5. This score is used to determine the likelihood of positive cluster hits.

Previous researchers in the Link lab used this approach to identify 79 putative gene clusters of 3,000+ microbe genomes, and ultimately discovered and fully characterized lasso peptides astexin-1, 2 and 3 from Gram-negative
freshwater organism *Asticcacaulis excentricus*. This precursor-centric approach is highly automated where it only requires simple genome sequences as input, and outputs putative lasso peptide gene clusters.

This chapter aimed to optimize this precursor-centric method by updating the precursor sequence pattern and maturation enzymes’ motifs. First, the precursor sequence pattern was modified based on sequence features of all characterized lasso peptides up to early 2017. Second, the motif matrix parameters for maturation enzymes were updated using the latest list of characterized B and C enzymes. By applying this optimized method, I identified 582 putative lasso peptide gene clusters with rank 7, and 366 clusters with rank 6 from 5203 complete genomes, 6971 drafted genomes, and 32000 sets of DNA contigs from bacteria and archaea.

2.2 Results

2.2.1 Updating precursor peptide pattern

The precursor-centric genome mining that centers around the identification of precursors has been developed and shown to be effective, resulting in the isolation of the astexins, caulonodins, caulsegnins, xanthomonins and many others. Since 2012, there are dozens of new lasso peptides have been characterized and new precursor sequence features have been revealed. First, some examples have shown that the amino acid at the first position of the core peptide can be different from glycine, where lasso peptide caulonodin starts with alanine, Class I lasso peptides, such as
sviceucin, start with cysteine, and streptomonomicin starts with serine.\textsuperscript{23,27,28} Thus, I expanded the amino acid selection at the first position of the core peptide from the only glycine to glycine, alanine, cysteine, and serine. Second, lactam cycles of lasso peptides have been confirmed to be as small as 7 aa membered ring.\textsuperscript{25} I relaxed the ring size constraint to be 7 to 9 amino acids. Third, besides AUG, prokaryotes use GUG and UUG as the alternative start codons. I incorporated all three possible start codons for ORFs searching, where AUG codes for Methionine, GUG codes for Valine, and UUG codes for Leucine.\textsuperscript{29} In addition, I expanded the sequence length ranges of both leader peptide and core peptide. The updated precursor pattern is shown in Figure 2.2.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{updated_pattern.png}
\caption{The updated lasso peptide precursor pattern. A conserved threonine residue is found in the penultimate position of the leader peptide. The first residue of core peptide could be glycine, alanine, cysteine, or serine. At the 7\textsuperscript{th} to 9\textsuperscript{th} position, the residue could be aspartate or glutamate.}
\end{figure}

\subsection{Updating maturation enzyme motifs}

In previous precursor-centric genome mining algorithm, Maksimov \textit{et al.} included 11 pairs of B and C enzymes from known or putative lasso peptide clusters as a training set to generate conserved motifs of maturation enzymes.
using MEME suite. Since then, many new lasso peptide clusters have been characterized. I updated the training set including all characterized lasso peptide clusters until the beginning of 2017. The full list containing the detail information of B and C maturation enzymes for this study is included in Table 2.1. For “split-B” lasso peptide clusters, I used the B2 proteins for motif generation.

<table>
<thead>
<tr>
<th>Cluster Number</th>
<th>Cluster Name</th>
<th>Organism</th>
<th>Lasso Peptide Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ADU13885.1/ADU13886.1</td>
<td>Asticcacaulis excentricus CB 48</td>
<td>Astexin I</td>
</tr>
<tr>
<td>2</td>
<td>ADU14097.1/ADU14096.1</td>
<td>Asticcacaulis excentricus CB 48</td>
<td>Astexin II, III</td>
</tr>
<tr>
<td>3</td>
<td>ABC39546.1/ABC39597.1</td>
<td>Burkholderia thailandensis E264</td>
<td>Capistin</td>
</tr>
<tr>
<td>4</td>
<td>ACL96260.1/ACL96259.1</td>
<td>Caulobacter crescentus ATCC 10059</td>
<td>Crescanodin</td>
</tr>
<tr>
<td>5</td>
<td>ADG1019.1/ADG11018.1</td>
<td>Caulobacter segnis ATCC 27756</td>
<td>Cauloseginin II</td>
</tr>
<tr>
<td>6</td>
<td>ABZ71113.1/ABZ71114.1</td>
<td>Caulobacter sp. K31</td>
<td>Caulonodin I</td>
</tr>
<tr>
<td>7</td>
<td>ABZ71367.1/ABZ71368.1</td>
<td>Caulobacter sp. K31</td>
<td>Caulonodin II</td>
</tr>
<tr>
<td>8</td>
<td>ABZ74312.1/ABZ74312.1</td>
<td>Caulobacter sp. K31</td>
<td>Caulonodin III</td>
</tr>
<tr>
<td>9</td>
<td>AAD29490.1/AS29496</td>
<td>Escherichia coli</td>
<td>Micorinic J25</td>
</tr>
<tr>
<td>10</td>
<td>ACG78808.1/ACG78807.1</td>
<td>Phenyllobacterium zucineum HK1</td>
<td>Zucinodin</td>
</tr>
<tr>
<td>11</td>
<td>EII93830.1/EII93832.1</td>
<td>Rhodanobacter thiooxydans L.CS2</td>
<td>Rhodanodin</td>
</tr>
<tr>
<td>12</td>
<td>BAL72549.1/BAL72550.1</td>
<td>Rhodococcus jostii</td>
<td>Ljafedit</td>
</tr>
<tr>
<td>13</td>
<td>BA97936.1/BA97937.1</td>
<td>Rubrivivax gelatinosus IL 144</td>
<td>Rubrivinodin</td>
</tr>
<tr>
<td>14</td>
<td>BA199047.1/BA179047.1</td>
<td>Sphingobium japonicum UT26S</td>
<td>Sphingonodin I</td>
</tr>
<tr>
<td>15</td>
<td>EKU73624.1/EKU73622.1</td>
<td>Sphingobium yanoikuyae ATCC 51230</td>
<td>Syanodon I</td>
</tr>
<tr>
<td>16</td>
<td>ABF54241.1/ABF54250.1</td>
<td>Spingopyxis alaskensis RB2256</td>
<td>Spingopyxin I</td>
</tr>
<tr>
<td>17</td>
<td>ABF54309.1/ABF54308.1</td>
<td>Spingopyxis alaskensis RB2256</td>
<td>Spingopyxin II</td>
</tr>
<tr>
<td>18</td>
<td>KI99639.1/KI99638.1</td>
<td>Streptomyces coelicolor</td>
<td>Streptomycinin</td>
</tr>
<tr>
<td>20</td>
<td>EFE74649.1/EFE74645.1</td>
<td>Streptomyces roseosporus NRRL 15998</td>
<td>SRO15-2005</td>
</tr>
<tr>
<td>21</td>
<td>EDSY509.1/EDSY508.1</td>
<td>Streptomyces sp. ATCC 29983</td>
<td>Sivaoxin</td>
</tr>
<tr>
<td>22</td>
<td>CGG37616.1/CCG37617.1</td>
<td>Xanthomonas citri pv. mangiferinaeindicae LMG 941</td>
<td>Xanthomonin III</td>
</tr>
<tr>
<td>23</td>
<td>EGD17356.1/EGD17357.1</td>
<td>Xanthomonas gardneri ATCC 19865</td>
<td>Xanthomonin I</td>
</tr>
</tbody>
</table>

Table 2.1 Lasso peptide gene clusters used for MEME training set. There were used to identify conserved motifs in the lasso peptide maturation machinery.

The updated motifs of B and C were then generated using MEME suite. There are four motifs in the B homologs (Figure 2.3) and three motifs in the C homologs (Figure 2.4). These motifs contain a RiPP recognition element (RRE) which binds to the leader peptide of the precursor peptide, a characteristic cysteine-histidine-aspartate catalytic triad in the B protein, and a serine-aspartate rich ATP-binding pocket in the C protein, homologous to those of...
asparagine synthetases.\textsuperscript{13,14} The motif elements then serve as useful features for the identification of new candidate clusters.

**Figure 2.3** The conserved motifs in lasso peptide B homologs as predicted by MEME software. Asterisks in the motifs refer to catalytic triad residues, Cys-His-Asp. Motif 2 at the N-terminus is the precursor peptide recognition domain.
Figure 2.4  The conserved motifs in lasso peptide C homologs as predicted by MEME software. The red boxed area in motif 2 refers to Ser-Asp rich ATP binding pocket.

2.2.3 Overview of the precursor-centric algorithm

This algorithm first searches small ORFs that match the updated precursor pattern. For each match, ORFs within 10,000 bases of the precursor was searched for B and C gene sequence motifs. The number of motifs matched, out of seven total (4 motifs for B, and 3 motifs for C), was termed into the rank, a score which was used to determine the likelihood of positive hits (Figure 2.5).
2.2.4 Analysis of lasso peptide cluster hits

In this study, 5203 complete genomes, 6971 draft genomes, and 32000 sets of DNA contigs from a variety of organisms have been analyzed using the precursor-centric genome mining method. I was able to identify 582 gene clusters with rank 7, 366 gene clusters with rank 6, Table 2.2. These putative clusters are from organisms including both bacteria and archaea.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Complete Genome</th>
<th>Draft Genome</th>
<th>Contig Set</th>
<th>Total Hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>5203</td>
<td>6970</td>
<td>32000</td>
<td>-</td>
</tr>
<tr>
<td>Rank 7</td>
<td>115</td>
<td>38</td>
<td>429</td>
<td>582</td>
</tr>
<tr>
<td>Rank 6</td>
<td>70</td>
<td>28</td>
<td>268</td>
<td>366</td>
</tr>
<tr>
<td>Rank 7+Rank 6</td>
<td>185</td>
<td>66</td>
<td>697</td>
<td>948</td>
</tr>
</tbody>
</table>

Table 2.2 The analysis of cluster hits in this study.

Rank 7 lasso peptide clusters are abundant in organisms like *Acinetobacter*, *Alphaproteobacter*, *Asticcacaulis*, *Blastomonas*, *Brevundimonas*, *Burkholderia*, *Caulobacter*, *Citrobacter*, *Citromicrobium*, *Enterobacter*, *Erythrobacter*, *Escherichia*, *Hyphomonas*, *Mycobacterium*, *Novosphingobium*, *Pandoraea*, *Phenylobacterium*, *Pseudomonas*, *Rhodanobacter*, *Sphingobium*, *Sphingomonas*, *Sphingopyxis*, *Stenotrophomonas*, *Streptomyces* and
Xanthomonas, Among rank 6 clusters, I found 144 clusters are from Streptomyces strains. In addition, I identified four clusters from archaea which are ranked 6. Interestingly, many newly mined clusters are worthy of further investigation where they exhibit novel features that have not seen in the known lasso peptide clusters. I analyzed several putative clusters as representatives in details below.

2.2.5 Lasso peptide clusters from Asticcacaulis benevestitus

Two putative lasso peptide clusters were identified from Asticcacaulis benevestitus, a cold-tolerant organism was isolated from a soil sample near the town of Vorkuta, Russia. The cluster architectures are shown in Figure 2.6. These two clusters exhibit many elements of conservation with the astexin family of lasso peptides. These gene clusters are associated with isopeptidases, GntR homologs and TonB-dependent transporters. Till now, there are only two isopeptidase-associated lasso peptide clusters have been characterized. Full characterization of the clusters from A. benevestitus in chapter 3 allows the further expansion of this subfamily.

I named the putative lasso peptides from the two clusters as benenodin-1 and benenodin-2. Based on their sequences, benenodin-1 has an 8-membered ring with an isopeptide bond between Gly-1 and Asp-8, while benenodin-2 has a 9-membered ring linked Gly-1 and Asp-9. Surprisingly, no bulkier aromatic or large positively charged amino acids at the tail of benenodin-1 could serve as steric lock residues to hold its threaded shape. This raises the two questions 1) which
residues keep benenodin-1 threaded? and 2) is benenodin-1 a stable lasso peptide under thermal and enzymatic stresses? An extensive study on the structure and stability of benenodin-1 in Chapter 3 provides new insights of biophysical and biochemical properties of lasso peptides.

![Diagram of lasso peptide clusters from Asticcacaulis benevestitus](image)

**Figure 2.6** Overview of the lasso peptide clusters from *Asticcacaulis benevestitus* (A) the cluster architectures of the two clusters. (B) Alignment of the precursor sequences of the benenodin-1 and benenodin-2. Benenodin-1 has an 8-member ring, while benenodin-2 has a 9-member ring.

### 2.2.6 Lasso peptide clusters from Sphingopyxis fribergensis

Previously Maskimov *et al.* predicted some bacteria can carry multiple lasso peptide biosynthetic clusters and produce three or more lasso peptides such as *Caulobacter sp* K31, *Sphingobium japonicum, Xanthomonas gardneri*. The Marahiel group conducted characterization studies and verified the existence of multiple lasso peptides from these organisms. These studies have revealed lasso peptides are abundant and indicated lasso peptides may play very important roles in their native organisms. From our analysis, four putative lasso peptide clusters carrying five precursor genes were identified from *Sphingopyxis fribergensis*. I named these family of lasso peptides as
fribergenodins. The cluster architectures and sequence alignment of the five lasso precursors are shown in Figure 2.7. All four gene clusters are associated with isopeptidases and TonB-dependent transporters. Interestingly, in the gene cluster of fribergenodin-3, genes encoding an autoinducer synthase and a DUF2285 domain-containing protein appear upstream of the precursor gene friA3. Previous studies reported these gene homologs are related to quorum sensing. Thus, fribergenodin-3 may act as a regulatory element for responding the change in population density of *Sphingopyxis fribergensis*. Future studies focusing on crosstalk between different fribergenodins and fribergenodin-3 effects quorum sensing may provide further insights into biological functions of lasso peptides.

![Figure 2.7](image)

**Figure 2.7** Overview of the lasso peptide clusters from *Sphingopyxis fribergensis* (A) the cluster architectures of the four clusters. (B) the sequence alignment of five fribergenodin precursors.
2.2.7 Lasso peptide cluster from *Streptomyces albus*

A cluster from *Streptomyces albus* has drawn our attention due to its unique gene cluster architecture: *albACBT* (Figure 2.8). Besides the conserved precursor gene *albA*, genes *albB* and *albC* encoding the lasso peptide maturation enzymes, it includes a gene *albT* encoding a protein annotated as a GNAT acetyltransferase downsteam of *albACB*. This cluster diverges from the lasso peptide clusters with either an isopeptidase or an ABC transporter. By speculating the sequence of AlbA, I found a single lysine at the 10th position on the predicted core peptide. I hypothesize AlbT may catalyze the acetylation of the Lys-10. Chapter 4 picks up with the investigation of this gene cluster and leads to the discovery of a new type of post-translational modification on lasso peptides.

![Figure 2.8](image)

*Figure 2.8* The putative lasso peptide gene cluster from *Streptomyces albus*. The A gene codes for the precursor peptide and the B and C genes encode maturation enzymes. The T gene encodes a putative acetyltransferase that may install acetyl group at lysine-10 (highlighted in green) of the predicted core peptide (highlighted in black).
2.2.8 Lasso peptide cluster from *Methanobacterium paludis*

![Diagram showing the lasso peptide clusters of paeninodin and paludinodin]

**Figure 2.9** Comparison of the lasso peptide clusters of paeninodin and paludinodin. For paeninodin, the *padeA* gene encodes the precursor peptide and the *padeB1/B2* and *padeC* genes encode maturation enzymes, *padeD* encodes for an ABC transporter. The *palK* gene encodes a putative HPr kinase that phosphorlates the C-terminal serine. The *palA* gene encodes the precursor peptide and the *palB1/B2* and *palC* genes encode maturation enzymes. The *palK* gene encodes a putative HPr kinase that may phosphorylate Ser-3, 11, or 15 (highlighted in orange) of the predicted core peptide of paludinodin.

Besides the cluster hits found in bacteria, I was able to identify a putative lasso peptide cluster from an archaea species, *Methanobacterium paludis* SWAN. This organism was isolated from a peatland in upstate New York. I named the putative lasso peptide as paludinodin, **Figure 2.9**. The paludinodin cluster features a "split-B" protein separate into two open reading frames (*palB1* and *palB2*). In addition, the gene *palK* encoding an HPr kinase\(^ {37} \) homolog is located between *palB2* and *palC*. In a previous study, Zhu *et al.* reported lasso peptide paeninodin gets phosphorylated at the C-terminal serine residue by an HPr like kinase.\(^ {38} \) I expected PalK works in the similar fashion where it is responsible for serine phosphorylation on paludinodin. Phosphorylated paludinodin could serve
as a signal molecule for important biological processes in archaea. By looking at the core peptide sequence of paludinodin, there are three serine residues (serine-3, serine-11 and serine-15) that potentially get phosphorylated by PalK. Further studies can be focused on pinning down the phosphorylation site(s) and investigating the role of phosphorylation in the biosynthesis of paludinodin. This will provide insights of the roles of lasso peptides in archaea.

2.3 Conclusion and Future Works

In this chapter, I optimized the precursor-centric genome mining algorithm using a new precursor sequence pattern and the latest motif features of maturation enzymes. By applying the updated algorithm, I identified 948 high-confidence (Rank 6+Rank 7) lasso peptide gene clusters from 5203 complete genomes, 6971 draft genomes, and 32000 sets of DNA contigs. These putative clusters are broadly distributed across organisms in bacteria and archaea. Many newly mined clusters exhibit novel precursor sequence features and post-translational elements. The recent studies from the Magarvey group\textsuperscript{7} and the Mitchell group\textsuperscript{17} reported that they were able to identify ~ 1300 putative lasso peptides. The difference in results are potentially due to the lack of full consideration of “split-B” lasso peptide gene clusters in our algorithm. Specifically, “split-B” lasso peptide gene clusters divide the B gene into two separated ORFs (\textit{B1} and \textit{B2}). The protein B1 containing the RRE domain (motif 2, \textbf{Figure 2.3}) binds to the leader peptide and delivers its peptide substrate to the cysteine protease B2 containing a catalytic triad (motif 1,3,4, \textbf{Figure 2.3}) for cleavage. In
our current algorithm, I only looked for the intact B protein. In the “split-B” case, without taking a count of B1/B2 proteins as separated ORFs, only B2 protein is identified by the algorithm. As a result, the B proteins in “split-B” lasso peptide gene clusters can only have the maximum of three motifs found, resulting a maximum rank of 6 if all three motifs of the C enzyme are found. Other variations in the “split-B” clusters can further drop the rank down to 5 or even lower.

To further improve this algorithm, I want to focus on diversifying maturation enzyme training sets, constructing a comprehensive precursor pattern, and minimization of redundant gene clusters. First, I used experimentally characterized B and C enzymes to train the MEME motif models in this study. Unfortunately, most of the characterized lasso peptide clusters are from Gram-negative bacteria. Thus, the training set is biased towards Gram-negative organisms, with only three Gram-positive examples. To generate unbiased motif models, it is necessary to include putative B and C enzymes from all types of bacteria. Moreover, I can divide lasso peptide gene clusters into subclasses based on their cluster architectures, such as isopeptidase associated clusters, ABC transporter associated clusters, and “split-B” associated clusters. I could generate a specific training set of maturation enzyme motifs for an individual subclass and then run the algorithm separately. For “split-B” associated clusters, I should search for co-localization of B1 and B2 proteins.

Second, researchers from our group found two lasso peptides through BLAST, fuscanodin and pandonodin. Fuscanodin is the first example of lasso peptide beginning with a tryptophan, a bulky aromatic residue. Pandonodin is the
longest known lasso peptide consist of 33 amino acids, with the loop and tail region containing 25 amino acids. The current algorithm would not identify these peptides. The discovery of these two lasso peptides informs me to adopt a more inclusive precursor pattern for small ORFs search.

Lastly, I found many gene cluster hits potentially produce either the same lasso peptide or closely-related lasso peptides from our study. For instance, I found 77 gene clusters containing capistruin or capistruin-like lasso peptides (one or two amino acid changes from capistruin). It is important to implement a downstream data processing system to reduce “me too” gene clusters and output clusters that contain truly novel elements.

By improving these aspects, the precursor-centric genome mining method could provide better specificity and resolution in the discovery of lasso peptide gene clusters.

![Peptide Sequences](image)

**Figure 2.10** The peptide sequences of fuscanodin and pandonodin. The first amino acid of fuscanodin is tryptophan. Fuscanodin is the first example of lasso peptide starts with a bulky aromatic residue. Pandonodin is the longest known lasso peptide containing total of 33 amino acids and the loop and tail region contains 25 amino acids.
2.4 References


(18) Bailey, T. L.; Gribskov, M. Combining evidence using p-values: application


(5), 527–542.


(34) Chekan, J. R.; Koos, J. D.; Zong, C.; Maksimov, M. O.; Link, A. J.; Nair, S. K. Structure of the Lasso Peptide Isopeptidase Identifies a Topology for


Chapter 3

A Naturally-Existing Thermal Switch:

Benenodin-1

Upon comparing the set of predicted lasso peptide clusters described in Chapter 2, the lasso peptide gene clusters from alpha-proteobacterium Asticcacaulis benevestitus has caught our attention. Since their cluster architectures are very similar to isopeptidase-associated lasso peptides, we aimed to further examine the clusters from A. benevestitus to expand our knowledge of isopeptidase-associated lasso peptides. In this chapter, we successfully characterized the lasso peptide benenodin-1 from A. benevestitus. To our surprise, benenodin-1 exhibits conformational switching between two distinct threaded conformers upon actuation by heat, thus is a natural example of a switchable mechanically-interlocked molecule (MIM). We also showed only the native conformer of benenodin-1 is cleaved by its associated isopeptidase, which may also be relevant to the biological function of these molecules. This study also shines lights on new directions of engineering lasso peptide based molecular motors.

This chapter is reproduced with permission from:

This work is a result of a collaboration with Michelle J. Wu and Jason Z. Qin. Wu discovered and cloned the lasso gene clusters of benenodins. Qin came up the hypothesis that benenodin-1 exists in two threaded conformers.

3.1 Introduction

Molecules containing mechanical bonds, such as catenanes, rotaxanes, knots and Borromean rings, have captured the attention of the scientific community, not only because of their intriguing architectures and topologies, but also as a result of the ability of their components to undergo controllable intramolecular movement.\(^1\) With the rapid development in the synthesis of mechanically-interlocked molecules (MIMs) in recent times, dozens of molecular switches and motors with new shapes and desired functions have been created by applying organic and polymer synthetic approaches.\(^2\) In addition to synthetic chemists, mother nature has built MIMs for biological usages including catenated circular DNA and protein chainmail.\(^3,4\)

Lasso peptides are naturally-occurring \([1]rotaxanes.\(^5\) The threaded slipknot structure is achieved via installation of an isopeptide bond between the N-terminal amine and either a glutamic acid or aspartic acid sidechain.\(^6\) The rings of lasso peptides range from 7 aa to 9 aa, or from 23 to 29 backbone atoms. Large amino acid sidechains on the thread serve as steric locks to keep lasso peptide in a thread form.\(^7\) This unique topology makes lasso peptides as attractive molecular building blocks for the development new MIMs.
By applying the precursor-centric genome mining method described in chapter 2, we identified two new lasso peptide gene clusters from the alpha-proteobacterium *Asticcacaulis benevestitus*. One of the clusters produces a lasso peptide named benenodin-1. Remarkably, benenodin-1 acts as a switchable mechanically-interlocked molecule (MIM), where it switches between two distinct [1]rotaxane conformers upon heating. This property has not been observed previously in naturally existing peptides. In addition, we have generated two variants of benenodin-1 with single amino acid substitutions, each of which is biased toward one of the two conformers upon heating. The benenodin-1 isopeptidase can only recognize and cleave one of the two conformers, providing new insights into isopeptidase substrate recognition.

3.2 Results

3.2.1 The lasso peptide gene clusters from *A. benevestitus*

There are two lasso peptide gene clusters from *A. benevestitus*. We named the peptides derived from the two clusters benenodin-1 and benenodin-2 accordingly. Like all isopeptidase-associated lasso peptide clusters described in Chapter 2, the benenodin-1 and -2 gene clusters include an isopeptidase which linearizes the threaded lasso peptide substrate. In addition, a putative TonB-dependent transporter and several putative regulatory proteins are located downstream of the isopeptidase gene (Figure 3.1A). The benenodin-1 and -2 core peptide sequences share little sequence identity. Whereas benenodin-1 has an 8 aa ring with an isopeptide bond between Gly and Asp (25 backbone
atoms), benenodin-2 has a 9 aa ring with a Gly-Asp isopeptide linkage. The sequences were also compared with the previously characterized astexin-3 (Figure 3.1B).

![Diagram of gene clusters and peptide sequences](image)

**Figure 3.1** The gene clusters and peptide sequences from *Asticcaucalis benevestitus*. A: Gene clusters of lasso peptides benenodin-1 and benenodin-2 showing the highly conserved architecture typical of lasso peptides produced in alpha-proteobacteria. The A, B, and C genes are required for lasso peptide biosynthesis while the E genes encode lasso peptide isopeptidase. B: Sequences of benenodins-1 and -2 aligned with astexin-3. The conserved threonine in the leader peptide is colored orange and the isopeptide-bonded glycine is colored blue. C: Engineered gene clusters for the heterologous expression of benenodins-1 and -2 in *E. coli*.

### 3.2.2 Heterologous expression of benenodin-1 lasso peptide

We first attempted to isolate benenondin-1 and benenodin-2 from its host organism, *Asticcaucalis benevestitus*. However, no masses corresponding to benenodin-1, benenodin-2 or related truncations were identified from cell lysate
or supernatant using MALDI-MS. Previous works have shown that lasso peptide gene clusters often can be heterologously expressed in *E. coli* and the production level of lasso peptide could reach at mg/L level. For heterologous expression of benenodin-1, we constructed a plasmid pCZ35 (Figure 3.1C), where the native benenodin-1 cluster was cloned into the pASK75 expression vector containing a strong *tet* promoter and a ribosome binding site upstream. Cells containing the pCZ35 plasmid were induced in M9 media, harvested and lysed for MALDI-MS and HPLC analysis. The data revealed a truncation of benenodin-1 lacking its 5 C-terminal aa (benenodin-1 ∆C5, retention time 14.7 min) (Figure 3.2A) was the dominant product. Smaller amounts of a ∆C4 truncation (retention time 14.4 min) were observed in HPLC, while a small signal for the full-length peptide was observed by mass spectrometry (Figure 3.2B).

**Figure 3.2** (A) HPLC and (B) MALDI-MS of cell extracts producing benenodin-1. The predominant form of benenodin-1 is the ∆C5 variant which lacks the final 5 C-terminal amino acids. Theoretical masses (singly protonated): benenodin-1: 2569.28 Da, benenodin-1 ∆C4: 2070.04 Da, benenodin-1 ∆C5: 2013.02.
For benenodin-2 expression, we constructed a plasmid pMW12 containing the benenodin-2 cluster in the similar fashion (Figure 3.1C). Using HPLC and MALDI-MS, we found the major product of pMW12 was the ∆C4 truncation of benenodin-2 with minor products of the ∆C5 truncation and full-length peptide (Figure 3.3). Attempts to dissolve the cell extract containing benenodin-2 and its truncation variants in water, acetonitrile, and mixtures of these solvents were unsuccessful, with the extract forming a gel-like phase in these solvents. This precluded further analysis of benenodin-2 in this study.

![Figure 3.3](A) HPLC and (B) MALDI-MS of cell extracts producing benenodin-2. None of the major peaks in the HPLC correspond to benenodin-2. However, peaks for a full-length peptide and its ∆C4 variant are present in the mass spectrum of the cell extract.

### 3.2.3 NMR structure of benenodin-1 ∆C5

Since benenodin-1 ∆C5 was the major product of heterologous expression, we chose this variant for NMR structural studies. The NOESY and TOCSY spectra are shown in Figure 3.S1, 3.S2, and the proton assignments are
We found NOEs between Gln-15 and each of the 8 aa in the ring (Figure 3.4A, Table 3.S2A). Similarly, Glu-14 exhibited NOEs to 6 of the 8 ring amino acids (Figure 3.4A, Table 3.S2A). These data strongly indicate Glu-14 and Gln-15 as the steric lock residues bracketing the ring. This is in contrast to most other lasso peptides which have bulkier aromatic or large positively-charged amino acids as steric lock residues. The 20 lowest energy structures of benenodin-1 ∆C5 are shown in Figure 3.4C. Overall, benenodin-1 ∆C5 has an 8 aa ring with 6 aa in the threaded loop and 5 aa in the tail.

Figure 3.4  NMR structure of benenodin-1 ∆C5. (A) Cartoon of the structure with grey residues representing the ring of the peptide, lime green residues in the loop, and blue residues in the tail. The steric lock residues Glu-14 and Gln-15 are shown in red. Arrows show NOEs observed between ring residues and steric lock residues. (B) The lowest energy conformer of benenodin-1 showing the backbone of the peptide and the sidechains of the steric lock residues. Coloring is the same as in the cartoon in part A. (C) As in part B, but showing the 20 lowest energy structures.
3.2.4 Thermally induced conformational switching of benenodin-1 ΔC5

Due to its small steric locks Glu-14 and Gln-15, we suspected benenodin-1 ΔC5 can be easily unthreaded with heat treatment. To test its thermal stability, benenodin-1 ΔC5 was incubated at 95 °C for up to 18 h, and then analyzed via HPLC and MALDI-MS at different time points. A second peak with the same mass but a different retention time (14.3 min) was observed. Interestingly, the area of the second peak plateaued after an incubation time of 2 hours (Figure 3.5A). This peak was collected and treated with carboxypeptidase B and Y, which reports on the threaded state of the peptide.5 No C-terminal degradation products were observed (Figure 3.6A), suggesting that the peptide in this new peak is still threaded.

Figure 3.5 Benenodin-1 establishes an equilibrium between two conformers upon heating. (A) Chromatograms of native benenodin-1 ΔC5 heated at 95 °C
for the indicated times (grey boxes). About 60% of the native peptide (conformer 1, retention time 14.7 min) is converted into a new species (conformer 2, retention time 14.3 min). (B) Heating of conformer 2 of benenodin-1 at 95 °C leads to the same equilibrium ratio of the two conformers.

The peptide in the 14.3 min retention time peak was isolated and subjected to heat treatment at 95 °C. The same ratio of the two peaks was achieved upon heating for 2 h or more (Figure 3.5B). After treating with carboxypeptidase B and Y, there were no mass shifts of the two peaks (Figure 3.6B). These data suggest that, upon heating, benenodin-1 ΔC5 achieves an equilibrium between two threaded conformers: conformer 1, the native state, and conformer 2, a putative partially unthreaded state. In this way, benenodin-1 functions as a thermally actuated rotaxane switch.

Figure 3.6  Carboxypeptidase assays on benenodin-1 ΔC5, conformers 1 (A) and 2 (B). Top panel: mass spectrum of purified peptide before digestion with carboxypeptidase. Bottom panel: mass spectrum of peptide that has been heated at 95 °C for 3 h and subsequently treated with carboxypeptidase. Neither conformer 1 or 2 exhibit any cleavage, suggesting that both conformers are
threaded. The +16 Da adduct on peptide is due to the oxidation of Met 19 of benenodin-1.

Figure 3.7 Conformer 2 of benenodin-1 ∆C5 is partially unthreaded. A: Cartoon representation of the benenodin-1 ∆C5 conformer 2 structure showing NOEs between the ring residues (grey) and the new steric lock residues (Ala-16 and Lys-17, turquoise). B: 20 lowest energy structures of benenodin-1 ∆C5 conformer 2 with the same coloring as part A and showing the sidechains of Glu-14, Gln-15, Ala-16, and Lys-17. C: Comparison of the lowest energy structures of conformer 1 (left) and conformer 2 (right) of benenodin-1 ∆C5.

To gain further insight into this conformational switch, we solved the structure of conformer 2 of benenodin-1 ∆C5. The NOESY and TOCSY spectrum are shown in Figure 3.S3, 3.S4, and the proton assignments are presented in Table 3.S1. Comparing to the native conformer 1 structure, conformer 2 has no NOEs between Glu-14 and the ring amino acids. The number of NOEs between
Gln-15 and the ring is also reduced in conformer 2. Instead, the ring is now in close proximity to Lys-17, as evidenced by NOEs between this residue and 6 of the 8 ring residues (Figure 3.7A). The energy minimized structure of conformer 2 (Figure 3.7B) shows that the Ala-16 sidechain is nearly completely buried within the ring of the peptide, and that Lys-17 is functioning as a secondary steric lock/stopper that prevents further unthreading of the peptide. Whereas conformer 1 has loop length of 6 aa and a tail of 5 aa, conformer 2 has a loop of 8 aa (including Ala-16) and a tail of only 3 aa (Figure 3.7C).

3.2.5 Kinetics and energetics of conformational switching

To probe the kinetics and energetics of the switch between the two benenodin-1 conformers, we carried out a series of heating experiments at temperatures ranging from 35-95 °C (Figure 3.8). The kinetics of the conformational switch from 55-95 °C fit well to a simple two state model. The solutions (Eq1 and Eq2) to the differential equations of the model are shown below, where the $k_1$ and $k_2$ are kinetic constants of the forward and reverse reactions.

Conformer 1 $\xrightarrow{k_1} \xleftarrow{k_2} \text{conformer 2}$

Fraction of conformer 1 $= \frac{k_1 e^{-(k_1+k_2)t} + k_2}{k_1 + k_2}$ (Eq1)

Fraction of conformer 2 $= \frac{k_1 - k_1 e^{-(k_1+k_2)t}}{k_1 + k_2}$ (Eq2)
Figure 3.8  Kinetics and energetics of conformational switching in benenodin-1. Native benenodin-1 ∆C5 (conformer 1) was heated at each of the indicated temperatures and the extent of switching was determined using HPLC. Red points represent conformer 1, blue points are conformer 2, and the solid lines represent fits to a two-state model (see Supporting methods for details). Error bars represent the standard deviation of two measurements, goodness of fit $R^2$ is greater than 0.97 for all fits. These data were used to calculate rate constants of switching and equilibrium ratios of the two conformers (Table 3.1). The temperature dependence of the rate constants was fit to an Arrhenius model in order to estimate the activation energy of switching, giving an estimate of the energy difference between the two conformers (lower right panel).
Figure 3.9  Arrhenius plots for forward kinetic constant, $k_1$ (A) and reverse kinetic constant, $k_2$ (B) and van’t Hoff plot (C) of conformational switching.
By fitting the rate constants to an Arrhenius model, we estimated the forward activation energy to be 27 kcal/mol while the reverse activation energy as 23 kcal/mol (Figure 3.9A, B). Assuming no hysteresis in the conformational switch, this implies that conformers 1 and 2 have comparable stability, differing by only about 4 kcal/mol (Figure 5). The equilibrium ratio [conformer 2]/[conformer 1] also varies with temperature such that ~63% of the peptide switches to conformer 2 at 95 °C, but only ~45% of the peptide is switched to conformer 2 at 55 °C (Table 3.1). The equilibrium data fit reasonably well to a van’t Hoff model (Figure 3.9C), which can be extrapolated to estimate the ratio of the peptide in each conformer at the temperature at which A. benevestitus lives. At equilibrium at the average summer temperature of Vorkuta, Russia (13 °C), ~35% of the peptide will exist in conformer 2. However, based on extrapolations of our kinetic data, it would take months for this equilibrium to be achieved. These fits also explain why we did not observe any benenodin-1 conformer 2 in our expressions of the peptide, which were carried out at 20 °C for 20 hours. At this temperature, only 0.2% of the peptide is expected to exist in conformer 2 during the time course of the experiment.

### 3.2.6 Conversion of benenodin-1 conformers into [2]rotaxanes

It has been demonstrated previously that the lasso peptide microcin J25 can be converted into a [2]rotaxane via protease cleavage of its loop region. Microcin J25 has large steric lock residues (Phe and Tyr), so its thread is locked firmly in place within the ring. We were interested in whether similar [2]rotaxanes could be generated from the two conformers of benenodin-1 given its small steric
lock residues. To this end, we generated a T12R variant of benenodin-1, installing a trypsin cleavage site within the loop of the peptide (Figure 3.10A). This variant was successfully produced by the benenodin-1 maturation machinery, though at ~10% of the yield of the wild-type peptide.

![Conformer 1 ↔ Conformer 2]

\[ K = \frac{[\text{Conformer 2}]}{[\text{Conformer 1}]} \]

<table>
<thead>
<tr>
<th>Temperature (°C)</th>
<th>( k_1 ) (h(^{-1}))</th>
<th>( k_2 ) (h(^{-1}))</th>
<th>( K_{\text{model}} )</th>
<th>( K_{\text{observed}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>95</td>
<td>1.822</td>
<td>1.131</td>
<td>1.61</td>
<td>1.68</td>
</tr>
<tr>
<td>85</td>
<td>0.420</td>
<td>0.285</td>
<td>1.47</td>
<td>1.57</td>
</tr>
<tr>
<td>75</td>
<td>0.169</td>
<td>0.119</td>
<td>1.42</td>
<td>1.50</td>
</tr>
<tr>
<td>65</td>
<td>0.079</td>
<td>0.068</td>
<td>1.16</td>
<td>1.24</td>
</tr>
<tr>
<td>55</td>
<td>0.016</td>
<td>0.020</td>
<td>0.81</td>
<td>0.80</td>
</tr>
</tbody>
</table>

**Table 3.1** Kinetic data of benenodin-1 conformational switching. Data were fit to a two-state model (see Methods for details) to determine \( k_1 \) and \( k_2 \). \( K_{\text{model}} \) was calculated as \( k_1/k_2 \). The experimental values (\( K_{\text{observed}} \)) as presented for comparison to \( K_{\text{model}} \).
Figure 3.10 Generating a protease-cleavable variant of benenodin-1. (A) The T12R benenodin-1 ΔC5 variant establishes an equilibrium between two conformers upon heating with a similar ratio to the wild-type peptide. These
Conformers can be cleaved by trypsin to generate [2]rotaxanes. (B) HPLC traces of T12R benenodin-1 ΔC5 as produced by E. coli (red trace), after heating for 3 h at 95 °C (green trace), and after heating and trypsin cleavage at 37 °C for 16 h (blue trace). While conformer 2 is cleaved completely by trypsin under these conditions, only ~25% of conformer 1 is cleaved. (C) LC-MS spectra of peaks collected from the blue trace in part B.

After heating the peptide at 95 °C for 3h, it exhibited switching similar to the wild-type peptide (Figure 3.10B). We subjected the mixture of conformer 1 and conformer 2 of this variant to trypsin cleavage for 16 h at 37 °C to generate potential [2]rotaxanes. Under these conditions, conformer 2 was cleaved completely while only ~1/4 of the conformer 1 peptide was cleaved (Figure 3.10B). This finding is consistent with our structures for the two conformers since the Arg-12 residue is more accessible to the protease in conformer 2 than in conformer 1 (Figure 3.10A). LC-MS analysis of these cleaved peptides shows that they are indeed [2]rotaxanes with the threaded structure remaining intact (Figure 3.10C).

3.2.7 Mutagenesis of steric lock residues in benenodin-1

Alanine scanning mutagenesis is commonly employed in studies of lasso peptide stability, as it can confirm proposed roles of steric lock residues. We generated the K17A, Q15A, and E14A variants of benenodin-1 ΔC5 and studied their thermostability at 95 °C. After heating the K17A variant (retention time 15.6 min) for 3 h at 95 °C, nearly all of the material shifted to a new peak with retention time 13.5 min (Figure 3.11). This large change in retention time suggested that the peptide may have fully unthreaded. To confirm this, the material at 13.5 min was collected and subjected to carboxypeptidase treatment.
After carboxypeptidase treatment, a new peak at 9.5 min appeared, and MALDI-MS analysis revealed a m/z of 873.39, corresponding to removal of 10 residues from the C-terminal tail of benenodin-1 ∆C5. Collectively, these data provide a strong argument that the Lys-17 residue is indeed serving as a secondary steric lock residue in conformer 2 as its removal allows for further irreversible unthreading of the lasso.

Figure 3.11 The K17A variant of benenodin-1 ∆C5 unthreads upon heating at 95 °C. A: HPLC traces of untreated benenodin-1 ∆C5 K17A (red), benenodin-1 ∆C5 K17A heated at 95 °C for 3 h (green), and heated benenodin-1 ∆C5 K17A that has been treated with carboxypeptidase (blue). B: mass spectra corresponding to intact benenodin-1 ∆C5 K17A (top) and heated, carboxypeptidase-cleaved peptide (bottom). Theoretical mass (MH+) of benenodin-1 ∆C5 K17A: 1955.96 Da, benenodin-1 ∆C14 K17A 873.41 Da. The fact that carboxypeptidase can cleave the C-terminal residues of the peptide is an indication that the peptide has unthreaded.
Based on both HPLC and the carboxypeptidase assay (Figure 3.12), the Q15A construct remained in a single conformation after heating for 3 h at 95 °C and thus did not display the thermally induced switching of the wild-type peptide. This indicates that the Q15A variant is biased toward conformer 1.

Figure 3.12 The Q15A variant of benenodin-1 ΔC5 does not undergo a conformation change upon heating at 95 °C. Traces and spectra are as described in the Figure 3.11 caption. Neither heating nor carboxypeptidase treatment has an effect on Q15A benenodin-1 ΔC5, showing that the peptide remains in a native-like threaded state. The +16 Da adduct on peptide is due to the oxidation of Met 19 of benenodin-1. Theoretical mass of benenodin-1 ΔC5 Q15A (MH⁺): 1955.99 Da.

The E14A variant had unusual behavior on the HPLC, running as a broad peak with a retention time of 17.5 min (Figure 3.13A), suggesting that this peptide may form a non-specific aggregate at the concentrations used for HPLC analysis. Heat treatment of the E14A variant resulted in a new, sharp peak at 14.4 min, back in the range of other benenodin variants (Figure 3.13C). We
propose that this new peak corresponds to a conformer 2 state. Carboxypeptidase treatment of the material at 17.5 min resulted in the cleavage of a single amino acid from the C-terminus, generating benenodin-1 E14A ΔC6 (Figure 3.13B). In contrast, carboxypeptidase did not cleave the material eluting at 14.4 min (Figure 3.13D), providing further support for the proposal that this peak corresponds to a peptide resembling conformer 2.

Figure 3.13 The E14A benenodin-1 ΔC5 variant switches to conformer 2 upon heating. (A) E14A variants of benenodin-1 run as broad peaks on the HPLC (red trace). Carboxypeptidase treatment of the mixture of E14A benenodin-1 ΔC4 and E14A benenodin-1 ΔC5 results in a new peak (turquoise trace). (B) Mass spectra of the mixture of E14A benenodin-1 ΔC4 and E14A benenodin-1 ΔC5 before (top spectrum) and after (bottom spectrum) treatment with carboxypeptidase. The major product after carboxypeptidase treatment is ΔC6 variant, suggesting that the peptide is still threaded. Theoretical masses (MH\(^+\))
of benenodin-1 ΔC4 E14A: 2012.03 Da, benenodin-1 ΔC5 E14A: 1955.01 Da, benenodin-1 ΔC6 E14A: 1823.97 Da. (C) As in part A, but the mixture of E14A benenodin-1 ΔC4 and E14A benenodin-1 ΔC5 has been heated to 95 °C for 3 h. The peaks for E14A benenodin-1 ΔC4 and E14A benenodin-1 ΔC5 are sharpened (red trace) and carboxypeptidase does not cleave the peptide (turquoise trace), suggesting that these benenodin-1 variants are in a state resembling conformer 2. (D) Mass spectra showing that carboxypeptidase does not cleave heated E14A benenodin-1 ΔC4 and E14A benenodin-1 ΔC5. The +16 Da adduct on the peptide is due to the oxidation of Met 19 of benenodin-1.

3.2.8 Substrate recognition by benenodin-1 isopeptidase

As discussed above, a hallmark of alpha-proteobacterial lasso peptide gene clusters is an isopeptidase (Figure 3.1) that specifically cleaves the threaded [1]rotaxane form of the lasso peptide into a linear product. Based on our recently published structure of the astexin-2/astexin-3 isopeptidase AtxE2\textsuperscript{11}, we generated a homology model of the benenodin-1 isopeptidase BenE using I-TASSER\textsuperscript{12, 13} (Figure 3.14).

![Homology modeling of the benenodin-1 isopeptidase BenE.](image)

**Figure 3.14** Homology modeling of the benenodin-1 isopeptidase BenE. (A) Isopeptidase AtxE2, which hydrolyzes astexin-2 and astexin-3 as substrates. Drawn from PDB coordinates 5TXC. The active site serine is highlighted in orange at the bottom of the substrate cleft. (B) Homology model of BenE derived
from the coordinates of AtxE2. The active site serine is highlighted in green. The substrate cleft in BenE is smaller than that of AtxE2. (C) Alignment of the crystal structure of AtxE2 with the homology model for BenE. The active site serine is shown as sticks. Figures generated using Pymol.

Overall, the predicted structure of BenE is very similar to that of AtxE2 with a slight narrowing of the substrate cleft, consistent with the smaller ring size of the benenodin-1 substrate relative to astexin-3. We were interested in whether the two different conformers of benenodin-1 ΔC5 could be cleaved by the BenE enzyme associated with the benenodin-1 gene cluster.

Figure 3.15 Substrate specificity of benenodin-1 isopeptidase BenE. BenE can only hydrolyze conformer 1 of benenodin-1. Benenodin-1 ΔC5, conformer 1 (red trace) is completely hydrolyzed by BenE (green trace). In a mixture of the two conformers generated by heating (turquoise trace), only conformer 1 is hydrolyzed (purple trace).
Reactions were set up under conditions that led to complete cleavage of the lasso peptide astexin-3 by its cognate isopeptidase\textsuperscript{14,15} with the benenodin-1 substrate present at 100 µM and the BenE enzyme at 100 nM. After a 16 h reaction at 20 °C, all of benenodin-1 conformer 1 was hydrolyzed, as evidenced by a change in retention time on HPLC, and confirmation by MALDI. However, when an equilibrium mixture of conformers 1 and 2 was treated under the same conditions, conformer 2 was not cleaved (Figure 3.15 and 3.16). This finding is consistent with our previous mutagenesis and structural studies on the astexin-3 isopeptidase.\textsuperscript{15} In this work we showed that the loop region of the lasso peptide was most important for its recognition by the isopeptidase. Since conformer 2 has a larger loop than the native form of the peptide (conformer 1), we propose that conformer 2 is no longer able to be accommodated within the active site of the enzyme.

**Figure 3.16** Mass spectra of isopeptidase cleavage reactions on benenodin-1 ΔC5. These spectra correspond to the HPLC traces in Figure 3.15. A:
Benenodin-1 ΔC5 conformer 1 is hydrolyzed completely by BenE. B: A mixture of the two conformers of benenodin-1 ΔC5 is only partially hydrolyzed by BenE.

Figure 3.17 A chimeric substrate comprised of the loop residues of astexin-3 (red) grafted onto benenodin-1 (green) is successfully hydrolyzed by both the benenodin-1 isopeptidase BenE and the astexin-3 isopeptidase AtxE2.

To probe the importance of the loop size further, we generated a chimeric lasso peptide in which the first five loop residues of astexin-3 (SVSGQ, Figure 3.1B) replaced the same 5 loop residues of benenodin-1 (SILTQ). This "loop swap" only requires 3 relatively conservative amino acid substitutions (Figure 3.17), and was correctly processed into a lasso peptide by the benenodin maturation machinery. We also attempted to produce the opposite "loop swap" chimera in which the benenodin-1 loop replaced the astexin-3 loop. Unfortunately, the astexin-3 maturation machinery did not process this substrate into a lasso peptide. While the loop size of astexin-3 and benenodin-1 is
identical, it should be noted that astexin-3 has a ring of 9 aa while benenodin's ring is only 8 aa. The benenodin-1/astexin-3 chimera was successfully hydrolyzed by both the astexin-3 isopeptidase AtxE2 and the benenodin-1 isopeptidase BenE (Figure 3.18). The chimeric peptide is actually a better substrate for AtxE2 than it is for BenE, despite the fact that all of the chimera except for the loop is derived from benenodin-1.

Figure 3.18 Hydrolysis of a benenodin-1/astexin-3 chimera by BenE and AtxE2. The chimera consists of the loop of astexin-3 grafted on to benenodin-1 \( \Delta C5 \). (A) The intact lasso peptide (red trace) is cleaved partially by BenE over a 16 h reaction time (green trace) and is cleaved completely by AtxE2 (blue trace). (B) Mass spectra corresponding to the HPLC traces in panel A. Theoretical mass of benenodin-1/astexin-3 chimera (MH\(^+\)): 1928.92.

In control experiments (Figure 3.19), we found that AtxE2 did not cleave wild-type benenodin-1 \( \Delta C5 \) (conformer 1) and that BenE did not cleave astexin-3. These data further support the idea that the loop segment of lasso peptides serves as the primary recognition element for the isopeptidase. In the case of
AtxE2, the chimeric substrate has the “wrong” ring size (i.e. one aa smaller than its native substrate, Figure 3.20), but the enzyme is still able to cleave the chimera due to the shape complementarity of the loop.

**Figure 3.19** Control isopeptidase hydrolysis experiments. (A) Benenodin-1 ∆C5 is not cleaved by AtxE2 within 16 h. (B) Astexin-3 is not cleaved by BenE within 16 h. The enzyme concentration in these experiments is 100 nM and substrate concentration is 100 μM, conditions that lead to full hydrolysis of the correct substrate within 16 h.

Since the BenE isopeptidase can discriminate between the two conformers of wild-type benenodin-1, we tested each of the alanine variants described above, both before and after heating, as substrates for BenE. As produced by the bacteria, in a putative conformer 1 state, each of the alanine variants (E14A, Q15A, and K17A) was hydrolyzed by the isopeptidase (Figure 3.21). As expected, the unthreaded K17A variant, generated by heating, was not a substrate for BenE (Figure 3.21).
Figure 3.20 Comparison of the backbone of the isopeptide-bonded rings of benenodin-1 and astexin-3 (PDB: 2M8F). The ring sequences are GVGFRPD for benenodin-1 and GPTPMVGLD for astexin-3 (see Figure 3.1B for full peptide sequences). Figures generated using Pymol.

We proposed above that the Q15A variant remains in a conformer 1 state after heating, and this was borne out by isopeptidase treatment experiments that showed complete hydrolysis of the heated Q15A peptide by BenE (Figure 3.22). Similarly, we proposed that, after heating, the E14A variant was biased toward conformer 2. In accordance with this, the heated E14A peptide was not cleaved by BenE (Figure 3.22). These isopeptidase assays confirm that the Q15A variant of benenodin-1 is biased toward conformer 1 upon heating while the E14A variant displays the opposite behavior and switches exclusively to conformer 2.
**Figure 3.21** Isopeptidase reactions on benenodin-1 ∆C5 alanine variants. (A) As produced by the bacteria in the native conformer 1 state, each of the three alanine variants E14A, Q15A, and K17A are hydrolyzed completely by BenE. (B)
Mass spectra of the HPLC traces in panel showing complete hydrolysis of the three alanine variants.

**Figure 3.22** Isopeptidase reaction on alanine variants of benenodin-1 after heating. (A) After heating, E14A benenodin-1 switches to conformer 2 and is not
cleaved by the isopeptidase. (B) After heating, Q15A benenodin-1 remains in a conformer 1 state and is completely cleaved by the isopeptidase. (C) After heating, most of K17A benenodin-1 has unthreaded and cannot be cleaved by the isopeptidase. A small fraction of the K17A variant remains threaded and is successfully hydrolyzed by the isopeptidase.

3.3 Conclusion

Here we describe the isolation of a new lasso peptide natural product from *A. benevestitus* that we have named benenodin-1. This name, which translates roughly to “good knot,” turned out to be particularly apt, because the peptide remains in its knotted state upon extensive heating. In fact, upon heating, the peptide establishes an equilibrium between two distinct slip-knotted conformers, a property never before observed for lasso peptides. Fitting of the conformational switching data at different temperatures to an Arrhenius model allowed us to determine the energy landscape of the two conformers. This work shows that, in addition to synthetic chemists, nature has also built MIMs that can act as thermally actuated switches. Since lasso peptides are directly gene-encoded, it means that these MIMs are subject to evolutionary pressures in nature as well as being readily engineered in the laboratory. We have shown here that single amino acid substitutions in the steric lock/stopper amino acids can lead to three different outcomes upon heating: irreversible unthreading, bias toward the native conformer, and bias toward the switched conformer.
3.4 Future Direction

In other recent work from our laboratory, we demonstrated that the lasso peptide microcin J25 could be cleaved in its loop region, generating a [2]rotaxane, and reassembled using disulfide bonding into radial [3] and [4]catenanes. Due to the bulky steric lock residues of microcin J25, however, these catenanes were rigid and thus unable to exhibit any switching behavior. The benenodin-1 peptide, with its built-in switching capability and its ability to be cleaved within its loop, is a prime candidate for building conformationally complex peptide-based MIMs such as catenanes, daisy chains that can exhibit switching behavior and more sophisticated properties (Figure 3.23). For instance, we could generate the daisy chain using click chemistry (Figure 3.25)

![Figure 3.23 Potential beneondin based MIMs.](image-url)
We also got inspired by a recent work from Giuseppone group, in which they used bistable [c2] daisy chain rotaxanes as the starting materials to generate reversible muscle-like macroscale polymers.\textsuperscript{16} By applying bio-conjugation methods and incorporation of unnatural amino acids in benendin-1, we believe we can use benenodin-1 to generate thermal actuated biopolymers (\textbf{Figure 3.24, 3.25 and 3.26}).

\textbf{Figure 3.24} Potential benenodin based polymers. The synthetic schemes of these molecules are shown in \textbf{Figure 3.25} and \textbf{3.26}. 
Figure 3.25 The synthetic scheme of a daisy chain using lasso peptides as the starting materials.
Figure 3.26 Schematic representation for the synthesis of a linear chain polymer based on daisy chain monomers and a cross-linked polymer network based on daisy chain monomers.
3.5 Materials and methods

3.5.1 Strains and reagents

*A. benevestitus* was purchased from the German Collection of Microorganisms and Cell Cultures (DSMZ), and whole genomic DNA was extracted using QIAGEN DNeasy Blood & Tissue Kit (Qiagen). PicoMaxx DNA polymerase (Agilent Technologies) was used for DNA amplification with primers purchased from Integrated DNA Technologies. The restriction enzymes *Hind*III, *Eco*RI, *Xba*I, and *Bgl*II (New England Biolabs) were used to clone the gene clusters into the vector. XL-1 blue *E. coli* cells were used for recombinant DNA steps, while BL-21 was used for peptide expression.

3.5.2 Plasmid construction

Molecular cloning was done according to standard protocols. Gene clusters for benenodin-1 were cloned into pASK75, which harbors a tetracycline-inducible promoter. The gene cluster was designed such that the *benA* gene was flanked on its 5' end by an *Xba*I site and on its 3' end by an *Eco*RI site, allowing for facile swapping of the *benA* gene. The *benB* and *benC* genes were inserted downstream between *Eco*RI and *Kpn*I sites. Benenodin-1 variants were generated using overlap PCR. The *benE* gene, encoding the benenodin-1 isopeptidase, was cloned into pQE60 (Qiagen) which installs a C-terminal His-tag for protein purification. The gene including the His-tag was cut out of this plasmid and ligated into pQE80 (Qiagen) to generate the final expression plasmid. Primers, restriction enzymes, and vectors used for each plasmid are
shown in Table 3.S3, 3.S4.

3.5.3 Lasso peptide expression and purification

For all peptide expressions, *E. coli* BL21 cells were transformed with the plasmid containing the lasso peptide gene cluster. Overnight cultures were grown in 5 mL of LB with ampicillin (100 μg/mL) at 37 °C. Cells were subcultured at an OD$_{600}$ of 0.02 into M9 minimal media, supplemented with 20 amino acids (0.04 g/L each) and ampicillin. Upon reaching an OD$_{600}$ of 0.2 to 0.3, cultures were induced with 200 μg/L anhydrotetracycline. The cultures were induced for 18 hours at 20 °C. The cells were harvested by centrifugation at 8,000 x g at 4 °C for 10 minutes, then lysed in methanol. The insoluble fraction was removed by centrifugation, and the methanol-soluble fraction, which contains the peptide of interest, was dried under reduced pressure and reconstituted in ultrapure water. The sample was further cleaned up using solid-phase extraction columns (Strata C$_8$, 6mL). The crude peptide was eluted from the column using methanol, dried under reduced pressure, and then resuspended in 50% acetonitrile (ACN) and 50% water solution. The peptide sample was injected onto a Zorbax 300SB-C18 Semi-Prep HPLC Column (9.4 by 250 mm, Agilent Technologies). Solvent gradient 1 was applied to the column at a flow rate of 4.0 mL/min: 10% ACN for 1 min, increase to 50% ACN over 19 min, increase to 90% ACN over 5 min, remain at 90% ACN for 5 min, decrease to 10% ACN in 2 min. Detection was done at 215 nm. HPLC fractions containing the peptide of interest were collected and lyophilized (Labconco Freezone 4.5/−105 °C). The yield of wild-type benenodin-1ΔC5 was 0.28 mg/L after all purification steps. Peptide identity was confirmed
using MALDI-MS. Peptide samples were diluted in a 2.5 mg/mL solution of α-cyano-4-hydroxycinnamic acid matrix, and spotted onto an Applied Biosystem(ABI) 384 Opti-TOF 123 mm × 81 mm SS plate for MALDI-MS analysis.

3.5.4 Benenodin-1 ΔC5, conformer 2 production

Lyophilized benenodin-1 ΔC5 (conformer 1) was reconstituted in ultrapure water, and then incubated at 95 °C in a thermocycler (Bio-Rad DNAEngine) for 3 h allowing the conversion to benenodin-1 ΔC5 conformer 2. The heated sample was injected on HPLC using solvent gradient 1 described above. The benenodin-1 ΔC5 conformers differ in retention time by ~0.5 min. Based on the retention times, the fraction containing benenodin-1ΔC5 conformer 2 was collected. The sample was then freeze-dried and reinjected on HPLC to ensure its purity.

3.5.5 Heating assay on lasso peptides

Lasso peptides were resuspended to a final concentration of 100 μM in ultrapure H$_2$O. 200 μL of each peptide was placed in a PCR tube and heated in a BioRad DNA Engine Thermal Cycler using the heated lid to minimize evaporation. Temperatures used ranged from 35-95 °C, and time points (each time point was an aliquot of 20 μL) were taken between 0 minutes and 13 days. Samples were cooled to 4 °C and subjected to HPLC. The area under the curve was calculated from the chromatogram and fit to Eq 1 and Eq2 to determine the $k_1$ and $k_2$. The values of $k_1$ and $k_2$ were estimated by fitting the experimental data to these equations using the SOLVER function of Microsoft Excel.
3.5.6 NMR structure analysis of benenodin-1 ΔC5, conformers 1 and 2

The TOCSY (60 ms mixing time, and NOESY (100 ms mixing time) spectra of benenodin-1 ΔC5, conformer 1 (2.5 mg/ml, 100 μL sample, 2.5 mm NMR tube) and benenodin-1ΔC5, conformer 2 (6.5 mg/ml, 360 μL, 5 mm NMR tube) in 95% H2O and 5% D2O solution were acquired on a Bruker Avance III 800 MHz spectrometer at 288 K. Structural modeling and energy minimization were performed using CYANA and GROMACS as described previously.\textsuperscript{14,18} Briefly, both TOCSY and NOESY spectrum were used for assigning proton chemical shifts. The volumes of cross peaks from NOESY spectrum were integrated using the MestReNova software package (Mestre Lab Research, S.S.L., Santiago de Compostella, Spain). The integrated peaks were used as input for automated NOE assignment using \textit{noeassign} macro in CYANA 2.1, where the ambiguous interactions were resolved and assigned.\textsuperscript{19} The refined cross peak volumes were calibrated using \textit{caliba} macro in CYANA 2.1, and then converted to upper distance restraints for stimulated annealing and final structure calculation.\textsuperscript{5} The top 20 structures were generated from initial 200 random structure seeds. The structures were refined and energy-minimized following the GROMACS structure refining protocol described by Spronk \textit{et al.}\textsuperscript{20,21} The atomic coordinates, distance constraints and related NMR experimental conditions have been uploaded in the PDB with accession numbers: conformer 1 (PDB: 5TJ1), and conformer 2 (PDB: 6B5W).
3.5.7 BenE isopeptidase expression and purification

*E. coli* BL-21 ΔslyD cells bearing the plasmid, pCZ67, encoding BenE isopeptidase were grown overnight in LB media with ampicillin at 37 °C. The cells were subcultured into LB media with ampicillin at OD<sub>600</sub> 0.02, which was grown at 20 °C with shaking. When the culture reached OD<sub>600</sub> of 0.6 to 0.8, 1 mM IPTG was added to induce BenE isopeptidase production. The culture was then grown for 5 hours. Cells then were harvested by centrifugation at 8,000 x g for 10 min at 4 °C. The cells were lysed via sonication, and the isopeptidase was purified under native conditions on Ni-NTA resin according to the manufacturer’s recommendation (Qiagen). Purified protein was analyzed by SDS-PAGE.

3.5.8 BenE and AtxE2 isopeptidase assay

For a typical isopeptidase assay, 100 μM of peptide substrate and 100 nM of enzyme were mixed in the assay buffer (1 x PBS, pH 7.5) with a total volume of 100 μL. The reaction was incubated at 20 °C for 16 h. 95 μL of the reaction mixture was injected on a Zorbax 300SB-C18 Semi-Prep HPLC Column using solvent gradient 1 described above in the Lasso Peptide Expression and Purification section. Based on the difference in retention time of lasso peptide and linearized lasso peptide, isopeptidase digestion can be monitored by integrating the peaks corresponding to lasso peptide and linearized lasso peptide. To further confirm linearization, 1 μL of reaction mixture was diluted 10 fold with 2.5 mg/mL solution of α-cyano-4-hydroxycinnamic acid matrix, and
spotted onto an Applied Biosystem(ABI) 384 Opti-TOF 123 mm × 81 mm SS plate for MALDI-MS analysis.

3.5.9 Carboxypeptidase assay

For a typical carboxypeptidase assay, 50 μM of the peptide substrate was digested with 1 U carboxypeptidase B (Sigma-Aldrich) and 1 U carboxypeptidase Y (Affymetrix) in carboxypeptidase digestion buffer (50 mM sodium acetate, pH = 6.0) with a total volume of 20 μL for 16 hours at 20 °C. 18 μL of the digested mixture was injected on HPLC as described above. 1 μL of the digested sample was further analyzed by MALDI-MS as described above.

3.5.10 Trypsin assay

Trypsin digestion was conducted in trypsin digestion buffer (50 mM NH₄HCO₃). T12R benenodin-1 ΔC5 was treated with 0.1 μg of sequencing grade trypsin (Promega) in a total volume of 80 μL. The reaction was carried out for 16 hours at 37 °C. 60 μL was analyzed on HPLC and peaks of interest were collected. These peaks were subsequently analyzed using an LC-MS (Agilent 6530 Accurate-Mass Q-TOF LC/MS).
3.6 Supplementary materials

3.6.1 Supplementary figures

Figure 3.S1 NOESY spectrum of benenodin-1 ΔC5, conformer 1 (native state).
Figure 3.S2  TOCSY spectrum of benenodin-1 ∆C5, conformer 1 (native state).

Figure 3.S3  NOESY spectrum of benenodin-1 ∆C5, conformer 2.
Figure 3.S4 TOCSY spectrum of benenodin-1 ΔC5, conformer 2.

3.6.2 Supplementary tables

<table>
<thead>
<tr>
<th>Benenodin-1 ΔC5 Conformer 1</th>
<th>Benenodin-1 ΔC5 Conformer 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residue</td>
<td>Hydrogen</td>
</tr>
<tr>
<td>GLY-1</td>
<td>H</td>
</tr>
<tr>
<td>GLY-1</td>
<td>HA2</td>
</tr>
<tr>
<td>GLY-1</td>
<td>HA3</td>
</tr>
<tr>
<td>VAL-2</td>
<td>H</td>
</tr>
<tr>
<td>VAL-2</td>
<td>HA</td>
</tr>
<tr>
<td>VAL-2</td>
<td>HB</td>
</tr>
<tr>
<td>VAL-2</td>
<td>QQG</td>
</tr>
<tr>
<td>GLY-3</td>
<td>H</td>
</tr>
<tr>
<td>GLY-3</td>
<td>HA2</td>
</tr>
<tr>
<td>GLY-3</td>
<td>HA3</td>
</tr>
<tr>
<td>PHE-4</td>
<td>H</td>
</tr>
<tr>
<td>PHE-4</td>
<td>HA</td>
</tr>
<tr>
<td>PHE-4</td>
<td>HB2</td>
</tr>
<tr>
<td>PHE-4</td>
<td>HB3</td>
</tr>
<tr>
<td>PHE-4</td>
<td>QD</td>
</tr>
<tr>
<td>PHE-4</td>
<td>QE</td>
</tr>
<tr>
<td>PHE-4</td>
<td>HZ</td>
</tr>
<tr>
<td>GLY-5</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-----</td>
</tr>
<tr>
<td>GLY 5</td>
<td>HA2</td>
</tr>
<tr>
<td>GLY 5</td>
<td>HA3</td>
</tr>
<tr>
<td>ARG-6</td>
<td>H</td>
</tr>
<tr>
<td>ARG-6</td>
<td>HA</td>
</tr>
<tr>
<td>ARG-6</td>
<td>QB</td>
</tr>
<tr>
<td>ARG-6</td>
<td>HG2</td>
</tr>
<tr>
<td>ARG-6</td>
<td>HG3</td>
</tr>
<tr>
<td>ARG-6</td>
<td>HD2</td>
</tr>
<tr>
<td>ARG-6</td>
<td>HD3</td>
</tr>
<tr>
<td>ARG-6</td>
<td>HE</td>
</tr>
<tr>
<td>ARG-6</td>
<td>QH1</td>
</tr>
<tr>
<td>ARG-6</td>
<td>QH2</td>
</tr>
<tr>
<td>PRO-7</td>
<td>HA</td>
</tr>
<tr>
<td>PRO-7</td>
<td>HB2</td>
</tr>
<tr>
<td>PRO-7</td>
<td>HB3</td>
</tr>
<tr>
<td>PRO-7</td>
<td>HG2</td>
</tr>
<tr>
<td>PRO-7</td>
<td>HG3</td>
</tr>
<tr>
<td>PRO-7</td>
<td>HD2</td>
</tr>
<tr>
<td>PRO-7</td>
<td>HD3</td>
</tr>
<tr>
<td>ASP-8</td>
<td>H</td>
</tr>
<tr>
<td>ASP-8</td>
<td>HA</td>
</tr>
<tr>
<td>ASP-8</td>
<td>HB2</td>
</tr>
<tr>
<td>ASP-8</td>
<td>HB3</td>
</tr>
<tr>
<td>SER-9</td>
<td>H</td>
</tr>
<tr>
<td>SER-9</td>
<td>HA</td>
</tr>
<tr>
<td>SER-9</td>
<td>QB</td>
</tr>
<tr>
<td>ILE-10</td>
<td>H</td>
</tr>
<tr>
<td>ILE-10</td>
<td>HA</td>
</tr>
<tr>
<td>ILE-10</td>
<td>HB</td>
</tr>
<tr>
<td>ILE-10</td>
<td>HG12</td>
</tr>
<tr>
<td>ILE-10</td>
<td>HG13</td>
</tr>
<tr>
<td>ILE-10</td>
<td>QG</td>
</tr>
<tr>
<td>ILE-10</td>
<td>QD</td>
</tr>
<tr>
<td>LEU-11</td>
<td>H</td>
</tr>
<tr>
<td>LEU-11</td>
<td>HA</td>
</tr>
<tr>
<td>LEU-11</td>
<td>QB</td>
</tr>
<tr>
<td>LEU-11</td>
<td>HG</td>
</tr>
<tr>
<td>LEU-11</td>
<td>QQD</td>
</tr>
<tr>
<td>THR-12</td>
<td>H</td>
</tr>
<tr>
<td>THR-12</td>
<td>HA</td>
</tr>
<tr>
<td>THR-12</td>
<td>HB</td>
</tr>
<tr>
<td>THR-12</td>
<td>QG</td>
</tr>
<tr>
<td>Residue</td>
<td>Proton</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>GLN-13</td>
<td>H</td>
</tr>
<tr>
<td>GLN-13</td>
<td>HA</td>
</tr>
<tr>
<td>GLN-13</td>
<td>QB</td>
</tr>
<tr>
<td>GLN-13</td>
<td>QG</td>
</tr>
<tr>
<td>GLN-13</td>
<td>HE21</td>
</tr>
<tr>
<td>GLN-13</td>
<td>HE22</td>
</tr>
<tr>
<td>GLU-14</td>
<td>H</td>
</tr>
<tr>
<td>GLU-14</td>
<td>HA</td>
</tr>
<tr>
<td>GLU-14</td>
<td>QB</td>
</tr>
<tr>
<td>GLU-14</td>
<td>HG2</td>
</tr>
<tr>
<td>GLU-14</td>
<td>HG3</td>
</tr>
<tr>
<td>GLN-15</td>
<td>H</td>
</tr>
<tr>
<td>GLN-15</td>
<td>HA</td>
</tr>
<tr>
<td>GLN-15</td>
<td>QB</td>
</tr>
<tr>
<td>GLN-15</td>
<td>HG2</td>
</tr>
<tr>
<td>GLN-15</td>
<td>HG3</td>
</tr>
<tr>
<td>GLN-15</td>
<td>HE21</td>
</tr>
<tr>
<td>GLN-15</td>
<td>HE22</td>
</tr>
<tr>
<td>ALA-16</td>
<td>H</td>
</tr>
<tr>
<td>ALA-16</td>
<td>HA</td>
</tr>
<tr>
<td>ALA-16</td>
<td>QB</td>
</tr>
<tr>
<td>LYS-17</td>
<td>H</td>
</tr>
<tr>
<td>LYS-17</td>
<td>HA</td>
</tr>
<tr>
<td>LYS-17</td>
<td>HB2</td>
</tr>
<tr>
<td>LYS-17</td>
<td>HB3</td>
</tr>
<tr>
<td>LYS-17</td>
<td>HG2</td>
</tr>
<tr>
<td>LYS-17</td>
<td>HG3</td>
</tr>
<tr>
<td>LYS-17</td>
<td>QD</td>
</tr>
<tr>
<td>LYS-17</td>
<td>QE</td>
</tr>
<tr>
<td>LYS-17</td>
<td>HZ</td>
</tr>
<tr>
<td>PRO-18</td>
<td>HA</td>
</tr>
<tr>
<td>PRO-18</td>
<td>QB</td>
</tr>
<tr>
<td>PRO-18</td>
<td>HG2</td>
</tr>
<tr>
<td>PRO-18</td>
<td>HG3</td>
</tr>
<tr>
<td>PRO-18</td>
<td>QD</td>
</tr>
<tr>
<td>MET-19</td>
<td>H</td>
</tr>
<tr>
<td>MET-19</td>
<td>HA</td>
</tr>
<tr>
<td>MET-19</td>
<td>HB2</td>
</tr>
<tr>
<td>MET-19</td>
<td>HB3</td>
</tr>
<tr>
<td>MET-19</td>
<td>HG2</td>
</tr>
<tr>
<td>MET-19</td>
<td>HG3</td>
</tr>
<tr>
<td>MET-19</td>
<td>QE</td>
</tr>
</tbody>
</table>

Table 3.S1  Proton assignments for Conformers 1 and 2
Table 3.S2  Lists of NOEs observed between ring residues of benenodin-1 ΔC5 and steric lock residues Glu-14 (blue), Gln-15 (yellow), Ala-16 (green), and Lys-17 (gray) for conformer 1 (A) and conformer 2 (B). Distance restraints are classified as in CYANA as medium range (2.4 – 5.5 Å) or long range (greater than 5.5 Å).
<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primer 1</td>
<td>TCTCTAGATGGTAGCGTGAGGAGACC</td>
</tr>
<tr>
<td>Primer 2</td>
<td>TGGGTACCTTAGGTAGGGGC</td>
</tr>
<tr>
<td>Primer 3</td>
<td>CGCGATTTGAGATTCCATTGGGCGTACAAACCC</td>
</tr>
<tr>
<td>Primer 4</td>
<td>CGCCCCCAATGGAATTCTCAATCAGGCATCAAGG</td>
</tr>
<tr>
<td>Primer 5</td>
<td>GCTTCGCTTGGCGCTTGC</td>
</tr>
<tr>
<td>Primer 6</td>
<td>GACGCAGGCGCAAGCGAAGC</td>
</tr>
<tr>
<td>Primer 7</td>
<td>CGGCTTGGGCCGTCTTCTGGB</td>
</tr>
<tr>
<td>Primer 8</td>
<td>CGCAGGAAGCGGCGAAGCCG</td>
</tr>
<tr>
<td>Primer 9</td>
<td>CCCATCGGCGCCGCCTTGTTC</td>
</tr>
<tr>
<td>Primer 10</td>
<td>GAACAGCCGCGCGATGG</td>
</tr>
<tr>
<td>Primer 11</td>
<td>CCCGACAGCATCCTGCGTCAGAACAAGCGAAG</td>
</tr>
<tr>
<td>Primer 12</td>
<td>CTTCGCTTGGTGCTGACGCAGATCGCTGCGG</td>
</tr>
<tr>
<td>Primer 13</td>
<td>GAGTTATTTTACACTCCCTATCAG</td>
</tr>
<tr>
<td>Primer 14</td>
<td>GCAAGTCGAGGCAAGCGATC</td>
</tr>
<tr>
<td>Primer 15</td>
<td>GACAGCGTGGCCGCAAGGCAAGGCAAG</td>
</tr>
<tr>
<td>Primer 16</td>
<td>TTCTTGGCGCTACGCTGTCGGGAGCGG</td>
</tr>
<tr>
<td>Primer 17</td>
<td>TGACCCCTTTCGAAATTCCAGGAGGAAATGA</td>
</tr>
<tr>
<td>Primer 18</td>
<td>TGAGATCTCTGTCCCCGGGAGGCACTC</td>
</tr>
</tbody>
</table>

**Table 3.S3** Sequences of primers used in this study
<table>
<thead>
<tr>
<th>Plasmid Name</th>
<th>Description</th>
<th>Primers</th>
<th>Restriction Sites</th>
<th>Vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCZ35</td>
<td><em>benABC</em></td>
<td>1,2,3,4</td>
<td>Xba, EcoRI, KpnI</td>
<td>pASK75</td>
</tr>
<tr>
<td>pCZ31</td>
<td><em>benA</em>(E14A)<em>BC</em></td>
<td>3,4,5,6</td>
<td>Xba, EcoRI</td>
<td>pASK75</td>
</tr>
<tr>
<td>pCZ32</td>
<td><em>benA</em>(Q15A)<em>BC</em></td>
<td>3,4,7,8</td>
<td>Xba, EcoRI</td>
<td>pASK75</td>
</tr>
<tr>
<td>pCZ33</td>
<td><em>benA</em>(K17A)<em>BC</em></td>
<td>3,4,9,10</td>
<td>Xba, EcoRI</td>
<td>pASK75</td>
</tr>
<tr>
<td>pCZ85</td>
<td><em>benA</em>(T12R)<em>BC</em></td>
<td>3,4,11,12</td>
<td>Xba, EcoRI</td>
<td>pASK75</td>
</tr>
<tr>
<td>pCZ79</td>
<td><em>benA</em>(atxA22-loop)<em>BC</em></td>
<td>13,14,15,16</td>
<td>Xba, EcoRI</td>
<td>pASK75</td>
</tr>
<tr>
<td>pMW19</td>
<td>BenE with C-terminal His tag</td>
<td>17,18</td>
<td>EcoRI, BglII</td>
<td>pQE60</td>
</tr>
<tr>
<td>pCZ67</td>
<td>BenE with C-terminal His tag</td>
<td>-</td>
<td>EcoRI, HindIII</td>
<td>pQE80</td>
</tr>
</tbody>
</table>

*Table 3.S4* Plasmids constructed in this study
3.7 References


Chapter 4

Albusnodin: An Acetylated Lasso Peptide

By applying bioinformatics tools shown in Chapter 2, we have revealed a new lasso peptide gene cluster carrying out novel post-translational modifications from the genome of *Streptomyces albus* DSM 41398. Here, we described a lasso peptide from this cluster, albusnodin, that is post-translationally modified with an acetyl group, the first example of a lasso peptide with this modification. We show through genetic and biochemical studies that gene *albT*, encoding an acetyltransferase within the lasso peptide gene cluster, is essential for the biosynthesis of albusnodin. Our study provides new insights into diverse post-translational modifications on lasso peptide and illustrates a heterologous expression system for lasso peptide from Gram-positive bacteria.

This chapter is adapted from a research article submitted to Chemical Communications. This article is currently under review. This work is result of a collaboration with Wai Ling Cheung-Lee at Princeton University and Hader E. Elashal at Seton Hall Univeristy. Cheung-Lee conjugated the gene cluster of albusnodin into *Streptomyces* species. Elashal conducted the chemical cleavage of albusnodin.

4.1 Introduction

Actinobacteria have long been a rich source for lasso peptides, a family of RiPPs typified by their isopeptide-bonded slipknot structure.\textsuperscript{1,2} Many of the early examples of lasso peptides isolated by conventional natural product isolation methods were derived from Actinobacteria.\textsuperscript{3,4} With the quickening pace of genome sequencing and the advancement of tools for identifying lasso peptide gene clusters from these genomes, many additional examples of lasso peptides encoded in Actinobacterial genomes have been revealed.\textsuperscript{5,6} The lasso peptide topology is installed via a specific post-translational modification (PTM), an isopeptide bond between the N-terminus of the peptide and an Asp or Glu sidechain. Both genome mining and natural product isolation studies have also revealed lasso peptides with further PTMs including phosphorylation, citrullination, C-terminal methylation, and disulfide bond formation.\textsuperscript{7,8,9,10}

Lasso peptides are often cryptic with minimal to zero expression in their host organisms. Thus biosynthetic gene cluster engineering and heterologous expression strategies are commonly used for isolating lasso peptides. Up to date, more than twenty lasso peptides from Gram-negative species have been characterized using \textit{E. coli} as a heterologous host.\textsuperscript{11} While it has been recognized that Actinobacteria are prolific producers of lasso peptides for some time, tools for the heterologous expression of lasso peptides from these organisms have lagged behind those developed for proteobacteria. There is one notable exception: the lasso peptide svinceucin from \textit{Streptomyces sviceus} was
produced in mg/L yields in a *S. coelicolor* heterologous expression system. Heterologous expression has been a particularly useful tool in the study of lasso peptides since many of these peptides are not produced under standard culture conditions. In addition, heterologous expression can allow for tests of the function of genes involved in lasso peptide biosynthesis and post-translational modification.

Post-translational modifications (PTMs) play a fundamental role in the dynamic regulation of protein functions and modulating various of biological processes. Among the wide range of PTMs, Nε-lysine acetylation is one of the prominent modifications that affects countless cellular events involving in metabolism and regulation. Nε-lysine acetylation is typically implemented by a group of enzymes: lysine acetyltransferases (KATs). One of most known classes of KATs is GCN5-related N-acetyltransferase (GNAT) superfamily. GNATs catalyze the transfer of acetyl groups from acetyl-CoAs to the side chains of lysine residues (Figure 4.1).

![Figure 4.1](image)

**Figure 4.1** Acetylation reaction catalyzed by GCN5-related N-acetyltransferase (GNATs). The acetyl group is provided by acetyl-coenzyme A.
The most well-studied Nε-lysine acetylation events are histone acetylations in eukaryotes.\textsuperscript{15} Once the lysine residues become acetylated, the overall positive charge on histone decreases. This alters the affinity between histones and DNAs, further affecting gene expression.\textsuperscript{16} Recent efforts have revealed Nε-lysine acetylation also prevalent in bacteria, such as \textit{Escherichia coli}, \textit{Salmonella enterica}, and \textit{Bacillus subtilis}.\textsuperscript{17,18,19} These examples including chemotaxis, cell shape, RNA metabolism, and DNA regulation are initial evidence that Nε-lysine acetylation acts as an effective modulator of diverse biological processes in prokaryotes.\textsuperscript{20}

In this study, we demonstrate the heterologous expression of a novel lasso peptide, albusnodin, encoded in the genome of \textit{Streptomyces albus} DSM 41398 in the hosts \textit{S. coelicolor} and \textit{S. lividans}. The albusnodin gene cluster includes a gene for a putative acetyltransferase. Only monoacetylated albusnodin was produced upon heterologous expression. We further show that the acetyltransferase gene is indispensable for the biosynthesis of albusnodin. Thus albusnodin is the first example of a lasso peptide with an obligate post-translational modification.

\section*{4.2 Results}

\subsection*{4.2.1 Discovery of the lasso peptide gene cluster of albusnodin}

In Chapter 2, we described an optimized genome mining method for lasso peptide discovery that uses pattern matching to identify potential lasso peptide
precursors (A genes) and then searches for adjacent B and C genes. We used this method to find a cluster in *Streptomyces albus* DSM 41398 with a gene cluster architecture unique from all previous experimentally characterized lasso peptides (Figure 4.2). Namely, the gene cluster from *S. albus* DSM 41398 includes a putative acetyltransferase giving it an ACBT architecture where the T gene refers to the putative acetyltransferase. In addition, whereas the vast majority of B genes in Actinobacteria are split between two open reading frames (ORFs), the B gene encoded in *S. albus* is a single ORF (Figure 4.2).

![Figure 4.2](image)  

**Figure 4.2** The gene cluster of albusnodin. The A gene encodes the precursor peptide and the B and C genes encode maturation enzymes. The T gene encodes a putative acetyltransferase installs the acetyl group on the Lys-10 of the core peptide on albusnodin.

In a recent genome mining study, Mitchell and colleagues identified this cluster as well, and noted that it belonged to a larger family of lasso peptides with 18 distinct precursors found in 29 strains. In further BLAST searches for homologs to the *S. albus* acetyltransferase gene, we identified 43 additional
distinct lasso peptide precursors that are found within gene clusters with an ACBT architecture (Figure 4.3).

### Table: Distinct Lasso Peptide Precursors

<table>
<thead>
<tr>
<th>Strain</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Streptomyces albus DSM 41398</td>
<td>---H6S---LGlST---ETRIVDELLPVEOQGALRPGQ---GOQKEDRRAYNC---</td>
</tr>
<tr>
<td>Streptomyces atratus</td>
<td>---MVHNC---REATPHEDLLPVEOQGALRPGQ---GOQKEDRRAYNC---</td>
</tr>
<tr>
<td>Streptomyces cellulosavus strain NRRL B-2493</td>
<td>---MVHPLQ---RAERTPHEDLLPVEOQGALRPGQ---GOQKEDRRAYNC---</td>
</tr>
<tr>
<td>Streptomyces griseoviolae strain NRRL ISP-5360</td>
<td>---MVHTPQ---AETDRENNLPEOQGALRPGQ---GOQKEDRRAYNC---</td>
</tr>
<tr>
<td>Streptomyces griseus NRRL WC-3480</td>
<td>---MRKLT---TIMQDRQDAAELIEOQDAAZLGQ---GOQKEDRRAYNC---</td>
</tr>
<tr>
<td>Streptomyces lauritii</td>
<td>---MRKLH---TIMQDRQDAAELIEOQDAAZLGQ---GOQKEDRRAYNC---</td>
</tr>
<tr>
<td>Streptomyces noursei ATCC 11455</td>
<td>---MDLVLV---TDTPHEIRLVQDAAZLGQ---GOQKEDRRAYNC---</td>
</tr>
<tr>
<td>Streptomyces xiamenensis</td>
<td>---MDVCT---ITLGEDAAALIEOQDAAZLGQ---GOQKEDRRAYNC---</td>
</tr>
<tr>
<td>Streptomyces sp. AA4</td>
<td>---MHEPLR---REDITPTROMVDGQIDAEQLZLGQ---GOQKEDRRAYNC---</td>
</tr>
<tr>
<td>Streptomyces sp. AToxAB-D23</td>
<td>---MTGLR---REDITAEQPTOVDGQIDAEQLZLGQ---GOQKEDRRAYNC---</td>
</tr>
<tr>
<td>Streptomyces sp. CB01249</td>
<td>---MIDCNE---REDTTPERKVNGDQAELZLGQ---GOQKEDRRAYNC---</td>
</tr>
<tr>
<td>Streptomyces sp. DpondAA-D4</td>
<td>---MIITNE---NTSTTPKREWAVLGAELZLGQ---GOQKEDRRAYNC---</td>
</tr>
<tr>
<td>Streptomyces sp. DvalAA-A9</td>
<td>---NTLP---RTETAPAGAELDQAELZLGQ---GOQKEDRRAYNC---</td>
</tr>
<tr>
<td>Streptomyces sp. MJM8645</td>
<td>---MSQRTWEMDQGQIDAEQLZLGQ---GOQKEDRRAYNC---</td>
</tr>
<tr>
<td>Streptomyces sp. MnntMP-M17</td>
<td>---MENYPI---QWETDAPRQDAAELIEOQDAAZLGQ---GOQKEDRRAYNC---</td>
</tr>
<tr>
<td>Streptomyces sp. MP311-18</td>
<td>---MEHVPENGADRAEKARGVEPVLQGETAVQAQOGLM---KGSEKDRAYN---</td>
</tr>
<tr>
<td>Streptomyces sp. NRRL B-1347</td>
<td>---KSTSPKHGRMD---KN---NTDTPKEPDDPVPVLQGETAVQAQOGLM---KGSEKDRAYN---</td>
</tr>
<tr>
<td>Streptomyces sp. NRRL F-5123</td>
<td>---MHEPLC---REDLTVSPDQVQVLKZLGQ---GOQKEDRRAYNC---</td>
</tr>
<tr>
<td>Streptomyces sp. NRRL F-4491</td>
<td>---ML---QTTQTPDPEEITQFMR---ERSTQDGFPYPYVVQVHGQGAAVTLQ---KGSEKDRAYN---</td>
</tr>
<tr>
<td>Streptomyces sp. OspMP-M45</td>
<td>---MV---TVLQDSSE---GPEKEDVLPVLQGETAVQAQOGLM---KGSEKDRAYN---</td>
</tr>
<tr>
<td>Streptomyces sp. XY431</td>
<td>---MV---TVLQDSSE---GPEKEDVLPVLQGETAVQAQOGLM---KGSEKDRAYN---</td>
</tr>
<tr>
<td>Actinomeliotichus sp. AD127-7</td>
<td>---MV---TVLQDSSE---GPEKEDVLPVLQGETAVQAQOGLM---KGSEKDRAYN---</td>
</tr>
<tr>
<td>Actinomadura formosensis</td>
<td>---MV---TVLQDSSE---GPEKEDVLPVLQGETAVQAQOGLM---KGSEKDRAYN---</td>
</tr>
<tr>
<td>Alloactinosynema iranicum</td>
<td>---MV---TVLQDSSE---GPEKEDVLPVLQGETAVQAQOGLM---KGSEKDRAYN---</td>
</tr>
<tr>
<td>Alloactinosynema iranicum strain IBRC-M 10403</td>
<td>---MV---TVLQDSSE---GPEKEDVLPVLQGETAVQAQOGLM---KGSEKDRAYN---</td>
</tr>
<tr>
<td>Actinosynema mirum DSM 43827</td>
<td>---MV---TVLQDSSE---GPEKEDVLPVLQGETAVQAQOGLM---KGSEKDRAYN---</td>
</tr>
<tr>
<td>Amycolatopsis alba DSM 44262</td>
<td>---MV---TVLQDSSE---GPEKEDVLPVLQGETAVQAQOGLM---KGSEKDRAYN---</td>
</tr>
<tr>
<td>Frankia sp. Allo2</td>
<td>---MV---TVLQDSSE---GPEKEDVLPVLQGETAVQAQOGLM---KGSEKDRAYN---</td>
</tr>
<tr>
<td>Frankia sp. BR</td>
<td>---MV---TVLQDSSE---GPEKEDVLPVLQGETAVQAQOGLM---KGSEKDRAYN---</td>
</tr>
<tr>
<td>Frankia sp. Cc49</td>
<td>---MV---TVLQDSSE---GPEKEDVLPVLQGETAVQAQOGLM---KGSEKDRAYN---</td>
</tr>
<tr>
<td>Frankia sp. Cc6</td>
<td>---MV---TVLQDSSE---GPEKEDVLPVLQGETAVQAQOGLM---KGSEKDRAYN---</td>
</tr>
<tr>
<td>Frankia sp. R43</td>
<td>---MV---TVLQDSSE---GPEKEDVLPVLQGETAVQAQOGLM---KGSEKDRAYN---</td>
</tr>
<tr>
<td>Lechevalieria aerocolonialis strain NBRC 13195</td>
<td>---MV---TVLQDSSE---GPEKEDVLPVLQGETAVQAQOGLM---KGSEKDRAYN---</td>
</tr>
<tr>
<td>Kitaatospora sp. MBE76</td>
<td>---MV---TVLQDSSE---GPEKEDVLPVLQGETAVQAQOGLM---KGSEKDRAYN---</td>
</tr>
<tr>
<td>Nocardiosis dassonvillii alburalbida NBRC 13392</td>
<td>---MV---TVLQDSSE---GPEKEDVLPVLQGETAVQAQOGLM---KGSEKDRAYN---</td>
</tr>
<tr>
<td>Nocardiosis dassonvillii dassonvillii DSM 43111</td>
<td>---MV---TVLQDSSE---GPEKEDVLPVLQGETAVQAQOGLM---KGSEKDRAYN---</td>
</tr>
<tr>
<td>Nocardiosis halofericans DSM 44410</td>
<td>---MV---TVLQDSSE---GPEKEDVLPVLQGETAVQAQOGLM---KGSEKDRAYN---</td>
</tr>
<tr>
<td>Nocardiosis kusunakensis DSM 45524</td>
<td>---MV---TVLQDSSE---GPEKEDVLPVLQGETAVQAQOGLM---KGSEKDRAYN---</td>
</tr>
<tr>
<td>Nocardiosis sp. NRRL B-16309</td>
<td>---MV---TVLQDSSE---GPEKEDVLPVLQGETAVQAQOGLM---KGSEKDRAYN---</td>
</tr>
<tr>
<td>Promicromonospora kroppenstedtii DSM 19349</td>
<td>---MV---TVLQDSSE---GPEKEDVLPVLQGETAVQAQOGLM---KGSEKDRAYN---</td>
</tr>
<tr>
<td>Pseudonocardia sp. Ae715_Ps2</td>
<td>---MV---TVLQDSSE---GPEKEDVLPVLQGETAVQAQOGLM---KGSEKDRAYN---</td>
</tr>
</tbody>
</table>

Figure 4.3 Complete list of distinct lasso peptide precursors found in actinobacteria with an ACBT gene cluster architecture. Asterisks from left to right highlight 1) the universally conserved Thr residue in the penultimate position of the leader peptide, the Gly residue at the N-terminus of the core peptide, the universally conserved Glu residue that forms the isopeptide bond, and a universally conserved Lys that is the site of acetylation. Note that some precursors include additional Lys residues and may be polyacetylated.

We used I-TASSER to generate a homology model of the acetyltransferase AlbT. The closest hit to AlbT with a structure in the PDB is a member of the GNAT family from *Campylobacter jejuni*. This enzyme has low

---

104
homology to AlbT (16 % identity/32 % similarity) and is larger than AlbT (163 aa vs. 147 aa for AlbT). While the acetyl-CoA binding pocket of the enzymes appears to be well-conserved structurally, other sections of the enzyme are more structurally divergent (Figure 4.4).

Figure 4.4 Structure prediction of the acetyltransferase found in the albusnodin gene cluster. A: Homology model of the AlbT acetyltransferase with catalytically important Tyr residue shown in stick representation. B: Alignment of the homology model in A with the C. jejuni PseH crystal structure (PDB 4XPL). The catalytically important Tyr residue is shown in sticks and acetyl-CoA is shown in space filling. The acetyl-CoA binding pocket is structurally well-conserved between the AlbT homology model and PseH though other regions of the protein are more poorly conserved.

4.2.2 Heterologous expression of albusnodin

To attempt to isolate the product of the S. albus gene cluster, we grew S. albus DSM 41398 in liquid media, but found no trace of the predicted lasso peptide in either the cell pellet or the culture media. We also attempted heterologous expression of the cluster in E. coli, which has been very successful.
in producing lasso peptides from proteobacteria. However, no lasso peptide was produced by *E. coli*. We turned next to heterologous expression in *Streptomyces* hosts. The *S. albus* gene cluster including the *A*, *C*, *B*, and *T* genes was cloned downstream of a constitutive promoter *ermEp* in the plasmid plJ10257\(^24\) generating the plasmid pCZ68. This plasmid was transformed into *E. coli* which was subsequently conjugated with *Streptomyces coelicolor* M1146\(^25\), *Streptomyces lividans* 66, and *Streptomyces albus* J1074, all strains that have been used previously as heterologous hosts.\(^26\)

### 4.2.3 Characterization of albusnodin

We tested each of these strains for lasso peptide production using MALDI-MS and found putative lasso peptide products in the *S. coelicolor* and *S. lividans* exconjugants, but surprisingly not in the *S. albus* J1074 exconjugant. The peptide was detected both in culture supernatants and lysate extracts of *S. coelicolor* and *S. lividans*. We have named this peptide albusnodin, representing a knotted peptide (nodin) from *S. albus*. The mass observed, MH\(^+\) = 1646.7, is consistent with a lasso peptide that is correctly cyclized with a single acetylation, but with the final amino acid of the peptide truncated (*Figure 4.5, 4.6, and 4.7*). Lasso peptides are commonly truncated from their C-termini when heterologously expressed.\(^27\) This is the only lasso peptide product observed from fermentation of the heterologous strains; no full-length product was observed nor was any non-acetylated product. This observation led to the question of whether the acetyltransferase is critical for the production of albusnodin.
Figure 4.5  MALDI-MS spectra of partially purified supernatants from *S. lividans* 66 cultures heterologously expressing albusnodin. Only one product is observed: albusnodin missing its C-terminal Cys with a single acetylation. When the acetyltransferase (*T*) gene is not included, no products are observed.
Figure 4.6  MALDI-MS spectra of partially purified supernatants from cultures heterologously expressing albusnodin in *S. coelicolor M1146*. When the acetyltransferase (T) gene is not included, no production of albusnodin is observed.
Figure 4.7  MALDI-MS spectra of cell lysates of S. coelicolor M1146 or S. lividans 66 harboring either the entire albusnodin gene cluster or a cluster lacking the acetyltransferase (T) gene. Removal of the T gene abrogates production of albusnodin. The same result is found in peptide isolated from the culture supernatant.

To address this, we constructed a plasmid, pCZ66, which contains the A, C, and B genes of the albusnodin cluster, but lacks the acetyltransferase. This cluster was introduced to the same three heterologous hosts described above, but we observed no lasso peptide products from these exoconjugants (Figure 109)
4.5). This data, combined with the observation of only acetylated peptide produced from the intact cluster, strongly suggests that the acetylation of albusnodin is an obligate PTM.

We next turned our attention to characterization of albusnodin. Using a C₁₈ column and a gradient that we have previously used to separate lasso peptides, albusnodin eluted exceptionally early, within the first two minutes of the HPLC run. In contrast, the well-studied lasso peptide microcin J25 elutes at around 17 minutes on this gradient and the more polar astexin-1, -2, and -3 peptides elute at around 14 minutes. The presence of albusnodin was confirmed by MALDI-MS analysis of the material collected from the column, but no obvious peak for albusnodin was observed using a UV detector at 215 nm (Figure 4.8). This low yield of albusnodin precluded structural analysis by NMR, the gold standard for establishing the threaded topology of lasso peptides. A distinct peak for albusnodin was visible when using an HPLC with a shallower gradient and a sensitive mass detector (Figure 4.9).
Figure 4.8  HPLC traces of albusnodin fractionation. Top: Using gradient 1 (Table 4.3), a mass corresponding to albusnodin elutes between 1 and 2 minutes. Middle: using a shallower gradient, gradient 2 (Table 4.4), the peptide elutes between 31.9 and 33.4 min. Bottom: Zoom-in of the middle HPLC trace showing no obvious peak for albusnodin.
Figure 4.9  LC trace (total ion chromatogram) of partially purified albusnodin using qTOF mass detector. The doubly fractionated sample shown in Figure 4.8 was injected (13 µL). Gradient 3 (Table 4.5) was used to generate this trace. Intensity on the y-axis refers to the total ion counts. The position of the albusnodin peak is indicated.
4.2.4 Locating the acetylated residue

Given that there is a single lysine, Lys-10, within albusnodin, we expected the acetyl group to be installed at this position. Fragmentation of albusnodin either using MALDI-TOF/TOF or LC-MS/MS confirmed this (**Figure 4.10** and 4.11).

**Figure 4.10** MS/MS analysis of albusnodin. Observed fragments are labeled on the cartoon. Fragment b11 includes the acetyl group while fragment b12 is missing the acetyl group.
Figure 4.11 Additional MS/MS spectra showing fragmentation of albusnodin. Top: MALDI-TOF/TOF, bottom three panels: qTOF LC-MS/MS.
It is noteworthy that while albusnodin has only a single potential acetylation site, several of the putative peptides homologous to albusnodin may be polyacetylated, (Figure 4.2 and Figure 4.3). To provide evidence that albusnodin exists in a threaded conformation, we carried out two protease digestions. Because of their knotted structure, lasso peptides are often resistant to proteases. Albusnodin contains two arginine residues, Arg-11 and Arg-12, which are potential cleavage sites for trypsin. Treatment of albusnodin with trypsin overnight led to no cleavage of the peptide, suggesting that the arginine residues fall within the loop region of albusnodin (Figure 4.12). We also carried out carboxypeptidase treatment of albusnodin, which can be used to report on the threaded state of lasso peptides. Albusnodin was resistant to carboxypeptidase cleavage, providing strong evidence that the peptide is threaded (Figure 4.12 and 4.13).
Figure 4.12 LC traces of partially purified albusnodin, albusnodin after 16 h of incubation with trypsin and albusnodin after 16 h of treatment with carboxypeptidase. Neither trypsin nor carboxypeptidase cleaves albusnodin. Gradient 3 was used here.

Figure 4.13 Mass spectra from the samples in Figure 4.12. Top: Spectra showing doubly and triply charged species. Bottom: Deconvoluted spectra.
4.2.5 The cyclization residue of albusnodin

Another potential ambiguity with the albusnodin structure is the fact that it has two acidic residues, Glu-8 and Asp-9, which can serve as the isopeptide bond location. There is precedence for lasso peptides with isopeptide-bonded ring sizes of 7-9 aa. Based on the bioinformatic analysis carried out above (Figure 4.3), we suspected that albusnodin was cyclized at Glu-8 since Glu in the 8th position of the peptide is universally conserved, but Asp-9 is not conserved. To provide support for this assertion, we used a recently described chemical cleavage method that is able to cleave peptides N-terminal to Ser, Cys, and Glu residues, but not Asp residues. We observed a peptide mass of 1689.75 consistent with a singly cleaved product that is presumably ring-opened at Ser-7 (Figure 4.14). If Glu-8 was not participating in the isopeptide bond, we would expect cleavage N-terminal to it, but we did not observe any masses consistent with cleavage at Glu-8. Combining the data, we can predict the overall structure of albusnodin as having an isopeptide-bonded ring between Gly-1 and Glu-8, acetylation on Lys-10, and the Arg-11 and Arg-12 residues within the loop of the peptide. It is unknown exactly where the C-terminal tail of the peptide threads through the ring, but there is highly-conserved tyrosine at position 14 which may serve as a steric lock residue that helps maintain the threaded structure (Figure 4.15).
Figure 4.14 Chemical cleavage of albusnodin. Top: Cleavage is only observed at Ser-7, suggesting that the peptide is cyclized at Glu-8. Bottom: Mass spectrum of intact albusnodin for comparison.

4.2.6 Conclusion and future direction

Here we have characterized a novel lasso peptide, albusnodin, from Actinobacteria using heterologous expression in *Streptomyces* strains. The peptide is only produced upon coexpression of a tailoring acetyltransferase enzyme, suggesting that acetylation of albusnodin may be an obligate PTM for the peptide. This is the first experimental demonstration of a lasso peptide with
an acetylation PTM, and adds to the growing list of lasso peptides that are tailored by PTMs. As many lasso peptides exhibit antimicrobial activity, it is possible that the acetylation is a resistance mechanism to protect the producing cells from poisoning themselves. Another possibility is that the acetylation occurs prior to lasso formation, and that the acetyl group assists in the formation of the lasso structure with the lasso cyclase enzyme. Further improvements in the heterologous expression of albusnodin or other acetylated lasso peptides are expected to enable further structural and functional characterization of these natural products.

Figure 4.15 Cartoon model of the likely topology of full length albusnodin. The blue amino acids, G1 and E8 are isopeptide bonded. The site of acetylation, K10, is in red. C16 is encoded in the gene cluster but is cleaved off in the heterologously expressed peptide. The exact position of threading is unknown; one possibility is shown here.
4.3 Materials and methods

4.3.1 Bioinformatics

The genome sequences in fasta format from bacteria and archaea were downloaded from NCBI genome database. An updated version of the original genome mining script developed by Maksimov et al. was used in this study. In this version, a more flexible lasso peptide precursor pattern was used, allowing for variation beyond glycine in the first amino acid of the core peptide. Additionally, a new MEME training set containing 28 known lasso peptide clusters was used to generate four motifs from B enzymes and three motifs from C enzymes by applying the MEME searching algorithm. The updated motif information was then incorporated into the script for maturation enzyme motif matching using MAST. Putative lasso peptide clusters were ranked from 0 to 7 based on the numbers of motifs found in maturation enzymes.

4.3.2 Growth of S. albus and genomic DNA extraction

The freeze dried culture of Streptomyces albus DSM 41398 was purchased from DSMZ. The culture pellet was rehydrated in GYM medium (glucose 4.0 g, yeast extract 4.0 g, malt extract 10.0 g per liter of distilled water, pH = 7.2) and grown on a GYM plate (glucose 4.0g, yeast extract 4.0g, malt extract 10.0g, CaCO₃ 2.0g, agar 12.0 g per liter of water, pH = 7.2) at 30 °C for 6 days. Spores were harvested and subcultured into 10 mL of GYM medium for genomic DNA extraction. The whole genomic DNA of Streptomyces albus DSM 41398 was extracted using Qiagen DNeasy Blood & Tissue Kit (Qiagen).
PicoMaxx DNA polymerase (Agilent Technologies) was used for albusnodin gene cluster amplification with primers purchased from Integrated DNA Technologies.

4.3.3 Plasmid construction and conjugation of albusnodin gene clusters

Molecular cloning was done according to standard protocols. The plasmids used in this study is shown in Table 4.1. The fragments containing the native RBS and \textit{albACB} or \textit{albACBT} gene clusters were amplified using primers from Table 4.2. This PCR amplification also introduced the restriction sites EcoRI and XhoI at 5' end and HindIII at 3' end of the fragments. Both \textit{albACB} and \textit{albACBT} were cloned into the multiple cloning sites (EcoRI and HindIII) of pQE80 (Qiagen) to create pCZ64 and pCZ65 accordingly. The plasmids pCZ64 and pCZ65 were then digested with XhoI and HindIII to generate fragments \textit{XhoI-albACB-HindIII} and \textit{XhoI-albACBT-HindIII}. The two fragments were inserted and ligated into the multiple cloning sites (XhoI and HindIII) of \textit{Streptomyces} expression vector pIJ10257\textsuperscript{24} to generate pCZ66 and pCZ68.

<table>
<thead>
<tr>
<th>Plasmid Name</th>
<th>Description</th>
<th>Primers</th>
<th>Restriction Sites</th>
<th>Vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCZ64</td>
<td>\textit{albACB}</td>
<td>1,2</td>
<td>EcoRI, XhoI, HindIII</td>
<td>pQE80</td>
</tr>
<tr>
<td>pCZ65</td>
<td>\textit{albACBT}</td>
<td>1,3</td>
<td>EcoRI, XhoI, HindIII</td>
<td>pQE80</td>
</tr>
<tr>
<td>pCZ66</td>
<td>\textit{albACB}</td>
<td>-</td>
<td>XhoI, HindIII</td>
<td>pIJ10257</td>
</tr>
<tr>
<td>pCZ68</td>
<td>\textit{albACBT}</td>
<td>-</td>
<td>XhoI, HindIII</td>
<td>pIJ10257</td>
</tr>
</tbody>
</table>

\textbf{Table 4.1} Plasmids used in this study
Table 4.2 Primers used in this study

The resulting plasmids were transformed into *S. albus* J1074 and *S. lividans* 66 by conjugation with *E. coli* S17-1 using standard protocols. Due to methyl-specific restriction mechanisms of *S. coelicolor* M1146, methylation deficient strain *E. coli* ET12567/pUZ8002 was used for transformation and conjugation with *S. coelicolor* M1146. Exconjugants were selected with hygromycin (50 μg/mL) and nalidixic acid (25 μg/mL). Successful exconjugants were restreaked onto fresh SFM plates (20 g soy flour, 20 g mannitol, 20 g agar, 10 mM MgCl$_2$ per liter of water) with selection for an additional 7 days before harvesting spores.

4.3.4 Albusnodin expression and purification

For a typical heterologous production of albusnodin, *Streptomyces* strains were cultured in GYM medium with total volume of 1L. All strains were cultured in 5 x 1000 mL Erlenmeyer flasks (each flask contained 200ml of GYM medium) that have been treated with 1mL dimethyldichlorosilane in toluene to minimize the culture from sticking to the glass wall and a stainless steel spring was placed at the bottom of each flask to improve aeration. Starting cultures were inoculated...
with $2 \times 10^9$ spores in 150 mL 2YT medium (tryptone 16.0g, yeast extract 10.0g, sodium chloride 5.0g per liter of water, pH = 7.0) and incubated at 30 °C with orbital shaking (350 rpm) for 8 h. The germlings were harvested by centrifugation at 2,000 x $g$, resuspended in 25 mL of GYM medium, dispersed by a quick sonication pulse and used to inoculate production cultures at a starting OD$_{450}$ of 0.005. Production cultures were incubated at 30 °C with orbital shaking (250 rpm) for additional 7 days.

The cell lysate and supernatant of the samples were separated by centrifugation at 8,000 x $g$ at 4 °C for 20 minutes. The cells were lysed in methanol. The insoluble fraction was removed by centrifugation at 8,000 x $g$ at 4 °C for 10 minutes, and the methanol-soluble fraction was dried using rotary evaporation and then resuspended in 500 μL of 50% acetonitrile (ACN) and 50% water solution. The supernatant was extracted using Strata C8 column (Phenomenex 6 mL size). The crude extract was eluted from the column using methanol, and dried using rotary evaporation, and then reconstituted in 500 μL of 50% acetonitrile (ACN) and 50% water solution.

4.3.5 HPLC analysis

The peptide samples from cell lysate and supernatant were injected onto a Zorbax 300SB-C18 Analytical-Prep HPLC Column (3.0 x 150mm, Agilent Technologies) using solvent gradient 1 (see Table 4.3). An HPLC fraction from 1 to 2 minutes contained the peptide of interest and was collected. The fraction was dried under reduced pressure, resuspended into 100 μL of water, and then
subjected onto a Zorbax 300SB-C18 Semi-Prep HPLC Column (9.4 x 250mm, Agilent Technologies) using solvent gradient 2 (See Table 4.4). Solvent A consisted of H₂O with 0.1 % trifluoroacetic acid (TFA) and solvent B consisted of acetonitrile with 0.1% TFA.

<table>
<thead>
<tr>
<th>Flow Rate</th>
<th>0.750 mL/min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (min)</td>
<td>% Solvent B</td>
</tr>
<tr>
<td>0.00</td>
<td>10%</td>
</tr>
<tr>
<td>1.00</td>
<td>10%</td>
</tr>
<tr>
<td>20.00</td>
<td>50%</td>
</tr>
<tr>
<td>25.00</td>
<td>90%</td>
</tr>
<tr>
<td>30.00</td>
<td>90%</td>
</tr>
<tr>
<td>32.00</td>
<td>10%</td>
</tr>
<tr>
<td>36.00</td>
<td>10%</td>
</tr>
</tbody>
</table>

Table 4.3 Solvent gradient 1

<table>
<thead>
<tr>
<th>Flow Rate</th>
<th>2.667 mL/min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (min)</td>
<td>% Solvent B</td>
</tr>
<tr>
<td>0.00</td>
<td>1%</td>
</tr>
<tr>
<td>1.00</td>
<td>1%</td>
</tr>
<tr>
<td>60</td>
<td>20%</td>
</tr>
<tr>
<td>70</td>
<td>90%</td>
</tr>
<tr>
<td>80</td>
<td>1%</td>
</tr>
<tr>
<td>100</td>
<td>1%</td>
</tr>
</tbody>
</table>

Table 4.4 Solvent gradient 2

4.3.6 MALDI analysis

The fraction from 29.5 to 31.5 min (gradient 2) was collected, and 1 μL of the collected peak was spotted on a MTP 384 target plate ground steel BC plate.
Subsequently, 1 μL of 2.5 mg/mL solution of α-Cyano-4-hydroxycinnamic acid (Sigma) in 50% acetonitrile/water was mixed with the spotted sample and allowed to dry. Molecular weight determination was performed in the $m/z$ 300–3,000 range using an UltraFlextreme MALDI TOF/TOF (Bruker). MS/MS product ion analysis was performed with collision energies of 20 to 100 eV. Both MS and MS/MS analyses were performed in positive-ion mode.

### 4.3.7 LC-MS analysis

LC-MS analysis was performed on an Agilent 1260 Infinity II HPLC system coupled to an Agilent 6530 Accurate-Mass Q-TOF spectrometer with an electrospray ionization (ESI) source operated in positive ion mode. The analytes were separated with an Agilent Zorbax 300SB C18 column with the solvent gradient 3 (Table 4.5). Solvent A consisted of H$_2$O with 0.1 % formic acid (FA) and solvent B consisted of acetonitrile with 0.1% FA. The analytes from the column were sent to the MS and spectra were acquired in profile mode. When running tandem MS/MS, the +2 and +3 charge states (Z) of albusnodin were used as the target ion with a specific retention time of 11.5 minutes +/- 0.5 minutes. Collision energies of 25-90 eV were used to obtain MS/MS spectra.

<table>
<thead>
<tr>
<th>Flow Rate</th>
<th>0.400 mL/min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post Time</td>
<td>8 min</td>
</tr>
<tr>
<td>Time (min)</td>
<td>% Solvent B</td>
</tr>
<tr>
<td>0.00</td>
<td>1%</td>
</tr>
<tr>
<td>1.00</td>
<td>1%</td>
</tr>
</tbody>
</table>
Table 4.5  Solvent gradient 3

<table>
<thead>
<tr>
<th>%</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>60.00</td>
<td>20%</td>
</tr>
<tr>
<td>70.00</td>
<td>90%</td>
</tr>
<tr>
<td>80.00</td>
<td>1%</td>
</tr>
<tr>
<td>100.00</td>
<td>1%</td>
</tr>
</tbody>
</table>

4.3.8 Carboxypeptidase assay

The peptide substrate was digested with 1 μL of 0.1 μg/μL carboxypeptidase B (Sigma-Aldrich) and 1 μL of 0.1 μg/μL carboxypeptidase Y (Affymetrix) in carboxypeptidase digestion buffer (50 mM sodium acetate, pH = 6.0) with a total volume of 20 μL for 16 hours at 20 °C. The digested sample was further analyzed using an LC-MS (Agilent 6530 Accurate-Mass Q-TOF LC/MS).

4.3.9 Trypsin assay

Trypsin digestion was conducted in trypsin digestion buffer (50 mM NH₄HCO₃). The peptide substrate was treated with 1μl of 0.1 μg/μL of sequencing grade trypsin (Promega) in a total volume of 20 μL. The reaction was carried out for 16 hours at 37 °C. The digested sample was subsequently analyzed using an LC-MS (Agilent 6530 Accurate-Mass Q-TOF LC/MS).

4.3.10 Chemical cleavage

To activate Ser and accounting for the possible activation of Glu, excess N,N’-disuccinimidyl carbonate (50 mg), diisopropylethylamine (40 μl of a 20% vol/vol solution in dimethylformamide (DMF)), and a crystal of dimethylaminopyridine were added to partially purified, lyophilized albusnodin in
450 µl of DMF in a 5 mL round bottom flask. The reaction mixture was left on a shaker for 5 minutes to ensure solubility of the reagents followed by the addition of 450 µl of water to allow for simultaneous amino acid modification and amide bond cleavage. The reaction was left on a shaker at room temperature for 48 hours which was followed by the analysis of the reaction mixture using LC-MS as described above.
4.4 References


(25) Gomez-Escribano, J. P.; Bibb, M. J. Engineering *Streptomyces coelicolor*


(34) Maksimov, M. O.; Pelczer, I.; Link, A. J. Precursor-centric genome-mining

Chapter 5

Construction of Lasso Peptide Fusion Proteins

As described in Chapter 1, lasso peptides not only have a broad spectrum of bioactivities, but also often exhibit intrinsic resistance against proteolytic and thermal degradation. This makes lasso peptides are attractive starting points for therapeutic peptide development. Despite progress made in the field, a high-throughput screening system on lasso peptide libraries and a platform for engineering lasso peptides as a robust reprogrammable molecular scaffold are yet to be established. Here, we demonstrated a fusion of two model proteins, the artificial leucine zipper A1 and the superfolder variant of GFP, to the C-terminus of the lasso peptide astexin-1. The ability to fuse lasso peptides to a protein of interest opens the doors for the development of lasso peptide display and screening systems. Moreover, we combined this fusion approach with cysteine bio-conjugation chemistry to graft two anticancer epitopes on astexin-1. This work opens new avenues for engineering lasso peptides as a stable scaffold for bioactive peptide sequences with promising therapeutic applications.

This chapter is reproduced with permission from:
The work in the section of lasso peptide tail extension and epitope grafting was mainly conducted by Kenneth Hubbell and Shubham Chatterjee. Hubbell constructed the lasso peptide epitope targeting VEGF, and Shubham created the lasso peptide epitope targeting TEAD4. The detailed information was provided in their senior theses.


5.1 Introduction

Because of their diverse bioactivities and protease resistance, lasso peptides are attractive candidates for peptide therapeutics. Previous work demonstrated the construction of saturation mutagenesis libraries of the lasso peptide microcin J25 (MccJ25).\textsuperscript{1} In addition, the integrin-binding tripeptide epitope RGD has been inserted into MccJ25.\textsuperscript{2} These studies provide the first glimpses at using lasso peptides as reprogrammable scaffolds. In order to perform high-throughput screens on lasso peptide libraries and create more reprogrammable scaffolds, it would be desirable to fuse lasso peptides to an appropriate carrier protein for any display systems. This chapter describes the construction of the first ever lasso peptide fusions.

Since lasso peptides do not have a free N-terminus, fusions can only be constructed by adding a peptide or protein to the C-terminus of the peptide. Previous work attempted to append sequences at the tail of lasso peptides, such as MccJ25. However, the production of peptides diminished to levels just above MS detection.\textsuperscript{3} With the development of the lasso peptide field, more than 20 lasso peptide structures were released in PDB database. Among these characterized lasso peptides, astexin-1 (PDB 2N68), has a shorter loop region and longer tail than MccJ25 (Figure 5.1A). The longer C-terminal tail region of astexin-1 makes it as an excellent candidate for fusion with proteins.\textsuperscript{4}

Here, we have developed a lasso peptide fusion system where we fused two model proteins, the artificial leucine zipper A1 and the superfolder variant of
GFP, to the C-terminus of the lasso peptide astexin-1. This system opens up the possibility of grafting peptide-based epitopes of therapeutic interest into a hyperstable lasso peptide scaffold. Combining with cysteine bio-conjugation chemistry, we successfully grafted two epitopes, VEGF binding epitope and TEAD4 binding epitopes, onto astexin-1.

5.2 Results

5.2.1 Astexin-1 as the model fusion peptide

Astexin-1 consists of a nine aa ring, a five aa loop, and a nine aa tail (Figure 5.1A). Compared other structurally characterized lasso peptides, the flexible long tail of astexin-1 makes it more accessible to protein fusions at C-terminus. In the meantime, astexin-1 does not have any antibiotic activities against heterologous expression host *E. coli*. These combined characters make astexin-1 a great model peptide to construct the lasso peptide fusion protein.

![Lasso peptide structure and gene cluster architecture. (A) Astexin-1 biosynthetic gene cluster](image)

Figure 5.1 Lasso peptide structure and gene cluster architecture. (A) Astexin-1 biosynthetic gene cluster.
(PDB: 2N68) with the ring (salmon), loop (blue), and tail (green) sections indicated. The steric lock residues are shown in pink. (B) Gene clusters for astexin-1 production (top) and astexin-1-fusion production (bottom). The fusion protein gene is added to the 3’ end of the astexin-1 precursor gene atxA1.

5.2.2 Leucine zipper protein A1 as the model fusion protein

We initially selected an artificial leucine zipper protein A1 (A1 for abbreviation) as a model protein. A1 is a small helical protein (62 aa, 6.9 kDa). It is not only well-folded but also expressed at mg/L level in E. coli. 5,6

5.2.3 Direct fusion to astexin-1

![Electrospray mass spectra of direct fusions of astexin-1 to A1 leucine zipper.](image)

The biosynthesis of astexin-1 involves three essential genes, atxA1, atxB1 and atxC1. The atxA1 encodes the precursor peptide AtxA1 which consists of an N-terminal leader peptide and a C-terminal core peptide. The leader peptide on
AtxA1 is cleaved off by AtxB1, and AtxC1 catalyzes the isopeptide bond formation and folds the core peptide into a final lasso shape. To make lasso peptide fusion proteins, we first attempted to link A1 directly to the C-terminus of astexin-1. We constructed a plasmid, pCZ1, where the A1 leucine zipper gene was placed right after the precursor gene atxA1. The genes atxB1 and atxC1 encode for maturation enzymes AtxB1 and AtxC1 were located downstream of the fusion genes (Figure 5.1B). In addition, we designed two control plasmids pCZ5 with the deletion of atxC1 and pCZ2 lacks both atxB1 and atxC1.

After analysis of the fusion proteins using mass spectrometry, we found proteins produced from both pCZ1 and pCZ5 shared the same mass of 10,349 Da. This mass corresponds to the astexin-1-A1 fusion with its leader peptide removed, but without dehydration due to cyclization. As we expected, pCZ2 produced the unprocessed fusion protein (Figure 5.2). The lack of lasso peptide formation in pCZ1 indicated that A1 leucine zipper sterically blocked AtxC1 to catalyze the cyclization reaction on Atx1-A1 fusion. Thus, we a flexible linker might be necessary for the proper function of the cyclization enzyme AtxC1. The production scheme is shown in Figure 5.3. However, it is noteworthy that AtxB1 is able to completely cleave the leader peptide from the precursor in both the presence and absence of AtxC1, indicating that AtxB1 is not sterically hindered from accessing the cleavage site between the leader and core peptide. This is in contrast to an observation in an in vitro study of MccJ25 biosynthesis in which McjB, homologous to AtxB1, was unable to cleave the leader peptide from the precursor McjA in the absence of McjC, the homolog of AtxC1.
Figure 5.3  Maturation of astexin-1 precursor fusion proteins. (A) Schematic of lasso peptide fusion protein maturation. The precursor to astexin-1 consists of a leader peptide (purple) and the core peptide (orange) that is converted into the lasso. Linkers (green), when present, are placed between the C-terminus of the core peptide and the fusion protein (red). The cysteine protease AtxB1 cleaves the leader peptide while the asparagine synthetase homologue AtxC1 generates the isopeptide bond that defines the lasso fold. (B) Constructs tested in this study and average masses (expected) of the various processed forms. A1: A1 artificial leucine zipper. sfGFP: superfolder GFP. (GSSG)$_x$: Gly-Ser-Ser-Gly repeat. Thb: thrombin cleavage site.
5.2.4 Introduction of Flexible Linkers

We first designed two flexible linkers, \((\text{GSSG})_2\) with two repeats of (Glycine-Serine-Serine-Glycine) peptide, and \((\text{GSSG})_5\) with five repeats of (Glycine-Serine-Serine-Glycine) peptide. We then generated two plasmids pCZ25 and pCZ16 where the \((\text{GSSG})_2\) and \((\text{GSSG})_5\) linkers were inserted between AtxA1 and A1 leucine zipper accordingly. By introducing the flexible linkers, the lasso forms of fusion proteins were generated from pCZ25 and pCZ16. Interestingly, the lasso conversion was approximately 50% by intact protein MS analysis, where both lassoed and linear fusion proteins coexisted as a mixture in the final product (Figure 5.4). We also constructed several controls by deleting B and C enzymes, the MS results were as expected (Figure 5.5).

Figure 5.4 Mass spectrometry of astexin-1-A1 fusion proteins with flexible linkers. (A) Deconvoluted electrospray mass spectrum of the astexin-1- \((\text{GSSG})_2\)-A1 construct showing a roughly 50:50 split between the lassoed and linear core peptide forms of the protein. (B) As in A, but for the astexin-1-(GSSG)_5-A1 construct. a.u.: arbitrary units.
Figure 5.5 Deconvoluted mass spectra of astexin-1-(GSSG)x-A1 fusion proteins. (A) Spectra of the astexin-1-(GSSG)2-A1 protein product of gene clusters lacking atxB1 (left) or both atxB1 and atxC1 (right). (B) As in panel A, but for the astexin-1-(GSSG)5- A1 protein.

5.2.5 Introduction of a thrombin digestion site

To accurately determine the conversion level of the lassoed fusion protein, we inserted a thrombin protease cleavage site (LVPR|GS, the peptide gets cleaved between R and G) after the (GSSG)2 and (GSSG)5 linkers to make pCZ29 and pCZ43. Treating lassoed/linear fusion protein mixture with thrombin,
the lasso peptide/linear peptide can be removed from the fusion proteins. Since the lasso peptide and linear peptide have different retention times on HPLC, we can quantify the abundance of lasso peptide and the extent of lassoed fusion protein formation.

We first analyzed the fusion proteins encoded on pCZ29 and pCZ43 using SDS-PAGE. The lassoed and linear fusion proteins were resolvable on the gel, where the lassoed product ran at lower molecular weight than its linear counterpart (Figure 5.6A). The masses of lasso and linear protein were confirmed using ESI-MS (Figure 5.6B). After thrombin digestion of the fusion proteins, three peaks corresponding to the lasso peptide, linear peptide, and the cleaved A1 protein appeared on HPLC (Figure 5.6C). The masses of lasso peptide and linear peptide were confirmed using MALDI-MS (Figure 5.7).
Figure 5.6  Quantification of the extent of lasso formation in astexin-1 fusion proteins. Left panels correspond to astexin-1-(GSSG)$_2$-thrombin-A1 protein; right panels correspond to astexin-1-(GSSG)$_5$-thrombin-A1 protein. (A) SDS-PAGE gel of purified astexin-1 fusion proteins: lane 1, intact astexin-1 gene cluster; lane 2, gene cluster lacking $atxC_1$; lane 3, gene cluster lacking $atxB_1$ and $atxC_1$. When both maturation enzymes are present, a mixture of lassoed and linear core peptide forms of the protein is produced (lane 1). If only the AtxB1 enzyme is present, the leader peptide is removed completely, but no lasso formation occurs (lane 2). If neither AtxB1 nor AtxC1 is present, the leader peptide is not cleaved. (B) Electrospray mass spectra of the mixture of lassoed and linear protein corresponding to the protein in lane 1 of panel A. (C) HPLC analysis of thrombin-digested astexin-1 fusion proteins. The area under the curve was computed to determine the ratio of lassoed to linear protein.
Figure 5.7  MALDI-MS analysis of thrombin-cleaved lasso and linear peptides.  (A) Spectrum of purified linear peptide (top) and lasso peptide (bottom) after thrombin digestion of the astexin-1-(GSSG)$_2$-thrombin-A1 construct. (B) As in A, but for the astexin-1-(GSSG)$_5$-thrombin-A1 protein. The lower intensity peaks 18 mass units below the labeled peaks are from MALDI-induced dehydration due to the high laser intensity needed to analyze these high molecular weight species.
By integration of the peaks for the lasso and linear peptides, the astexin-1-(GSSG)$_2$-thrombin-A1 construct obtained 49.7 ± 0.3% lasso peptide conversion, whereas the astexin-1-(GSSG)$_5$-thrombin-A1 construct with 53.1 ± 0.4% of the peptide lassoed. To further confirm the lassoed peptides were indeed in their threaded forms, we treated the lassoed peptides to carboxypeptidases. A mixture of carboxypeptidases B and Y removes amino acids from the C-terminus of the lasso peptide but is usually unable to cleave beyond the steric lock residues of the lasso peptide due to steric hindrance.\(^9,^{10}\) When we subjected the astexin-1-(GSSG)$_2$ and astexin-1-(GSSG)$_5$ lasso peptides to carboxypeptidase B and Y treatment, we observed astexin-1 (1–20) as a major product in MALDI-MS along with some products resulting from incomplete cleavage (Figure 5.8). The fact that the carboxypeptidase cocktail cannot cleave beyond this residue is strong confirmation that the astexin-1-(GSSG)$_2$ and astexin-1-(GSSG)$_5$ peptides are indeed lassoed.

### 5.2.6 Fusing Astexin-1 to GFP

To this point, we have successfully fused a small helical leucine zipper protein to the C-terminus of astexin-1, we then aimed to evaluate this fusion approach with a larger and structurally more complex model protein. We selected the superfolder variant of GFP (sfGFP) for such experiments. sfGFP\(^{11}\) has a \(\beta\)-barrel structure with a mass of 26.8 kDa, making it bulkier than the A1 protein. Since the (GSSG)$_5$ linker gave the highest conversion to lasso peptide for the A1
protein, an astexin-1-(GSSG)$_5$-thrombin-sfGFP construct was designed and expressed in *E. coli* (Figure 5.9).

**Figure 5.8** MALDI-MS analysis of carboxypeptidase-digested lasso peptides. Following thrombin cleavage and HPLC purification, the tail-extended astexin-1 peptides were digested with carboxypeptidase B and Y to confirm formation of
the lasso structure. A: Lasso peptide derived from the astexin-1-(GSSG)$_2$-thrombin-A1 protein B: Lasso peptide derived from the astexin-1-(GSSG)$_5$-thrombin-A1 protein. In both cases, the carboxypeptidase digestion stops at residue Ile-20 due to steric hindrance of the lasso fold.

Figure 5.9 Fusion of astexin-1 to superfolder GFP (sfGFP). (A) Gene cluster for astexin-1 sfGFP fusion production and model of the astexin-1- (GSSG)$_5$-thrombin-sfGFP fusion protein. The (GSSG)$_5$-thrombin-sfGFP portion of the protein was modeled in I-TASSER and then connected to the NMR structure of astexin-1. (B) Mass spectrum of the mixture of lassoed and linear protein produced by the gene cluster in A. (C) HPLC analysis of thrombin-digested astexin-1- (GSSG)$_5$-thrombin-sfGFP protein.

To test whether fusion of the lasso peptide has any effect on the fusion protein function, we measured the inherent fluorescence intensity of three different concentrations of purified astexin-1-(GSSG)$_5$-thrombin-sfGFP and
compared these measurements to the fluorescence of purified sfGFP alone (Figure 5.10). We saw no statistically significant difference in fluorescence between the two proteins, indicating that the lasso peptide fusion does not interfere with the folding or chromophore formation of sfGFP. The percentage of lasso peptide formed was 44.4 ± 0.6%, slightly lower than what was observed for the A1 protein (Figure 5.9).

![Comparison of fluorescence](image.png)

**Figure 5.10** Comparison of the fluorescence of astexin-1-(GSSG)₅-thrombin-superfolder GFP (red) to native superfolder GFP (blue).

### 5.2.7 Tail extension and epitope grafting

To this point, we have successfully appended model protein leucine zipper A1 and superfolder GFP to the C-terminus of the lasso peptide astexin-1. The ability to fuse lasso peptides to a protein of interest is an important step toward
peptide display systems for the high-throughput screening of lasso peptide libraries for new functions.

Previous work\textsuperscript{12} in the Link lab had attempted to insert bioactive sequences into the tail region of a lasso peptide. However, the inserted sequences are often susceptible to proteolytic degradation during expression.\textsuperscript{13,14} Our fusion approach opens the doors to design and engineer tail-extended lasso peptides. By treating thrombin-cleavable fusion protein with thrombin, we generated the longest lasso peptide (astexin-1-(GSSG)$_5$-thrombin), with a total of 45 amino acids and a tail of 33 amino acids. Combining this tail extension with other bioconjugation methods, our approach affirms the possibility of engineering lasso peptides as a stable scaffold for epitope grafting.

\textbf{Figure 5.11} Production schematic of tail extended lasso epitope. A cysteine is substituted into the lasso ring region, and a bioactive epitope is inserted between the tail of the lasso and the fusion protein. This linker region includes the
thrombin recognition site LVPRGS enabling the fusion protein to be removed in vitro. The cysteine in the ring and a cysteine in the epitope region oxidize to form a stabilizing disulfide bond.

To graft small bioactive sequences onto the tail of astexin-1, we introduced disulfide chemistry in the lasso peptide protein fusion system. Briefly, we first inserted the bioactive sequences with a cysteine end at the C-terminus of astexin-1, followed by a (GSSG)ₙ linker, a thrombin digestion site, A1 protein. Second, a cysteine substitution (V6C) in the ring of the astexin-1 reacted with the cysteine of the bioactive sequence to form a disulfide bond “tying off” the end of the lasso. Third, the final lasso epitope was generated by removal of A1 protein using thrombin. The production scheme is shown in Figure 5.11.

<table>
<thead>
<tr>
<th>Epitope</th>
<th>Binding Target</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>VEGF</td>
<td>IHVMEWECFRLC</td>
</tr>
<tr>
<td>2</td>
<td>TEAD4</td>
<td>ITGSVDDHFAKALGDTWLQIKAAAC</td>
</tr>
</tbody>
</table>

Table 5.1 Sequences of bioactive epitopes that are grafted on lasso peptide astexin-1(V6C)

To develop the “proof-of-concept” lasso epitopes, we have successfully grafted two bioactive sequences on astexin-1. The first sequence (CIHVMWEWECFERL) is an inhibitor for vascular endothelial growth factor (VEGF), where VEGF acts as a key mediator of angiogenesis in cancer (Table 5.1, Figure 5.12).₁⁵,₁⁶
The second sequence (ITGSVDHFAKALGDTWLQIKAAC) is the TDU2 domain of VGLL4, where VGLL4 is a competitive inhibitor to the oncoprotein Yes-Associated Protein (YAP), that prevents the interaction of YAP to the transcription factor TEAD4 (Table 5.1, Figure 5.13). TEAD4 is a transcriptional enhancer factor overexpressed in colorectal cancer. The detailed studies were included in Kenneth Hubbell’s and Shubham Chatterjee’s senior theses. Together we have demonstrated that lasso peptide can be modified to accommodate pharmaceutically-relevant sequences into disulfide-constrained loops between the lasso tail and the lasso ring. This is one of the first examples of an epitope being engineered into a lasso peptide, and the largest therapeutic sequence ever inserted into a lasso.
5.3 Conclusion

Here we developed astexin-1 based lasso peptide protein fusion system. First, we fused two model proteins, the artificial leucine zipper A1 and the superfolder variant of GFP, to the C-terminus of the lasso peptide astexin-1, with a flexible linker in between. Second, by conduction thrombin digestion and HPLC experiments, we determined lasso peptide fusion conversion can reach \( \sim 50\% \). Finally, by combining the fusion approach with cysteine bio-conjugation chemistry, we laid the foundation for using lasso peptide scaffold for epitope grafting. We successfully grafted two bioactive peptide sequences related to cancer therapeutics onto astexin-1. Further analysis related to binding properties of lasso epitopes towards their targets, stability against enzymatic, chemical and thermal degradation would be conducted to affirms this grafting approach. We are also seeking new therapeutic peptides sequences for the expanding lasso epitopes. In addition, we are currently seeking the possibility for lasso peptide display is display on *Caulobacter crescentus* via its S-layer protein RsaA. This protein is secreted via a C-terminal secretion sequence and is thus compatible
with fusion of a lasso peptide to the N-terminus of RsaA or one of its fragments. The fusion of lasso peptides to these bacterial protein scaffolds is expected to expand the toolbox for high-throughput screening of constrained peptide libraries.

5.4 Materials and methods

5.4.1 Strains and reagents

PicoMaxx DNA polymerase (Agilent) was used for PCR amplification with primers purchased from IDT. T4 ligase and restriction enzymes EcoRI, BglII, and KpnI for cloning were purchased from New England Biolabs. The XL-1 Blue E. coli strain was used for all recombinant DNA steps, while BL-21 ΔslyD was used for protein expression.

5.4.2 Plasmid construction

Table 5.S1 includes a list of all plasmids used in this paper. All new plasmids were derived from pQE-80, which includes a T5 promoter, a copy of lacIq, and an ampicillin resistance marker. The primers used to generate each construct are listed in Table 5.S2, and the sequences of all primers appear in Table 5.S2. The pCA14 plasmid includes the atxB1C1 operon encoding the astexin-1 maturation enzymes AtxB1 and AtxC1. This operon was amplified using PCR from the pMM32 plasmid9 and digested with BglII and HindIII. This product was ligated into similarly digested pQE80. To generate fusions between astexin-1 and the A1 leucine zipper,11 overlap PCR was employed. The atxA1 gene encoding the astexin-1 precursor was amplified from pMM32 with a 5′
The gene encoding the A1 leucine zipper includes a C-terminal 6x-His tag and was amplified from pQE9-A111 with a 3' *Bgl*II site. These two fragments were joined by overlap PCR, digested with *EcoRI* and *Bgl*II, and ligated into similarly digested pCA14, resulting in the plasmid pCZ1. The pCZ5 plasmid is a control for pCZ1, and it consists of the gene encoding the astexin-1-A1 fusion and just the AtxB1 maturation enzyme. A PCR product spanning the astexin-1-A1 fusion gene and AtxB1 was generated with a 3' end *Kpn*I site. This product was digested with *EcoRI* and *Kpn*I and ligated into pQE-80 to generate pCZ5. The pCZ2 plasmid is an additional control for pCZ1 and lacks both the AtxB1 and AtxC1 maturation enzymes. It was constructed in the same way as pCZ5.

The pCZ16/pCZ20/pCZ21 plasmid series is analogous to pCZ1/pCZ5/pCZ2 described above except that a flexible glycine-serine linker, (GSSG)₅, was inserted between astexin-1 and the A1 leucine zipper. The gene encoding astexin-1-(GSSG)₅-A1 was assembled by overlap PCR. The (GSSG)₅ fragment was amplified from pJP47 while the astexin-1 and A1 gene fragments were amplified from pCZ1. Once assembled, the PCR product was digested as above for pCZ1 and ligated into pCA14 to generate pCZ16. The pCZ20 and pCZ21 plasmids were generated in an analogous fashion to pCZ5 and pCZ2. The pCZ43 plasmid is similar to pCZ16 except it also includes a thrombin cleavage site between the glycine-serine linker and A1. pCZ43 was assembled in the same way as pCZ16, except that primers encoding the thrombin cleavage site were used. Control plasmids pCZ44 and pCZ45 were made in an analogous
fashion to pCZ20 and pCZ21. A set of plasmids with shorter glycine-serine linkers, \((GSSG)_2\), were also constructed, both without (pCZ25, pCZ26, pCZ27) and with (pCZ29, pCZ40, pCZ39) thrombin cleavage sites. These plasmids were derived from the constructs with the \((GSSG)_5\) linker.

Finally, plasmids with genes encoding for astexin-1-(GSSG)_5-superfolder GFP (pCZ22) and astexin-1-(GSSG)_5-thrombin-superfolder GFP (pCZ46) were constructed in the same way as pCZ16 and pCZ43, respectively. The superfolder GFP fragment was amplified from pDA38, which is a pQE60 derivative with the C-terminally 6x His-tagged superfolder GFP gene inserted in the multiple cloning site. All plasmid sequences were verified by Genewiz. The information of plasmids and primers are shown in Table 5.S1, 5.S2.

### 5.4.3 Purification of astexin-1

Expression and purification of astexin-1 were carried out in a similar fashion to the purification of astexin-321 with some minor changes. Briefly, BL21 E. coli cells harboring the plasmid pMM32 (contains the \(atxA1B1C1\) operon) were grown in M9 minimal medium and induced at an OD600 of 0.2 to 0.3 to produce astexin-1. The cells were lysed in methanol, and the insoluble fraction was removed by centrifugation. The methanol-soluble fraction, which contains astexin-1, was dried in a rotary evaporator, reconstituted in minimal water, and cleaned up using solid-phase extraction (Strata C8, 6 mL). Crude astexin-1 was eluted from the column using methanol which was subsequently removed using a rotary evaporator. The peptide was purified using HPLC (Zorbax 300SB-C18
Semi-Prep HPLC Column, 9.4 by 250 mm, Agilent Technologies) by applying the following solvent gradient at a flow-rate of 4.0 mL/min: 10% acetonitrile (ACN) in water for 1 min, ramp up to 50% ACN over 19 min, ramp up to 90% ACN over 5 min, 90% ACN for 5 min, ramp down to 10% ACN in 2 min. HPLC fractions containing astexin-1 were pooled and lyophilized (Labconco Freezone 4.5/−105 °C), resulting in a white powder.

5.4.4 Fusion protein expression

Plasmids were introduced into BL-21 ΔslyD for protein expression. Cells harboring plasmids encoding either astexin-1-A1 fusions or astexin-1-sfGFP fusions were grown in LB medium (10 g/L tryptone, 5 g/L yeast extract, and 10 g/L NaCl) with 100 mg/L ampicillin (Affymetrix) for plasmid maintenance. In a typical experiment, an overnight culture was diluted to an OD600 of 0.02, and the culture was grown at 37 °C until reaching an OD600 of 0.5. At this time, the culture was shifted to RT (∼21 °C), induced with 1 mM isopropyl-β-D-1-thiogalactopyranoside (Gold Biotechnology), and allowed to express protein for 16 h. Cells then were harvested by centrifugation at 8000g for 10 min at 4 °C, and cell pellets were frozen at −80 °C until needed for purification.

5.4.5 Fusion protein purification

Following protein expression, astexin-1-A1 fusion proteins and astexin-1-sfGFP fusion proteins were purified under native conditions. Cell pellets were resuspended in lysis buffer (50 mM NaH2PO4, 300 mM NaCl, 10 mM imidazole,
pH 8). The cells were incubated for 30 min on ice with the addition of 0.25 mg mL\(^{-1}\) lysozyme (Sigma-Aldrich) and lysed by sonication. Following lysis, cells were centrifuged at 10 000g for 10 min at 4 °C, and the supernatants containing fusion proteins were purified on Ni-NTA resin according to the manufacturer's recommendation (Qiagen). Purified proteins were analyzed by SDS-PAGE.

5.4.6 ESI-MS analysis of intact fusion protein

Electrospray ionization mass spectrometric (ESI-MS) analysis of fusion proteins was performed on a Bruker Daltonics MicroTOF-Q II mass spectrometer. Purified fusion proteins were first purified by HPLC using the same gradient and conditions described in the “Purification of astexin-1” section. Fusion proteins were collected based on their retention times on HPLC and subsequently lyophilized. Lyophilized samples were reconstituted in MS buffer (50% acetonitrile, 0.1% formic acid in H2O) and injected into the mass spectrometer. The observed average molecular weight of each fusion protein sample was obtained by deconvolution of the acquired intensity vs m/z data using the Maximum Entropy algorithm (Spectrum Square Associates). Theoretical average molecular weights for the fusion proteins were calculated using mMass.

5.4.7 Thrombin digestion of fusion proteins and HPLC analysis

Prior to thrombin digestion, proteins were purified via HPLC as described in the previous paragraph. The peak corresponding to the intact protein was collected and the protein solution was lyophilized. This dried protein was
reconstituted in thrombin cleavage buffer 1 (1× PBS, pH 7.3) for the A1 fusion proteins or thrombin cleavage buffer 2 (50 mM Tris, pH 8.0) for the sfGFP fusion proteins. Thrombin digestion assays were performed with 1.5−2 U thrombin (GE Healthcare) to cleave 450−750 μg of astexin-1-A1 fusion protein in 150−200 μL of thrombin cleavage buffer 1 for 16 h at RT (−21 °C). For each astexin-1-sfGFP fusion protein sample, thrombin digestion assays were performed with 3 U thrombin (GE Healthcare) to cleave 900−1300 μg protein in 300 μL of thrombin cleavage buffer 2 (50 mM Tris, pH 8) for 16 h at RT (−21 °C). Following thrombin digestion, digested fusion proteins were analyzed on HPLC using the same solvent gradient described above in the ESI-MS section. The typical injection volume of astexin-1-A1 fusion protein digests was 75–100 μL, while the injection volume of astexin-1-sfGFP fusions was 290 μL. To quantify the percentage of lasso peptide, the area under the peaks of the lasso peptide and the linear were determined using the HPLC software, and the percentage of lasso peptide was calculated as (area under lasso peak)/(area under lasso peak + area under linear peak). The percentage of lasso formation was calculated as the average of three or four biological replicates, and the error reported is the standard deviation of those measurements. The cleaved peptides (both lasso and linear) were also collected based on their retention time, dried under reduced pressure (DyNAVap, LabNet), and analyzed by MALDI mass spectrometry as well as carboxypeptidase digestion.
5.4.8 Carboxypeptidase digestion assays

Dried thrombin-cleaved lasso peptides and linear peptides were digested with 0.5 U carboxypeptidase B (Sigma-Aldrich) and 0.5 U carboxypeptidase Y (Affymetrix) in carboxypeptidase digestion buffer (50 mM sodium acetate, pH 6.0) for 3 h at RT. Digested samples were further analyzed by MALDI-MS as described above.

5.4.9 sfGFP fluorescence intensity measurement

For a typical fluorescence intensity measurement, 100 μL of purified sfGFP or astexin-1-(GSSG)$_5$-Thb-sfGFP at 2.5 μM, 5 μM, and 10 μM in elution buffer (50 mM NaH$_2$PO$_4$, 300 mM NaCl, 250 mM imidazole, pH 8) was added to a Costar 96-well black plate (Fisher Scientific) and equilibrated at 24 °C for 5 min. Protein concentrations were determined using the BCA assay according to the manufacturer’s recommendations (Pierce). Maximum emission at 510 nm was measured upon excitation with 485 nm light. The fluorescence level was measured in a BioTek Synergy 4 plate reader. Three technical replicates were performed for each concentration.
<table>
<thead>
<tr>
<th>Plasmid Name</th>
<th>Description</th>
<th>Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCA14</td>
<td>AtxB1C1</td>
<td>NA</td>
</tr>
<tr>
<td>pMM32</td>
<td>AtxA1-B1C1</td>
<td>NA</td>
</tr>
<tr>
<td>pQE9-A1</td>
<td>A1</td>
<td>NA</td>
</tr>
<tr>
<td>pDA38</td>
<td>sfGFP</td>
<td>NA</td>
</tr>
<tr>
<td>pJP47</td>
<td>(GSSG)$_5$</td>
<td>NA</td>
</tr>
<tr>
<td>pCZ1</td>
<td>Atx1-A1-BC</td>
<td>1,2,3,4</td>
</tr>
<tr>
<td>pCZ5</td>
<td>Atx1-A1-B</td>
<td>1,5</td>
</tr>
<tr>
<td>pCZ2</td>
<td>Atx1-A1-woBC</td>
<td>1,6</td>
</tr>
<tr>
<td>pCZ16</td>
<td>Atx1-(GSSG)$_6$-A1-BC</td>
<td>1,4,7,8,9,10</td>
</tr>
<tr>
<td>pCZ20</td>
<td>Atx1-(GSSG)$_6$-A1-B</td>
<td>1,5</td>
</tr>
<tr>
<td>pCZ21</td>
<td>Atx1-(GSSG)$_6$-A1-woBC</td>
<td>1,6</td>
</tr>
<tr>
<td>pCZ43</td>
<td>Atx1-(GSSG)$_6$-Thb-A1-BC</td>
<td>1,4,11,12</td>
</tr>
<tr>
<td>pCZ44</td>
<td>Atx1-(GSSG)$_6$-Thb-A1-B</td>
<td>1,5</td>
</tr>
<tr>
<td>pCZ45</td>
<td>Atx1-(GSSG)$_6$-Thb-A1-woBC</td>
<td>1,6</td>
</tr>
<tr>
<td>pCZ25</td>
<td>Atx1-(GSSG)$_6$-A1-BC</td>
<td>1,4,13,14</td>
</tr>
<tr>
<td>pCZ26</td>
<td>Atx1-(GSSG)$_6$-A1-B</td>
<td>1,5</td>
</tr>
<tr>
<td>pCZ27</td>
<td>Atx1-(GSSG)$_6$-A1-woBC</td>
<td>1,6</td>
</tr>
<tr>
<td>pCZ29</td>
<td>Atx1-(GSSG)$_6$-Thb-A1-BC</td>
<td>1,4,15,16</td>
</tr>
<tr>
<td>pCZ40</td>
<td>Atx1-(GSSG)$_6$-Thb-A1-B</td>
<td>1,5</td>
</tr>
<tr>
<td>pCZ39</td>
<td>Atx1-(GSSG)$_6$-Thb-A1-woBC</td>
<td>1,6</td>
</tr>
<tr>
<td>pCZ22</td>
<td>Atx1-(GSSG)$_6$-sfGFP-BC</td>
<td>1,4,17,18</td>
</tr>
<tr>
<td>pCZ46</td>
<td>Atx1-(GSSG)$_6$-Thb-sfGFP-BC</td>
<td>1,4,18,19</td>
</tr>
</tbody>
</table>

**Table 5.S1**  Plasmids used in this study
**Table 5.S2**  Primers used in this study

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primer 1</td>
<td>5'-AAT CGA ATT CTA AAT GAA AGG AGA GTT CGA AAT GC -3</td>
</tr>
<tr>
<td>Primer 2</td>
<td>5'-GTC CTG GTT GAT GCG AGA CTC T -3</td>
</tr>
<tr>
<td>Primer 3</td>
<td>5'-TCG CAT CAA CCA GGA CGG ATC CAT GGC TAG CGG TG -3</td>
</tr>
<tr>
<td>Primer 4</td>
<td>5'-GGA TAG ATC TTC AGT GAT GGT GAT GAT TAA GCT TGG CTG CAG GTC G -3</td>
</tr>
<tr>
<td>Primer 5</td>
<td>5'-CTA CCG ATC GGT ACC TCA GAC GAC CAG -3</td>
</tr>
<tr>
<td>Primer 6</td>
<td>5'-CCG ATC GGT ACC TCA GTG ATG GTG ATG -3</td>
</tr>
<tr>
<td>Primer 7</td>
<td>5'-GCA TCA ACC AGG ACG GCA GCT CCG GCG G -3</td>
</tr>
<tr>
<td>Primer 8</td>
<td>5'-CCG CCG GAG CTT CGGTCCTGGTTG ATG C -3</td>
</tr>
<tr>
<td>Primer 9</td>
<td>5'-GGG GTA GCT CAG GCA TGG CTA GCG GTG AC -3</td>
</tr>
<tr>
<td>Primer 10</td>
<td>5'-GTC ACC GCT AGC CAT GCC TGA GCT ACC CC -3</td>
</tr>
<tr>
<td>Primer 11</td>
<td>5'-CTG GTT CCG CTG GGA TCT ATG GCT AGC GG -3</td>
</tr>
<tr>
<td>Primer 12</td>
<td>5'-AGA TCC ACG CCG AAC CAG GCC TGA GCT AC -3</td>
</tr>
<tr>
<td>Primer 13</td>
<td>5'-GCA GCT CCG CGG GTA GCA GCG GAA TGG CTA GCG -3</td>
</tr>
<tr>
<td>Primer 14</td>
<td>5'-CCG CTG CTA CCG CGG GAG CTG CCG TCC TGG TTG -3</td>
</tr>
<tr>
<td>Primer 15</td>
<td>5'-CAG CGG ACT GTG TCC CGG TGG ATC TAT GGC TAG CGG -3</td>
</tr>
<tr>
<td>Primer 16</td>
<td>5'-CTA GCC ATA GAT CCA CGC GGA ACC AGT CCG CTG CTA C -3</td>
</tr>
<tr>
<td>Primer 17</td>
<td>5'-CCG GGG TAG CTC AGG CAT GAG CAA AGG AGA AG -3</td>
</tr>
<tr>
<td>Primer 18</td>
<td>5'-CTT CTC CTT TGC TCA TGC CTG AGC TAC CCC CG -3</td>
</tr>
<tr>
<td>Primer 19</td>
<td>5'-CTG GTT CCG CTG GGA TCT ATG AGC AAA GGA GAA G -3</td>
</tr>
</tbody>
</table>
5.5 References


(10) Iwatsuki, M.; Tomoda, H.; Uchida, R.; Gouda, H.; Hirono, S.; Omura, S. Lariatin, antimycobacterial peptides produced by Rhodococcus sp. K01-


Chapter 6

Concluding Remarks

Lasso peptides, characterized by a slipknot topology, are an emerging class of ribosomally-synthesized and post-translationally modified natural products. The unique conformation, ultra-stability, and diverse biological functions make many lasso peptides attractive candidates for generating new molecular shapes and peptide-based therapeutics. To explore these possibilities, this thesis presented a systematic investigation of lasso peptides in the areas of discovery (Chapter 2), characterization (Chapter 3 and Chapter 4) and engineering (Chapter 5).

The work presented in Chapter 2 sought to optimize and expand a precursor-centric genome mining algorithm and apply the updated algorithm to discover novel lasso peptide gene clusters. Through the analysis of precursor sequences of known lasso peptides, I was able to identify a comprehensive precursor peptide pattern considering the length of the leader and the core peptide, the conserved threonine at the penultimate position of the leader peptide, starting residues of the core peptide, and ring size. I updated hidden Markov models for motifs of lasso peptide maturation enzymes by including recently characterized B and C enzymes in the training set. Applying the improved algorithm to survey both complete genomes, draft genomes and DNA contig sets from bacteria and archaea, I identified 948 high-confidence lasso peptide gene
clusters. Among the newly discovered clusters, some clusters with novel precursor sequence features and post-translational elements diverge from currently known lasso peptide clusters. I then carried out detailed investigations on lasso peptide gene clusters from *Asticcacaulis benevestitus* (Chapter 3) and *Streptomyces albus* (Chapter 4) as representatives.

*Asticcacaulis benevestitus* carries two isopeptidase-associated lasso peptide gene clusters. In Chapter 3, I successfully engineered the two gene clusters and produced both lasso peptides, named benenodin-1 and benenodin-2, through heterologous expression in *E. coli*. Through the NMR structural and thermal stability studies of benenodin-1, I found benenodin-1 switches between two distinct threaded conformers upon heating. This makes benenodin-1 the first natural example of a peptide based molecular machine. I also showed that only the native conformer of benenodin-1 is cleaved by its associated isopeptidase BenE1. In addition, I generated a variant of benenodin-1 which can be converted from a [1]rotaxane into a [2]rotaxane via trypsin cleavage of its loop region. These findings make benenodin-1 a potential building block to generate new molecular shapes and new molecular motors.

Chapter 4 presented a study of a novel lasso peptide, albusnodin, from *Streptomyces albus*. The gene cluster of albusnodin contains four genes *albACBT*. The gene *albT* encodes a protein annotated as an acetyltransferase downstream of the genes encoding the maturation enzymes. To produce albusnodin, I established a heterologous expression system using host organisms *Streptomyces coelicolor* and *Streptomyces lividans*. I found that
albusnodin is post-translationally modified with an acetyl group on Lys-10 of the core peptide. This makes albusnodin the first example of a lasso peptide with this modification. By conducting a knock-out study of the gene *albT*, I showed that albusnodin is only produced in the presence of *albT*, suggesting the acetylation of albusnodin may be a required post-translational modification for the biosynthesis and maturation of the peptide. Furthermore, I identified 43 closely related Actinobacteria strains also carry albusnodin-like gene clusters. The new type of post-translational modification on albusnodin further expands the known chemical and structural diversity of lasso peptides.

Finally, in Chapter 5, I laid out the foundations for developing a platform for high-throughput lasso peptide screening. To establish such a platform, it is essential to fuse lasso peptides to carrier proteins for display systems including phage, bacterial and yeast display. As a proof-of-concept, I was able to fuse two model proteins, A1 leucine zipper and superfolder GFP, to the C-terminus of the lasso peptide astexin-1 with a linker in between. The success in constructing a lasso peptide fusion protein is a stepping stone for establishing a high-throughput screening of lasso peptide libraries for new functions. In addition, this fusion approach allowed us to extend the tail portion of lasso peptides, where the well-folded model protein serves as a protecting group to prevent the degradation of the lasso tail during heterologous expression. Combining with thrombin cleavage and cysteine bio-conjugation chemistry techniques, I applied this fusion approach to graft two anti-cancer epitopes on the tail of astexin-1 and connect the extended tail back to the ring of astexin-1 through a disulfide bond. This work
opens new avenue for engineering lasso peptides as a stable scaffold for bioactive peptide sequences with promising therapeutic applications.

In conclusion, the investigations presented in this thesis shed light on a genome-led approach for the discovery and characterization of lasso peptides. By doing so, I hope that I have illustrated the value of using bioinformatics tools to guide natural product discovery. In addition, the discovery of beneondin-1 as a molecular machine has opened new possibilities using lasso peptides as molecular “Legos” for constructing more complex mechanically interlocked machines and biopolymers. Moreover, I hope that my work in construction of lasso peptide fusion proteins have inspired researchers to come up new ways of engineering lasso peptides and other RiPPs.