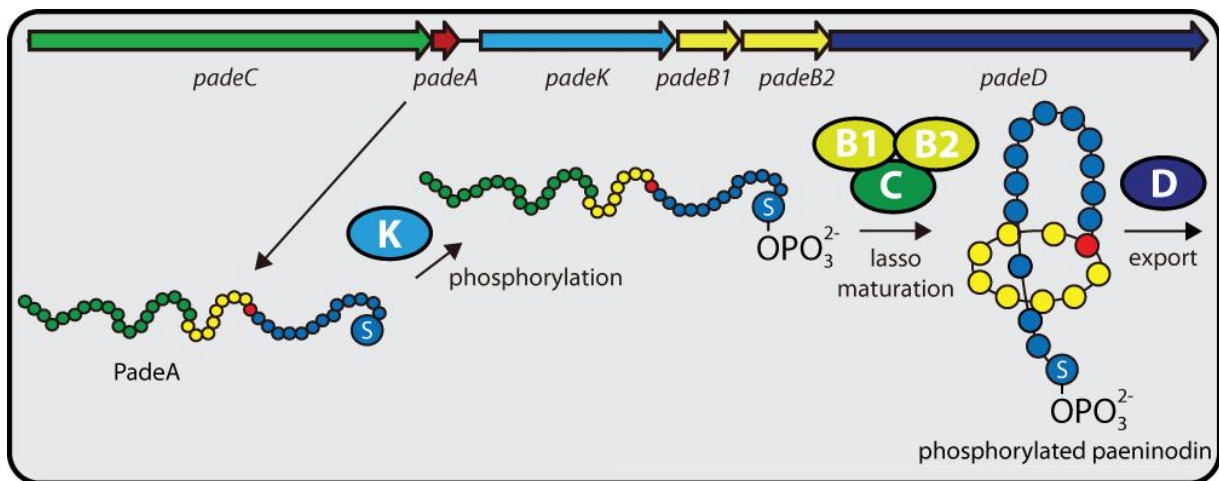# Discovery and Insights into the Unique Tailoring of the Paeninodin Lasso Peptide from *Paenibacillus dendritiformis* C454

## Dissertation

## Shaozhou Zhu



**Marburg an der Lahn 2016**

## Entdeckung und Mechanismen der einzigartigen Modifikation des Lassopeptids Paeninodin aus *Paenibacillus dendritiformis* C454

**Dissertation**

zur Erlangung des Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

dem Fachbereich Chemie

der Philipps-Universität Marburg

(Hochschulkennziffer 1180)

vorgelegt von

**Shaozhou Zhu**

aus Shandong, China

Marburg an der Lahn 2016

Vom Fachbereich Chemie

Der Philipps-Universität Marburg als Dissertation

am _____ angenommen

Erstgutachter:      Prof. Dr. M. A. Marahiel

(Philipps-Universität Marburg)

Zweitgutachter:     Prof. Dr. Peter Graumann

(Philipps-Universität Marburg)

Tag der Dissertation:

III

*Dedicated to my family and my wife Huimei Ren*

**Publications:**

**Zhu, S.**, Hegemann, J. D., Fage, C.D., Zimmermann, M., Xie, X., Linne, U., & Marahiel, M. A. Insights into the Unique Phosphorylation of the Lasso Peptide Paeninodin, *The Journal of Biological Chemistry*, doi: 10.1074/jbc.M116.72210

**Zhu, S.**, Fage, C.D.,Hegemann, J. D., Yan, D., & Marahiel, M. A. Dual substrate-controlled kinase activity leads to polyphosphorylated lasso peptides, *(submitted to FEBS letters)*

Hegemann, J. D., Fage, C. D., **Zhu, S**., Harms, K., Di Leva, F. S., Novellino, E.,Marinelli, L& Marahiel, M. A. (2016). The ring residue proline 8 is crucial for the thermal stability of the lasso peptide caulosegnin II. *Molecular BioSystems*. 2016,**12**, 1106-1109

Hegemann, J. D., Zimmermann, M., **Zhu, S**., Steuber, H.,Harms, K., Xie, X& Marahiel, M. A. 2013.Xanthomonins I-III are a New Class of Lasso Peptides Featuring a Seven-Membered Macrolactam Ring. *Angew. Chem. Int. Ed. Engl*. 50, 8714-8717

Hegemann, J. D., Zimmermann, M., **Zhu, S**., Klug, D., & Marahiel, M. A. 2013. Lasso peptides from proteobacteria: Genome mining employing heterologous expression and mass spectrometry. *Peptide Science*, 100(5): 527-542.

# Table of contents

# Abstract

Lasso peptides, such as microcin J25, BI-32169, lariatin and capistruin, are a structurally unique and pharmacologically relevant class of RiPPs (ribosomally synthesized and posttranslationally modified peptides) natural products. Compared with other intensively modified RiPPs, such as lantibiotics, lasso peptides only have a unique knotted topology in which the tail of the peptide is threaded through an N-terminal macrolactam ring and trapped by steric hindrance of bulky side chains stabilizing the entropically disfavored lasso structure. Except for this unusual knot structure, further posttranslational modifications on lasso peptides are very rare. Besides, lasso peptides have so far only been isolated from Proteo- and Actinobacterial sources. In this thesis, the lasso gene cluster from the Firmicute *P. dendritiformis* was investigated. Paeninodin, a new lasso peptide with an unusual phosphorylation at the side chain of the last serine was discovered by expression of this cluster in a heterologous host. The Paeninodin lasso peptide was isolated from a culture pellet. Mass spectrometric, carboxypeptidase Y assays and IM-MS studies proved paeninodin to be a new representative of lasso peptides. Morever, the biosynthetic pathway of modified lasso peptide was delineated through *in vivo* and *in vitro* studies. The kinase turned out to be a novel lasso peptide precursor kinase with wide substrate specificity. These results provide a way for the generation of novel lasso peptide analogs and, thereby, would facilitate lasso peptide engineering in the future.

# Zusammenfassung

Lassopeptide, wie Microcin J25, BI-32169, Lariatin und Capistruin, sind eine strukturell einzigartige und pharmakologisch relevante Gruppe von RiPP-Naturstoffen (ribosomal-synthetisierte und post-translational modifizierte Peptide). Verglichen mit anderen stark modifizierten RiPPs, wie Lantibiotika, haben Lassopeptide nur eine einzige verknotete Topologie, in der der Schwanz des Peptids durch einen N-terminalen Makrolactamring gefädelt ist und dort durch sterische Wechselwirkungen von sperrigen Seitenketten festgehalten wird, wodurch die entropisch ungünstige Lassofaltung stabilisiert wird. Mit Ausnahme dieser ungewöhnlichen Knotenstruktur sind weitere post-translationale Modifikationen an Lassopeptiden sehr selten. Des Weiteren wurden Lassopeptide bisher nur aus proteo- und actinobakteriellen Quellen isoliert. In dieser Doktorarbeit wurde der Gencluster aus dem Firmicutes *P. dendritiformis* untersucht. Es wird die Entdeckung von Paeninodin, einem neuen Lassopeptid mit ungewöhnlicher Phosphorylierung an der Seitenkette des letzten Serins, durch Expression des zugehörigen Genclusters in einem heterologen Wirt beschrieben. Paeninodin konnte aus den Zellen der Expressionskultur isoliert werden. Massenspektrometrische und Carboxypeptidase Y Assays sowie IM-MS Studien bewiesen, dass Paeninodin ein neuer Vertreter der Lassopeptide ist. Zusätzlich wurde die Biosynthese von modifizierten Lassopeptiden beschrieben auf Basis von *in vivo* und *in vitro* Studien. Es zeigte sich, dass die Kinase eine neuartige Lassopeptid-Vorläufer-Kinase mit breiter Substratspezifität ist. Diese Ergebnisse ermöglichen eine neue Route zur Generierung von neuartigen Lassopeptidanalogas und fördern das Engineering von Lassopeptiden in der Zukunft.
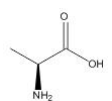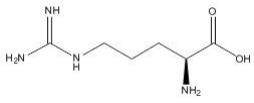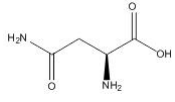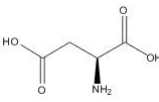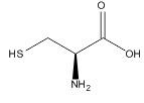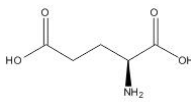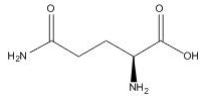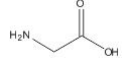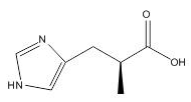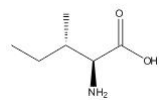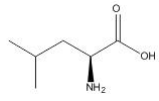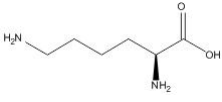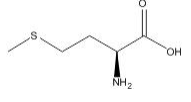
# List of abbreviations

| | |
|---|---|
| aa | Amino acid |
| ABC | ATP-binding-casette |
| ATCC | American Type Culture Collection |
| ATP | Adenosin-5'-triphosphat |
| BLAST | Basic local alignment search tool |
| bp | Base pairs |
| CDS | Coding sequence |
| CID | Collision-induced dissociation |
| ClpC1 | ATP-dependent Clp protease ATP-binding subunit |
| COSY | Correlation spectroscopy |
| Da | Dalton |
| ddH$_2$O | double-distilled water |
| DMSO | Dimethylsulfoxide |
| DNA | Deoxyribonucleic acid |
| dNTP | 2'-Deoxynucleoside-5'-triphosphate |
| DSMZ | German collection of microorganisms and cell cultures GmbH |
| *E. coli* | *Escherichia coli* |
| EDTA | *N,N,N',N'*- Ethylenediaminetetraacetic acid |
| EIC | Extracted ion chromatogram |
| ESI | electron-spray ionization |
| EtOH | Ethanol |
| FP | Forward Primer |
| FPLC | Fast performance liquidchromatography |
| FT | Fourier transformation |
| h | Hour |
| HEPES | 2-[4-(2-Hydroxyethyl)-1-piperazinyl]ethanesulfonic acid |
| HMBC | Heteronuclear multiple bond coherence |
| HPLC | High performance liquid chromatography |
| HR-MS | High-resolution mass spectrometry |
| IPTG | Isopropyl-β-D-thiogalactopyranoside |
| IMAC | immobilized metal affinity chromatography |
| IM-MS | ion mobility-mass spectrometric |
| Kan | Kanamycin |
| Kb | Kilobase pairs |
| L | Liter |

XIII
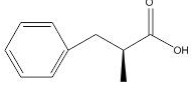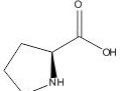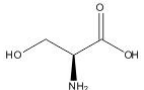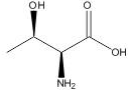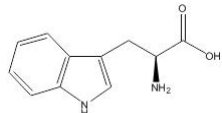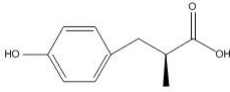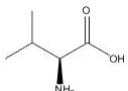
| | |
|---|---|
| LB-Medium | Lysogeny Broth Medium |
| LC-MS | liquid chromatography-mass spectrometry |
| M | Mol pro Liter |
| MCS | multiple cloning site |
| MccJ25 | Microcin J25 |
| MeOH | Methanol |
| MEME | Multiple Em for Motif Elicitation |
| min | Minutes |
| MRSA | methicillin-resistant Staphylococcus aureus |
| MS | mass Spectrometry |
| MS2 | Tandem mass spectrometry |
| Ni-NTA | Ni-nitriloacetic acid |
| NMR | Nuclear magnetic resonance |
| NRP | non-ribosomal peptide |
| NRPS | non-ribosomal peptide synthetases |
| OD | optical density |
| PAGE | Polyacrylamidgelelektrophorese |
| PCR | Polymerase chain reaction |
| PDB | Protein Data Bank |
| *P. dendritiformis* | Paenibacillus dendritiformis |
| PTM | post-translational modification |
| RiPP | ribosomally synthesized and post-translationally modified peptides |
| RP | ribosomal peptide |
| rpm | rounds per minute |
| SDS-PAGE | sodium dodecyl sulfate polyacrylamide gel electrophoresis |
| SLIM | site-directed ligation independent mutagenesis |
| *T. composti* | *Thermobacillus composti* |
| T7 | T7-Promotor |
| TFA | trifluoroacetic acid |
| Tris | Tris (hydroxymethyl) aminomethane |
| UV | Ultraviolet |
| v/v | Volume per volume |
| w/v | Weight per volume |

**Table 1. Overview of the proteinogenic amino acids.**

| Amino acid | 3 letter code | 1 letter code | M [g/mol] | Structure |
|---|---|---|---|---|
| Alanine | Ala | A | 89 | |
| Arginine | Arg | R | 174 | |
| Asparagine | Asn | N | 132 | |
| Aspartic acid | Asp | D | 133 | |
| Cysteine | Cys | C | 121 | |
| Glutamic acid | Glu | E | 147 | |
| Glutamine | Gln | Q | 146 | |
| Glycine | Gly | G | 75 | |
| Histidine | His | H | 155 | |
| Isoleucine | Ile | I | 131 | |
| Leucine | Leu | L | 131 | |
| Lysine | Lys | K | 146 | |
| Methionine | Met | M | 149 | |

| | | | | |
|---|---|---|---|---|
| Phenylalanine | Phe | F | 165 | |
| Proline | Pro | P | 115 | |
| Serine | Ser | S | 105 | |
| Threonine | Thr | T | 119 | |
| Tryptophan | Trp | W | 204 | |
| Tyrosine | Tyr | Y | 181 | |
| Valine | Val | V | 117 | |

# 1. Introduction

## 1.1 Natural Products

Natural products, as the name implies, are chemical compounds of natural origin[1, 2]. They are generally biologically functional compounds isolated from all kinds of natural sources, such as plants, microorganisms, fungi and animals[3]. Due to the diverse pharmacological or biological activity, natural products have played significant roles in science and medicine[4]. Humanity has already learned to prepare "mixer crude extract" to protect themselves against some diseases since ancient times. Quinine, for example, is such a medicine that has been used to prevent and treat malaria and to treat babesiosis for a long time[5]. With the great development of biological and chemical research in the 19th century, the start of the wide use of natural products marked the birth of modern medicine[6]. It has been suggested that more than 60 % of drugs used in clinical applications are derived from natural products or synthetic analogs[1]. The acetylsalicylic acid (aspirin, prepared in 1897) is an example of one of the first semisynthetic drugs based on a natural product which is still widely used today[7]. Another milestone for modern medicine was the discovery of penicillin as the first antibiotic in 1929[8]. Its large-scale fermentative production in the 1940s was one of the most important steps in modern medicine's use of natural products of microbial origin[8]. Later on, more and more natural products or derivatives from them were approved for clinical applications[2, 6]. Even today, natural products are still the main source or are used as starting points for drug discovery. The Nobel Prize in Physiology or Medicine, for example, was awarded to natural products again last year. Avermectin discovered by William C. Campbell and Satoshi Ōmura and Artemisinin discovered by Youyou Tu have revolutionized the treatment of the most devastating parasitic diseases[9].

Notably, microbial natural products are the origin of most of the drugs on the market today[10]. This is due to the huge number of microbial species and their short generation times, which endows the natural products with a rich variety of evolution[11]. However, it is believed that more than 99 % of the microbial species extant have not yet been discovered. The so-called "secondary metabolites" produced by microorganisms have a broad range of functions[12]. These include pheromones that act as social signaling molecules, agents that solubilize and transport nutrients (e.g. siderophores), and antibiotics that are used against competitors[13]. In fact, most antibiotics are produced by microorganisms. Microorganisms are in constant competition for resources and ecological niches, therefore, they need to develop all kinds of defense mechanisms[14]. Meanwhile, microorganisms are also developing all kinds of protection mechanisms in order to survive[14, 15]. However, these protection strategies also represent a major problem for modern medicine which is termed "antibiotic resistant"[16]. *Streptococcus*

*pyogenes*, for example, and staphylococci organisms that cause respiratory and cutaneous infections are now resistant to all of the older antibiotics[16]. Therefore, searching for new antibiotic agents and, thus, the isolation and identification of new natural substances are still necessary.

Natural products can generally be classified into alkaloids, terpenes, polyketides, carbohydrates, lipids and peptides, based on their chemical aspects[1]. Their high structural complexity is derived from the intensively evolutionary development of their biosynthesis machinery. It, therefore, requires broad basic research to isolate new natural substances, and characterize and understand their biosynthetic mechanisms. We are particularly interested in natural products from peptides. Therefore, several classes of peptide natural products, especially ribosomally synthesized and posttranslationally modified peptides (RiPPs), and their biosynthesis routes are described below.

## 1.1.1 Nonribosomal Peptides

Nonribosomal peptides, as the name implies, are synthesized in an mRNA-independent way[17-19]. They are a very diverse family of natural productsof extraordinary pharmacological importance. They are often toxins, siderophores or pigments produced by microorganisms and can be used as antibiotics, cytostatics and immunosuppressants in commercial ways[20, 21]. Figure 1.1 shows examples of the structural diversity of the bioactive compounds of nonribosomal origin[22-24].



Fengycin (antibiotic)                    Cyclosporin A (immunosuppressive agent)

Bleomycin (antitumor)                    Bacillicactin (siderophore )

**Figure 1.1:**Examples of the structural diversity of bioactive compounds of NRPs.

The assembling of these compounds is achieved by large multimodular enzymes called nonribosomal peptide synthetases (NRPSs)[20]. Each module contains at least three domains: the adenylation domain (A domain), the condensation domain (C-domain) and the peptidyl carrier protein (PCP domain) [25], and is responsible for the incorporation of an amino acid building block in the final product. The function of the A domain is the recognition and activation of the respective amino acid as the aminoacyl-adenylate under ATP hydrolysis[25]. The activated amino acid is then transferred to the thiol group of the phosphopantetheinyl bound in the PCP domain and forms the peptide bond in the C domain.Upon the amide bond formation, the elongated peptide is then transferred into the next downstream PCP domain and serves as a donor substrate in the following condensation step[25]. In addition to these three essential domains, there is also a fourth essential domain called the thioesterase (TE) domain. It is usually located in the termination module of the assembly line. The TE domain is responsible for the release of the product from the NRPS machinery by cyclization[26].

Besides these four domains required in a NRPS assembling machine, other optional domains can also be found, such as the epimerization (E) domain, the cyclization (Cy) domain, the methylation (MT) domain and the formylation (F) domain[27]. These domains can catalyze diverse in cis modification reactions, thus, can increase the diversity of the NRPS natural products. What is more, in addition to all the domains mentioned above, there are also a lot of tailoring enzymes which can be found in the NRPS gene cluster, such as methytransferase and monooxygenases. These enzymes can catalyze diverse transmodification reactions which can further increase the chemical diversity of the NRPS natural products[27].



**Figure 1.2:** The structure and biosynthetic gene cluster of teixobactin. (a) The NRPS gene cluster contains two genes – txo1 and txo2. The T domain here is equal to the PCP domain. (b) Schematicstructure of teixobactin. The N-methylation ofthe first phenylalanine is catalyzed by themethyltransferase (MT) domain in module 1. The ring closure between the last isoleucine and threonine (Thr) is catalyzed by the TE domain. (c) Structure of teixobactin (Figure from ref. [29]).

3

Figure 1.2 shows one example of the biosynthesis of the NRPS natural product teixobactin[28, 29], a new antibiotic which was recently isolated from the uncultured bacteria *Eleftheria terrae.* Teixobactin was synthesized by a two-mega enzyme complex consisting of 11 modules. Eleven amino acids from each module were incorporated into the final antibiotic[28].

## 1.1.2 Cyclodipeptides

Cyclodipeptides and their derivatives, called the diketopiperazines (DKPs), constitute a large class of secondary metabolites synthesized predominantly by microorganisms[30-33]. Similar to other natural products, DKPs are also a very diverse family of natural products with important biological activities, such as antitumor, antifungal, immunosuppressive and antibacterial activities[30-32]. Figure 1.3 shows examples of the structural diversity of bioactive DKPs.
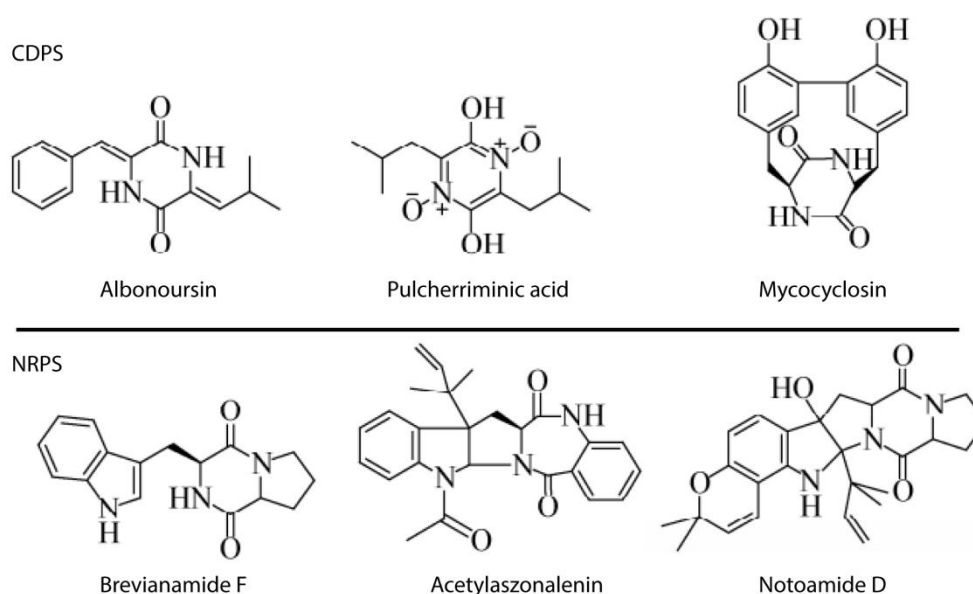


**Figure 1.3:**Examples of the structural diversity of DKPs.

The biosynthesis of the cyclodipeptides was previously thought to be mainly synthesized by NRPS[30]. However, an NRPS-independent biosynthetic pathway was discovered in 2002[34]. It turns out that AlbC, a novel cyclodipeptide synthease (CDPS), was involved in the biosynthesis of albonourisin from *Streptomyces noursei*[34].

Cyclodippetide syntheases are very small enzymes (≈ 300 aa) compared with NRPS, and are able to form two successive peptide bonds by using aminoacyl-tRNA as a substrate. The 2,5-DKP moiety can be synthesized after the free tRNA has been released, as is shown in Figure 1.4.
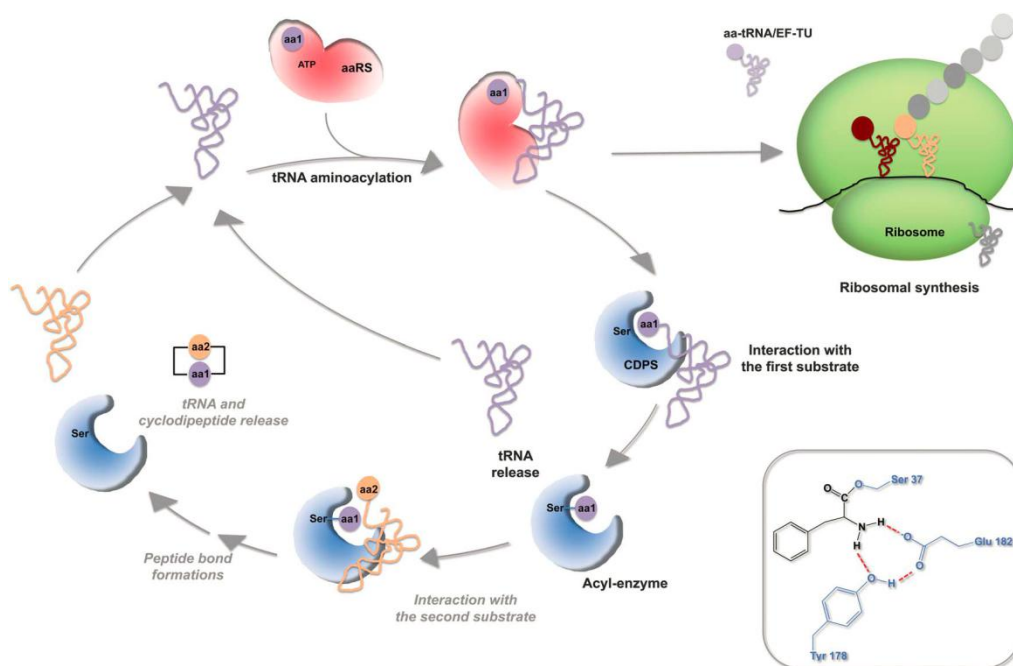
**Figure 1.4:** Biosynthesis of DKPs by CDPs. The CDPs hijack the aa-tRNA to produce the 2,5-DKP moiety (Figure from ref. [30]).

The DKPs synthesized by CDPs can be modified further by tailoring enzymes[30, 35-37]. Similar to other natural products, the CDPS genes in prokaryotes are generally organized into operon-like structures. These enzymes are probably involved in modifying the cyclodipeptide. There are couples of tailoring enzymes that have been experimentally characterized and shown to have different activities, namely $\alpha$, $\beta$-dehydrogenation, DKP ring oxidation, methylation and C–C aryl coupling[30, 37].

## 1.1.3 Ribosomally synthesized and posttranslationally modified peptides

Another major class of peptide natural products is RiPPs[38]. These molecules have diverse structures not directly accessible to natural ribosomal peptides[39, 40]. It was believed in the past that these compounds were mainly synthesized by NRPS. However, the genome sequencing efforts of past decades have revealed that they are RiPPs[40]. With the known biosynthetic gene cluster for these compounds and genome mining methods, it was shown that these molecules could be produced in all kinds of life, and their biosynthetic genes clusters exist widely in the genomes sequenced currently[41-43]. The extensive modification endows these peptides with unique structures that can increase their chemical stability, make them better for target recognition and, thus, have diverse bioactivity[40].
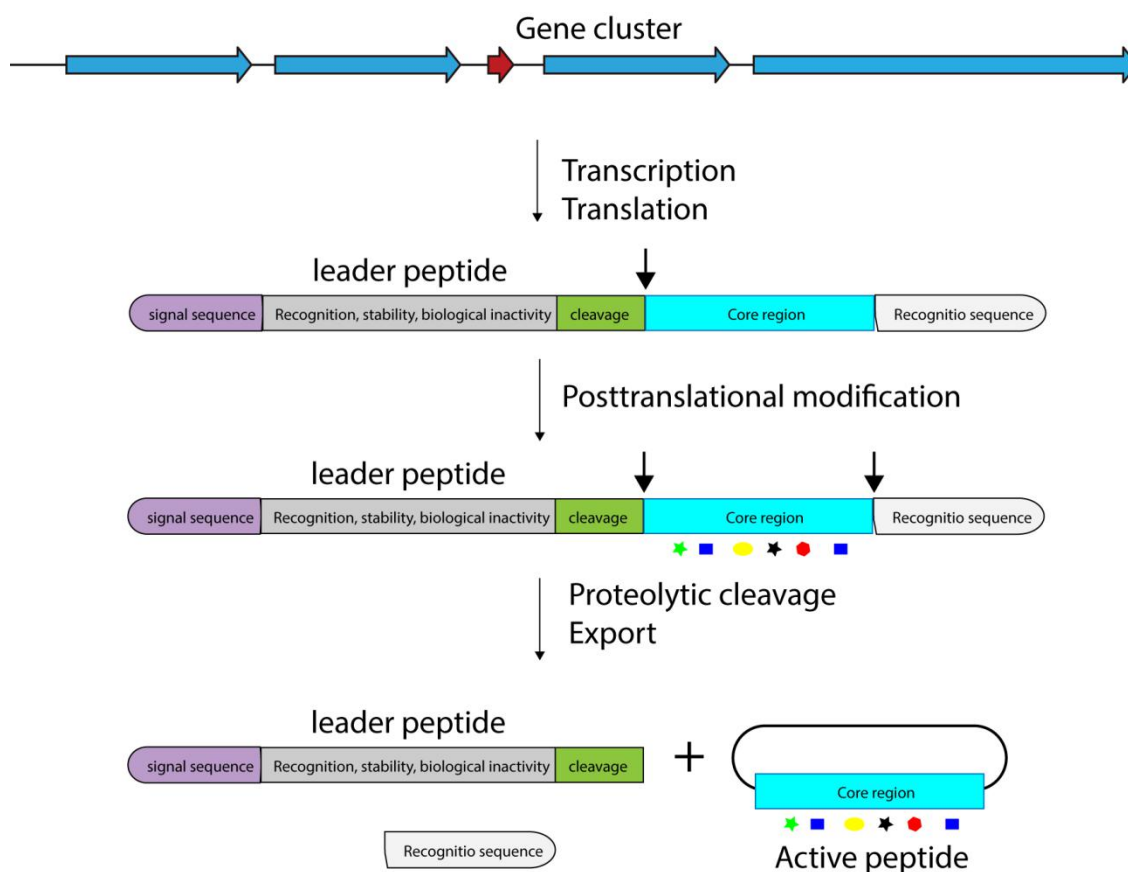
**Figure 1.5:** General biosynthetic pathway for RiPPs.

The RiPPs are initially synthesized as a precursor peptide, typically 20–110 residues in length, which is a structure gene located in the gene cluster[40, 41]. As shown in Figure 1.5, a typical precursor peptide contains several segments, such as a signal peptide, leader peptide and core peptide.A signal sequence linked to the N-terminal of the leader peptide is used to direct the peptide to the specific cellular compartments where the posttranslational modifications will take place. The leader peptide, which is biologically inactive, is usually important for recognition by many of the posttranslational modification enzymes and for export[44]. It is usually appended to the N-terminus of the core peptide. However, in some cases, it can also be found at the C-terminus of the core peptide, such as bottromycins[45-47]. A cleavage sequence can be found between the leader peptide and the core peptide, which is used for cleavage by some protease or ABC transporters.The core peptide is the segment of the precursor peptide that will be transformed into the final natural product[40].

After the precursor peptide was generated, extensive posttranslational modifications (PTM) were performed by the tailoring enzymes in the gene cluster[40]. After all the PTMs were incorporated into the core sequence, the leader peptide and signaling peptide were then proteolytically cleaved. Some PTMs also take place only after the proteolytic cleavage of the precursor peptide[40]. These are mainly modifications that require the released N-terminus, such as cyclizations. Even though the precursor peptide is of ribosomal origin, the posttranslational

6

modification could significantly enhance the structural diversity of the mature products.This allows the natural products better target recognition or an increase in the chemical, proteolytic and metabolic stability[48].

## 1.2 Important Classes of Ribosomally Synthesized Peptides

Some important classes of ribosomally synthesized peptides, which play an important role in the pharmaceutical industry, are presented below.

### 1.2.1 Lanthipeptides

Lanthipeptides (peptides containing Lanthionine) are a family of RiPPs containing (methyl) lanthionine residues[49-51]. The first lanthipeptide, nisin, was discovered in 1927 and has become the best understood lanthipeptide. It is widely used as a food preservative in processed cheese, meats and beverages[51].

Lanthipeptides are normally small peptides that undergo extensive posttranslational modifications. The common posttranslational modifications involve the dehydration of Ser and Threonine (Thr) residues in the precursor peptide to yield 2,3-didehydroalanine (Dha) and (Z)-2,3-didehydrobutyrine (Dhb), respectively[52]. The stereospecific intramolecular addition of a Cys residue onto Dha or Dhb is then introduced to form a lanthionine (Lan) or methylanthionine (MeLan) bridge (Figure 1.6)[49-52].
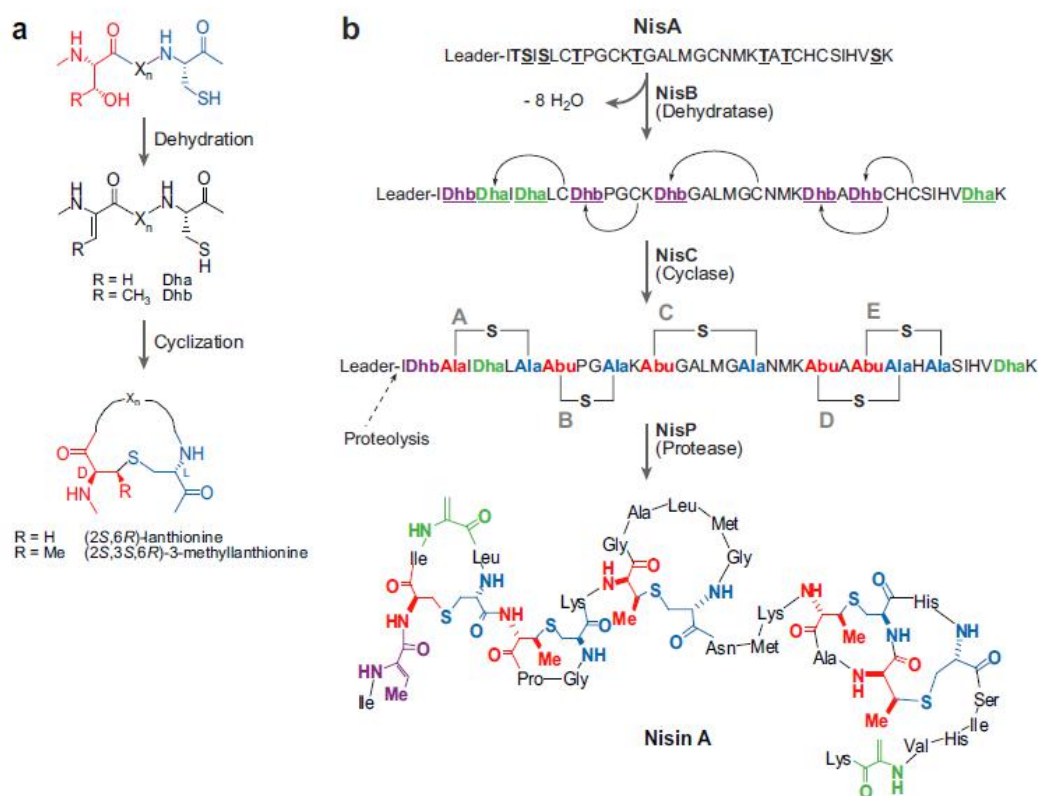
**Figure 1.6:**Biosynthesis of nisin lantibiotic. (a) Installation of lanthionine (Lan) or methyllanthionine (MeLan) residues into prepeptide. (b) The posttranslational maturation process of nisin (Figure from ref. [49]).

Lanthipeptides can be generally classified into four classes based on the biosynthetic enzymes that install the Lan and MeLan motif[40, 49-52]. Class I lanthipeptides are synthesized by two different enzymes: a dehydratase LanB and a cyclase LanC, as shown in Figure 1.7. Class II lanthipeptides are synthesized by a single lanthipeptide, synthetase-LanM. LanM contains two distinguished domains: an N-terminal dehydratase domain that bears no homology to LanB, and a C-terminal LanC-like cyclase domain. Class III lanthipeptides are also synthesized by a single lanthipeptide synthetase, LanKC. LanKC contains three different domains: an N-terminal lyase domain, a central kinase domain and a C-terminal cyclase domain. Class IV lanthippetides are synthesized in a similar synthetase with three domains. The only difference is that the C-terminal is a LanC-like Cyclase domain[40, 48-52].
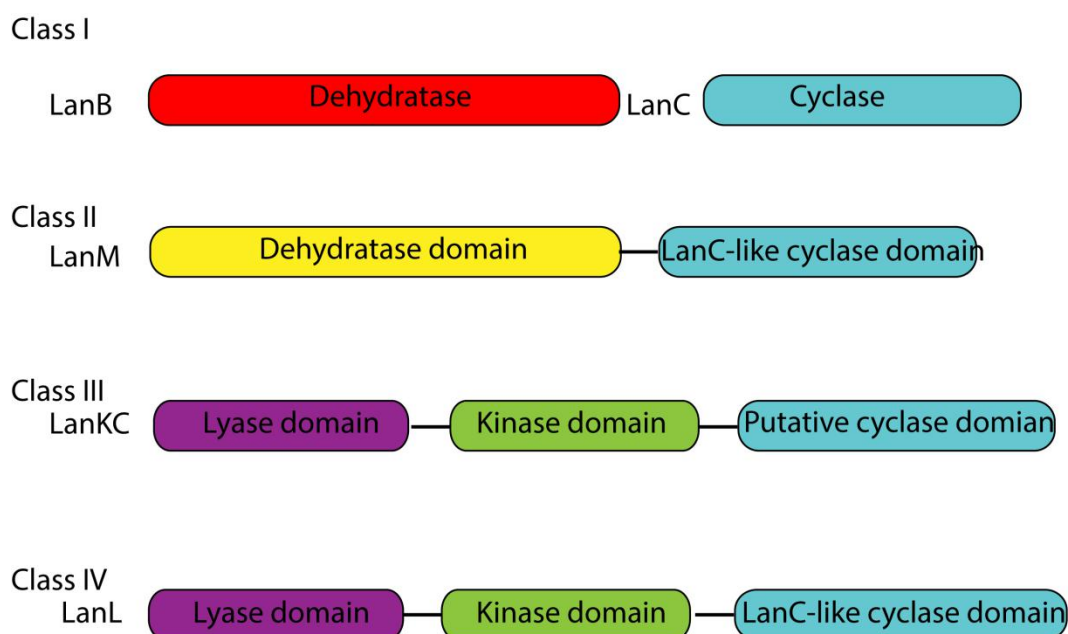


**Figure 1.7:** Overview of Lanthipeptide classification.

Lanthipeptides have showed diverse functions. Most of them are antimicrobial compounds such as Nisin, which has been used in the food industry for more than 40 years to combat food-borne pathogens[53-55]. Another example is duramycin, which is being evaluated for treatment of cystic fibrosis[56-59]. In addition to the antimicrobal activity, Lanthipeptides can also show other functions. Two recently new isolated lanthipeptides called pinensin A and pinensin B, for example, were found to be highly active against many filamentous fungi and yeasts, but show only weak antibacterial activity[60]. Another example of lanthipeptides that show very interesting bioactivity are the morphogenetic peptides SapB and SapT from streptomycetes. These peptides are believed to function as biosurfactants during the formation of aerial hyphae[49, 61, 62].

## 1.2.2 Bottromycins

Bottromycins are a family of RiPPs with intensive modifications[63-65].The structure of bottromycin contains an unusual macrocyclic amidine and a thiazole ring. In addition, four β-methylated amino acids could be found (Figure 1.8). Bottromycin was first discovered in 1957 as an antibiotic isolated from *Streptomyces bottropensis*[63-65]. It has been shown to inhibit methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant *Enterococci* (VRE) among other Gram-positive bacteria and mycoplasma[63-65].
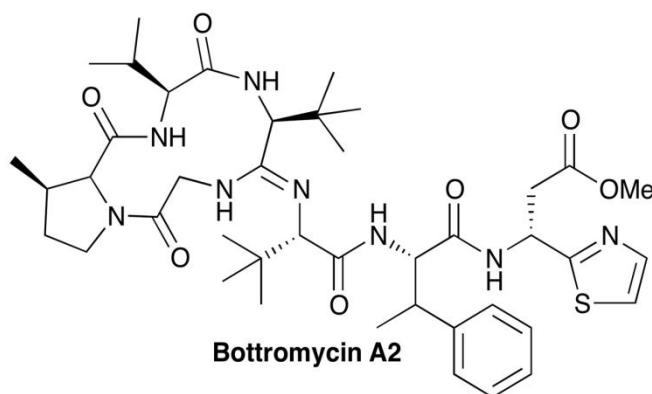


**Bottromycin A2**

**Figure 1.8:**Structure of Bottromycin A2.

The biosynthesis of bottromycin was studied independently by three groups in 2012. It was shown that bottromycin was synthesized as a RiPP[45-47].
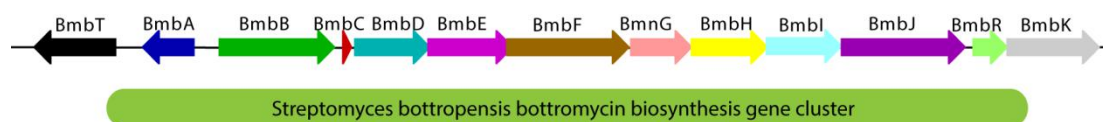


Gene annotation and proposed function in bottromycin producer

| S. bottropensis | Predicted function |
| --- | --- |
| bmbT | Major facilitator superfamily/transporter |
| bmbA | o-Methyltransferase |
| bmbB | Radical SAM methyltransferase |
| bmbC | Precursor peptide |
| bmbD | YcaO-domain |
| bmbE | YcaO-domain |
| bmbF | Radical SAM methyltransferase |
| bmbG | α/β Hydrolase |
| bmbH | Metallo-dependent hydrolase |
| bmbI | Cytochrome P450 |
| bmbJ | Radical SAM methyltransferase |
| bmbR | Transcriptional regulator |
| bmbK | M17 aminopeptidase |

BmbC:     MGPVVVFDCMTADFLNDDPNNAELSALEMEELESWGAWDGEATS

**Figure 1.9:** Bottromycin gene cluster in *S. bottropensis*. The gene encoding the bottromycin precursor peptide is shown in red.

9

 The gene cluster from *Streptomyces bottropensis* is shown in Figure 1.9. The gene cluster contains 13 open reading frames (ORFs). The predicted function of each gene is summarized in Figure 1.9. The precursor peptide, termed BmbC, has 44 amino acids. The amino acids forming the bottromycin core peptide are marked in red in Figure 1.9. The sequence is Gly-Pro-Val-Val-Val-Phe-Asp-Cys. Unlike other RiPPs, bottromycin has no leader peptide except a Met at the N-terminal. Instead, it has a 35 residues following peptide function as a recognition peptide for posttranslational modification[45-47].

Precursor    MGPVVVFDCMTADFLNDDPNNAELSALEMEELESWGAWDGEATS
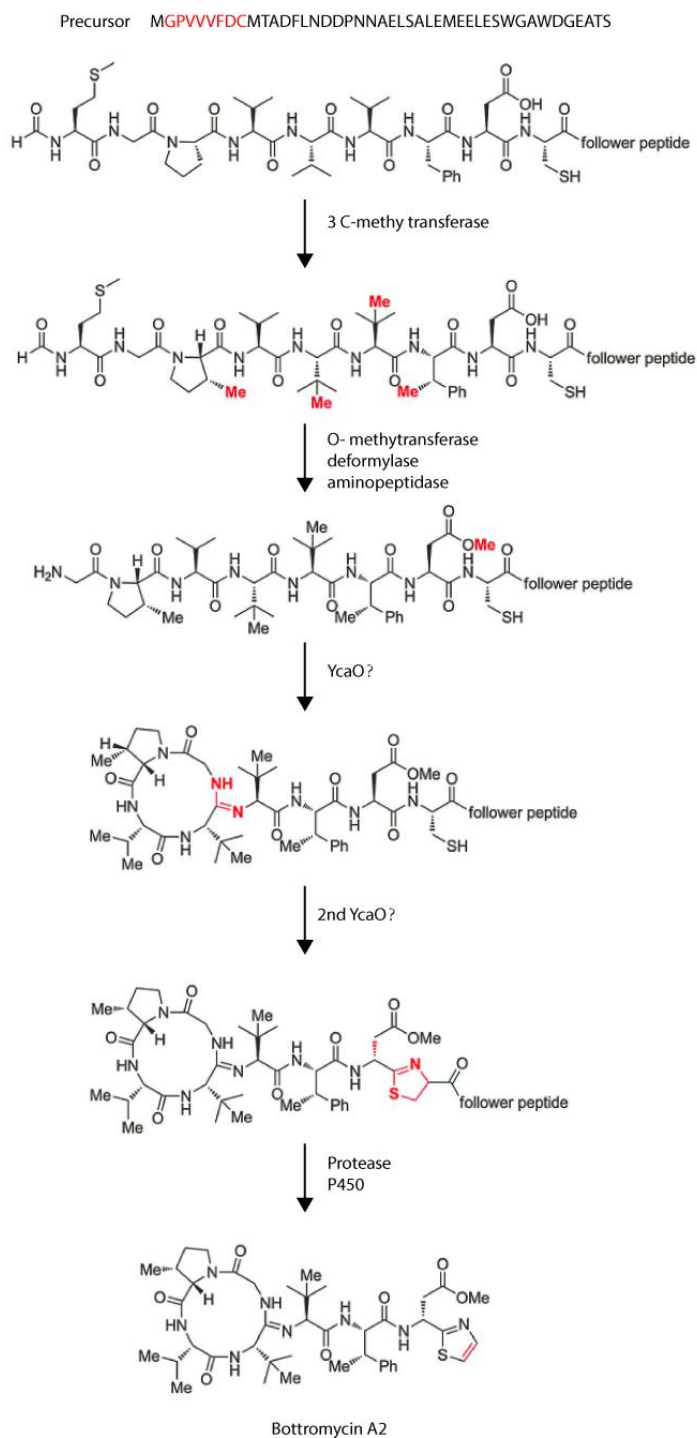


**Figure 1.10:** Potential biosynthetic pathway to bottromycin A2 (Figure modified based on ref [45-47]).

10

A biosynthetic pathway for bottromycin has been proposed by a gene-inactive experiment *in vivo* (Figure 1.10). Three radical S-adenosyl methionine (SAM) methyltransferases are believed to form β-methylate amino acid residues[45-47]. While the O-methyltransferase is responsible for the formation of anester, the aminopeptidases are predicted to cleave the N-terminal methionine residue. There are two YcaO-like proteins and it is hypothesized that one catalyzes macrocyclic amidine formation, while the other catalyzes thiazoline formation. There are two hydrolases which may catalyze follower peptide hydrolysis. However, the order of the posttranslational modifications is currently not known and, hence, the order depicted in Figure 1.10 is arbitrary[45-47].

## 1.2.3 Thiopeptides

Thiopeptides are a growing class of sulfur-rich, highly modified heterocyclic peptide antibiotics which also belong to RiPPs[66-71]. These peptides possess a characteristic macrocyclic core that consists of a monoaza six-membered ring central to multiple thiazoles and dehydroamino acids[66, 67]. Similar to bottromycin, many members in this family exhibit interesting activity against various pathogens, including methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant Enterococci (VRE) [67]. Figure 1.11 shows one example of Thiopeptides called Thiostrepton.Thiostrepton was discovered by Donovick et al., who described its antibacterial properties in 1955, but the biosynthesis of these compounds was unclear until 2009[68].
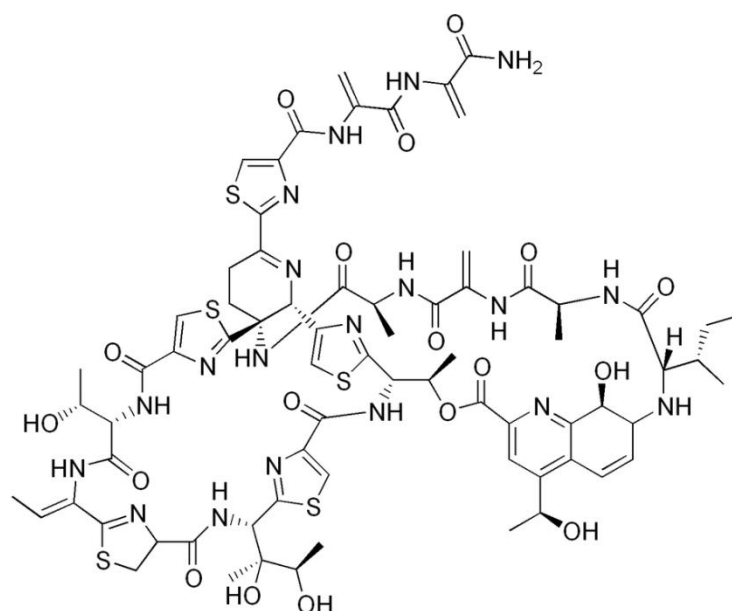


**Figure 1.11:**Structure of Thiostrepton.

The biosynthetic pathway for thiostrepton contains 21 ORFs. The TsrH geneencoding for precursor peptide contains 58 amino acids, as shown is Figure 1.12[68]. The first 41 aa is the leader peptide and the last 17 aa is the core peptide (IASASCTTCICTCSCSS). After the

11

precursor is produced, TsrO and TsrM, which are dehydratase, catalyze the formation of thiazole or thiazoline from every cysteine. After this step, a further three dehydratases, TsrJ, TsrK and TsrS, convert all the serine into dehydroalanines. TsrN and TsrL are suggested to be responsible for the hetero Diels-Alder cyclization. In 2010, tclM, a homologue of TsrL, was shown to be responsible for the transannular heteroannulation at the core of this class of molecules[72].
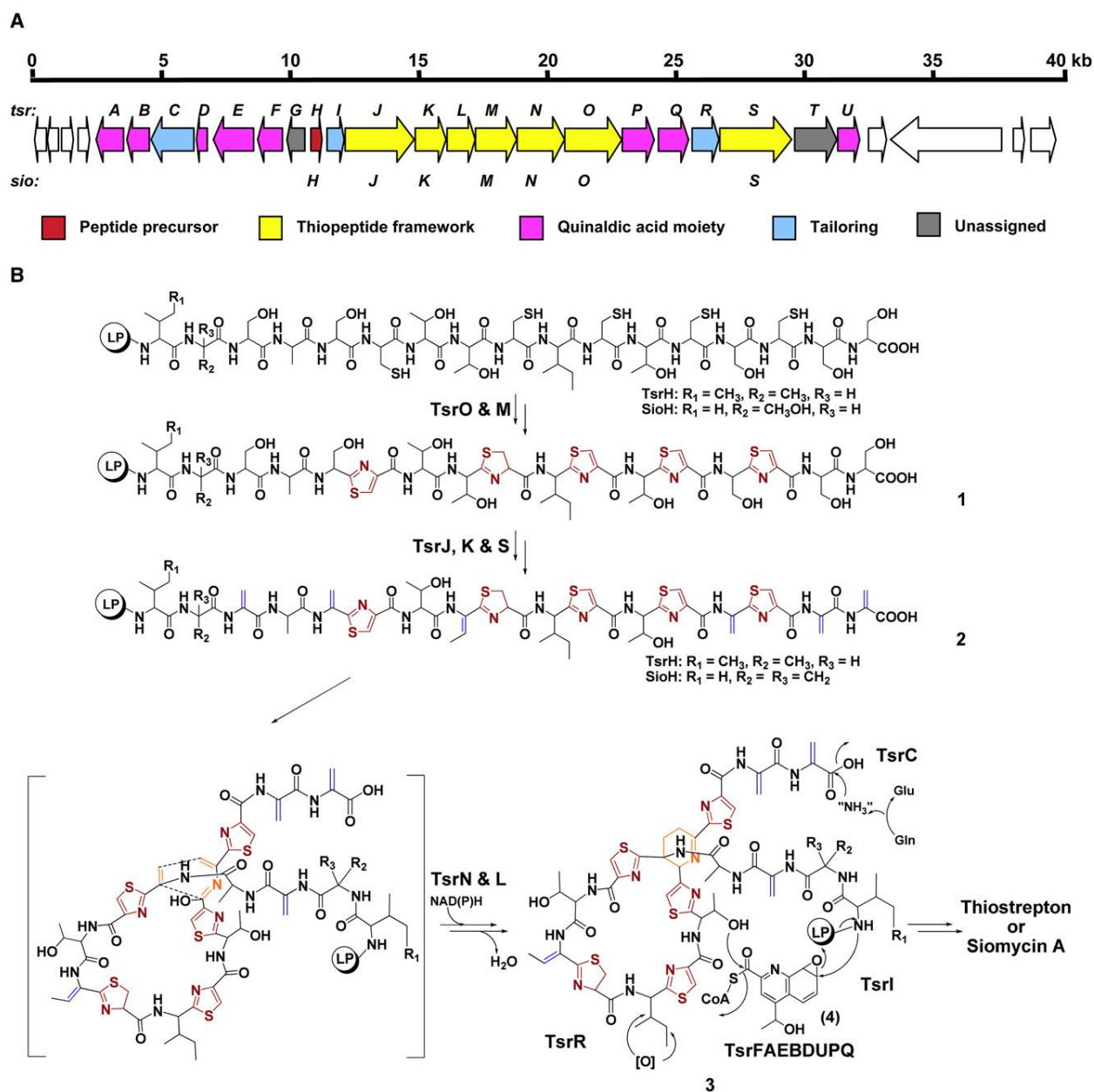


**Figure 1.12:** Potential biosynthetic pathway to Thiostrepton (Figure from ref. [68]).

## 1.2.3 Cyanobactins

Cyanobactins are small cyclic peptides that are RiPPs (Figure 1.13)[73-78]. They are mainly produced by cyanobacteria living in symbioses, as well as terrestrial, marine or freshwater

environments. These peptides possess intensive modifications, including azole/azoline rings, D-sterocenters and, in some cases, prenyl groups[75, 77]. Many cyanobactins exhibit interesting activity, such as antimalarial, antitumor and multidrug reversing activities, and potential to be used as pharmaceutical leads[73-75].



Patellamide A            Trichamide            Trunkamide
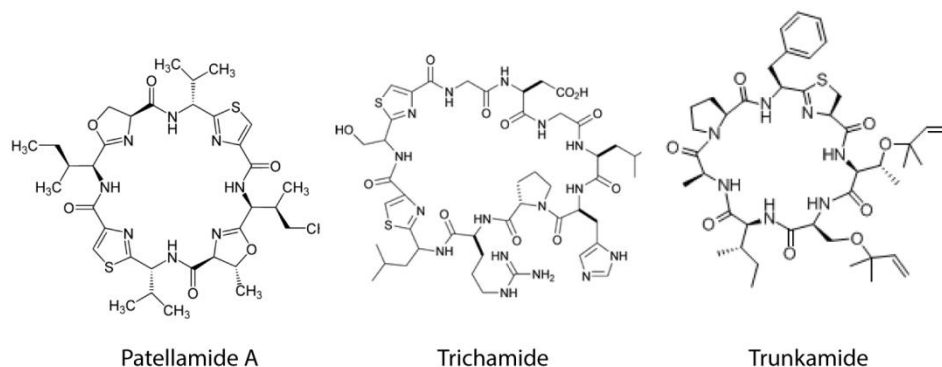
**Figure 1.13:**Examples of the structural diversity of Cyanobactins.

Patellamide is the most well-known compound of this family. It was first isolated in 1981 and turned out to be produced by *Prochloron didemni*, a cyanobacterial symbiont of *Lissoclinum patella*[79-81]. Patellamide shows moderate cytotoxicity and activity against multidrug-resistant cancer cell lines. The biosynthetic pathway for Patellamide was determined after genome sequencing of *P. Didemi*. The gene encoding the precursor peptide contains a leader peptide and two core peptides (for patellamides A and C), as shown in Figure 1.14[79-81]. There is also a protease recognition sequence before each core peptide and a N-terminal recognition sequence with a length of four to five amino acids. In the first biosynthetic step, the Cydodehydratase patD catalyzes the formation of dihydrothiazol or dihydrooxazol ring from Cys or Ser/Thr groups. Subsequently, the protease patG truncates the C-terminal recognition sequence and cyclizes the peptide backbone. patG is also responsible for the dehydrogenation of dihydrothiazol or dihydrooxazol ring to thiazole or oxazole. patF has no function in the biosynthesis of patellamides[78, 80, 81].

MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASVTACITFCAYDGVEPSITVCISVCAYDGE
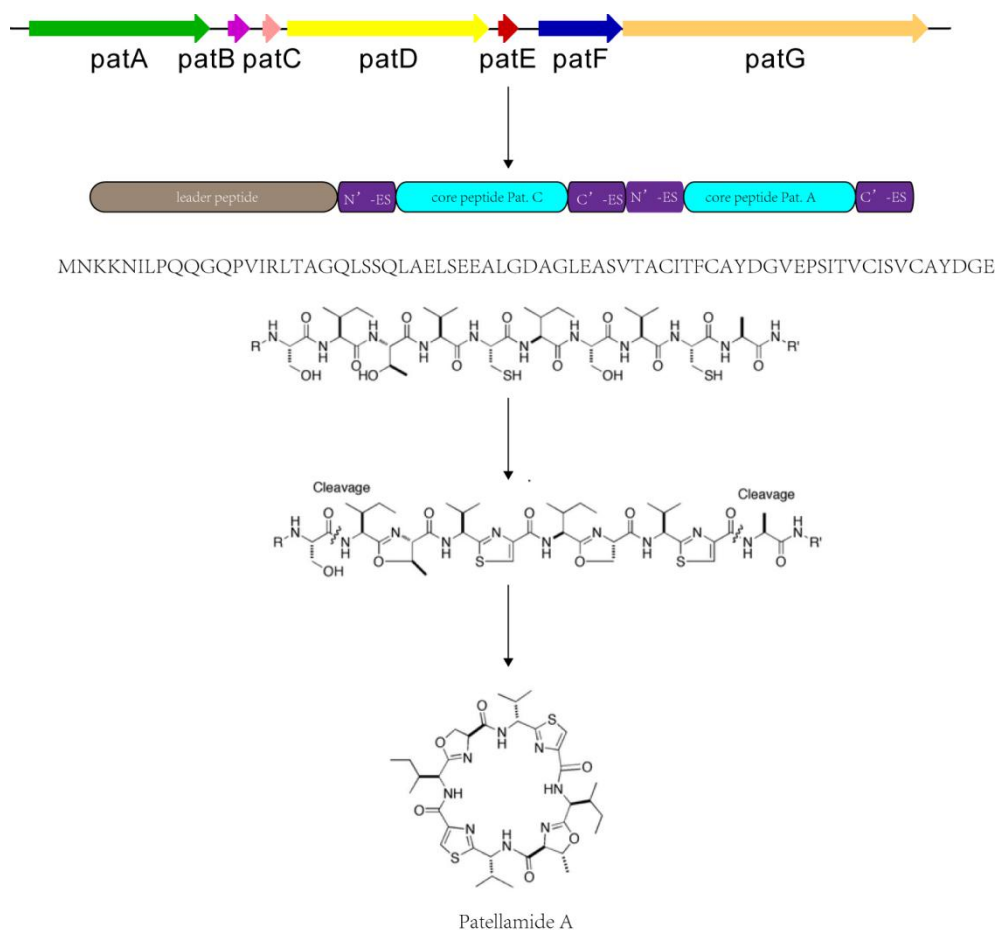
Patellamide A

**Figure 1.14:** Potential biosynthetic pathway to patellamide A.

## 1.2.4 Proteusins

Proteusins, known previously as polytheonamides, were isolated from the marine sponge *Theonella swinhoei* in 1994[82, 83]. These compounds are extensively modified. Each is a 48 aa peptide, but 19 different amino acids are nonproteinogenic[84, 85]. The modifications include an unprecedented N-acyl moiety, many tert-leucines, C-methylated amino acids and, most interestingly, a lot of D-amino acids (Figure 1.15). Proteusins were assumed for a long time to be products of NRPS, which can generate peptides with unusual residues[84, 85].

14
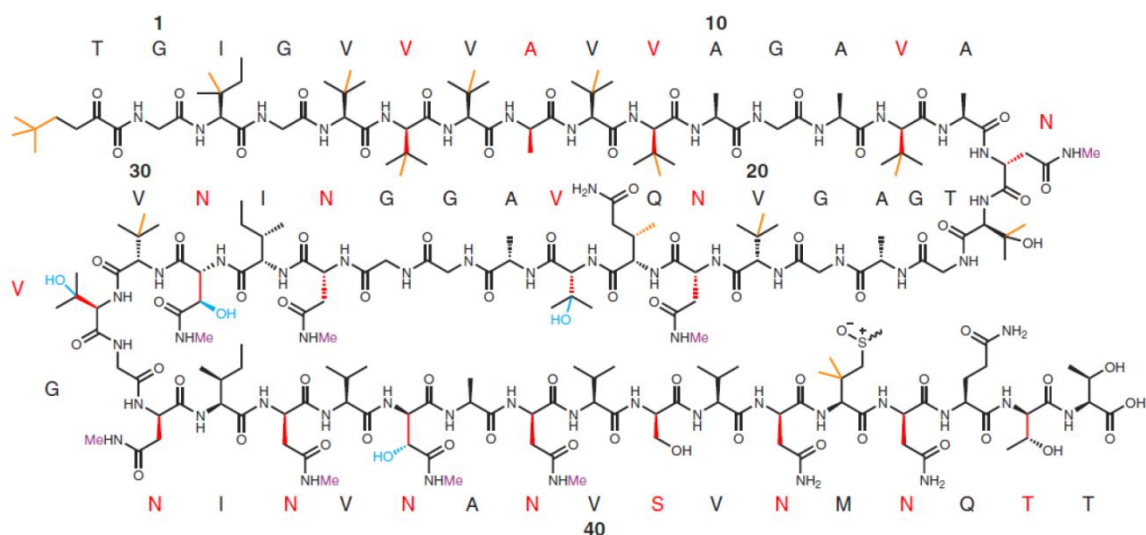
**Figure 1.15:**Structure of polytheonamide (Figure from ref. [86]).

However, in 2012, Freeman et al. identified a biosynthetic pathway that contains a precursor peptide for polytheonamide by mining the sponge metagenome[86]. It turned out that this gene cluster is from a bacterial endosymbiont of *Theonella swinhoei.* The gene cluster also contains six other tailoring enzymes that carry out 48 different posttranslational modifications, which makes polytheonamides the most extensively modified RiPPs known to date (Figure 1.16)[40, 86]. It was proposed that after the precursor peptide is produced, PoyD, a single epimerase, generates most, and possibly all, D-residues in the peptide. The second step in polytheonamide biosynthesis is dehydration of a Thr residue by PoyF, which seems to be responsible for the unusual N-acyl residue. Further conversion to the N-acyl unit is probably catalyzed by radical-SAM methyltransferases (PoyB and/or C are candidates) [40, 86]. PoyE, a single N-methyltransferase, was shown to be responsible for the generation of eight N-methylated Asn residues. The functions of other proteins in the gene cluster have been elusive so far[40, 86].

**Leader peptide**  MADSDNTPTSRKDFETAIIAKAWKDPEYLRRLRSNPREVLQEELEALHPGAQLPDDLGISIHEEDENHVHLVMPRHPQNVSDQTLTDDDLDQAAGG

**Core peptide**  TGIGVVVAVVAGAVANTGAGVNQVAGGNINVVGNINVNANVSVNMNQTT

| Gene | Predicted function |
|------|--------------------|
| poyA | Precursor |
| poyB | Radical SAM methyltransferase |
| poyC | Radical SAM methyltransferase |
| poyD | Epimerase |
| poyE | SAM Methyltransferase |
| poyF | Dehydratase |
| poyG | Unknown |
| poyH | Unknown |
| poyI | Oxidoreductase |
| poyJ | Unknown |
| poyK | Unknown |

**Figure 1.16:** Biosynthetic pathway to polytheonamide.

## 1.2.5 Sactipeptides

Sactipeptides are a new growing class of RiPPs that show diverse bioactivities (Figure 1.17)[87-90]. The characteristic structure of a sactipeptide is an intramolecular thioether bond that crosslinks the sulfur atom of a cysteine residue with the a-carbon of another residue[40, 87-90]. The first compound of this family, subtilosin A, was isolated from *Bacillus subtilis 168* in 1985. The three-dimensional solution structure of subtilosin A was solved by NMR in 2003[91-94].



Subtilosin A

Thurincin H
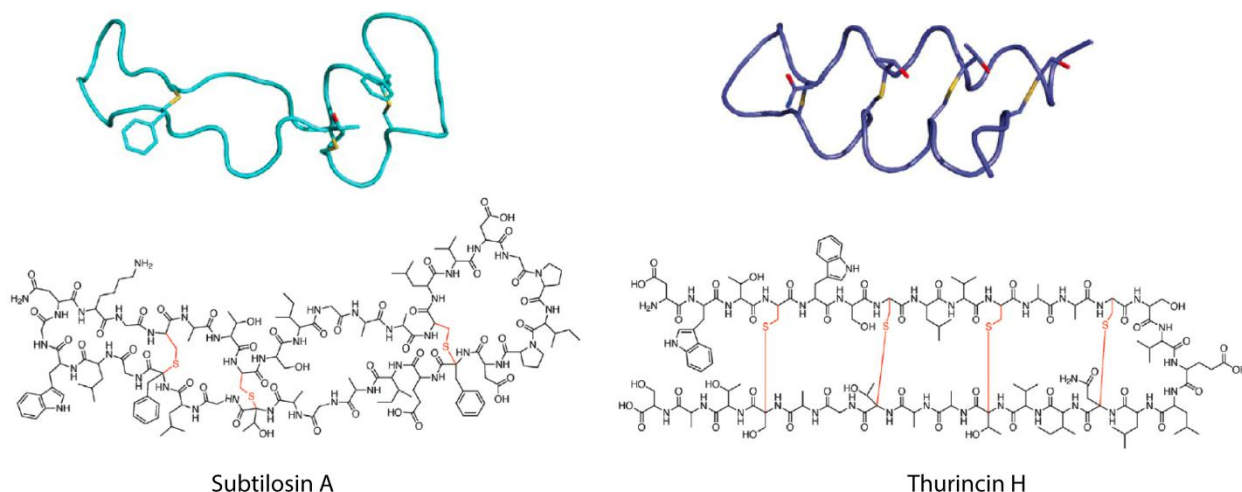
**Figure 1.17:** Examples of the structural diversity of sactipeptides (Figure from ref. [87]).

Most of the sactipeptides isolated thus far show some antimicrobial activity. Another example is the sporulation killing factor (SKF) produced by *Bacillus subtilis*, which plays a key role in sporulation[89, 95, 96]. It is interesting that most of these compounds only have a narrow specificity

16

of antimicrobial activity, which suggests that these peptides require recognition of a specific receptor molecule in the target organisms[40, 87, 91, 92, 95, 96].
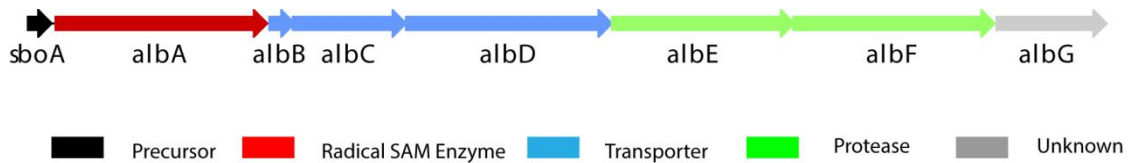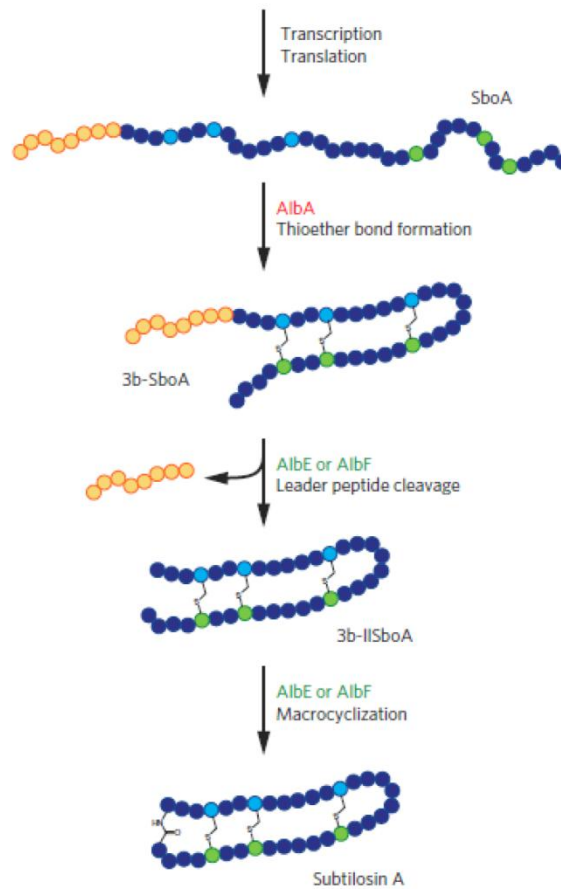


**Figure 1.18:** Biosynthetic pathway to subtilosin A (Figure from ref. [90]).

Several gene clusters responsible for the biosynthesis of sactipeptides have been reported (Figure 1.18)[87-90]. It has been shown for the biosynthesis of subtilosin A that the characteristic sulfur to a-carbon crosslink was generated by a single radical SAM enzyme AlbA. Other cases, such as SKF and thurincin, also use an AlbA homology enzyme to generate the unusual sulfur bridge[87-90].

## 1.3 Lasso peptides

Lasso peptides are a growing family of intriguing RiPPs[40, 97-99]. They contain about 20 residues and the characteristic structure of these molecules is a unique interlocked topology that involves an N-terminal seven- to nine-residue macrolactam ring where the C-terminal tail is threaded, stabilizing the entropically disfavored lasso structure[40, 97-99]. The first lasso peptide was discovered in 1991. Since then, more than 40 lasso peptides have been described[97].

### 1.3.1 Classification of lasso peptides

Lasso peptides normally use bulky amino acids located in the tail below and above the ring to trap the tail. However, they also sometimes use a disulfide bridge to further stabilize the structure. Therefore, lasso peptides are subdivided into three subtypes, depending on the absence (class II) or presence of one (class III) or two (class I) disulfide bridges (Figure 1.19)[100].



Siamycin
Class I

Mccj25
Class II
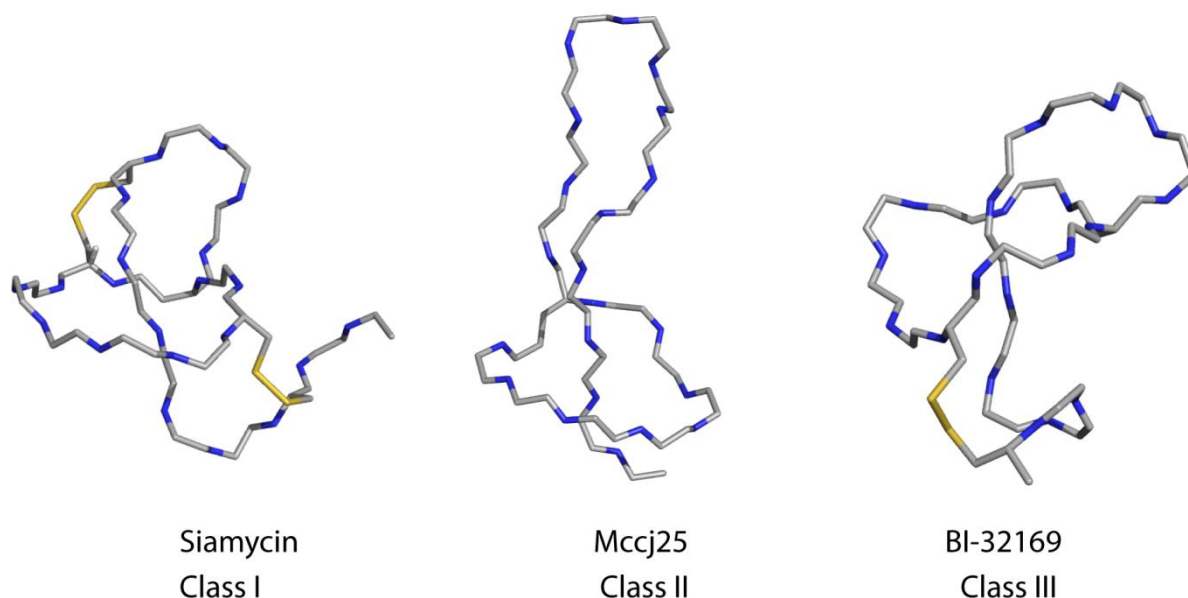
BI-32169
Class III

**Figure 1.19:** Three classes of lasso peptide. Siamycin as an example of a class I lasso peptide (Protein Data Bank [PDB]: 1RPB); BI 32169 is the only class III lasso peptide (PDB: 3NJW); Mccj25 is an example of a class II lasso peptide (PDB: 1Q71).

Class I lasso peptides have two disulfide bonds: one between the N-terminal cysteine and the loop, and the other between the ring and the tail. Currently, there are four examples of class I lasso peptides which have been investigated. It was shown in the biosynthesis of Sviceucin that no gene was involved in the biosynthesis of the disulfide bonds. There is only one example of a class III lasso peptide called BI-32169. This lasso peptide contains a single disulfide bond between a cysteine at the C-terminus of the peptide and the ring. Most of the lasso peptides isolated belong to class II, which lacks disulfide bonds[100]. There are 33 examples of class II lasso peptide, which are summarized in Table 1.1. Most of them were discovered after 2008[97, 98].

**Table1.1:**Overview of all known lasso peptides sorted by class.

18

| Name | Sequence | Host strain |
|---|---|---|
| **ClassI** | | |
| RP 71955/aborycin | CLGIGSCNDFAGCGYAVVCFW | *Streptomyces sp.* |
| Siamycin I/MS-271/NP-06 | CLGVGSCNDFAGCGYAIVCFW | *Streptomyces sp.* |
| Siamycin II | CLGIGSCNDFAGCGYAIVCFW | *Streptomyces sp.* |
| Sviceucin | CVWGGDCTDFLGCGTAWICV | *Streptomyces sviceus* |
| **ClassII** | | |
| Anantin | GFIGWGNDIFGHYSGDF | *Streptomyces coerulescens* |
| Capistruin | GTPGFQTPDARVISRFGFN | *Burkholderia thailandensis* |
| Lariatin A | GSQLVYREWVGHSNVIKP | *Rhodococcus sp. K01-B0171* |
| Lariatin B | GSQLVYREWVGHSNVIKPGP | *Rhodococcus sp. K01-B0171* |
| Microcin J25 | GGAGHVPEYFVGIGTPISFYG | *Escherichia coli AY25* |
| Propeptin | GYPWWDYRDLFGGHTFISP | *Microbispora sp. SNA-115* |
| RES-701-1 | GNWHGTAPDWFFNYYW | *Streptomyces sp. RE-701* |
| SRO15-2005 | GYFVGSYKEYWSRRII | *Streptomyces roseosporus* |
| Astexin-1 | GLSQGVEPDIGQTYFEESRINQD | *Asticcacaulis excentricus CB 48* |
| Astexin-2 | GLTQIQALDSVSGQFRDQLGLSAD | *Asticcacaulis excentricus CB 48* |
| Astexin-3 | GPTPMVGLDSVSGQYWDQHAPLAD | *Asticcacaulis excentricus CB 48* |
| Caulosegnin I | GAFVGQPEAVNPLGREIQG | *Caulobacter segnis* |
| Caulosegnin II | GTLTPGLPEDFLPGHYMPG | *Caulobacter segnis* |
| Caulosegnin III | GALVGLLLEDITVARYDPM | *Caulobacter segnis* |
| Sungsanpin | GFGSKPIDSFGLSWL | *Streptomyces sp.* |
| Burhizin | GGAGQYKEVEAGRWSDRIDSDDE | *Burkholderia rhizoxinica HKI454* |
| Caulonodin I | GDVLNAPEPGIGREPTGLSRD | *Caulobacter sp. K31* |
| Caulonodin II | GDVLFAPEPGVGRPPMGLSED | *Caulobacter sp. K31* |
| Caulonodin III | GQIYDHPEVGIGAYGCEGLQR | *Caulobacter sp. K31* |
| Zucinodin | GGIGGDFEDLNKPFDV | *Phenylobacterium zucineum HLK1* |
| Rhodanodin | GVLPIGNEFMGHAATPGITE | *Rhodanobacter thiooxydans LCS2* |
| Rubrivinodin | GAPSLINSEDNPAFPQRV | *Rubrivivax gelatinosus IL44* |

| | | |
|---|---|---|
| Sphingonodin I | GPGGITGDVGLGENNFGLSDD | *Sphingobium japonicum UT26* |
| Sphingonodin II | GMGSGSTDQNGQPKNLIGGISDD | *Sphingobium japonicum UT26* |
| Syanodin I | GISGGTVDAPAGQGLAGILDD | *Sphingobium yanoikuyae XLDN2-5* |
| Sphingopyxin I | GIEPLGPVDEDQGEHYLFAGGITADD | *Sphingopyxis alaskensis RB2256* |
| Sphingopyxin II | GEALIDQDVGGGRQQFLTGIAQD | *Sphingopyxis alaskensis RB2256* |
| streptomonomycin | SLGSSPYNDILGYPALIVIYP | *Streptomonospora alba* |
| Chaxapeptin | GFGSKPLDSFGLNFF | *Streptomyces leeuwenhoekii C58* |
| Xanthomonin I | GGPLAGEEIGGFNVPG | *Xanthomonas gardneri* |
| Xanthomonin II | GGPLAGEEMGGITT | *Xanthomonas gardneri* |
| Xanthomonin III | GGAGAGEVNGMSP | *Xanthomonas citri* |
| Lassomycin | GLRRLFADQLVGRRNI | *Lentzea kentuckyensis* |
| **ClassIII** | | |
| BI-32169 | GLPWGCPSDIPGWNTPWAC | *Streptomyces sp.* |

## 1.3.2 Function of lasso peptides

Lasso peptides were discovered prior to 2008 in the course of activity-driven compound isolations. Their activities include antimicrobial, anti-viral and anti-metastatic activity[101-108]. Others also function as receptor antagonists, enzyme inhibitors and so on (Table 1.2)[40, 97-100, 109, 110]. However, the activity of most of the lasso peptides discovered since 2008 through genome mining are still unknown. They are isolated mainly from Proteobacteria, and their biosynthetic gene clusters lacks an immunity-conferring ABC-transporter[99, 111-118]. These lasso peptides are believed to function as a type of scavenging molecule[113, 118]. In addition to the native bioactivity, lasso peptides can also be promising molecular scaffolds for epitope grafting. In the case of Mccj 25, it has an unusually long loop that could be used for drug design. The integrin-binding motif RGD was recently successfully grafted onto Mccj25 by substituting a tripeptide sequence in the loop region with Arg-Gly-Asp[119]. The Mccj 25-RGD displayed nanomolar affinity towards $a_v\beta_3$, $a_v\beta_5$, $a_5\beta_1$, and $a_{IIb}\beta_3$ integrins. An optimization of a grafted lasso peptide, named MccJ25(RGDF), was recently shown to be a highly potent and selective $\alpha_v\beta_3$ integrin inhibitor[120].

**Table1.2:** Biological activities of known lasso peptides

| Name | Inhibitor/ antagonist of | Antimicrobial activity |
|---|---|---|
| Anantin | atrial natriuretic factor | - |

| BI-32169 | glucagon receptor | - |
|---|---|---|
| Capistruin | Gram-negative RNA polymerase | Yes |
| Lariatin | | Yes |
| Lassomycin | ClpC1 | Yes |
| Microcin J25 | Gram-negative RNA polymerase | Yes |
| Propeptin | prolyl endopeptidase | Yes |
| RES-701 type | endothelin type-B receptor | - |
| Siamycin type | HIV fusion and replication, myosin light chain kinase | Yes |
| streptomonomycin | - | Yes |
| Chaxapeptin | lung cancer cell line | - |
| Sungsanpin | lung cancer cell line | - |

### 1.3.3 Biosynthesis of lasso peptides

Lasso peptides have thus far only been isolated from Proteo- and Actinobacterial sources[40, 97-99]. The biosynthetic gene cluster for lasso peptide could be classified into three classes prior to the research carried out in this work[40, 97-99]. The biosynthesis of the Mccj 25 lasso peptide represents one class and has been well studied in the past[101-103, 121-126]. This biosynthetic cluster contains four genes: mcjA encodes the precursor peptide, mcjB is an ATP-dependent cysteine protease, mcjC is an ATP-dependent asparagine synthetase homologue and mcjD is an ABC transporter. While McjD is only needed to export the mature antimicrobial compound, McjB and McjC participate directly in the maturation of McjA into MccJ25[101-103, 121-126]. Other examples belonging to this family include Capistruin[127]. There is another class of lasso peptide gene cluster which also contains an ABC-transporter, but the B protein is split into two separate ORFs homologous to the N- and C-terminal domains of McjB. Examples of members of this family includes lariatin, lassomycin, streptomonomicin, sviceucin and chaxapeptin[104-108, 128]. They are mainly isolated from Actinobacteria. Regarding Astexin and other recently new isolated lasso peptides from Proteobacteria, the gene clusters lack an ABC-transporter and instead feature other highly conserved adjacent genes, such as putative peptidases, which have recently been shown to act as lasso peptide-specific isopeptidases (Figure 1.20)[99, 111-118].
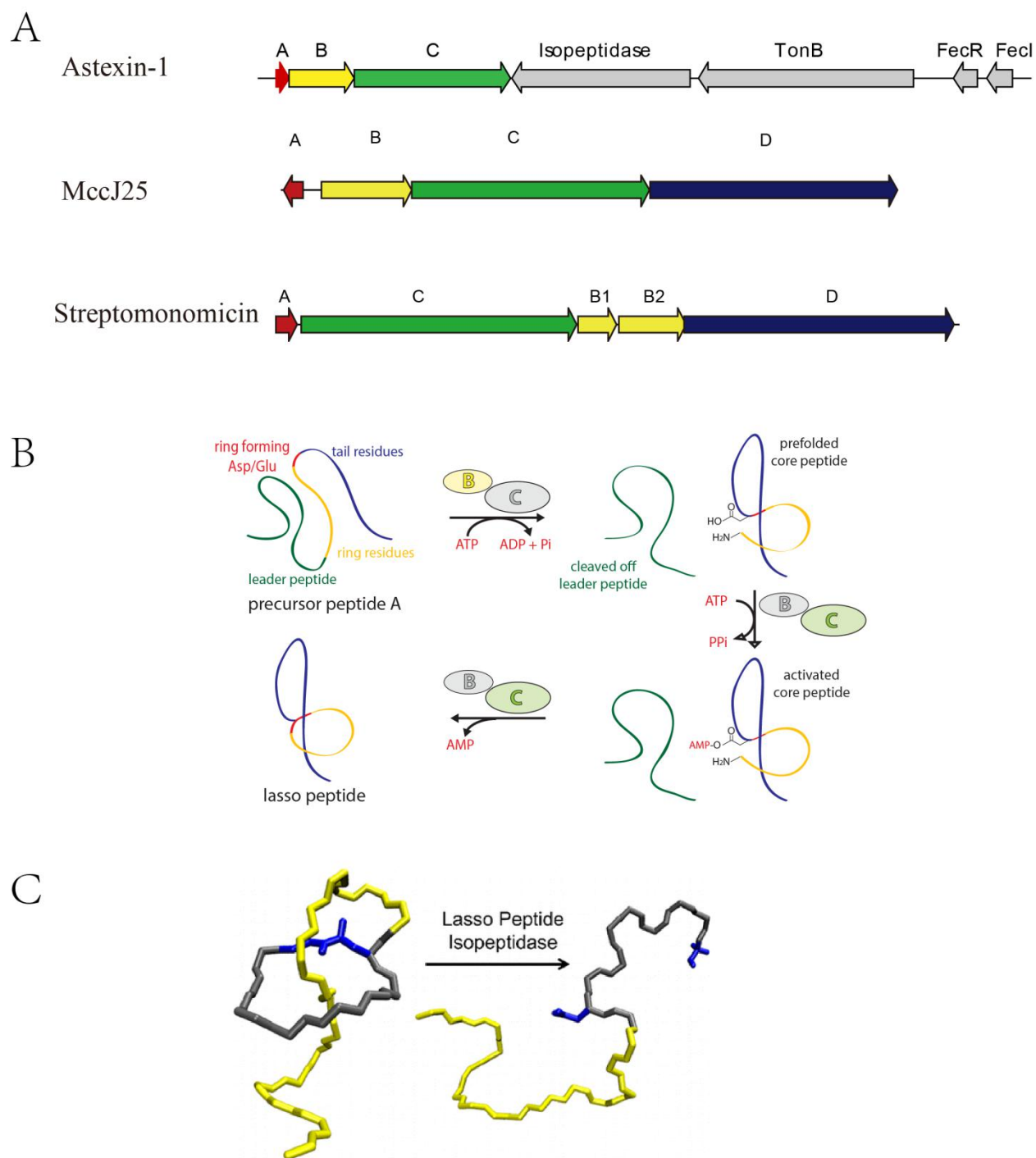
**Figure 1.20:** Biosynthesis of Lasso peptide. (A) Three types of biosynthetic gene cluster for lasso peptide. (B) Scheme of the suggested mechanism of lasso peptide biosynthesis. (C) Isopeptidase linearizes the folded lasso peptides ( Figure from ref [113]).

## 1.3.4 Discovery of lasso peptides

All the lasso peptides found before 2008 were discovered in the course of activity-driven compound isolations. Some lasso peptides, such as lassomycin, sungsanpin and streptomomycin, were also discovered in the same way after 2008[97, 104, 105, 107]. A genome mining method was applied in 2008 to discover new lasso peptides. As a result, capistruin was

isolated and showed moderate antimicrobial activity[127]. After this example, the same method was applied and a lot of lasso peptides lacking any bioactivity were isolated from Proteobacterial sources, including caulosegnin, xanthomonin and rubrivinodin[111, 112, 114-116]. Link et al. developed a precursor-centric genome-mining approach for lasso peptide discovery and several lasso peptides from Proteobacteria including astexin were found[117]. They also discovered that these lasso peptides are produced by a lasso peptide gene cluster featuring a lasso peptide Isopeptidase[113, 118]. Except for all the methods mentioned above, a mass spectrometry-guided genome mining approach was also applied for lasso peptide discovery[42]. This led to the discovery of sviceucin and SRO15-2005 lasso peptide[42, 108]. All the genome mining methods are generally powerful in lasso peptide discovery, but the bioactivity of these compounds is still elusive.

## 1.4 Aim of this work

The aim of this thesis was to further exploit the lasso peptide biosynthetic system. All the lasso peptides were isolated from either Proteobacterial or Actinobacterial sources[97-99, 112]. Lasso peptides from Firmicutes have not been reported previously. Thus, the first part of this thesis was an attempt to isolate the first lasso peptide, paeninodin, from Firmicutes. The lasso peptide was successfully heterologously expressed in *E. coli* and the topology of the lasso peptide was proved by MS2, ion mobility-mass spectrometry (IM-MS) and stability studies.

Bioinformatics analysis showed that the lasso peptide gene clusters from Firmicutes have an unusual CAKB1B2D organization, which is different from the gene clusters from Proteobacteria and Actinobacteria. It contains not only the conserved genes encoding the precursor peptide, B, C protein, and ABC-transporter, but an additional gene coding for a putative kinase. So far, descriptions of lasso peptide tailoring in the literature are notably rare (examples include C-terminal methylation of lassomycin and Trp hydroxylation of some RES-701 family lasso peptides)[104]. Therefore, the second part of this thesis investigates the role of the unusual kinase in the gene cluster. We successfully isolated a phosphorylated lasso peptide by heterologous expression of the whole gene cluster in *E. coli*. We identified the kinase by knock-out and in trans experiments as a novel tailoring enzyme for lasso peptide for the first time.

The third part of the thesis investigates how the kinase tailors the lasso peptide. We clearly showed by *in vitro* experiment that the kinase first modified the precursor peptide and then the modified precursor peptide was transformed into a modified lasso peptide. Moreover, a plausible catalytic mechanism for the kinase is proposed based on mutagenesis studies. Finally, it could be unambiguously determined by substrate specificity studies and estifrication assays combined with NMR studies that this kinase exclusively transfers a phosphate group from ATP to the side chain of the C-terminal Ser. This thesis reveals how lasso peptides are chemically

diversified and establishes the foundation for rational engineering. The main results of this thesis were published in *The Journal of Biological Chemistry (JBC)*[129].

# 2. Material

## 2.1 Equipments

**Table 2.1:** List of devices used in this work.

| Device | manufacturer and type |
|---|---|
| Analytical balance | Sartorius |
| Autoclave | Tuttnauer 5075 ELV, Fedegari Autoclavi SPA FVA3/A1 |
| Centrifuges | Sorvall RC 5B Plus and RC6+ (SS‑34, SLC‑300,SLC‑4000 rotors), Heraeus Minifugue RF and Megafugue 1.0R, Eppendorf 5415 D, 5415 Rand 5702 R |
| Clean-Bench | Antair BSK |
| Documentation system for DNA‑electrophoresis gels | Cybertech CS1 camera, Mitsubishi video copy thermo printer |
| Electrophoresis chamber | Agarose gel chambers manufactured in-house (PUMa, Marburg), Bio-Rad Mini-PROTEAN 3 gel chamber |
| Electroporation | Bio-Rad Gene-Pulser II |
| Fast protein liquid chromatography (FPLC) | Amersham Pharmacia Biotech Äktaprime and Äktapurifier: (pH/C-900,UV-900, P-900 and Frac-900 modules) |
| French-Press | SLM Aminco French-Pressure Cell Press 5.1, Thermo SpectronicStandard Cell 40 KP |
| LTQ-FT-ICR-MS | Thermo Fischer Scientific |
| HPLC-systems | Agilent series 1100 (HPLC-system withvacuum degasser, quarternary pump, auto samplerpreparative fraction collector, column thermostat, HP-ChemStationsoftware) |
| HPLC-MS-System (High Resolution) | Thermo Fisher Scientific LTQ-FTAgilent 1100 HPLC system:Vacuum degasser, DAD detector, quaternary pump, autosampler |
| HPLC-MS-System | Hewlett-Packard MSD 1100Agilent 1260 HPLC system: Vacuum degasser, DAD detector, quaternary pump,autosampler, fraction collector |

25

| Device | manufacturer and type |
|---|---|
| Incubators | New Brunswick Scientific Series 25, Innova 4300 Shaker, Infors HTMultitron II and Unitron |
| Lyophilizer | Christ Alpha 2-4 LSC |
| NMR | Bruker AV600 |
| pH meter | Schott CG 840 |
| Nanodrop-Photometer | PEQLab ND-1000 |
| Pipettes | Eppendorf Research series |
| Spectrophotometer | PEQLab Nanodrop ND-1000; Pharmacia Ultrospec 3000 |
| Thermal cycler | Eppendorf Mastercycler Personal |
| Thermomixer | Eppendorf Thermomixer comfort |
| Vortexer | Scientific Industries VortexGenie2 |
| Water deionizer | Seral Seralpur Pro90CN |

## 2.2 Chemicals, enzymes and consumables

All chemicals not listed in Table 2.2 were purchased from Sigma-Aldrich (Steinheim, Germany), Fluka (Steinheim, Germany) or Merck (Darmstadt, Germany) in p.a. quality and were used without further purification.

**Table 2.2:** Chemicals, enzymes and general materials and consumables.

| Manufacturer | Products |
|---|---|
| Applichem (Darmstadt, Germany) | ampicillin, kanamycin, media components |
| Biomol (Ilvesheim, Germany) | DTT |
| Eppendorf (Hamburg, Germany) | 1.5 and 2.0 mL reaction tubes |
| Eurogentech (Seraing, Belgium) | agarose, electroporation cuvettes |
| GE Healthcare (Freiburg, Germany) | IPTG, FPLC Ni-NTA and Superdex 200 5/150 GL SEC columns, yeast extract, aldolase, ovalbumin, ribonuclease, aprotinin protein standards |
| Invitrogen (Karlsruhe, Germany) | *E. coli* strains (BL21, TOP10) |

| Manufacturer | Products |
| --- | --- |
| Macherey & Nagel (Düren,Germany) | C18-HPLC columns (Nucleosil, Nucleodur) |
| Macherey-Nagel (Düren, Deutschland) | 125/2 Nucleodur 100-3 C18ec, 250/12 Nucleodur C18ec, 250/4.6 Nucleodur C18 HTec |
| Merck Biosciences - Novagen (Nottingham, UK) | pET Vector |
| Millipore (Schwalbach, Germany) | Dialysis membranes (pore size: 0.025 µM), Amicon Ultra‐15 concentrators |
| New England Biolabs (Frankfurt, Germany) | desoxyribonucleotides (dATP, dTTP, dGTP, dCTP), DNA markers, protein size markers, restriction endonucleases,Phusion Hi-Fidelity DNA polymerase, T4 DNA ligase |
| Oxoid (Cambridge, UK) | agar nr. 1, tryptone |
| QIAgen (Hilden, Germany) | QIAquick Gel Extraction kit, Ni-NTA IMAC resin |
| Sarstedt (Nümbrecht, Germany) | Pipette tips, Falcon tubes (15 and 50 mL) |
| Schleicher & Schüll (Dassel, Germany) | Sterile filters (0.20 and 0.45 µm), Whatmann-3MM paper |

## 2.3 Oligonucleotide

Oligonucleotide primers used in this work are summarized in Table below.They are used either for the amplification of target genes or for specific mutagenesis.

**Table 2.3.**Primers used for cloning of the *padeCAKB1B2D* gene cluster. Genomic DNA of *P. dendritiformis* C454 was used as a template for PCR. Introduced restriction sites are underlined.

| construct | name | sequence |
| --- | --- | --- |
| pET41a-*padeCAKB1B2D* | Pade_C_NdeI_FP | GG GAA TTC CAT ATG GGT GCA GTA GCA GGC ATT TAT CAC CTT CAA CAC G |
| | Pade_D_XhoI_RP | AT ATCTC GAG CTA CAG CGT ATA GAC GTT AAC AGG CTG CGC CTC |

**Table 2.4.** Primers for mutagenesis of the *padeCAKB1B2D* gene cluster and cloning of the *padeK* gene. All mutations were introduced using SLIM. SLIM overhang regions are underlined. Mutated positions are highlighted in bold. pET41a-*padeCAKB1B2D* was used as template for all PCR reactions. Primers P1 and P2 were used in several PCR reactions.

| constructs | name | sequence |
|---|---|---|
| pET41a-*padeCAB1B2D* | Pade_deHpr_P1 | AGTTTGTAAAATGAGAACAGGGGAACCCTTATACCCGTTG |
| | Pade_deHpr_P2 | CCGGCAAGGAGAGAATGACCA ATGAGCAAAC |
| | Pade_deHpr_P3 | <u>GTAATGTTTCTCTCCAAG</u>AGTTTGTAAAATGAGAACAGGGGAAC |
| | Pade_deHpr_P4 | <u>CTTGGAGAGAAACATTAC</u>CCGGCAAGGAGAGAATGAC |
| pET41a-*padeCAKB1B2D*(H20A) | Pade H20A P1 | GTCTTCATCCGGGTCCGGCTGGAAAGCATC |
| | Pade H20A P2 | ACAATTGGCTTAATGAATGCGGTTACTGGTG |
| | Pade H20A P3 | <u>CTAGCTATCGTA**CGC**CAC</u>GTCTTCATCCGGGTCCGGCTGGAAAG |
| | Pade H20A P4 | <u>GTG**GCG**TACGATAGCTAG</u>ACAATTGGCTTAATGAATGCGGTTAC |
| pET41a-*padeCAKB1B2D*(Y21A) | Pade Y21A P3 | <u>CTAGCTATC**CGC**ATGCAC</u>GTCTTCATCCGGGTCCGGCTGGAAAG |
| | Pade Y21A P4 | <u>GTGCAT**GCG**GATAGCTAG</u>ACAATTGGCTTAATGAATGCGGTTAC |
| pET41a-*padeCAKB1B2D*(Y21F) | Pade Y21F P3 | <u>CTAGCTATC**AAA**ATGCAC</u>GTCTTCATCCGGGTCCGGCTGGAAAG |
| | Pade Y21F P4 | <u>GTGCAT**TTT**GATAGCTAG</u>ACAATTGGCTTAATGAATGCGGTTAC |
| pET41a-*padeCAKB1B2D*(D22A) | Pade D22A P3 | <u>CTAGCT**CGC**GTAATGCAC</u>GTCTTCATCCGGGTCCGGCTGGAAAG |
| | Pade D22A P4 | <u>GTGCATTAC**GCG**AGCTAG</u>ACAATTGGCTTAATGAATGCGGTTAC |
| pET41a-*padeCAKB1B2D*(S23A) | Pade S23A P3 | <u>CTA**CGC**ATCGTAATGCAC</u>GTCTTCATCCGGGTCCGGCTGGAAAG |
| | Pade S23A P4 | <u>GTGCATTACGAT**GCG**TAG</u>ACAATTGGCTTAATGAATGCGGTTAC |
| pET41a-*padeCAKB1B2D*(S23) | Pade de1 P2 | ATTGGCTTAATGAATGCGGTTACTGGTGCGGTAC |
| | Pade de1 P3 | <u>TGTCTAATCGTAATGCAC</u>GTCTTCATCCGGGTCCGGCTGGAAAG |
| | Pade de1 P4 | <u>GTGCATTACGATTAGACA</u>ATTGGCTTAATGAATGCGGTTAC |
| pET41a-*padeCAKB1B2D*(D22 S23) | Pade de2 P2 | GGCTTAATGAATGCGGTTACTGGTGCGGTAC |
| | Pade de2 P3 | <u>AATTGTCTAGTAATGCAC</u>GTCTTCATCCGGGTCCGGCTGGAAAG |
| | Pade de2 P4 | <u>GTGCATTACTAGACAATT</u>GGCTTAATGAATGCGGTTAC |
| pET41a-*padeCAKB1B2D*(T-2A) | Pade T-2A-P1 | AACATCCAACACTTCAAGGGACGGCTTGCTGTACTGCTTTTTC |
| | Pade T-2A-P2 | GGTACCTCTACCCCGGATGCTTTCCAGCCGGACCCG |
| | Pade T-2A-P3 | <u>CGGGCCAGCCAT**CGC**CTGATG</u>AACATCCAACACTTCAAGGGA C |
| | Pade T-2A-P4 | <u>CATCAG**GCG**ATGGCTGGCCCG</u>GGTACCTCTACCCCGGATG |
| pET41a-*padeCAKB1B2D*(T-2C) | Pade T-2C-P3 | <u>CGGGCCAGCCAT**GCA**CTGATG</u>AACATCCAACACTTCAAG GGAC |
| | Pade T-2C-P4 | <u>CATCAG**TGC**ATGGCTGGCCCG</u>GGTACCTCTACCCCGGATG |
| pET41a-*padeCAKB1B2D*(T-2V) | Pade T-2V-P3 | <u>CGGGCCAGCCAT**CAC**CTGATG</u>AACATCCAACACTTCAAGGGAC |
| | Pade T-2V-P4 | <u>CATCAG**GTG**ATGGCTGGCCCG</u>GGTACCTCTACCCCGGATG |

| | | |
|---|---|---|
| pET41a-*padeCAKB1B2D*(A1C) | PadeA1C-P3 | <u>CGGGCC</u>**GCA**CATGGTCTGATGAACATCCAACACTTCAAGGGAC |
| | PadeA1C-P4 | <u>CATCAGACCATG</u>**TGC**GGCCCG GGTACCTCTACCCCGGATG |
| pET41a-*padeCAKB1B2D*(A1G) | PadeA1G-P3 | <u>CGGGCC</u>**GCC**CATGGTCTGATGAACATCCAACACTTCAAG GGAC |
| | PadeA1G-P4 | <u>CATCAGACCATG</u>**GGC**GGCCCG GGTACCTCTACCCCGGATG |
| pET41a-*padeCAKB1B2D*(A1S) | PadeA1S-P3 | <u>CGGGCC</u>**GCT**CATGGTCTGATGAACATCCAACACTTCAAGGGAC |
| | PadeA1S-P4 | <u>CATCAGACCATG</u>**AGC**GGCCCGGGTACCTCTACCCCGGATG |
| pET41a-*padeCAKB1B2D*(D9E) | PadeD9E-P1 | AGGTACCCGGGCCAGCCATGGTCTGATGAACATCCAACAC |
| | PadeD9E-P2 | GCCGGACCCGGATGAAGACGTGCATTACGATAGCTAGAC |
| | PadeD9E-P3 | <u>TGGAAAGC</u>**TTC**CGGGGTAGAGGTACCCGGGCCAGCCATGG |
| | PadeD9E-P4 | <u>CTACCCCG</u>**GAA**GCTTTCCAGCCGGACCCGGATGAAGACGTG |
| pACYC-*padeK* | PadeK-NdeI-FP | GGAATTC<u>CATATG</u>ACCGAACGAGCAGCGGTAAG |
| | PadeK-XhoI-RP | CCG<u>CTCGAG</u>CATTGGTCATTCTCTCCTTGC |
| | pACYCD-NdeI | GGAATTC<u>CATATG</u>TATATCTCCTTATTAAAGTTAAAC |
| | pACYCD-XhoI | CCG<u>CTCGAG</u>TCTGGTAAAGAAACCGCTGC |

**Table 2.5.** Primers for cloning of *padeK* and *thcoK*. Overhang regions for *Gibson Assembly* are highlighted in bold. Introduced restriction sites are underlined.

| constructs | name | sequence |
|---|---|---|
| pETMBP-*padeK* | pETMBP-PadeK-NdeI-FP | GGA ATT C<u>CA TAT G</u>AC CGA ACG AGC AGC GGT AAG |
| | pETMBP-PadeK-XhoI-RP | CCG <u>CTC GAG </u>TCATTG GTC ATT CTC TCC TTG C |
| pET41-*thcoK* | KinaseThco FP | **ACTTTAAGAAGGAGATATACAT**ATGA CCA GGA CAA ATA CCG CTA TCG GTA CCG GGC CTT C |
| | KinaseThco RP | **GGTGGTGGTGGTGGTGCTCGAG**TCG GCT CAC CTC CCC ATC GGC ATG CGT CTC TAT G |

**Table 2.6.** Primers for cloning of the *padeA* gene into pET-48b(+). Overhang regions for *Gibson Assembly* are highlighted in bold.

| constructs | name | sequence |
|---|---|---|
| pET48b-*padeA* | PadeA V FP | CTC GAG GCT TAA TTA ACC TAG GCT GCT AAA CAA AGC C |
| | PadeA V RP | GGG TCC CTG AAA GAG GAC TTC AAG AGC CGC G |
| | PadeA FP | **AAGTCCTCTTTCAGGGACCC** ATG AAA AAG CAG TAC AGC AAG CCG TCC CTT GAA GTG |
| | PadeA RP | **CCATGGGACGGTGTAAACGG**TCATCGGCTCACCT |

C CCCATCGGCATGCGTCTCTATG

**Table 2.7.** Primers for exchanging *padeK* with *thcoK* in the pET41a-*padeCAKBBD* gene cluster. Overhang regions for *Gibson Assembly* are highlighted in bold.

| construct | name | sequence |
|---|---|---|
| pET41a-*padeCA-thcoK-padeB1B2D* | ThcoKexc V-FP | CCGTTTACACCGTCCCATGGTTGGATTCAGCACGCTTGATC |
| | ThcoKexc V-RP | GCTCATCTCTCCAAGAGTTTGTAAAATGAGAACAGG GGA ACC CTT ATA CC |
| | ThcoKexc K-FP | **CAAACTCTTGGAGAGATGAGC**ATGACCAGGACAAATA CCG GCT ATC GGT ACC GGG CCT TC |
| | ThcoKexc K-RP | **CCATGGGACGGTGTAAACGG**TCATCGGCTCACCTC CCCATCGGCATGCGTCTCTATG |

**Table 2.8.** Primers for the mutagenesis of *padeK* and *thcoK*. All mutations were introduced using SLIM. SLIM overhang regions are highlighted in bold. pET41a-*padeCAKB1B2D* and pET41a-*padeCA-thcoK-padeB1B2D* were used as templates for *padeK* and *thcoK* mutations, respectively.

| constructs | name | sequence |
|---|---|---|
| pET41a-*padeCAKB1B2D* _padeK*(D161A) | PadeK D161A P1 | CAG CGG ATA TCC CTC GGA CAC CAG GTG CAA GG |
| | PadeK D161A P2 | CCC GTC GTG ATG ACT CAA GGC TCT CCT TGG GTT G |
| | PadeK D161A P3 | **AATAACATCCGCGCT AAG** CAG CGG ATA TCC CTC GGA CAC CAG |
| | PadeK D161A P4 | **CTTAGCGCGGAT GTT ATT** CCC GTC GTG ATG ACT CAA GGC TCT CCT TG |
| pET41a-*padeCAKB1B2D* _padeK*(D162A) | PadeK D162A P3 | **AATAACCGCATC GCT AAG** CAG CGG ATA TCC CTC GGA CAC |
| | PadeK D162A P4 | **CTTAGCGATGCGGTT ATT** CCC GTC GTG ATG ACT CAA GGC TCT CCT TG |
| pET41a-*padeCAKB1B2D* _padeK*(D161A D162A) | PadeK D161AD162A P3 | **AATAACCGCCGCGC TAA G**CA GCG GAT ATC CCT CGG ACA CCA G |
| | PadeK D161AD162A P4 | **CTTAGCGCGGCGGTTATT** CCC GTC GTG ATG ACT CAA GGC TCT CCT TG |
| pET41a-*padeCAKB1B2D* _padeK*(H123A) | PadeK H123A P1 | TAA TTT TGC GCT GCA GCA AGA TAA TGC CCA TAC |
| | PadeK H123A P2 | CGT AGC TAT TGA CGG GAA GGC CTA TGC CAT TAT C |
| | PadeK H123A P3 | **GCGCTGCCCGCCAA AGG CA**T AAT TTT GCG CTG CAG CAA GAT AAT GC |
| | PadeK H123A P4 | **TGCCTTTGGCGGGC AGC GC**C GTA GCT ATT GAC GGG AAG |
| pET41a-*padeCAKB1B2D* _padeK*(K144A) | PadeK K144A P1 | CGG ATT CGC CGA TAA TGG CAT AGG CCT TCC CG |
| | PadeK K144A P2 | GGC CTT GCA CCT GGT GTC CGA GGG ATA TCC G |
| | PadeK K144A P3 | **AAGGTCGACGCGCC CGC CC**C GGA TTC GCC GAT AAT GGC ATA G |

| | | |
|---|---|---|
| | PadeK K144A P4 | **GGGCGGGCGCGTCG ACC TT**G GCC TTG CAC CTG GTG TCC GAG |
| pET41a-*padeCA-thcoK-padeB1B2D_thcoK*(H119A) | ThcoK H119A P1 | GAT GCG CCG CTG CAG CAG GAG CGC GC |
| | ThcoK H119A P2 | GTG GCG CGC GAC GGG CGC GCA TAC GCC ATC |
| | ThcoK H119A P3 | **GACGCTGCCCGCAAG CGG CAG** GAT GCG CCG CTG CAG CAG G |
| | ThcoK H119A P4 | **CTGCCGCTTGCGGGC AGC GTC** GTG GCG CGC GAC GGG C |
| pET41a-*padeCA-thcoK-padeB1B2D_thcoK*(K140A) | ThcoK K140A P1 | CGA TTC GCC GAC GAT GGC GTA TGC GCG |
| | ThcoK K140A P2 | TCC GCG GCG CTG CTG GAA CGC GGA TTC |
| | ThcoK K140A P3 | **CATCGTCGACGCGCC CGC GCC** CGA TTC GCC GAC GAT GG |
| | ThcoK K140A P4 | **GGCGCGGGCGCGTCG ACG ATG** TCC GCG GCG CTG CTG GAA C |
| pET41a-*padeCA-thcoK-padeB1B2D_thcoK*(D157A) | ThcoK D157A P1 | ACG GAA TCC GCG TTC CAG CAG CGC CGC |
| | ThcoKD157A P2 | GCC ATC GTC TTC GAT GAG CGC GGC ACG C |
| | ThcoK D157A P3 | **GGCGACGTCCGCCGT CAC CAG** ACG GAA TCC GCG TTC CAG CAG C |
| | ThcoK D157A P4 | **CTGGTGACGGCGGAC GTC GCC** GCC ATC GTC TTC GAT GAG CG |
| pET41a-*padeCA-thcoK-padeB1B2D_thcoK*(D158A) | ThcoK D158A P3 | **GGCGACCGCGTC CGT CAC CAG** ACG GAA TCC GCG TTC CAG CAG |
| | ThcoK D158A P4 | **CTGGTGACGGACGCGGTC GCC**GCC ATC GTC TTC GAT GAG CGC G |
| pET41a-*padeCA-thcoK-padeB1B2D_thcoK*(D157A D158A) | ThcoK D157AD158A P3 | **GGCGACCGCCGCCGT CAC CAG** ACG GAA TCC GCG TTC CAG CAG C |
| | ThcoK D157AD158A P4 | **CTGGTGACGGCGGCGGTC GCC**GCC ATC GTC TTC GAT GAG CGC G |

**Table 2.9.**Primers for mutagenesis of pET48b-*padeA*. All mutations were introduced using SLIM. SLIM overhang regions are highlighted in bold. pET48b-*padeA* was used as template for all PCR reactions.

| constructs | name | sequence |
|---|---|---|
| pET48b-*padeA*-Core | GP-PadeAcore P1 | AAGAGGACTTCA AGAGCCGCGGAGTGATG |
| | GP-PadeAcore P2 | CCCCGGATGCTTTCCAGCCGGACCC |
| | GP-PadeAcore P3 | **TAGAGGTACCGGGTCCCTGA**AAGAGGACTTC AAGAGCCGC |
| | GP-PadeAcore P4 | **TCAGGGACCCGGTACCTCTA**CCCCGGATGCTTTC CAGCCG |
| pET48b-*padeA*(S23A) | GP-PadeA-S23A P1 | CACGTCTTCATCCGGGTCCGGCTGGAAAGC |
| | GP-PadeA-S23A P2 | GCTTAATTAACCTAGGCTGCTAAACAA AGCCCG AAAGG |
| | GP-PadeA-S23A | **CTCGAGCTACGCATC GTA ATG**CACGTCTTCATC CGGGTCCGGC |

| | P3 | |
|---|---|---|
| | GP-PadeA-S23A P4 | **CATTACGATGCGTAGCTCGAG**GCTTAATTAACC TAGGCTGCTAAACAAAG |
| pET48b-*padeA*(7xA-HYDS) | GP-PadeA-7xA-HYDS P1 | CTGGAAAGCATCCGGGGTAGAGGTACCCGG |
| | GP-PadeA-7xA-HYDS P2 | CATTACGATAGCTAGCTCGAGGCTTAATTAACCTAG G |
| | GP-PadeA-7xA-HYDS P3 | **CGCTGCCGCTGCTGCCGCTGC**CTGGAAAGCATC CGGGGTAGAGGTACCCGG |
| | GP-PadeA-7xA-HYDS P4 | **GCAGCGGCAGCAGCGGCAGCG**CATTACGATAGC TAGCTCGAGGCTTAATTAACCTAGG |
| pET48b-*padeA*(3xA-S) | GP-PadeA P1 | 3xA-SCACGTCTTCATCCGGGTCCGGCTGGAAAGC |
| | GP-PadeA P2 | 3xA-SGCTTAATTAACCTAGGCTGCTAAACAAAGCCCGAA AGG |
| | GP-PadeA P3 | 3xA-S**CTCGAGCTAGCTTGCTGCCGC**CAC GTC TTC ATC CGG GTC CGG CTG GAA AGC |
| | GP-PadeA P4 | 3xA-S**GCGGCAGCAAGCTAG CTC GAG** GCT TAA TTA ACC TAG GCT GCT AAA CAA AG |

## 2.4 Vector

## 2.4.1 pET41a(+)

The expression vector pET41a(+) was used for the production of recombinant proteins in *E. coli.* It was also used for heterologous expression of the lasso peptide gene cluster. This vector has an N-terminal GST (glutathione-S-transferase) tag, which allows purification of target protein by glutathione affinity chromatography and may increase the solubility of the target protein. After the GST tag, there are a His tag, thrombin cutting site, S • Tag and aenterokinase cutting site. The vector also contains a C terminal His tag. The cloned gene is under the control of lac operator and a T7 promoter. In *E. coli* BL21 which has the T7 RNA polymerase,



**Figure 2.1:** Physical map of pET41a(+).

the expression of the gene can be induced with the addition of IPTG or lactose. The vector also contains a kanamycin resistance gene Kan[R] which was used as a selection marker. This vector

was used to clone the whole paeninodin lasso peptide gene cluster, the PadeK and ThcoK Kinases. *Nde*I was chosed as a cutting site, so only a His tag was added at the C terminal of the recombinant protein.

## 2.4.2 pACYDuet-1

The pACYDuet-1 vector was used for co-expression of PadeK kinase with the lasso peptide gene cluster without kinase. It has two multiple cloning sites (MCS), which are both under the control of lac operator and T7 promoter. The transcription is dependent on the T7 polymerase and is inducible by IPTG or lactose. The pACYDuet-1 has a different origins of replication site and a different resistance marker compared with pET41a(+). The vector also contains a chloramphenicol resistance gene $Cm^R$ which was used as a selection marker. Thus this vector could be co-expression with pET41a(+). *E. coli* BL21 (DE3) was used as a host for expression.
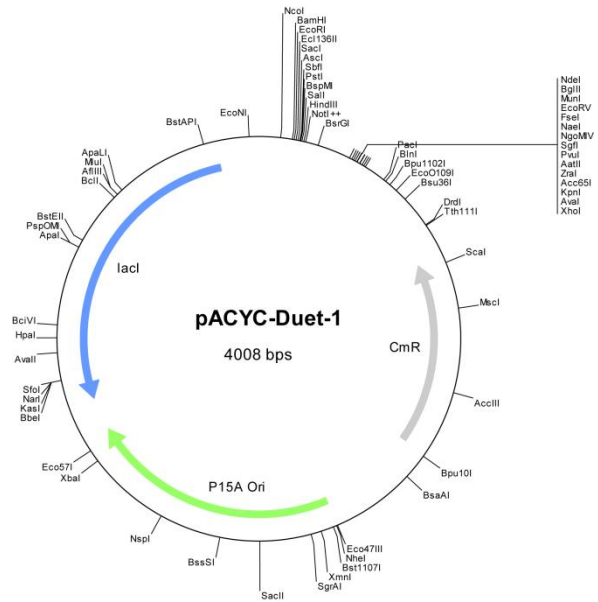


**Figure 2.2:** Physical map of pACYC-Duet-1 vector.

## 2.4.3 pET-48b(+)

pET-48b(+) was used for heterologous expression of all kinds of precursor peptide. The vector has a thioredoxin fusion tag at the N terminal. A tev protease cutting site was inserted between the thioredoxin fusion tag and the target protein. Thioredoxin fusion tag was commonly used to promote the solubility of its fusion partners. Particularly it is suitable for expression of small protein or peptide. The fusion tag could protect the partner from degradation and facilitate the purification by Ni-NTA affinity chromatography. The vector contains a kanamycin resistance gene $Kan^R$ which was used as a selection marker. This vector was used for
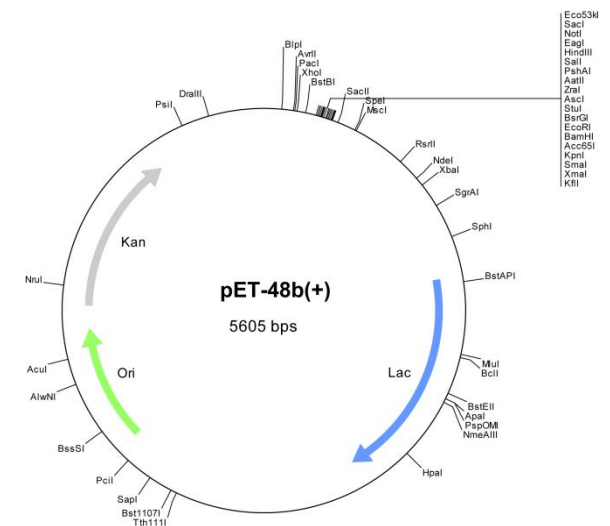


**Figuer 2.3:** Physical map of pET-48b(+) vector

33

expression of the precursor peptide PadeA and its analogues. *E. coli* BL21 (DE3) was used as a host strain.

### 2.4.4 pETMBP-1a

pETMBP-1a vector was used for expression of PadeK kinase. The vector has MBP (Maltose binding protein) fusion tag at the N terminal. MBP fusion tag like thioredoxin fusion tag was commonly used to promote the solubility of its fusion partners.

The MBP tag has a 6x his tag at N terminal, thus could be purified by Ni-NTA affinity chromatography. Besides, the MBP tag itself could also be served as an affinity tag for purification by means of a maltose matrix. Between the MBP fusion tag and the target protein, there is a Tev protease cutting site which could be used to remove the tag. The tev protease cutting site was changed to Hrv-3C protease cutting site in

**Figure 2.4:** Physical map of pETMBP-1a vector.

some cases by mutagenesis. The vector contains a kanamycin resistance gene Kan$^R$ which was used as a selection marker. This vector was used for expression of the PadeK kinase. *E. coli* BL21 (DE3) was used as a host strain.

## 2.5 Bacterial strains

### 2.5.1 *Paenibacillus dendritiformis* C454

*Paenibacillus dendritiformis* is a species of pattern-forming bacteria, which was first discovered in 90's by Ben-Jacob's group. It is a social microorganism that could change its morphotype under different environments. Moreover, the strain could produce self-toxins to inhibit sibling colonies. Bioinformatics studies showed that the strain contains diverse natural products gene clusters like NRPs, PKs and RiPPs. In this work it was cultivated for extracting chromosomal DNA and homologous fermentation for lasso peptide production.

### 2.5.2 *Thermobacillus composti* KWC4

*Thermobacillus composti* KWC4 is a moderately thermophilic bacterium isolated from a composting reactor. It could grow aerobically at 32-61 °C, with optimal growing occurring at 50 °C. The strain is moderately halotolerant and is able to grow at NaCl concentrations up to 4.4% (w/v). The DNA G+C content of this strain is about 60%. This strain contains a lasso peptide gene cluster that shows high similarity with the one from *P. dendritiformis* C454. Therefore, the

kinase from this strain was used to replace the insoluble kinase from the paeninodin gene cluster.

### 2.5.4 *Escherichia coli* TOP10

The *E. coli* TOP10 stored by our lab was used as a general host for cloning and plasmid preparation. The genotype of this strain is: F-*mcrA*Δ(*mrr-hsdRMS-mcrBC*) φ80*lacZ*ΔM15 Δ*lac*X74 *recA*1 *ara*D139 Δ(*ara-leu*)7697 *galU galKrpsL*(Str^R) *endA*1 *nupG*.

### 2.5.3 *Escherichia coli* BL21 (DE3)

The *E. coli* BL21(DE3), purchased from Invitrogen, was used as a general host for the heterologous expression of T7 promoter-driven vectors. The stain contains the T7 RNA polymerase gene which is controlled by the lacUV5 promoter. The T7 RNA polymerase is expressed upon induction by IPTG which induces high level expression of target protein from T7 promoter driven expression vectors.This strain is a derivative of *E. coli* B strain and lacks both the lon protease and the ompT membrane protease which may degrade expressed proteins. The genotype of this strain is: F-*ompT gal dcm hsdS*B(r~B~-m~B~-) λ(DE3).

## 2.6 Culture media

Culture media listed below were used for cultivation or fermentation of the bacterial strain. Solid agar plates in some cases were prepared by adding 1.5% (m/w) agar. All media were sterilized by autoclavation (121°C, 20 min). Antibiotics and other additional components were added after sterile-filtration.

### 2.6.1 LB-Medium

The lysogeny-broth (LB) medium was used for cultivation of *E. coli* species in this work. An exact composition is given in Table 2.10.

**Table 2.10:** Composition of LB medium used. The pH was adjusted to pH 7.0.

| Component | Concentration |
|---|---|
| yeast extract | 5 g/L |
| tryptone | 10 g/L |
| NaCl | 5 g/L |

## 2.6.2 CASO Medium

The CASO medium was used for cultivation of *P. dendritiformis*. The DSMZ number of this medium is 220. An exact composition is given in Table 2.11.

**Table 2.11:** Composition of CASO medium used. The pH was adjusted to pH 7.3.

| Component | Concentration |
| --- | --- |
| Peptone from casein | 15 g/L |
| Peptone from soymeal | 5 g/L |
| NaCl | 5 g/L |

## 2.6.3 Sulfolobus Medium

The Sulfolobus medium was used for cultivation of *Thermobacillus composti* KWC4. The DSMZ number of this medium is 88. An exact composition is given in Table2.12.

**Table 2.12:** Composition of Sulfolobus medium used. The pH was adjusted to pH 7.0.

| Component | Concentration |
| --- | --- |
| $(NH_4)_2SO_4$ | 1.3 g/L |
| $KH_2PO_4$ | 0.28 g/L |
| $MgSO_4$ x 7 $H_2O$ | 0.25 g/L |
| $CaCl_2$ x 2 $H_2O$ | 0.07 g/L |
| $FeCl_3$ x 6 $H_2O$ | 0.02 g/L |
| Allen's trace element solution (see below) | 10ml/L |
| Yeast extract | 1.0 g/L |

**Table 2.13:** Composition of Allen's trace element solution used. The pH was adjusted to pH 7.0.

| Component | Concentration |
| --- | --- |
| $MnCl_2$ x 4 $H_2O$ | 180mg/L |
| $Na_2B_4O_7$ x 10 $H_2O$ | 450mg/L |
| $ZnSO_4$ x 7 $H_2O$ | 22mg/L |

| | |
|---|---|
| $CuCl_2$ x 2 $H_2O$ | 5mg/L |
| $Na_2MoO_4$ x 2 $H_2O$ | 3mg/L |
| $VOSO_4$ x 2 $H_2O$ | 3mg/L |
| $CoSO_4$ | 1mg/L |

## 2.6.4 M9 Medium

M9 medium was used for the heterologous expression of lasso peptide gene cluster in *E.* coli as a standard medium of the heterologous peptide Lasso fermentations. The asterisked sterile stock solutions were added after autoclaving

**Table 2.14:** Composition of M9 medium used. The pH was adjusted to pH 7.0.

| Component | Concentration |
|---|---|
| $Na_2HPO_4 \cdot 12\ H_2O$ | 17.1 g/L |
| $KH_2PO4$ | 3.0 g/L |
| $NH_4Cl$ | 1.0 g/L |
| NaCl | 0.5 g/L |
| $MgSO_4$(2 M) | 1.0 ml/L |
| $CaCl_2$(0.5 M) | 0.2 ml/L |
| Glucose solution (40% w / v) * | 10.0 ml/L |
| Vitamin Mix (500x)* | 0.2 ml/L |

The 500x M9-vitamin mix was prepared from the following list. 10 M NaOH was added until all components were dissolved. The solution was then filter sterilized and stored until use at 4 °C.

**Table 2.15:** Composition of Vitamin Mix used. The pH was adjusted to pH 7.0.

| Component | Concentration |
|---|---|
| choline chloride | 1.0 g |
| folic acid | 1.0 g |
| pantothenic acid | 1.0 g |
| nicotinamide | 1.0 g |

| | |
|---|---|
| myo-inositol | 2.0 g |
| pyridoxal hydrochloride | 1.0 g |
| thiamine | 1.0 g |
| riboflavin | 0.1 g |
| disodium adenosine 5'-triphosphate | 0.3 g |
| biotin | 0.2 g |
| | ad 300 mL ddH$_2$O |

# 3. Methods

## 3.1 Bioinformatic Methods

### 3.1.1 Genome mining of new lasso gene cluster by PSI-BIAST

The biosynthesis of Mccj25 has been well studied in the past[121-123]. The biosynthetic cluster for Mccj25 contains four genes: mcjA encodes the precursor peptide; mcjB, an ATP-dependent cysteine protease; mcjC, an ATP-dependent asparagine synthetase homolog; and mcjD, an ABC transporter[122]. Among these four genes, mcjB homologs are found almost exclusively in lasso peptide biosynthetic gene cluster, thus we used the B protein for genome mining studies. The sequence of mcj25 (accession number YP_007443297.1) was downloaded from NCBI (National Center for Biotechnology Information) and was applied in the position-specific iterated basic local alignment search tool (PSI-BLAST) to start a homology search. The results of the first round was then chosed as starting points for a further search run based on a position-specific scoring matrix (PSSM). The following parameters were applied: maximum target sequences of 500, expect threshold of 10 and PSI-BLAST threshold of 0.005. BLOSUM62 algorithm was used for judgment of amino acid similarity. With these settings, the PSI-BLAST was repeated 4 times and provided the maximum number of results[112].

All the identified B proteins were checked individually. About 10 Kb was downloaded from NCBI and surrounding genes were analyzed for potential lasso peptide biosynthetic enzymes. Potential ORFs were created and directly analyzed by the basic protein BLAST. The genes which could be identified as an asparagine synthetase homolog were named as C protein. The genes which could be identified as an ABC-transport were named as D protein. It is very often to observe that a small protein (80-110 aa) is located beside the B protein which shows homology to PqqD protein. We classified it as a B1 protein. The B protein in a B1 containing lasso peptide gene cluster is thus renamed as a B2 protein. After the essential B and C protein were assigned for lasso peptide gene cluster, small proteins of 40 to 70 aa length were tested for their suitability as a possible precursor peptide. Generally, lasso peptides carry a Gly, Ala, Cys, or Ser residue at position 1 and a ring-forming Glu or Asp residue at positions 7 to 9. A conserved Thr at the penultimate position of the leader peptide was also highly conserved. Based on these standards, the samll ORFs were checked by hand to see if they are suitable for a precursor peptide.

### 3.1.2 Multiple sequence alignment

Multiple sequence alignment is a useful tool to identify conserved and variable regions in homologous proteins. These results could be used to further map the relationship of protein

39

sequences by generating phylogenetic tree and identify the evolution history of the target protein. The multiple sequence alignments shown in this work were created by several different tools. For generating phylogenetic tree of the kinases, ClustalX2 was used to do the multiple sequence alignment. For multiple sequence alignment of the precursor peptides, CLC Main Work bench was used.

### 3.1.3 Identification of conserved motifs with MEME algorithm

The kinase involved in the lasso peptide biosynthetic pathway is the first time reported and represent a new family of serine kinase, thus it is necessary to search the conserved motifs of the protein and gain some information about the catalytic mechanism. The Multiple EM for Motif Elicitation (MEME:http://meme-suite.org/tools/meme) was used to identify the conserved motits[130]. The parameters were set as follows. For the expected occurrence of motifs per sequence is selected as 5. The minimum size of the conserved motifs is set to 6-50 aa.

### 3.1.4 Phylogenetic tree analysis

Protein sequences were retrieved from the gene clusters and aligned by CLUSTAL X2 with standard settings as implemented in Mega 6. Phylogenetic trees were then constructed using the minimum evolution and neighbor-joining method, and the Poisson-correction amino acid model. Phylogeny was tested by bootstrapping with 1,000 replications. Phylogenetic trees were constructed in this manner for both the C proteins and the kinases.

## 3.2 Molecular biology techniques

### 3.2.1 General strains maintenance

*E. coli* Top 10 and *E. coli* BL21(DE3) were mantained on LB-agar plates containing corresponding antibiotic and incubated at 37 °C. *Paenibacillus dendritiformis* C454 was maintained on CASO-agar plates at 30 °C. *Thermobacillus composti* KWC4 was maintained on Sulfolobus-agar plates at 50 °C. Liquid cultures were incubated under continuous shaking at 200 rpm. Different antibiotics were added if necessary. For heterologous expression of the lasso peptide, M9 medium was used for fermentation. For long-term storage, 20% (v/v) sterile glycerol was added to a liquid culture and the resulting stocks were flash-frozen in liquid nitrogen and stored at -80 °C until further use.

### 3.2.2 Preparation of genomic DNA from *P. dendritiformis* and *T. composti* KWC 4

Preparation of genomic DNA from *P. dendritiformis* and *T. composti* were done by a FastDNA® SPIN Kit. 5 mL of liquid culture were harvested by centrifugation. The pellets were resuspended

in 200 μL ddH$_2$O. 1 mL CLS-TC lysis solution was then added. The mixture was then homogenized in the FastPrep® Instrument for 40 seconds at a speed setting of 6.0 and centrifuged at 14000 x g for 5-10 minutes. The supernatant was then transferred to a 2.0 mL microcentrifuge tube and added an equal volume of Binding Matrix. The mixture was incubated with gentle agitation for 5 minutes at room temperature on a rotator and half of the suspension was transferred to a SPIN™ filter and centrifuge at 14000 x g for 1 minute. Then add the remaining suspension and centrifuge as before. 500μL prepared SEWS-M was added and the pellet was gently resuspended using the force of the liquid from the pipet tip. Then centrifuge at 14000 x g for 1 minute and discard contents of catch tube. The tube was centrifuged a second time at 14000 x g for 2 minutes to remove left ethanol and replace the catch tube with a new tube. The DNA was eluted by gentally resuspending the binding matrix above the spin filter in 100 μL of DES and incubated for 5 minutes at 55 °C in a heat block. Centrifuge at 14000 x g for 1 minutes to bring eluted DNA into the catch tube. The DNA was stored at -20 °C until further use.

### 3.2.3 PCR-based gene amplification and Vector construction

Generally, target genes were amplified from genomic DNA by PCR (polymerase chain reaction) using the primers listed table 2.3-2.9. The Phusion High-Fidelity DNA Polymerase was used for all PCR reactions. HF buffer and GC buffer combined with different concentration of DMSO (2.5% or 5%) were used to screen optimum PCR conditions according to the manufacturer's manual. PCR products were purified from agarose gel slices with the QIAquick Gel Extraction Kit (QIAGEN).

For vector construction, two different methods were used. One method is that the PCR products and the plasmid are digested with restriction endonucleases and ligated with T4 DNA ligase. Another method is that the vector and insert DNA are ligated with *Gibson Assembly Master Mix* according to the manufacturer's manual. Ligation products were dialyzed with a MF-Millipore Membrane Filter and then transformed into *E. coli* TOP10 cells. Positive transformants were selected on LB-agar plates and used for plasmid preparation. The right constructs were verified by sequencing (GATC Biotech).

### 3.2.4 Preparation of electrocompetent *E. coli* cells

Both the *E. coli* Top 10 and *E. coli* BL21(DE3) electrocompetent cells were prepared by the following protocol:

— 300 ml of LB medium was inoculated with 3 ml of an *E. coli* overnight culture

—incubate culture to an OD600 0.5-0.6 (37 °C, 200 rpm)

—culture was transferred to 6 Falcon tubes (50 ml each) and centrifuge at 4000 rpm (4 °C，20min)

—Pellets were resuspended twice with 40 ml of ice-cold, sterile ddH$_2$O and then centrifuged again (approx 25 min)

— Pellets were resuspended in 1 ml of ice-cold, sterile ddH$_2$O and transferred into 2 ml Eppendorf tubes, and centrifuge at 13,000 rpm (4 °C, 5 min)

—Pellets were resuspended in 500 µl 10% glycerol and then centrifuge (13000 rpm, 4 °C, 5 min).

—Pellets were resuspended in 200 µl 10% glycerol and a 40 µl in 1.5 mL Eppendorf tubes in aliquots, flash-frozen in liquid nitrogen and stored at -80 °C until further use.

### 3.2.5 Preparation of plasmid DNA from *E. coli*

A single colony of *E. coli* carrying the right plasmid was inoculated into 5 mL fresh LB medium (with matching antibiotic) and incubated overnight at 37 °C (200 rpm). All the culture was centrifuged in a 2 mL Eppendorf reaction tube and wait for further use.

#### 3.2.5.1 *Quick and Dirty* Plasmid Preparation

For the quick and dirty method, all the pellets were resuspended in 300 µl Buffer P1 (50 mM Tris·HCl, 10 mM EDTA, 100 µg/L Rnase A, pH 8.0). Then 300 µl Buffer P2 (200 mM NaOH, 1%(w/v) SDS) was added and mixed by gently inverting the tube. The lysis was performed at room temperature for 5 minutes. After that, 300 µl Buffer P3 (2.55 M KOAc, pH 4.8) was added and mixed by gently inverting the tube again. The mixture was incubated on ice for 10 minutes and centrifuged for 10 minutes (13000 rpm, 4 °C). The supernatant was transferred to a clean 1.5 mL Eppendorf reaction tube and mixed with 600 µl isopropanol. The components were mixed by briefly inverting the reaction tube and centrifuged again (13000 rpm, 30 min, 4°C). After centrifugation, the plasmid DNA will precipitate in the pellet. The supernatant was discarded and the pellet was washed with 70% aq. ethanol and centrifuged again (13000 rpm, 10min, 4 °C). The supernatant was discarded and the pellet was dried at 37 °C. The pellet was then dissolved in 100 µl ddH$_2$O and the concentration was determined by Nanodrop. The solution was then stored at -20 °C until further use.

#### 3.2.5.1 GenElute$^{TM}$ HP Plasmid Miniprep Kit

The HP GenElute Plasmid Miniprep Kit (Sigma-Aldrich) was used to prepare high purity plasmid DNA for sequencing. By this method, the cell pellet was processed according to the

manufacturer's instructions. The concentration was measured by Nanodrop and the DNA solution was stored at -20 °C until further use.

## 3.2.6 SLIM-Mutagenesis

Mutations were introduced in a DNA sequence by the site-directed ligase independently Mutagenesis (SLIM). All the primers used were listed in Table 2.3-2.9. All the PCR reactions were performed using the Phusion Hi-Fidelity DNA polymeraseaccording to the manufacturer's manual.
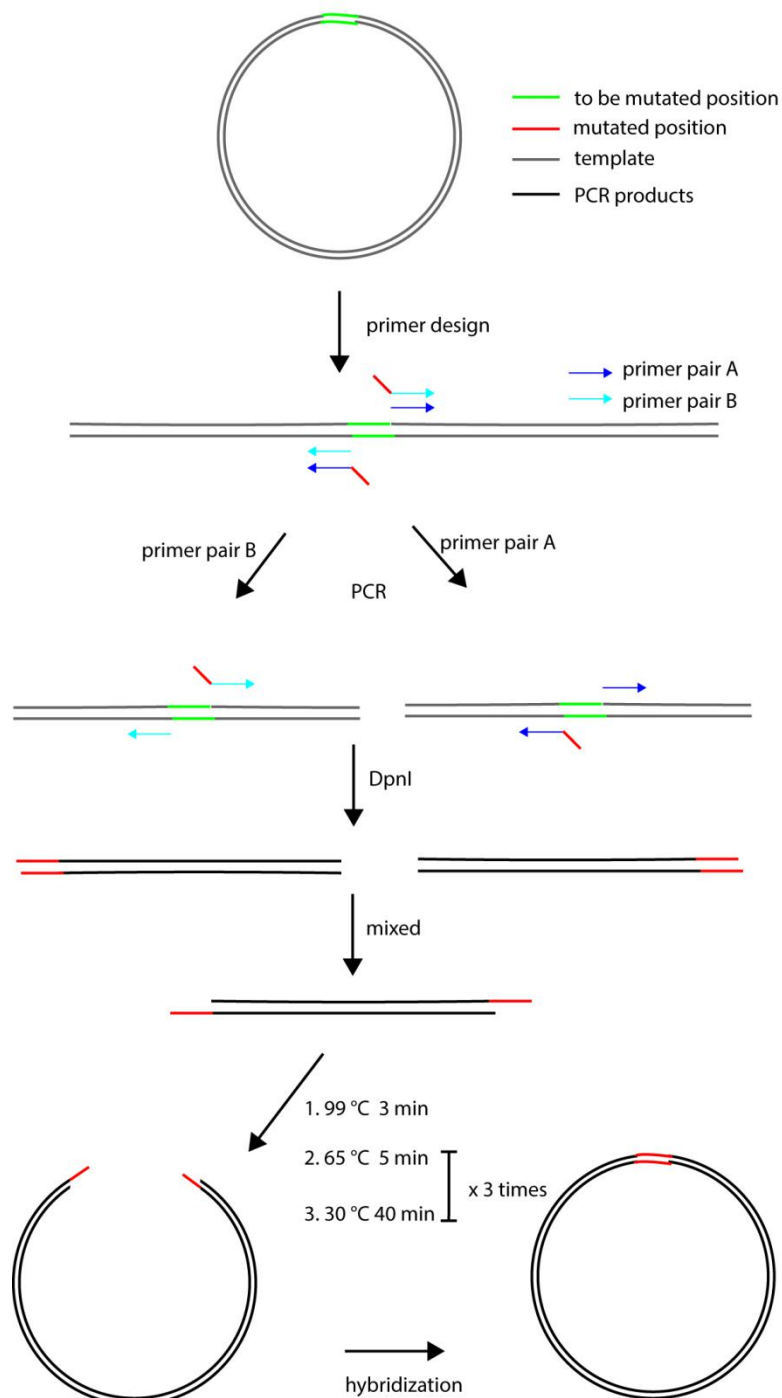
**Figure 3.1**: Schematic of the SLIM method. (A) For the SLIM method two different primer pairs (A and B) must be used in two different PCRs. (B) products of both PCRs and schematic representation of the subsequent hybridization reaction. The resulting plasmid, double-strand break can be transformed directly.

SLIM Mutagenesis, as the name implies, is a method that no ligation is needed for a mutagenesis reaction[131]. This mehod is generally better than the round-the-horn site-directed mutagenesis with high yield of correct construct. But one disadvantage of this method is that two pairs of primers are needed in one mutation. The two forward and two reverse primers each have the same annealing sequence, but differ in their 5 'overhang. The reverse and forward primer carring a complementary 5 'overhang have at least 20 base pairs. To generate the mutation, two PCR reactions were performed. The PCR products were treated with the restriction enzyme *Dpn*I according to the manufacturer (New England Biolabs) in order to remove the methylated template plasmid DNA. The reaction mixture was then incubated at 80 °C for 20 minutes to inactive the *Dpn*I enzyme. Subsequently, the so-called hybridization reaction was performed. 10 µl of each PCR products were mixed and treated under the following PCR conditions: 3 min 99 °C, 5 min 65°C, 40 min 30°C; the last two steps were repeated three times. After the hybridization step, 10 µl of the reaction solution was dialyzed with a MF-Millipore Membrane Filter and then transformed into *E. coli* TOP10 cells. Positive transformants were selected on LB-agar plates and used for plasmid preparation. The right constructs were verified by sequencing (GATC Biotech).

### 3.2.7 Preparation of expression constructs

All the right constructs verified by sequencing (GATC Biotech) were then retransformed into electrocompetent *E. coli* BL21 (DE3) cells. Positive transformants were selected on LB-agar plates with matching antibiotic. A single colony was picked for further expression.

## 3.3 Protein Chemical Method

### 3.3.1 Fermentation of *P. dendritiformis* C454 to produce paeninodin

For homologous fermentation of *P. dendritiformis* C 454, CASO Medium and M9 medium were used. A total of 500 mL of each medium was inoculated with 5 mL of *P. dendritiformis* C454 overnight cultures and cultivated for 5 days at 30 °C.

### 3.3.2 Heterologous production of paeninodin in *E. coli*

For heterologous production of Paeninodin in *E. coli,* 600 mL M9 medium was used to screen for the best fermentation conditions. pET41a-*padeCAKB1B2D* containing the lasso peptide gene cluster was first retransformed into electrocompetent *E. coli* BL21 (DE3) cells. Positive

transformants were selected on LB-agar plates with 50 µg/mL kanamycin. A single colony was picked and 60 mL LB medium supplemented with 50 µg/mL kanamycin was inoculated with this colony and grown overnight at 37 °C with shaking. From these cultures, 2 L baffled culture flasks, containing 600 mL of M9 fermentation media each, were inoculated to $OD_{600}\approx$ 0.01. For fermentation at 37 °C, IPTG (final concentration of 0.05 mM) was used to induce the expression at $OD_{600}\approx$ 0.5-0.6. For fermentation at 20 °C, cells were first grown at 37 °C to $OD_{600}\approx$0.4-0.5 in about 4.5 h, then slowly cooled down to 20 °C for another 1 h. At $OD_{600}\approx$ 0.5-0.6, the cells were induced with IPTG (final concentration of 0.05 mM) and cultivated at 20 °C for 3 days.

### 3.3.3 Protein expression

#### 3.3.3.1 Expression of PadeK kinase with pETMBPa-1 Vector

The expression of the PadeK kinase was facilitated by a maltose binding protein (MBP) phusion tag. pETMBP-padeK was retransformed into electrocompetent *E. coli* BL21 (DE3) cells. Positive transformants were selected on LB-agar plates with 50 µg/mL kanamycin. A single colony was picked to make an overnight culture. Then 500 mL of LB medium in a 2 L baffled culture flasks were inoculated to $OD_{600}\approx$ 0.01. The cells were grown at 37 °C to $OD_{600}\approx$ 0.4 in about 2 h. The cells were then cooled down to 16 °C for another 1 h and induced with 0.05 mM IPTG for 16 h. Cells were harvested by centrifugation (8000 rpm, 15 min, 4 °C), resuspended in 30 mL HEPES buffer A (50 mM HEPES, 300 mMNaCl, 30 mM imidazole, 5% glycerol; pH 8.0) and stored at -80 °C until further use.

#### 3.3.3.2 Expression of ThcoK kinase with pET41a(+) Vector

For expression of the ThcoK kinase, pET41a(+) vector was used and A $His_6$-tag was introduced at the C-terminus of the kinase to facilitate the purification. Large-scale protein expression was carried out in 10 times 600 mL LB medium. Initially, 60 mL LB medium (with 50 µg/mL kanamycin) was inoculated with *E. coli* BL21(DE3) cells carrying the pET41-thcoK vector and grown overnight at 37 °C with shaking. From this culture, 2 L baffled culture flasks containing each 600 mL of LB medium was inoculated to $OD_{600}\approx$ 0.01. The cells were then grown at 37 °C until the OD reach to 0.4 in about 2 h. The cells were then cooled down to 16 °C for another 1 h and induced with IPTG (0.05 mM) for 16 h. Cells were harvested by centrifugation (8000 rpm, 15 min, 4 °C), resuspended in 30 mL HEPES buffer A (50 mM HEPES, 300 mMNaCl, 30 mM imidazole, 5% glycerol; pH 8.0) and stored at -80 °C until further processing.

#### 3.3.3.3 Expression of PadeA as thioredoxin fusion protein with pET-48b (+) vector

For the expression of the precursor peptide PadeA, a thioredoxin (Trx) fusion tag was used to facilitate purification[88]. The plasmid pET48b-padeA was retransformed into electrocompetent *E.*

45

*coli* BL21 (DE3) cells. The resulting *E. coli* BL21 (DE3) with pET48b-padeA strain was then grown in LB medium containing kanamycin (50 µg/mL) at 37 °C with shaking for overnight. Heterologous expression of the Trx-PadeA protein was carried out with the same protocol as used for the kinases.

### 3.3.3.4 Expression of the Hrv3C protease

For the expression of the Hrv3C protease, a Glutathione S-transferase (GST) fusion tag was used to facilitate purification. GST-tagged Hrv3C protease thus can be purified based on the ability of GST to bind its substrate, glutathione (GSH).The Matrix (GSH Sepharose 4 Fast Flow) of the column used (GSTPrep FF 16/10 20 mL (GE Healthcare)) contains covalently bound, immobilized Glutathione which was used to bind the GST fusion proteins. After binding to the column, the GST-fusion proteins were eluted from the column by using reduced glutathione. Heterologous expression of the Hrv3C protein was carried out with the same protocol as used for the kinases. Cells were harvested by centrifugation (8000 rpm, 15 min, 4 °C), resuspended in 30 mL Lysis buffer A (50 mM Tris, 150 mMNaCl; pH 7.5) and stored at -80 °C until further processing.

### 3.3.4 Cell disruption

For cell disruption, a French Press (SLM Aminco) was used. The components of the system were put on ice to cold down. Meanwhile, the frozen cell suspension was thawed at room temperature. The cells were then pressed through the Standard Cell (Thermo) and lysed 2-3 times. After that, the mixture was centrifuged (30min, 170000 rpm, 4 °C) to remove the insoluble cell components. The supernatant was transferred to another clean tube and filtered for further purification.

### 3.3.5 Protein purification

### 3.3.5.1 Purification of the PadeK kinase

The initially attempt to get soluble PadeK kinase was failed. Therefore, Maltose-binding protein (MBP)-tag was used to increase the solubility of the PadeK kinase. Meanwhile, there is a 6x His tag at the N terminal of the MBP tag which could facilitate the purification of the kinase. After the purification, the MBP fusion tag could be removed by Hrv 3C protease. Thus, purification of PadeK kinase took place in four stages. First, a Ni-NTA affinity chromatography was performed and then a second size exclusion chromatography was performed using a Superdex 200 16/60 prep grade column. The pure proteins were then treated with Hrv3C protease. In the end, the Hrv3C protease and the MBP phusion tag were removed by revise Ni-NTA affinity chromatography. Pure padeK kinases were then obtained.

Ni-NTA affinity chromatography was performed using an ÄKTA purifier system (Amersham Pharmacia Biotech). A Ni-NTA column was equilibrated with buffer Hepes A (50 mM hepes, 300mM NaCl, 30 mM imidazole and 5% glycerol, pH 8.0) for 15 min with a flow rate of 1 mL/ min. Then the supernatant of the cell lysate (section 3.3.4) was applied to the column with a flow rate of 1 mL/ min. The column was continued wash with 5 % buffer Hepes A2 (50 mM hepes, 300mM NaCl, 250 mM imidazole and 5% glycerol, pH 8.0)  for 10 min. Proteins were then eluted using a linear gradient from 5 to 100% Hepes B buffer in 15 min at a flow rate of 1 mL/ min. The fraction size was 2 mL each. Purified proteins were identified by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), combined, concentrated and buffer exchanged into Hepes buffer B (10 mM hepes, 150 mM NaCl, and 5% glycerol, pH 7.5) using Amicon Ultra filters. Then a second size exclusion chromatography was performed using the same FPLC system equipped with a Superdex 200 16/60 prep grade column equilibrated with Hepe buffer B. The concentrated protein solution was then supplemented with Hrv-3C protease (final concentration 0.1 mg/mL) to carry out the cleavage reaction at 4 °C for 16 h. Final purification of PadeK was achieved by a reverse Ni-NTA affinity chromatography. The Ni-NTA column was equilibrated with Hepes buffer A and after the sample had been applied, the pass through proteins were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), combined and concentrated. Pure protein fractions were flash-frozen in liquid nitrogen and stored at -80 °C until further usage.

### 3.3.5.2 Purification of the ThcoK kinase

When we turned to the ThcoK kinase, we were pleasantly surprised to note that the ThcoK kinase is extremely soluble. Therefore, a C terminal 6x his tag was added at the C terminus and used to facilitate the purification. Thus for the purification of ThcoK kinase, a Ni-NTA affinity chromatography was performed first and then a second size exclusion chromatography was performed to obtain pure ThcoK kinase.

Ni-NTA affinity chromatography was performed using an ÄKTA purifier system (Amersham Pharmacia Biotech). The Ni-NTA column was equilibrated with Hepes buffer A and after the sample had been applied (flow rate 1 mL/min), proteins were eluted with increasing concentrations of imidazole in Hepes buffer A. Purified proteins were identified by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), combined, concentrated and buffer exchanged into Hepes buffer B (10 mM hepes, 150 mM NaCl, and 5% glycerol, pH 7.5) using Amicon Ultra filters. Then a second size exclusion chromatography was performed using the same FPLC system equipped with a Superdex 200 16/60 prep grade column equilibrated with Hepe buffer B. Pure protein fractions were flash-frozen in liquid nitrogen and stored at -80 °C until further usage. Notably, Even though the ThcoK kinase is super soluble, it tends to

form aggregates with itself when temperature changed drastically. We handled this enzyme at room temperature consistently.

### 3.3.5.3 Purification of the Trx-PadeA Phusion Protein

Thioredoxin fusion tag was used to facilitate the purification of PadeA precursor peptide. Between the thioredoxin fusion tag and the PadeA peptide, there is a 6xHis tag which is used to facilitate the purification and a Hrv 3C protease cutting site which is used to remove the tag by treating with Hrv3C protease. Thus for the purification of Trx-PadeA Phusion protein, two steps including a Ni-NTA affinity chromatography and a size exclusion chromatography were applied. Generally, the Trx-PadeA phusion protein was purified using the same method as the ThcoK kinase. The concentrated protein solution was then supplemented with Hrv-3C protease (final concentration 0.1 mg/mL) to carry out the cleavage reaction at 4 °C for 16 h. Final purification of PadeA was achieved by apreparative HPLC.

### 3.3.5.4 Purification of the Hrv 3C protease

Glutathione S-transferase (GST) is an enzyme found in most organisms. It consists of 220 amino acid residues and could be used as a fusion tag, making the tagged protein more soluble as well as simplifying its purification and detection. The purification of GST-tagged protein is based on the high affinity of GST for glutathione (GSH). When applied to the affinity medium, GST-tagged proteins bind to the glutathione ligand, and impurities are removed by washing with binding buffer. Tagged proteins are then eluted from the chromatography medium by using the solublereduced glutathione. The purity of the protein after purification is over 90% due to the high specificity of GSH to GST.

Table3.1: Composition of Buffer used for purification of Hrv3C protease

| lysis buffer | Wash buffer | Elution buffer | Storage buffer |
|---|---|---|---|
| 50 mM Tris pH=8.0 | 50 mM Tris pH=8.0 | 50 mM Tris pH=8.0 | 50 mM Tris pH=7.5 |
| 1 M NaCl | 1 M NaCl | 1 M NaCl | 150 M NaCl |
| 1 mM EDTA | 1 mM EDTA | 1 mM EDTA | 10 mM EDTA |
| 1 mM DTT | 1 mM DTT | 1 mM DTT | 1 mM DTT |
| 10% protease inhibitor | | 10 mM reduced glutathione | 20% glycerol |
| Benzonase | | | |

For the purification of the Hrv-3C protease with a GST-tag, The GSTPrep FF16/10 column was equilibrated with wash buffer for 15 min with a flow rate of 1 mL/ min (Table 3.1). Then the cell lysate was applied to the column with a flow rate of 1 mL/ min. The column was continued wash with wash buffer for 10 min. Proteins were then eluted using elution buffer at a flow rate of 1 mL/ min. The fraction size was 2 mL each. Purified proteins were identified by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), combinedand buffer exchanged into storage buffer. The protein was concentrated to 10-20 mg/mL and stored at -80 °C for further use.

### 3.3.6 Lasso peptide precursor peptide purification

After the Trx tag was removed, the protease added two GP residues to the PadeA resulting in a GP-PadeA peptide. Final purification of GP-PadeA wasthen achieved by two steps HPLC purification. For this purpose, a preparative HPLC (microbore 1100 HPLC system (Agilent) with a VP 250/21 Nucleodur C18 Htec 5 μm column (Macherey-Nagel) at room temperature with a flow rate of 18 mL/min) was used. For the first round, the following gradient of water/0.1% trifluoroacetic acid (solvent A) and MeCN/0.1% trifluoroacetic acid (solvent B) was used: Linear increase from 10% to 50% B in 30 min, followed by a linear increase from 50% to 95% B in 2 min and holding 95% B for an additional 3 min. GP-PadeA containing fractions were then evaporated at 40 °C and reduced pressure. The semipure precursor peptides were then combined to 10 mL and subsequently applied to a second round of purification using the following gradient of water/0.1% trifluoroacetic acid (solvent A) and MeCN/0.1% trifluoroacetic acid (solvent B): Linear increase from 20% to 50% B in 30 min, followed by a linear increase from 50% to 95% D in 2 min and holding 95% B for additional 3 min. GP-PadeA containing fractions were collected, confirmed by LC-MS and lyophilized.

## 3.4 Natural Products Isolation

### 3.4.1 Extraction of the cell pellets with Methanol

After the fermentation from section 3.3.2, the cell pellets were resuspended with 1 mL ddH$_2$O by vortex, and then extracted with 60 mL of MeOH for overnight. The mixture was then centrifuged (6,000 rpm for 20 min at 4 °C) and the pellet were discarded. The extract supernatant was then filtered and solvent was removed via evaporation at 40 °C and reduced pressure. Dry pellet extracts were then resuspended in a total of 900 μL of 50% MeOH, cleared by centrifugation (13,000 rpm for 60 min at 4 °C) and analyzed via LC-MS.

### 3.4.2 Extraction of culture supernatants by means of Amberlite XAD 16

On the other hand, culture supernatants were extracted by using XAD16 resin (Sigma Aldrich, $\approx$ 7 g/L culture supernatant). The XAD16 containing supernatant was stirred for 1h and was then removed by filtration. After washed several times with ddH$_2$O, the resin on the filter was eluted with a total of 60 mL of MeOH. Solvent was then removed via evaporation and reduced pressure. The dry pelltet were then resuspended in a total of 900 μL of 50% MeOH, cleared by centrifugation (13,000 rpm for 60 min at 4 °C) and analyzed via LCMS.

### 3.4.3 Lasso Peptide Purification

Large scale fermentation was performed to isolate the paeninodin and phosphorylated paeninodin lasso peptide. For each peptide, 10 times 10 L of M9 minimal medium was fermented at 37 °C for 3 days. Since the ABC transporter is inactive in *E. coli*, only the pellets were extracted for lasso peptide. The pellets from 10L fermentation were resuspended by vortex and extracted with 500 mL of MeOH and the solvent was evaporated at 40 °C and reduced pressure. Dried extracts were then resuspended in 10 mL of 50% MeOH, cleared by centrifugation (17000 rpm, 1h). The supernatant was filtered by a Filtropur S filter (0.2 μm) and were then subjected to preparative HPLC (same as the one used for precursor peptide).

The vector pET41a-PadeCAKB1B2D was used to produce paeninodin in *E. coli* BL21. For paeninodin, it was purified by two rounds of HPLC. For the first round, the following gradient of water/0.045% formic acid (solvent A) and MeOH/0.05% formic acid (solvent B) was used: Linear increase from 20% to 30% B in 5 min, followed by a linear increase from 30% to 95% B in 25 min and holding 95% B for an additional 3 min. The pooled fractions of paeninodin were evaporated at 40°C and reduced pressure. The semipure lasso peptide were combined and subsequently applied to a second round of purification using the following gradient of water/0.1% trifluoroacetic acid (solvent A) and MeCN/0.1% trifluoroacetic acid (solvent B): Linear increase from 10% to 35% B in 30 min, followed by a linear increase from 35% to 95% D in 2 min and holding 95% B for additional 3 min. The pooled fractions of purepaeninodin were were collected, confirmed by LC-MS and lyophilized.

The vector PadeCA (ThcoK) B1B2D was used to produce phosphorylated paeninodin in *E. coli* BL21. It was also purified by two rounds of HPLC. For the first round, the following gradient of water/0.045% formic acid (solvent A) and MeOH/0.05% formic acid (solvent B) was used: Linear increase from 20% to 30% B in 30 min, followed by a linear increase from 30% to 95% B in 2 min and holding 95% B for an additional 3 min. The pooled fractions of phosphorylated paeninodin were evaporated at 40°C and reduced pressure. The semipure lasso peptides were combined and subsequently applied to a second round of purification using the following

gradient of water/0.1% trifluoroacetic acid (solvent A) and MeCN/0.1% trifluoroacetic acid (solvent B): Linear increase from 20% to 50% B in 30 min, followed by a linear increase from 50% to 95% D in 2 min and holding 95% B for additional 3 min. The pooled fractions of purephosphorylated paeninodin were collected, confirmed by LC-MS and lyophilized.

## 3.5 Analytical Methods

### 3.5.1 HPLC_MS

Generally, Mass spectrometric analysis of each extracts was performed with a high-resolution LTQ-FT ultra instrument (Thermo Fisher Scientific) connected to an Agilent 1100 HPLC system equipped witha detector at 215 nmand with a Nucleosil 300-8 C18 column (125×2 mm, Macherey and Nagel). The following gradient of water/0.1% trifluoroacetic acid (solvent A) and MeCN/0.1% trifluoroacetic acid (solvent B) at a column temperature of 40 °C and a flow rate of 0.2 mL/min was applied: holding 2% B for 2 min, followed by a linear increase from 2% to 30% B in 18 min, a subsequent linear increase from 30% to 95% B in 15 min and holding 95% B for additional 2 min.

Mass spectrometric analysis was also done with a low-resolution 1100 series MSD (Hewlett-Packard) coupled with a micropore 1260 HPLC system (Agilent Technologies) for the thermal stability, protease assays of the purified lasso peptides and all kinds of in vitro assays. Separations of different compounds were achieved by using different C18 column and specific gradients. Typically, the following gradient was used at 40 °C column temperature at a flow rate of 0.3 mL/min: Linear increase from 5% to 30-60 % B in 35 min, followed by a linear increase from 30-60% to 95% B in 2 min and holding 95% B for additional 5 min.

To quantify each compounds measured, UV-peak areas or intensive of extracted ion chromatograms was integrated and relative production was determined by comparison between different samples.

### 3.5.2 MS-MS fragmentation

In case that we need to further fragment the target mass, collision-induced dissociation fragmentation studies within the linear ion trap were done using HPLC-MS. Based on the molecular weight of different compounds, different charged ions were selected for further fragmentation. Generally, for lasso peptide, $[M+2H]^{2+}$ ions were selected for fragmentation, as they were the dominant species in the spectra. For the precursor peptide, the $[M+4H]^{4+}$ or $[M+5H]^{5+}$ were selected for fragmentation, as they were the dominant species in the spectra. The energy for fragmentation was set to 35%.

### 3.5.3 IM-MS spectrometric

For IM-MS analysis of paeninodin, a hybrid quadrupole ion-mobility time-of-flight mass spectrometer (Waters Synapt G2-Si) equipped with an ESI source operated in positive-ion mode was employed. Two samples of paeninodin were prepared. One was incubated at 65°C for 4 h. then mixed with formic acid and the supercharging agent sulfolane. The other was not heat treated and directly mixed with formic acid and the supercharging agent sulfolane. The measurements were performed with previously described instrument.

### 3.5.4 NMR Spectroscopy

For NMR structure determination of paeninodin lasso peptide, paeninodin was dissolved in different solvents with concentrations varied from 4.5 to 10 mM: $H_2O/D_2O$ (9:1), $CD_3OH$, $CD_3OH/CDCl_3$ (1:1), $CD_3OH/CDCl_3$ (3:1), and DMSO-$d_6$. The temperature dependence was studied by recording $^1H$ spectra in the range from 278 to 318 K as long as the solvent allowed.

For NMR measurements of modified precursor peptide, 3.3 mg enzymatically generated PadeA (13-23)-Ser23-$OPO_3^{2-}$, 3.2 mg chemically synthesized PadeA (13-23)-Ser23-$OPO_3^{2-}$(Biomatik), and 3.0 mg PadeA (13-23) (Biomatik) were dissolved in 0.2 mL of $H_2O/D_2O$ (9:1), respectively. For $^1H$ signal assignment, and NOESYspectra were recorded on a Bruker Avance 600 MHz spectrometer equipped with an inverse triple resonance $^1H$-$^{13}C$-$^{15}N$ probe with z-gradient. Temperature effect on the spectra was surveyed by recording $^1H$ spectra in a temperature range from 283 to 308 K. Consequently, all 2D spectra were recorded at 298 K. TOCSY spectra were recorded with a mixing time of 80 ms, while for NOESY measurements mixing times of 150 and 300 ms were applied. Water suppression was achieved by using excitation sculpting. The $^1D$ $^{31}P$ and 2D $^1H$-$^{31}P$ long-range correlation spectra of the phosphorylated compounds were recorded on a Bruker Avance III 500 MHz spectrometer with a broadband observation probe with z-gradient. For the latter, a standard pulse program for HMQC (heteronuclear multiple quantum correlation) was used, with a gradient ratio of 70:30:80 adjusted to the observation of $^1H$-$^{31}P$ correlation and a long-range coupling constant of 7 Hz. $^1D$ $^1H$ spectra were acquired with 65536 data points, while 2D $^1H$-$^1H$ spectra were collected using 4096 points in the F2 dimension and 512 increments in the F1 dimension. For 2D spectra, 32 transients were used, with a relaxation delay of 3.0 s. The $^1D$ $^{31}P$ spectra were recorded with 12288 scans, while 2D $^1H$-$^{31}P$ spectra were recorded with 128 transients and 128 increments in the F1 dimension. Chemical shifts of $^1H$ were referenced to DSS. All spectra were processed with Bruker TOPSPIN 3.2.

## 3.6 Biochemical Methods

### 3.6.1 Thermal stability studies of Paeninodin and phosphorylated Paeninodin

For the thermostability studies of paeninodinand phosphorylated paeninodin, 10 µg of each lasso peptide was redissolved in 100 µl ddH$_2$O. The samples were incubated at 20 °C, 35 °C, 50 °C, 65 °C, 80 °C and 95 °C for 4 h, respectivelyand cooled down on ice. The samples and an untreated sample were then analyzed by LC-MS.

### 3.6.2 Carboxypeptidase Y assays

To investigate the topology of the paeninodin, a previously established assay in which pure lasso peptide was heated and then incubated with carboxypeptidase Y was performed[112]. 10 µg lasso peptide was first heated at 65 °C for 4h and then analyzed by LC-MS, the heat labile lasso peptide was then readily form two topologies: One is the lasso peptide; one is a branched cyclic peptide where the tail of the lasso peptide already escaped from the macrolactam ring. These two compounds have the same mass but a different retention time in LC-MS. After unthread the lasso peptide, the samples were directly applied into a carboxypeptidase Y assays. The assays were performed over the course of 4 h at room temperature. Carboxypeptidase Y assays were done with 0.5 U carboxypeptidase Y per reaction in a buffer containing 50 mM MES and 1 mM CaCl$_2$ at pH 7.0. After the reaction, the samples were analyzed via LC-MS.

### 3.6.3 In vitro phosphorylation assays

Generally, each kind of pure peptide was redissolved in ddH$_2$O to make a 2.5 mM solution stocks for in vitro studies. A Tris buffer (50 mM Tris •HCl, 5 mM 5 mM MgCl$_2$, 1 ATP and 1 mM ATP. pH 8.0) was used. The reaction was carried out at 37 °C for 4 h, stopped by the addition of TFA (final concentration of 5% (v/v)) and subjected to HPLC-MS analysis.

To evaluate the modification position of the phosphate group in the peptide, scale-up experiments were carried out to synthesize a modified precursor peptide. A truncated version of the precursor peptide-PadeA (13-23) peptide was used to perform the scale-up assays. The PadeA(13-23) peptide was chemically synthesized by Biomatik. A typical scale-up assay mixture (30 mL) contained contained Tris •HCl (50 mM, pH 8.0), MgCl2 (5 mM), ATP (1 mM), PadeA(13-23) peptide (20 mg, ≈517 µM/L) and recombinant ThcoK ( 2 mg, ≈1.67 µM/L). The reaction was carried out at 37 °C for 16 h shaking at 220 rpm.

Phosphorylated PadeA (13-23) was then purified by a preparative HPLC (same as the one used for lasso peptide purification). The following gradient of water/0.1% trifluoroacetic acid (solvent

A) and MeCN/0.1% trifluoroacetic acid (solvent B) was used: Linear increase from 5% to 25% B in 30 min, followed by a linear increase from 25% to 95% B in 2 min and holding 95% B for an additional 5 min. The pooled fractions of phosphorylated PadeA (13-23) were confirmed by LC-MS, collected and lyophilized. By the scale-up assays, $\approx$ 10 mg pure phosphorylated PadeA (13-23) was isolated as white powder and stored at -20 °C until further usage.

### 3.6.4 Esterification assays

A chemical modification assays which could selectively methylate all carboxylic acid moieties by use of methanolic HCl was applied to study the modification position of the peptide[132]. In a typically reaction, 10 μg peptide was redissolved in100 μL of 2 M methanolic HCl (prepared by dropwise addition of 160 μL of acetyl chloride with stirring to 1 mL of cold methanol). The reaction was performed at 25 °C (500 rpm, 4h). Solvent was then removed by lyophilization and the resulting sample was redissolved in 100 μL 20% acetonitrile for further HPLC-MS analysis. Three kinds of peptides were modified in these assays. They are the unmodified PadeA(13-23), PadeA(13-23)-S23 side chain phosphorylation peptide synthesized by Biomatik and phosphorylated PadeA(13-23) obtained from scale-up assays.

### 3.6.5 Antibacterial Assays.

8 different strains were used to investigate the antibacterial activity of paeninodin and phosphorylated paeninodin by spot-on-lawn assays. They are *Bacillus subtilis* MR168, *Micrococcus flavus* ATCC 10240, *Burkholderia rhizoxinica*, *Sphingobium japonicum*, *Xanthomonas citri*, *Bacillus cereus*, *Bacillus megaterium* WSH-002, *and Bacillus amyloliquefaciens.* Soft agar was prepared by adding 0.7% agar to liquid LB. Soft agar (10 mL, at 48 ° C) was then inoculated using 100 μL of an overnight culture of each organism (OD $\approx$ 0.01) and spread on an LB agar plate. Then, 10 ng of paeninodin and phosphorylated paeninodin were spotted on the soft agar plate followed by incubation at 30 °C or 37 °C for 1-3 days.

# 4. Results

## 4.1 Genome Mining for a Novel Lasso Peptide Biosynthetic Gene Cluster
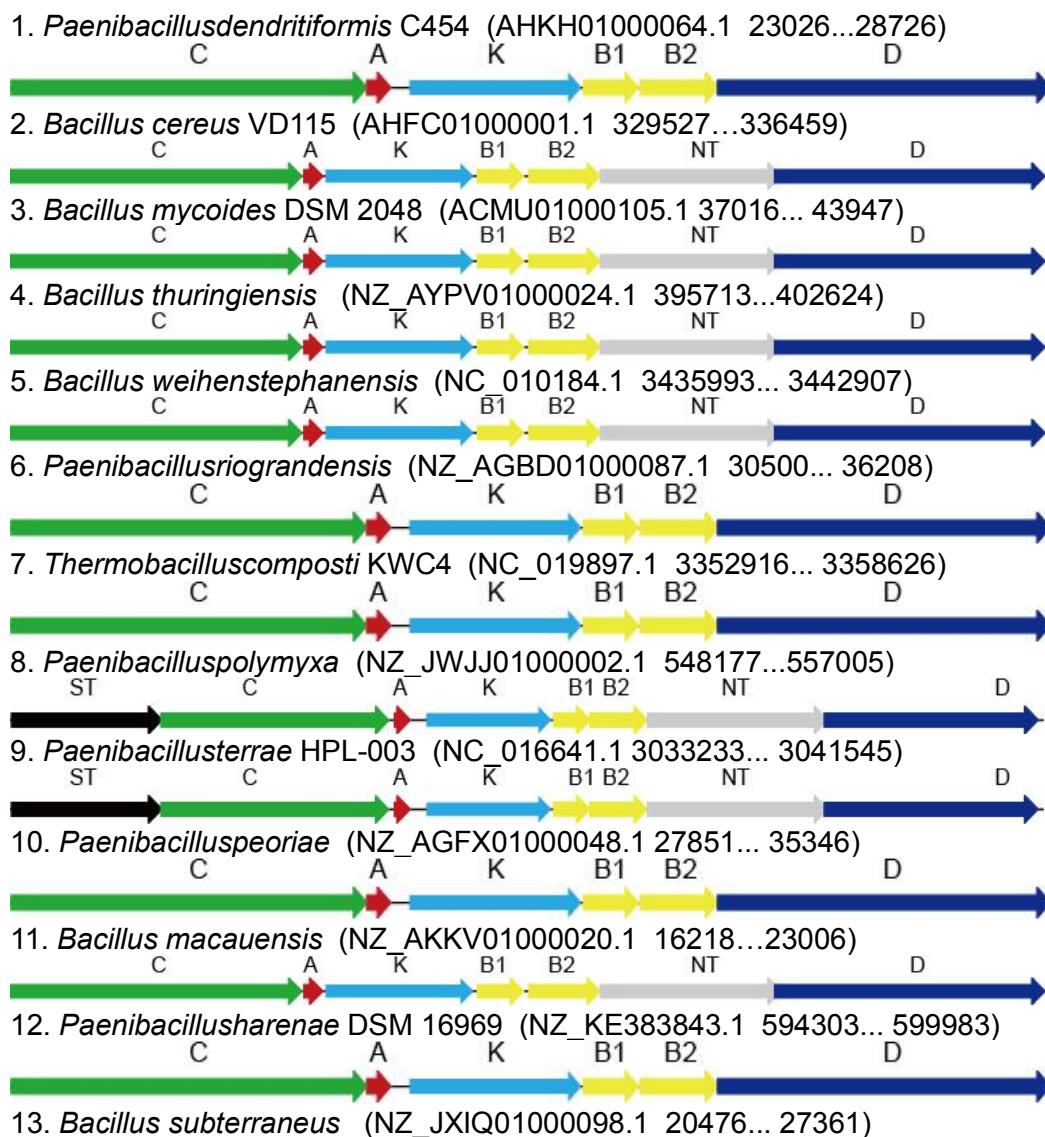
The number of RiPPs has grown rapidly in recent years[40]. One reason is due to the large genome sequencing projects in the past decade. Taking advantage of the available genome sequence, *in silico* mining of genomes combined with modern molecular biology has led to the discovery of a large number of new ribosomal natural products[41]. This method, called genome mining, provides a powerful tool to predict and then isolate new lasso peptides. The first example demonstrated for lasso peptide isolation was Capistruin, which was isolated in 2008[127]. Genome mining methods for lasso peptides can generally be divided into two classes. One is called the precursor-centric genome-mining approach, developed mainly by the A.J. Link group; the other is called B protein-centric genome mining and has been developed mainly by our group. Both methods take advantage of sequence homology among genes encoding precursor peptides or biosynthetic proteins, thus, leading to the identity of new biosynthesis gene clusters. Since the isolation of Capistruin in 2008, more than 30 different lasso peptides have been isolated by different genome-mining methods[97, 111, 112, 115, 117]. Notably, the A.J. Link group discovered a novel Isopeptidase in 2013 that linearizes lasso peptides, and this proved that the genome-mining method is also a powerful tool to identify novel enzymes[113, 118].

Our group is particularly interested in the B protein-centric genome-mining method because B protein in lasso peptide biosynthesis gene clusters has a very low homology similarity to other enzymes. It only has a very low similarity to eukaryotic transglutaminases. Using this method, our group exploited the lasso peptide biosynthesis world in proteobacteria in depth[112]. The discovery of novel lasso peptides included Xanthomonins I-III, a new class of lasso peptides with a seven-residue macrolactam ring[115], Caulonodin lasso peptides with unprecedented N-terminal residues and several Caulosegninslasso peptides derived from a single biosynthetic gene cluster[111, 116]. These lasso peptides broaden the existing knowledge of lasso peptides greatly and provide diverse scaffolds for further applications. However, lasso peptides have so far only been isolated from Proteo- and Actinobacterial sources[97-99]. No other phyla have been reported that can produce lasso peptide. What is more, lasso peptide biosynthetic gene clusters typically only encode enzymes for biosynthesis and export; tailoring of lasso peptide is very rare. Therefore, we here focus on the discovery of novel lasso gene clusters which are different from the ones already reported.

55

## 4.1.1 Genome Mining Reveals a Novel Lasso Peptide Biosynthetic Gene Cluster Type in Firmicutes

A B protein-centric genome-mining method was applied to identify a novel lasso peptide biosynthesis gene cluster. Most of the lasso peptide gene clusters identified have already been published in our genome-mining paper[112]. However, interestingly, a family of gene clusters predominantly found in Firmicutes was identified, which had been overlooked in the past. The clusters identified contain homologues to the core-conserved genes of known lasso peptide biosynthetic gene clusters, in addition to a gene coding for a putative kinase (Table 4.1). Other lasso peptide systems identified also feature a putative sulfotransferase-encoding gene at the beginning of the biosynthetic operon and/or a putative nucleotidyltransferase-encoding gene between those coding for the B2 and D homologues[129]. All the potential lasso peptide gene clusters identified are summarized in Table 4.1

**Table 4.1.** Kinase-harboring lasso peptide gene clusters from Firmicutes.

1. *Paenibacillusdendritiformis* C454 (AHKH01000064.1 23026...28726)

2. *Bacillus cereus* VD115 (AHFC01000001.1 329527…336459)

3. *Bacillus mycoides* DSM 2048 (ACMU01000105.1 37016... 43947)

4. *Bacillus thuringiensis* (NZ_AYPV01000024.1 395713...402624)

5. *Bacillus weihenstephanensis* (NC_010184.1 3435993... 3442907)

6. *Paenibacillusriograndensis* (NZ_AGBD01000087.1 30500... 36208)

7. *Thermobacilluscomposti* KWC4 (NC_019897.1 3352916... 3358626)

8. *Paenibacilluspolymyxa* (NZ_JWJJ01000002.1 548177...557005)

9. *Paenibacillusterrae* HPL-003 (NC_016641.1 3033233... 3041545)

10. *Paenibacilluspeoriae* (NZ_AGFX01000048.1 27851... 35346)

11. *Bacillus macauensis* (NZ_AKKV01000020.1 16218…23006)

12. *Paenibacillusharenae* DSM 16969 (NZ_KE383843.1 594303... 599983)

13. *Bacillus subterraneus* (NZ_JXIQ01000098.1 20476... 27361)

56

14. *Bacillus kribbensis* (NZ_KE387239.1 41261... 48195)

15. *Bacillus megaterium* WSH-002 (CP003017.1 3899438... 3905391)

16. *Bacillus flexus* (NZ_JANV01000106.1 11574... 17331)

17. *Bacillus oceanisediminis* (NZ_ALEG01000067.1 32511... 40721)

18. *Bacillus vietnamensis* (NZ_CCDN010000009.1 39263... 46139)

19. *Bacillus aryabhattai* (NZ_JPIE01000001.1 482105... 489018)

20. *Paenibacillusalginolyticus* (NZ_AUGY01000047.1 37379... 44267)

21. Bacillus endophyticus (NZ_ALIM01000014.1 895870... 901612)

22. Paenibacillusalgorifonticola (NZ_LAQO01000025.1 22190... 29253)

23. Paenibacilluswulumuqiensis (NZ_KQ040793.1 393438... 401239)

24. Paenibacillus borealis (NZ_CP009285.1 998071... 1003776)

25. Bacillus akibai (BAUV01000025.1 34686... 41595)

We selected four typical lasso peptide gene clusters, one each from *Paenibacillus dendritiformis* C454, *Thermobacillus composti* KWC4, *Bacillus cereus* VD115 and *Paenibacillus polymyxa*, to represent the diverse organization of the gene clusters in Firmicutes (Figure 4.1). Surprisingly, we noticed that even though there are different potential tailoring enzymes associated with the lasso peptide gene cluster, the precursor peptides among these systems are very similar. All the lasso peptides predicted bear an N-terminal Ala residue, a feature only recently observed, similar to Caulonodin lasso peptides[116, 129]. We hypothesized that these gene clusters originated from a common ancestor.
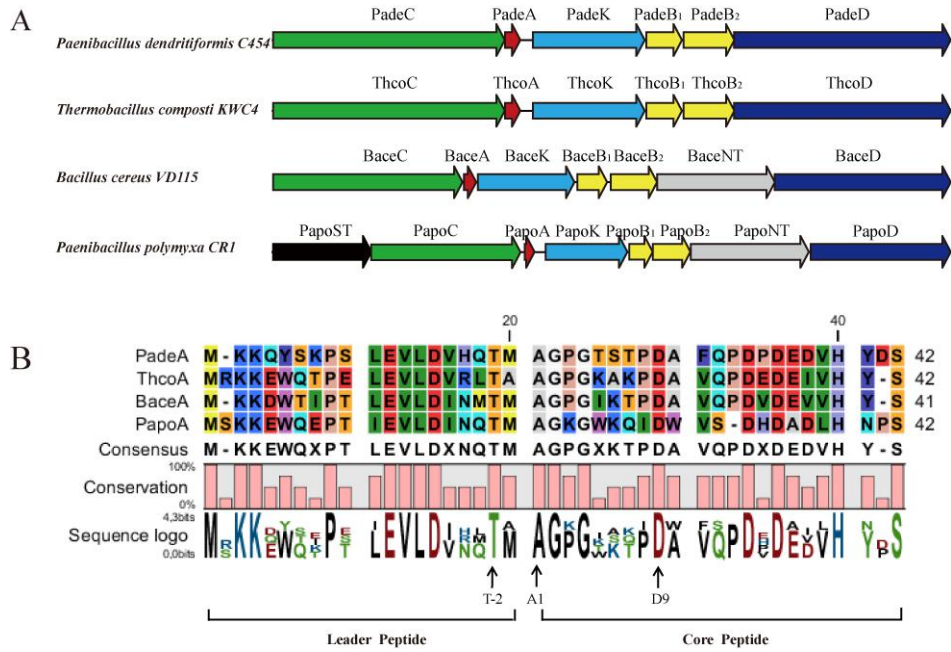
**Figure 4.1**:(a) Overview of the types of lasso peptide biosynthetic gene clusters identified in Firmicutes. Open reading frames (ORFs) ending in K, NT and ST encode putative kinases, nucleotidyltransferases and sulfotransferases, respectively. (b) Alignment of the corresponding precursor peptides

We then performed a blast analysis of potential tailoring enzymes surrounding the lasso peptide gene cluster. They indeed show some homology to known enzymes. Table 4.2 provides an overview of these genes, their putative roles and arrangements within the clusters

**Table 4.2.**Proposed functions of the ORFs in the putative lasso peptide biosynthetic gene clusters that were identified by genome mining in Firmicutes.

| Protein | length(aa) | GenBank accession no. | proposed function | conserved domain analysis |
|---|---|---|---|---|
| PadeA | 42 | - | | |
| ThcoA | 42 | AGA59244.1 | precursor peptides | |
| BaceA | 41 | EJR65203.1 | | - |
| PapoA | 42 | YP_008910696.1 | | |
| PadeB1 | 99 | WP_006678398.1 | | |
| ThcoB1 | 98 | AGA59242.1 | maturation enzymes, B1part of 'split B', protease | PqqD superfamily, coenzyme PQQ synthesis protein D |
| BaceB1 | 101 | EJR65205.1 | | |
| PapoB1 | 100 | YP_008910698.1 | | |
| PadeB2 | 142 | WP_006678399.1 | | |
| ThcoB2 | 163 | AGA59241.1 | maturation enzymes, B2part of 'split B', protease | transglutaminase-like superfamily |
| BaceB2 | 157 | EJR65206.1 | | |
| PapoB2 | 161 | YP_008910699.1 | | |
| PadeC | 649 | WP_006678396.1 | | |
| ThcoC | 645 | AGA59245.1 | maturation enzymes, adenylation, cyclization | glutamine amidotransferases class-II (GATase) asparagine synthase_B type |
| BaceC | 645 | EJR65202.1 | | |
| PapoC | 641 | YP_008910695.1 | | |
| PadeD | 607 | WP_006678400.1 | | |
| ThcoD | 607 | AGA59240.1 | export, immunity protein | ABC-type multidrug transport system |
| BaceD | 599 | EJR65208.1 | | |
| PapoD | 601 | YP_008910701.1 | | |

| | | | | |
|---|---|---|---|---|
| PadeK | 313 | WP_006678397.1 | | |
| ThcoK | 310 | AGA59243.1 | putative kinase | HPrK-related kinase, |
| BaceK | 328 | EJR65204.1 | | |
| PapoK | 351 | YP_008910697.1 | | |
| | | | | |
| BaceNT | 400 | EJR65207.1 | putative | uncharacterized |
| PapoNT | 506 | YP_008910700.1 | nucleotidyltransferase | nucleotidyltransferase |
| | | | | |
| PapoST | 422 | YP_008910694.1 | putative pulfotransferase | sulfotransferase family |

## 4.1.2 Genome Mining Reveals a Novel Lasso Peptide Biosynthetic Gene Cluster Type in Proteobacteria

After we identified the lasso peptide gene clusters in Firmicutes which had been overlooked and considering the rare organization of the gene cluster, we thought that there are group a of novel and cryptic lasso peptide gene clusters which have overlooked in the past. Therefore, we used the kinase from Firmicutes to further exploit the lasso peptide biosynthesis diversity by a kinase homology-based genome-mining method. This could give us a chance to perform a global survey of lasso peptide gene clusters featuring at least one tailoring enzyme. We found that these kinase homologues exist widely in nature. We could also identify several clusters from Proteobacteria preferring an AB1B2CKD organization, which was predicted in our B protein-centric genome-mining paper[112] (Table 4.3). All the B homologues in these clusters are split into B1 and B2. Remarkably, all the kinase-featured lasso peptide gene clusters from Proteobacteria are generally lacking in the nucleotidyltransferase and sulfotransferase found in Firmicutes. Several potential lasso peptide gene clusters from Proteobacteria are summarized in Table 4.3.

**Table 4.3.** Kinase-harboring lasso peptide gene clusters from Proteobacteria.



59

33. *Sphingomonas*sp. Y57 (NZ_LDES01000074.1  26756... 32158)

34. *Mesorhizobium* sp. WSM1293  (NZ_KI911320.1  331530... 337146)

35. *Altererythrobacteratlanticus*  (NZ_CP011452.1  1803258... 1808418)
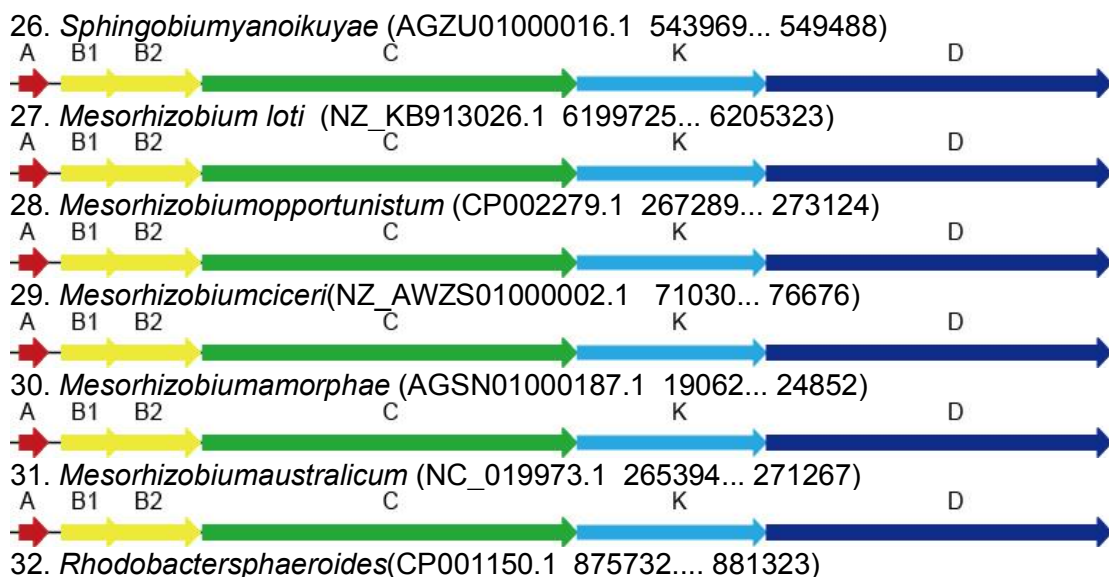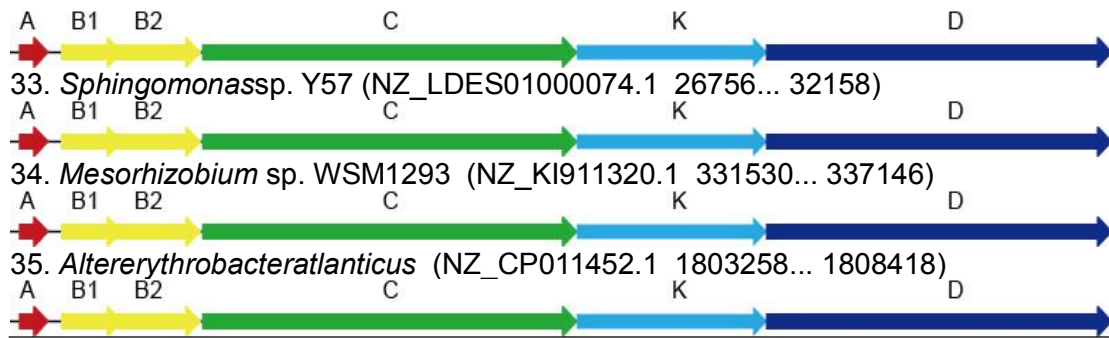
## 4.1.3 Phylogenetic tree of the 35 kinase-containing lasso peptide biosynthetic gene clusters identified

We then performed a phylogenetic analysis of the kinase and its homologies. The kinase sequences were retrieved from the gene clusters and phylogenetic trees were constructed using the minimum evolution and neighbor-joining method. Phylogenetic trees derived from all these 35 kinases clearly segregate into two clades (as shown in Fig. 4.2), even though it seems that these two different subfamilies are close to each other and probably originated from the same ancestor. The lasso peptide gene clusters from Firmicutes generally prefer an XCAKB1B2XD organization, while clusters from Proteobacteria prefer an AB1B2CKD organization. All the B homologues in these clusters are split into B1 and B2. Remarkably, 16 out of 25 gene clusters from Firmicutes also featured an uncharacterized nucleotidyltransferase. Two gene clusters featured a sulfotransferase, making the gene cluster more diverse. These transferases are highly likely to be novel tailoring enzymes for the lasso peptide, but, at the present time, the functions of these enzymes are still elusive. Further investigations are, thus, needed[129].
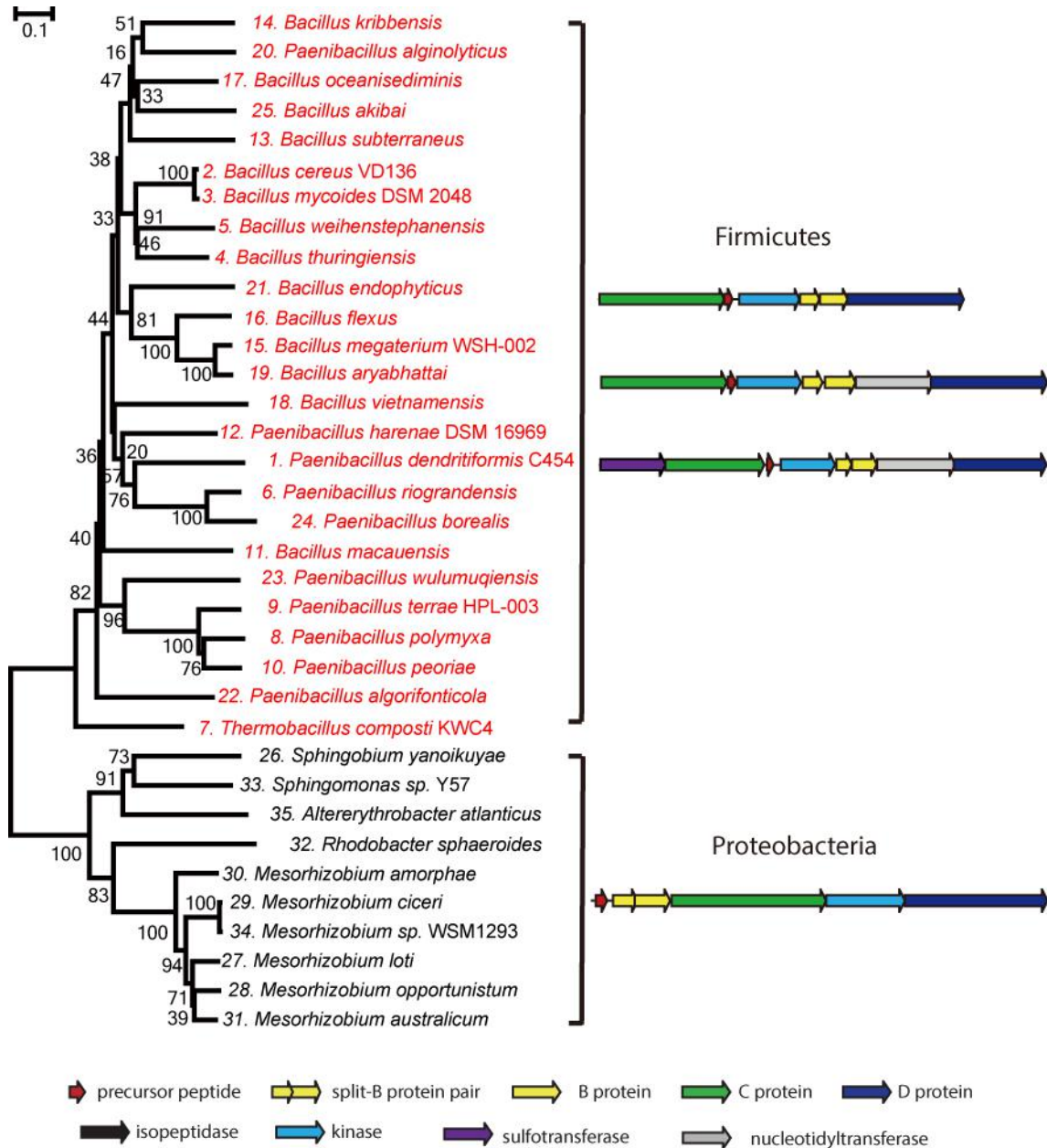
**Figure4.2:** Phylogenetic tree of the 35 kinase-containing lasso peptide biosynthetic gene clusters identified.

## 4.1.4 Comparison of precursor peptides from kinase-harboring lasso peptide biosynthetic gene clusters

The precursor peptide or the mature lasso peptide could be the substrate of the kinase. We subsequently turned our attention to the precursor peptide. We first identified the precursor gene manually in order to compare the precursor peptides. The precursors were generally predicated based on three conserved features of lasso peptide precursors: 1) a Thr residue at the penultimate position of the leader peptide, 2) N-terminal glycine, cysteine, serinine or alanine on mature lasso peptide and 3) aspartate or glutamate being the last residue in the ring at position 7, 8 or 9. After we had identified all the precursor peptides, the precursor peptide sequences were retrieved from the gene clusters and alignment was performed. Overall, the

61

precursors are highly conserved in their own clades (Figure 4.3). Regarding the precursors from Firmicutes, the core peptide prefers starting with an alanine, while the core peptide from Proteobacteria prefers starting with a glycine. However, they both like forming a nine-residue macrolactam ring with an aspartic acid. The Thr residue in the penultimate position of the leader peptide is also conserved in both clades. Precursors from Firmicutes are generally polar, negatively charged with lots of aspartic or glutamic acid residues in both the leader peptide and core peptide region, while precursors from Proteobacteria are mostly hydrophobic in the core peptide region and polar in the leader peptide region. We then chose one precursor from each clade and did an alignment. There is no similarity between these two precursors except the last serine (Figure 4.3).

This suggested that these gene clusters indeed segregate into two clades and we hypothesized that these kinases specifically recognize the last serine of the precursor peptide and have a wide substrate specificity[129].



**A** Precursors from Firmicutes

**B** Precursors from Proteobacteria

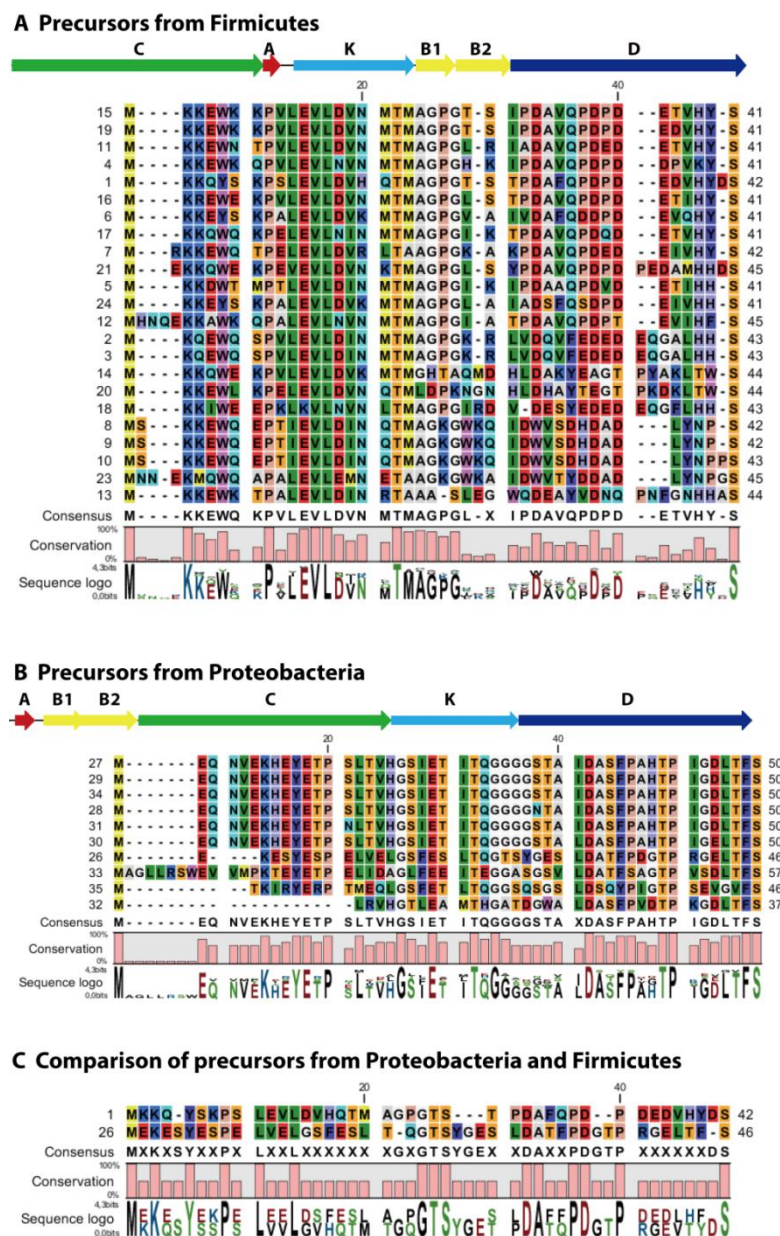**C** Comparison of precursors from Proteobacteria and Firmicutes

62

**Figure4.3:**Multiple sequence alignment of precursor peptides from kinase-harboring lasso peptide biosynthetic gene clusters. Alignment of the precursor peptides found in (A) Firmicutes and (B) Proteobacteria. (C) Comparison of precursor peptides from Firmicutes and Proteobacteria to each other.

## 4.2 Discovery and Isolation of Lasso Peptide Paeninodin and Phosphorylated Paeninodin

After we had identified these novel lasso peptide gene clusters, we turned our attention to look for the natural products. We chose the putative lasso peptide biosynthetic gene cluster from *P. dendritiformis* C454 as a target for further investigation. This gene cluster contains the additional kinase-encoding gene (padeK) compared to previously characterized systems. Moreover, it is the first lasso gene cluster investigated from Firmicutes[129].

We first worked on the wild strain of *P. dendritiformis* C454 to see if the native strain could produce the lasso peptide. For this purpose, *P. dendritiformis* C454 was fermented in various medium, such as M9, LB andCASO medium. The pellets and supernatant of each fermentation were extracted and analyzed by mass spectrometry. To our surprise, the fermentation extracts of native *P. dendritiformis* C454 failed to produce any lasso peptide. A heterologous approach had to be developed to solve this problem.

### 4.2.1 Heterologous production of the novel lasso peptide paeninodin and phosphorylated paeninodin in *E. coli*

It was found from the studies of the lasso peptide capistruin that lasso peptide gene clusters from Proteobacteria could be heterologously expressed in *E. coli*[112, 127]. Since then, more than 20 lasso peptides from Proteobacteria have been produced in this way[97]. For our studies, the gene cluster is from Firmicutes and a similar approach has been adopted, but the risk is very high. We used our previous method and the cluster was cloned into the pET41a, which is under the control of the T7 promoter. We then expressed these constructs and the empty pET41a vector as a control in *E. coli* BL21. An amount of 600 ml M9 minimal medium was chosen for the initial screens under 1, 2 and 3 days at 37 °C and 3 days at 20 °C.
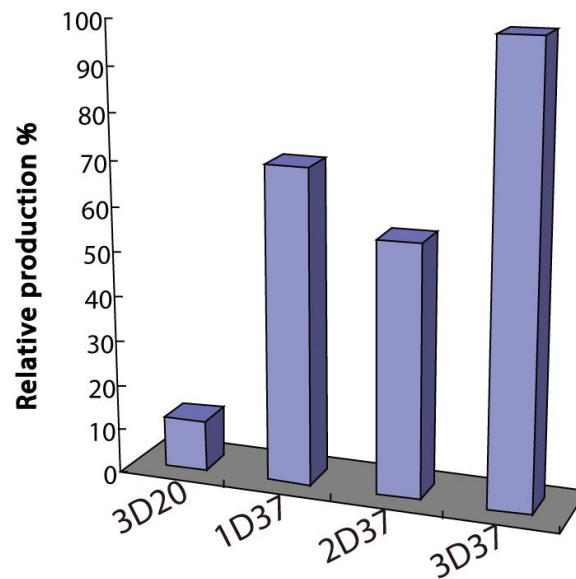
**Figure4.4:**Comparison of relative lasso peptide production for the different fermentation conditions. 3D20, 3 days at 20 °C; 1D37, 1 day at 37 °C; 2D37, 2 days at 37 °C; and  1D37, 3 days at 37 °C.

We then harvested the cultures, threw away the supernatants and extracted the cell pellets with 50 mL MeOH overnight. Next, solvent was removed via evaporation at 40 °C and reduced pressure. The dry pellet was then resuspended in a total of 900 µL 50 % MeOH, and cleared by centrifugation (13000 rpm, 1 h). Subsequently, the supernatant was analyzed by HPLC-MS. Results showed that under three days at 37 °C was the best condition for the production (Figure 4.4) and comparison with an empty vector control confirmed the presence of two additional major peaks ([M+2H]$^{2+}$ observed = 1200.5023, retention time = 21 min; [M+2H]$^{2+}$ observed = 1240.4796. retention time = 20 min) in the pET41a-CAKB1B2D constructs (Figure 4.5). These two UV peaks indicated that there truly was a modification happening to the wild lasso peptide. The first hypothesis was that this unknown kinase endows the lasso peptide with a phosphate group, and we named this novel lasso peptide paeninodin and phosphorylated paeninodin[129].
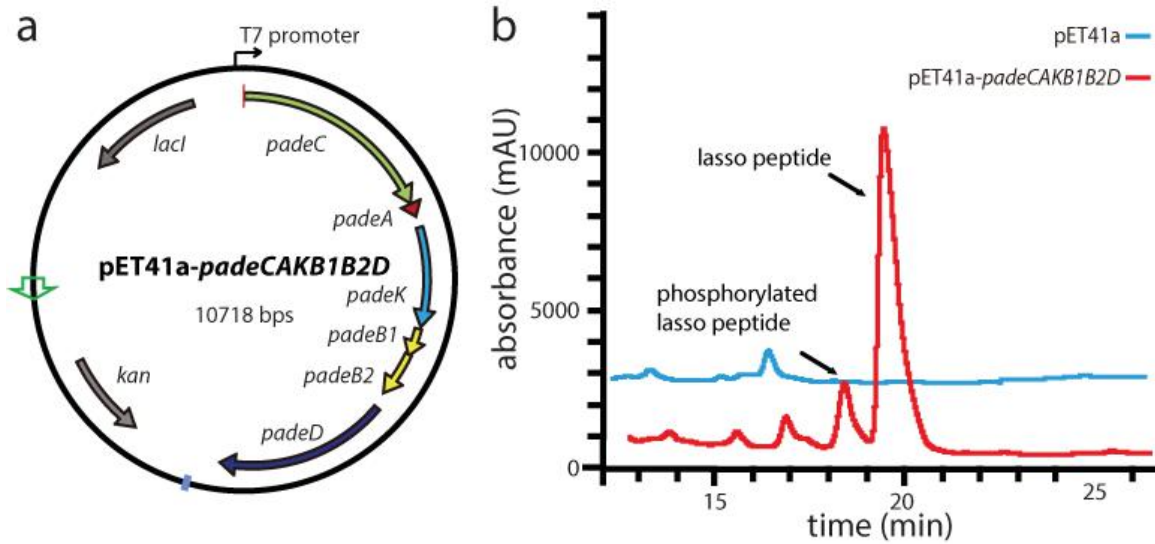
**Figure4.5:** (a) Vector map of the heterologous paeninodin production plasmid. (b) UV traces of the LC-MS analyses of pellet extracts of cultures after expression of an empty vector control, the complete cluster in pET41a, respectively.

## 4.2.2 Mass spectrometric analysis

In order to acquire additional information, especially of the modification position of the lasso peptide, extensive mass spectrometric studies on both paeninodin and phosphorylated paeninodin were carried out on a LTQ-FT instrument by collision-induced dissociation (CID). The MS2 fragmentation gives a clear y and b series of fragment ions of the lasso peptide, as shown in Figure 4.6. The smallest b fragment (b10) and the largest y series fragment (y14) of paeninodin and phosphorylated paeninodin observed correspond to the ring plus alanine and the tail, respectively. This proves that the internal backbone to side chain cyclization is between residue Ala1 and Asp9. Comparing the MS2 fragmentation of unmodified and modified lasso peptide, the main difference is the y series, and the difference between each series is about 80. This indicates that the modification happens at the tail of the lasso peptide and the tailoring enzyme could phosphorylate the lasso peptide. Furthermore, the smallest y series fragment (y4) from the paeninodin observed proves that the modification happens at the last four residues of the peptide. To further confirm the modification position of the phosphate group, we selected the y11 fragment of both paeninodin and phosphorylated paeninodin and carried out a further MS3 fragmentation experiment. The MS3 fragmentation spectra show the characteristic fragmentation pattern of each peptide clearly. Comparing these two spectra, the b series fragment (b11) observed plus a phosphate group demonstrates clearly that the phosphate group is attached at the last serine residue[129].

In conclusion, the mass spectrometric results suggest the peptide is a lasso peptide with a phosphate group at the last serine residue. However, since the serine has a hydroxyl group at the side chain and a carboxyl group at the C-terminus, it is unclear at which position the modification happens. Thus, further experiments are necessary to prove it.
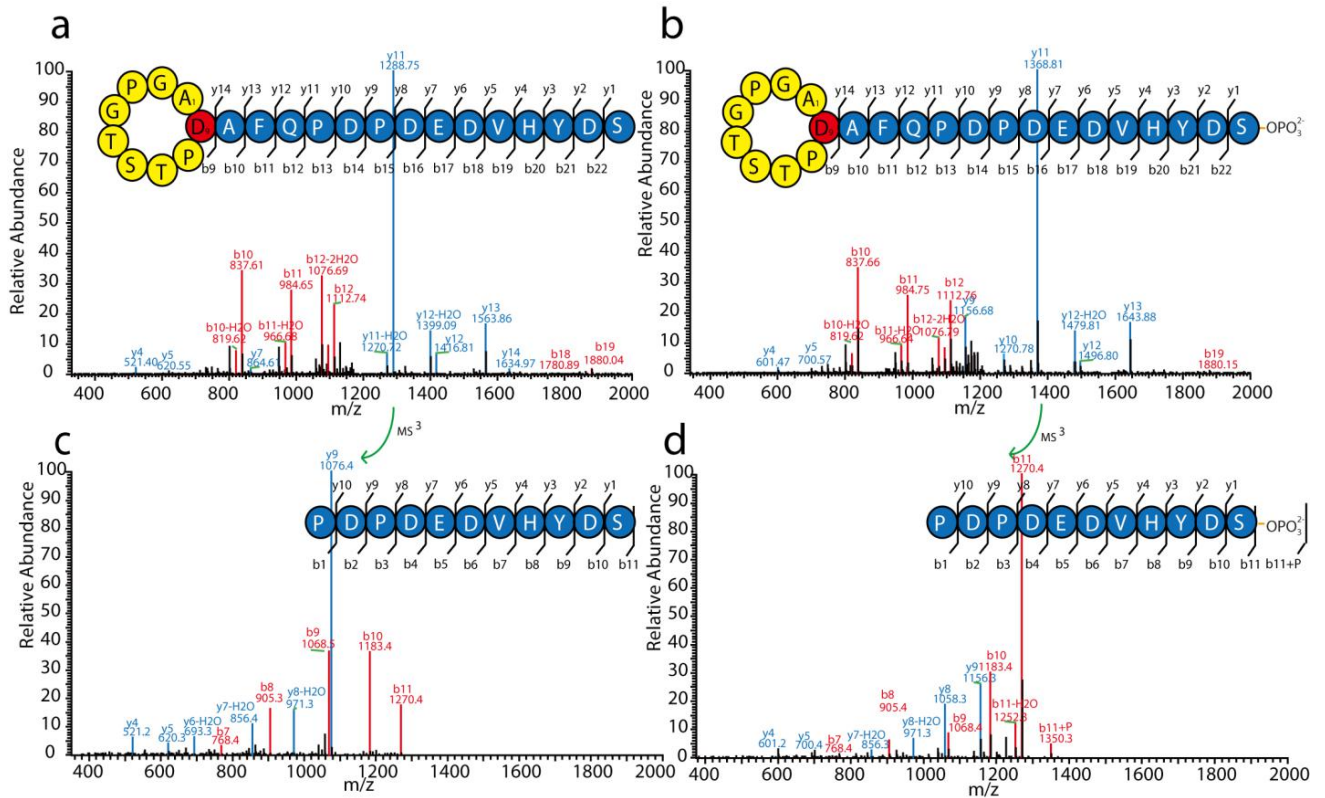
**Figure4.6:**MS$^2$ spectra of (a) paeninodin and (c) paeninodin-OPO$_3{}^{2-}$, and MS$^3$ spectra of the y11 fragments of (b) paeninodin and (d) paeninodin-OPO$_3{}^{2-}$.

## 4.2.3 The kinase is responsible for the modification of paeninodin

We hypothesized that this unknown kinase endows the lasso peptide with a phosphate group. To confirm this hypothesis, the kinase gene was knocked out by SLIM mutagenesis. The resulting plasmid was named pET41a-PadeCAB1B2D and transferred into *E. coli* BL21. Meanwhile, the plasmid pACYCDuet-1-PadeK was used to complement *E. coli* BL21 harboring pET41a-PadeCAB1B2D. By using the same procedure, extracts from these cultures were analyzed for the presence of paeninodin. Whereas the strain harboring pET41a-PadeCAKB1B2D could produce both paeninodin and phosphorylated paeninodin, no phosphorylated paeninodin could be detected in the extract from strains harboring pET41a-PadeCAB1B2D, as shown in Figure 4.7. On the other hand, after the plasmid pACYDuet-1-PadeK, which encodes the unknown kinase, was used to complement *E. coli* BL21 harboring pET41a-PadeCAB1B2D, the HPLC-MS experiments demonstrated clearly that phosphorylated paeninodin production was restored upon complementation, even though the overall yield decreased (Figure 4.7B). Altogether, these data showed that this kinase is responsible for the modification of lasso peptide, hence, we were able to identify a novel lasso peptide tailoring enzyme for the first time[129].
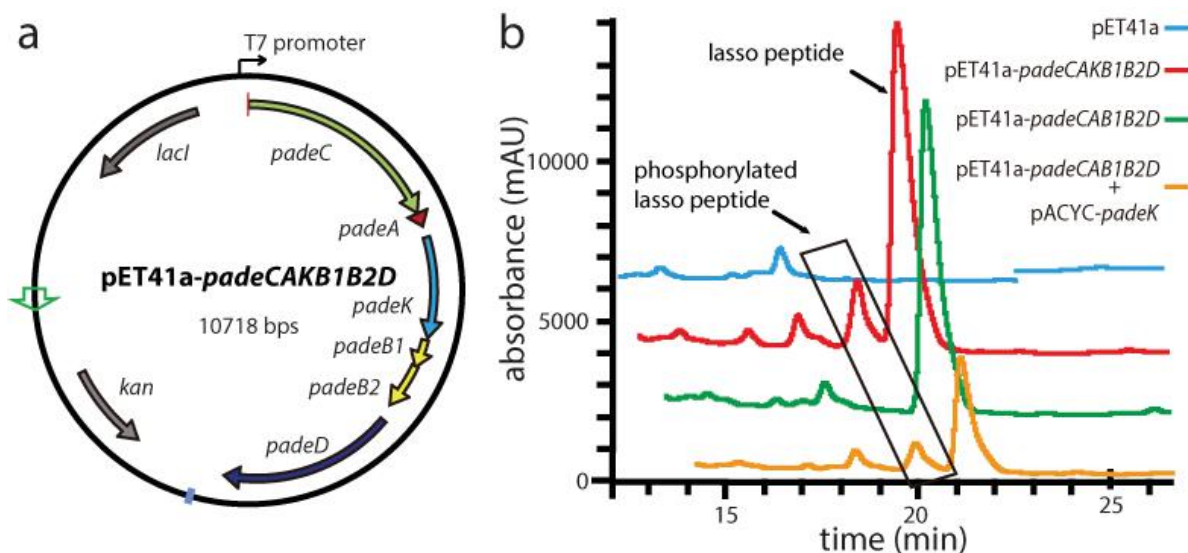
**Figure4.7:**(a) Vector map of the heterologous paeninodin production plasmid. (b) UV traces of the LC-MS analyses of pellet extracts of cultures after expression of an empty vector control, the complete cluster in pET41a, the padeK knock-out cluster and the supplementation of the knock-out plasmid with padeKpACYC, respectively.

## 4.3 Characterization of Lasso Peptide Paeninodin and Phosphorylated Paeninodin

### 4.3.1 Purification of paeninodin and paeninodin-OPO$_3^{-2}$

After we had initially observed that cultivation in M20 medium at 37 °C for three days is the best heterologous condition for the production of paeninodin and its analog, we tried to engineer the biosynthetic pathway to increase the yield. We use our past method to incorporate an artificial ribosomal binding site between genes and tried to rearrange the order of the genes[112]. However, none of them would work (data not shown). On the basis of these facts, we used this condition and performed large-scale fermentations to isolate our lasso peptides. The culture was scaled-up to 100 L volume in order to obtain sufficient amounts of lasso peptide for NMR structure elucidation studies and other assays. The cell pellets of the fermentation were extracted with MeOH and the solvent was evaporated at 40 °C and reduced pressure. The dry pellet was then resuspended in 50 % MeOH and cleared by centrifugation (17000 rpm, 1 h). Subsequently, the supernatant was used to purify the lasso peptide. Purification of a large amount was achieved by preparative HPLC (See methods; Figures 4.8 and 4.9). At the end, paeninodin could be purified with a yield of 0.1 mg/L culture, while phosphorylated paedenin could be purified with a yield of 0.05 mg/L culture. The unmodified and modified lasso peptide is enough for further studies[129].
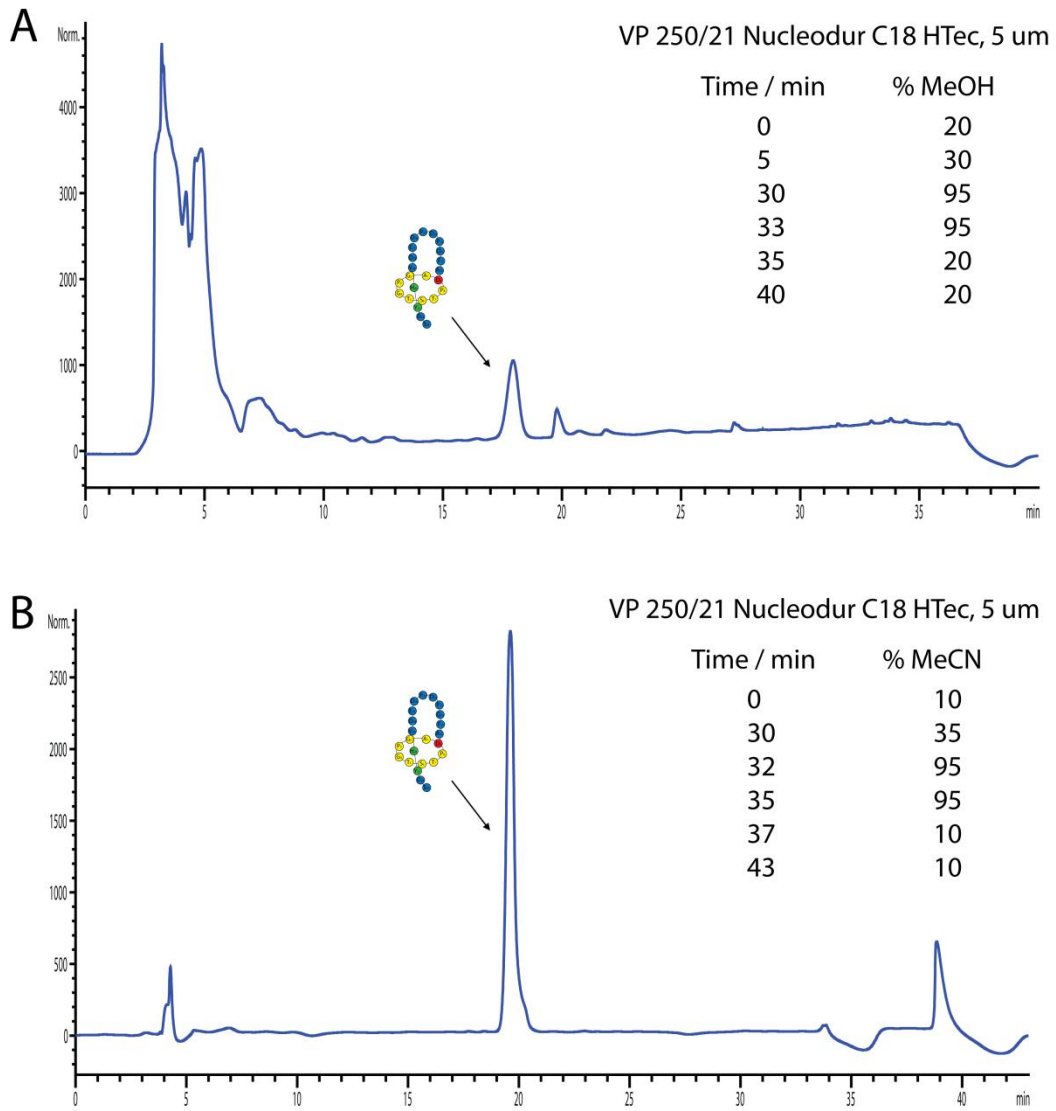
**Figure4.8:**Chromatograms of the paeninodin purification by preparative HPLC. (A) First purification with water/0.05 % formic acid and MeOH/0.045 % formic acid. (B) Second purification with water/0.1 % TFA and MeCN/0.1 % TFA.

**Figure4.9:** Chromatograms of the paeninodin-OPO$_3{}^{2-}$purification by preparative HPLC. (A) First purification with water/0.05 % formic acid and MeOH/0.045 % formic acid. (B) Second purification with water/0.1 % TFA and MeCN/0.1 % TFA.
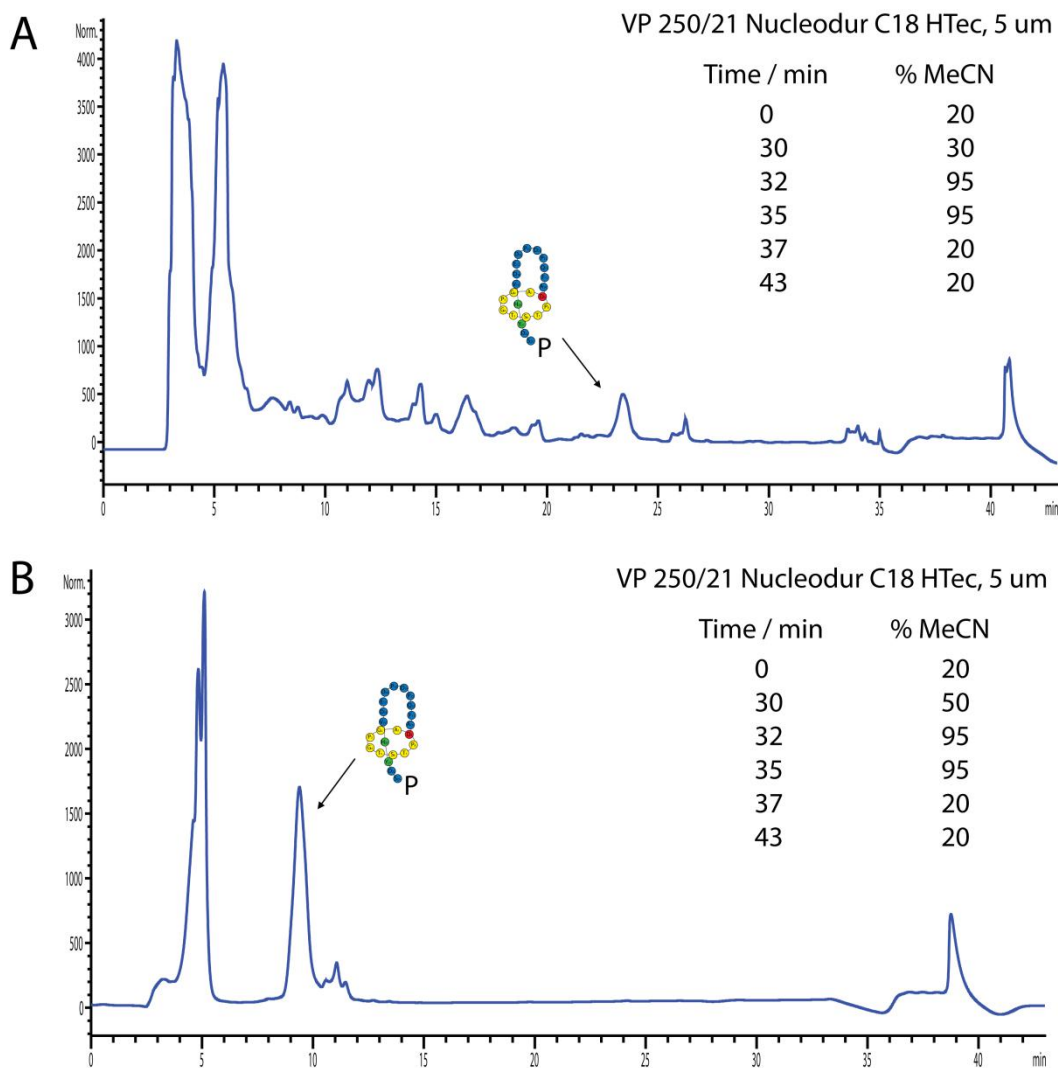
## 4.3.2 Confirmation of the lasso topology of paeninodin

### 4.3.2.1 Thermal stability

After producing paeninodin and phosphorylated paeninodin, we first wanted to test the thermal stability of the lasso peptide and especially the thermal stability of the modified lasso peptide. Normally, neither a phosphate ester nor a mixed carbonic-phosphoric anhydride is chemically stable in aqueous solution at physiological pH. Since the modified lasso peptide is only 20 % of the wild type, it is unclear whether this is due to the hydrolysis of itself or the kinase activity is just low. A solution of 10 µg of purified paeninodin and phosphorylated paeninodin were incubated at 20, 35, 50, 65, 80 and 95 °C for 4 h, respectively, to study the thermal stability. Samples were then cooled on ice and were, subsequently, analyzed via LC-MS using appropriate gradients. The paeninodin lasso peptide turns out to be a moderate thermostable lasso peptide (Figure 6A). It showed only minimal unthreading after 4 h at 50 °C, while under

69

65 °C for 4 h, not only does the lasso peptide start unthreading, it also tends to decompose, resulting in two main degradation products. One is the nine-amino acid macrolactam ring plus five amino acid and the nine-amino acids tail peptide[129]. This phenomenon was also observed in the thermal stability studies of caulosegnin lasso peptide[111]. The aspartate residue located between these two proline amino acids seems be able to catalyze the self-hydrolysis of the lasso peptide via formation of a cyclic imide intermediate, which proves the lasso peptide not a thermostable lasso peptide. The stability profile of phosphorylated paeninodin is similar to the wild type lasso peptide, but, surprisingly, the modification is quite stable and is still attached at the nine-amino acids tail peptide even after 4 h at 95 °C (Figure 4.10). This actually gives us a hint that the activity of the kinase in this gene cluster is just very low and *in vitro* studies might prove our hypothesis.
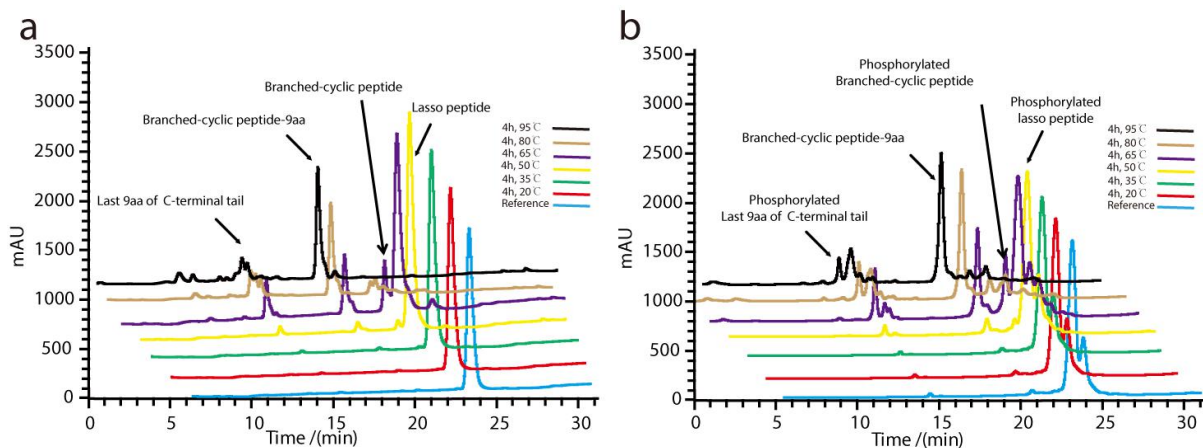


**Figure4.10:**(a) UV traces of LC-MS analyses of paeninodin samples after incubation for 4 h at different temperatures. (b) UV traces of LC-MS analyses of paeninodin-$OPO_3^{2-}$ samples after incubation for 4 h at different temperatures.

### 4.3.2.2 Proteolytic stability

Several strategies have been developed to distinguish the topology of lasso peptide in the past. Firstly, carboxypeptidase Y assays were performed to prove that the paeninodin is a true lasso peptide. This method has been widely applied to investigate the topology of lasso peptide in the past[111, 112, 115]. The carboxypeptidase Y can normally cleave off amino acids from the C-terminus of a peptide chain. We observed in the heat stability assays that the paeninodin is a heat-labile lasso peptide. The paeninodin forms two topologies readily at 65 °C for 4 h. One is the lasso fold and the other is a branched cyclic peptide. Theoretically, the carboxypeptidase Y could only degrade the tail of the lasso peptide that is not shielded by the surrounding macrolactam ring, while the carboxypeptidase Y could further degrade the branched cyclic peptide's tail. The branched cyclic analog of paeninodin was digested until only one additional amino acid after the macrolactam ring was present, as shown in Figure 4.11, thus, proving the topology proposed. By contrast, there is no degradation of the lasso peptide which proves that it has a lasso fold and the macrolactam ring is able to sterically shield amino acids positioned inside or close by

the ring. Moreover, this is also a hint that the lasso peptide has a very short tail, which is another proof that the H20 and Y21 act as the plug amino acids on opposite sides of the macrolactam[129].
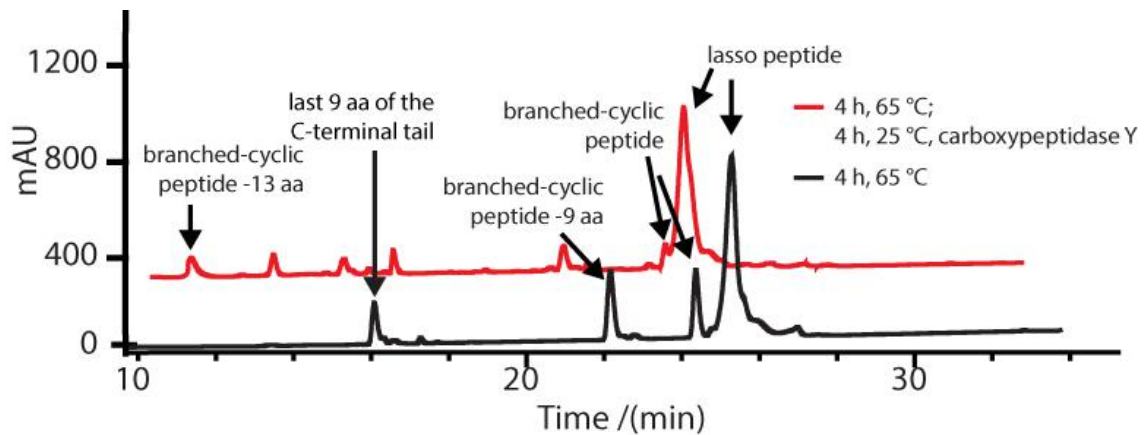


**Figure4.11:**Result of the carboxypeptidase Y treatment of paeninodin after prior incubation for 4 h at 65 °C.

## 4.3.2.3 IM-MS

IM-MS analysis of paeninodin was carried out to further substantiate these findings (Figure 4.12). IM-MS analysis of (a) untreated paeninodin and (b) a sample of paeninodin after incubation for 4 h at 65 °C gave two different results, as is shown in Figure 4.12. The peaks of the $[M+4H]^{4+}$ ions corresponding to paeninodin and its branched-cyclic analog are also shown in Figure 4.12[129] (theoretical value = 600.7534). It can be clearly seen that a new peak with a longer drift time (and, therefore, belonging to a compound with a larger collision cross-section) appears next to the double peak of the lasso peptide after thermal treatment of paeninodin, which is in accordance with previous IM-MS studies on lasso peptides and their branched-cyclic analogs[133]. This result provides further evidence that the new compound, sharing the lasso peptide's mass and observed after incubation at 65 °C, is a branched-cyclic peptide resulting from thermally-induced unthreading of paeninodin. The observation of a double peak for the lasso peptide suggests that paeninodin can adopt multiple, distinct and interchangeable conformations in its native state[129].
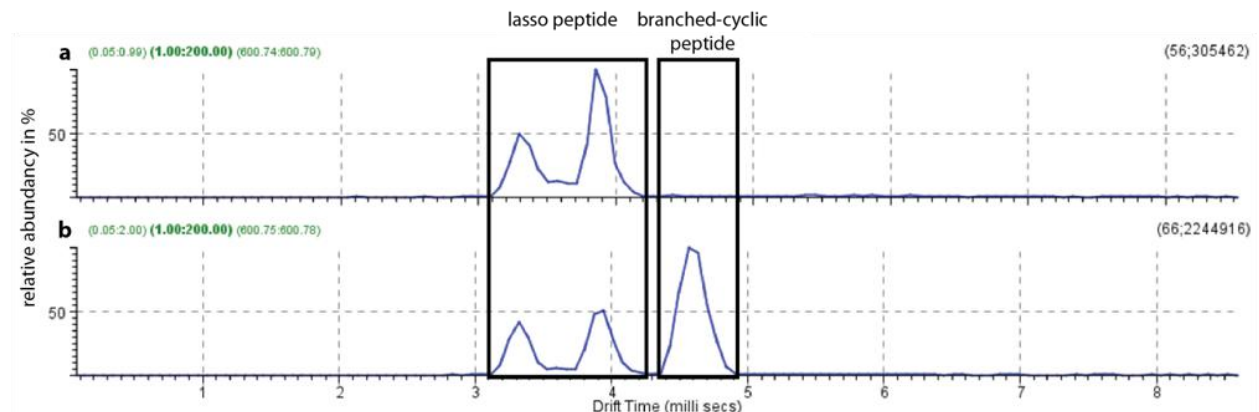
**Figure4.12:** IM-MS analysis of (a) untreated paeninodin and (b) a sample of paeninodin after incubation for 4 h at 65 °C. The peaks of the $[M+4H]^{4+}$ ions corresponding to paeninodin and its branched-cyclic analog (theoretical value = 600.7534) are shown.

### 4.3.2.4 NMR studies and structure elucidation

Efforts were taken to determine the three-dimensional structure of paeninodin. Firstly, a solution of paeninodin in water (~40 mg/mL) was subjected to sparse-matrix screening (Qiagen). As no crystal growth was observed even after 50 days, structure elucidation was attempted via NMR spectroscopy[129]. For this, purified paeninodin was dissolved in five solution variants containing four different solvents: H2O/D2O (9:1), CD3OH, CD3OH/CDCl3 (1:1), CD3OH/CDCl3 (3:1) and DMSO-d6. The lasso peptide concentration was varied from 4.5 to 10.0 mM, while temperature dependence was studied by recording 1H spectra in the range of 278 to 318 K. To our surprise, the 1H spectra obtained (Figure 4.13) showed resonance signals in the amide proton region (6.5 – 9.5 ppm), corresponding to more than one set of paeninodin structures (the primary sequence contains 19 NH) at reduced temperature in all systems tested. Changes in signal line width and number in this region were observed at higher temperatures, and are assumed to be due to chemical exchange in solution[129]. At first, we suspected sample degradation (e.g. unthreading or breaking of the Asp14-Pro15 bond, as shown in Figure 4.10) under the experimental conditions; however, LC-MS analysis, carried out on samples following NMR measurements, discounted this. Therefore, we hypothesize that paeninodin may adopt more than one stable conformation in solution. In addition, we propose that these coexisting conformations undergo rapid chemical exchange, which results in overlapping signals and hinders NMR structure determination. A few natural peptides were shown to assume multiple, stable and distinct conformations; in these cases, NMR structure solution was possible, albeit challenging. As the thermal instability of paeninodin prevented the recording of NMR spectra at higher temperatures, structure determination was impossible for this lasso peptide[129].

Nevertheless, in accordance with numerous previous publications[111,112], a series of analytical assays already show that paeninodin exhibits the typical characteristics of a lasso peptide.
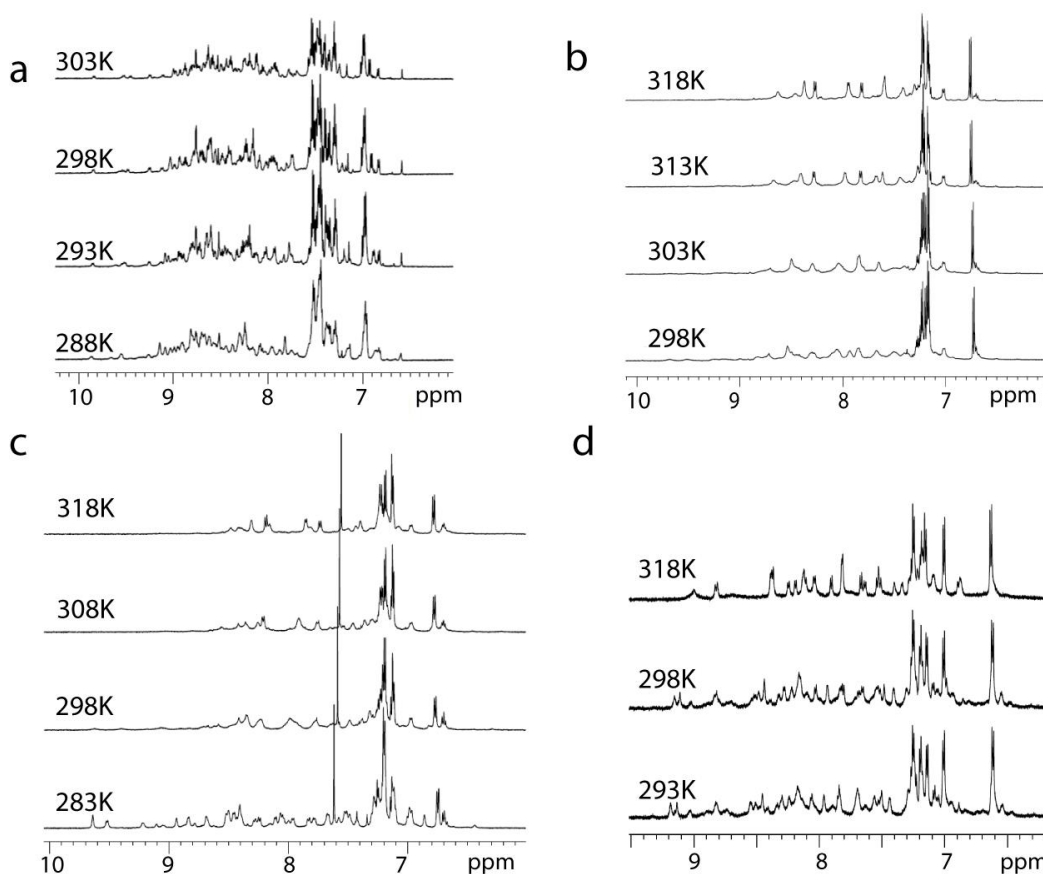
**Figure4.13:**^{1}H-NMR spectra of amide protonsin the 6-11 ppm region of paeninodin, collected at various temperatures in (a)$H_2O/D_2O$ (9:1), (b)$CD_3OH$, (c)$CD_3$ OH/$CDCl_3$ (1:1) and (d)DMSO-$d_6$.

## 4.3.3 Antibacterial Assays

Paeninodin is a lasso peptide originating from a gene cluster featuring an ABC transporter. Past research has shown that only Mccj25, capistruin, lariatin and streptomonomicin, which have an associated ABC transporter, are antimicrobial peptides[105, 125, 127]. This makes sense, since the ABC transporter serves as an exporter, immunity factor and self-protective mechanism. On the basis of that, paeninodin is a potential antimicrobial peptide. The antibacterial activities of paeninodin and phosphorylated paeninodin were evaluated by spot-on-lawn assays against *Bacillus subtilis* MR168, *Micrococcus flavus* ATCC 10240, *Burkholderia rhizoxinica*, *Sphingobium japonicum*, *Xanthomonas citri*, *Bacillus cereus*, *Bacillus megaterium* WSH-002 *and Bacillus amyloliquefaciens*. All the antibacterial assays turned out to be negative. This is probably due to a lack of high-throughput screening tools in our lab or these lasso peptides serve as signal molecules[129]. Further screening for the antimicrobial activity of paeninodin and phosphorylated paeninodin is on the way.

## 4.4 Mutagenesis of the Paeninodin Gene Cluster

Because paeninodin could be synthesized heterologously in *E. coli*, a mutational analysis of the PadeA gene *in vivo* was performed to gain first insights into the biosynthesis and modification position of this lasso peptide. This approach has been extensively applied to study microcin J25, capistruin and other lasso peptides discovered recently to gain insights into their biosynthesis or engineering[111, 114-116, 134]. Based on past experiences, a total of 14 mutants were generated based on carefully selected key residues in the precursor (Figure 4.14).
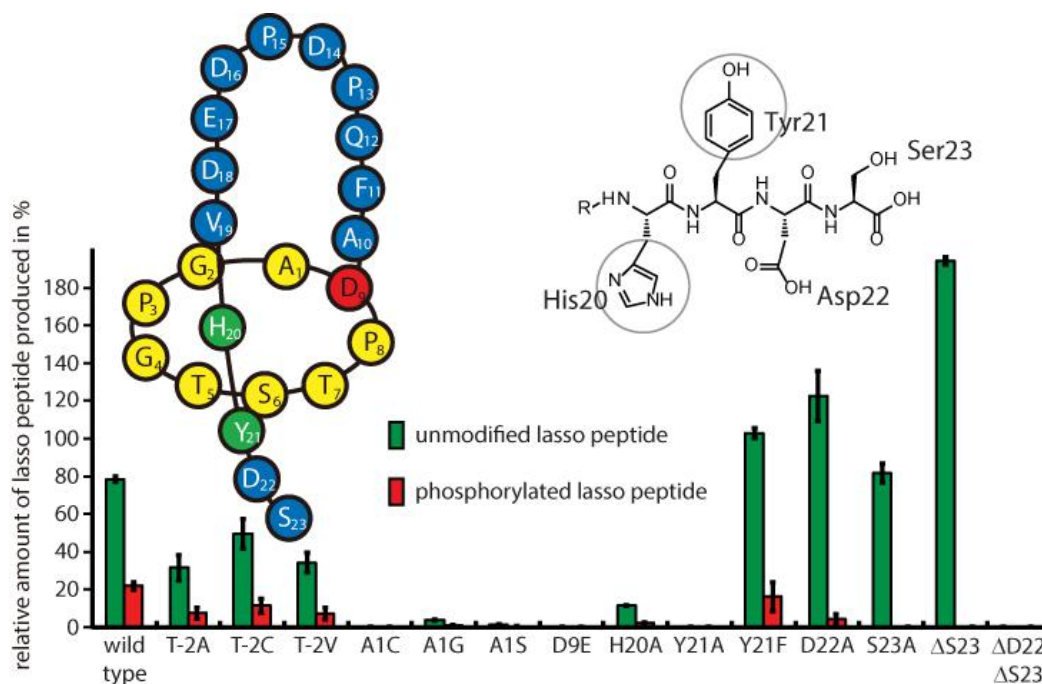


**Figure4.14:** Mutational analysis of padeA and its effect on paeninodin production levels. A production level of 100 % refers to the sum of modified and unmodified paeninodin observed for the wild type system. A schematic of the proposed primary structure of paeninodin is shown above.

Past research has shown that the conserved Thr residue in the penultimate position of the leader peptide of lasso peptides is a recognition element for the lasso peptide maturation machinery[126]. Our mutational analysis shows that the processing enzymes of the paedenin biosynthetic gene cluster tolerate substitutions of the Thr residue at the penultimate position of the leader peptide, but the yield of the lasso peptide production is indeed affected. The T-2A variant produced only 31 % of the wild type, while T-2V produced 33 % of the wild type. In agreement with the studies for caulosegnins lasso peptides, the best tolerated substitution was T-2C, showing a production that was about 50 % of the wild type. On the other hand, all these mutations did not affect the ratios of the modified and unmodified lasso peptide[129].

Paedenin is another rare example of lasso peptide starting with alanine. Although caulonodin and streptomonomicin, recent newly isolated lasso peptides, demonstrate that lasso peptide could feature a serine or alanine at position 1 of the core peptide, no lasso peptide with an Ala-

Asp isopeptide linkage has been reported yet[105, 116]. Before these two examples, the first residue of all lasso peptides reported was limited to Gly or Cys. In accordance with other mutational studies of caulonodin, which is the first example of lasso peptides featuring a serine or alanine residue at the N-terminus, the tolerance for substitution at this position of paedenin was very limited[116]. The A1C mutations abolished the production completely, while, in the cases of A1G and A1S mutations, at least trace amounts of the lasso peptides predicted were still detectable. It is worth noting that in our genome-mining survey, we noticed that most of the potential lasso peptide in bacilli prefers starting with an alanine, indicating that this is a key residue specifically recognized by the processing enzymes of this class of lasso peptide. Considering the highly conserved sequence of the precursor in bacilli, these lasso peptide gene clusters are probably originally from the same ancestor. The D8E variant was also not produced, suggesting that the macrolactam ring size is also important for enzymatic recognition[129].

We then turned our attention to the tail to get further information about the so-called plug amino acids for maintenance of the lasso fold. Since paeninodin is a nine-ring lasso peptide, the residue needed to trap the tail must be rather large, such as aromatic amino acids. It is possible to get hints by mutagenesis about which amino acid actually served as a plug. Hence, a set of mutants focusing on the last four amino acids was created *in vivo*. The alanine scan of the last four amino acids indicated that Y21 is probably responsible for the entrapment of the tail, since only this mutation Y21A abolished the production completely. Meanwhile, the mutation H20A was only produced at 11 % of the wild type. On the other hand, the Y21F, D22A and S23A variants are produced at a level more or less the same as the wild type[129]. We guess that the H20 and Y21 act as the plug amino acids on opposite sides of the ring, forming a sandwich-like structure, as it is the case with microcin J25[125].

Another advantage of these mutational studies is that we could have a first glance at the substrate specificity of the kinase. The kinase could modify the H20A, Y21F and D22A variants, indicating it could tolerate substitutions of this region. Moreover, we already know from the mass spectrometric results that the phosphate group is at the last serine residue, but we do not know whether it is at the side chain hydroxyl group or at the carboxyl group at the C-terminus. Mutational studies could give us a first clue. All the mutations at the S23 position could produce unmodified lasso peptide at the same level as the wild type, but no modified lasso peptide variants could be observed on the UV level. These mutations almost abolished the production of the modified lasso peptide; it is also highly likely that the modification happens at the side chain hydroxyl group of the serine[129]. However, further experiments are still needed to prove the modification position.

## 4.5 Modification Pathway

### 4.5.1 Recombinant production and purification of active PadeK kinase

After purifying paeninodin and phosphorylated paeninodin and figuring out their topology, we come to the question, how does the modification happen? Due to the unusual knot structure of lasso peptide, the first hypothesis is that lasso peptide is, firstly, synthesized by the mature enzymes and then modified by the kinase to yield the final products. In fact, a new protease-AtxE2 has recently been shown to be a novel lasso peptide isopeptidase that hydrolyzes astexins-2 and -3 specifically and converts them to linear peptides[113]. This enzyme works by recognizing the lasso topology rather than a specific amino acid sequence. Inspired by this example, we first investigated whether this enzyme had any activity toward paeninodin lasso peptide.

Consequently, an *in vitro* study was required. Unfortunately, initial attempts to get soluble PadeK kinase failed. Therefore, a Maltose-binding protein (MBP)-fusion tag was used to facilitate the purification (Figure 4.15). The gene was cloned into pETMBP_1a vector and introduced into *E. coli* BL21 (DE3) cells for protein overproduction (Figure 4.15).
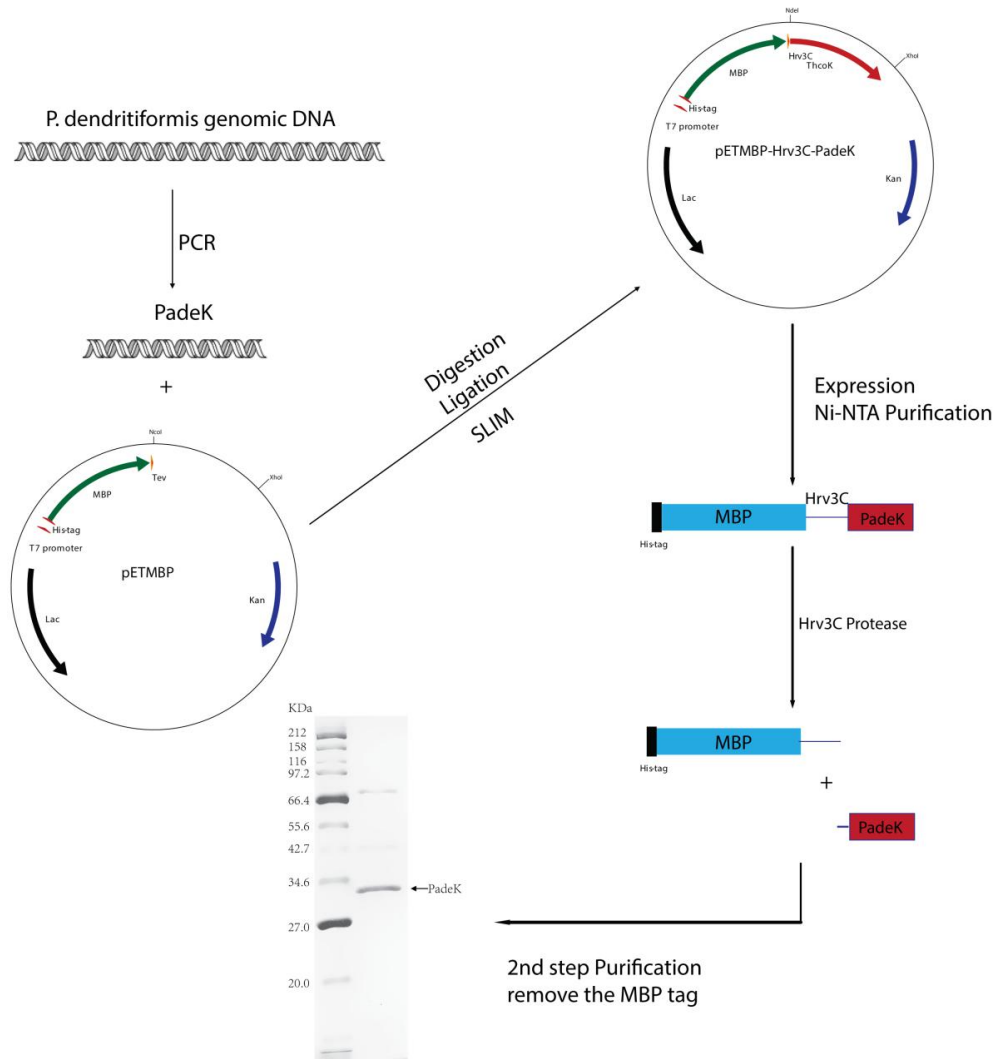
**Figure4.15:**Methods for the production of PadeK. A Maltose-binding protein (MBP)-fusion tag (with N-terminal His$_6$-tag) was used to facilitate purification. An HRV-3C protease site was introduced between the MBP-fusion tag and PadeK.

The MBP-PadeK was then purified by a Ni-NTA column and treated with Hrv-3C protease for tag removal. PadeK is notably difficult to handle, because it tends to form aggregates with itself and MBP[129] (Figure 4.16).
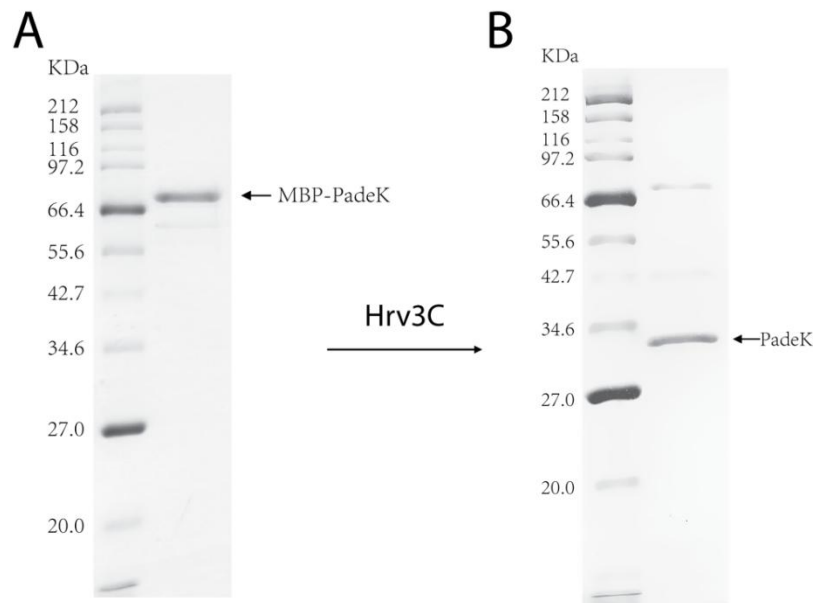
**Figure4.16:** (A) SDS-PAGE gel of purified MBP-PadeK (lane 2, 67 kDa) and protein marker (lane 1). (B) SDS-PAGE gel of purified PadeK (lane 2, 32 kDa) and protein marker (lane 1).

## 4.5.2 Recombinant production and purification of GP-PadeA precursor peptide

We hypothesized that the kinase, firstly, modifies the precursor to form modified precursor, then the mature enzymes transfer the modified precursor into the final lasso peptide. To prove this hypothesis, the precursor peptide PadeA should be produced first. The precursor peptide is hard to synthesize by chemical methods, therefore, we produced the PadeA via a heterologous expression method[88]. The PadeA gene was, firstly, amplified from the chromosomal DNA of *P. dendritiformis* C454. The Pade A gene was then cloned into a pET-48b(+) vector, resulting in a pET-48b(+)-PadeA vector (Figure 4.17). In this way, the PadeA is fused with a thioredoxin gene, which is connected by the sequence of the Hrv-3C protease cleavage site. Subsequently, *E. coli* BL21 (DE3) cells were transformed with the resulting pET-48b(+)-PadeA plasmid. The subsequent expression of the thioredoxin PadeA fusion protein was carried out in LB medium.
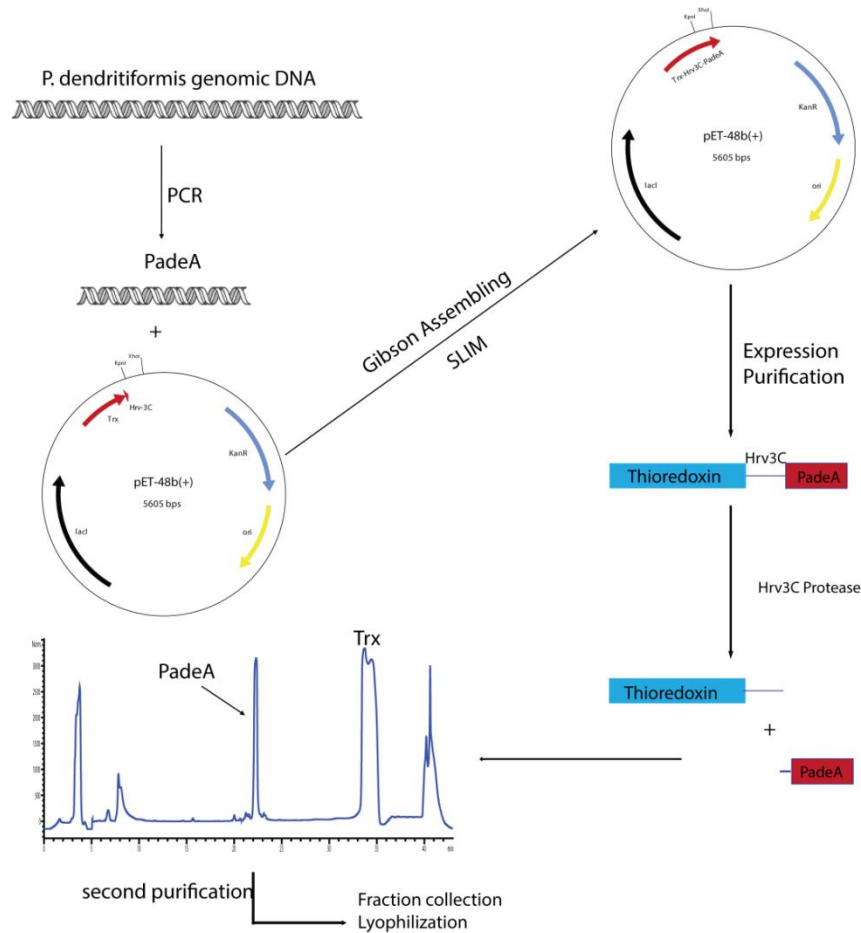
**Figure4.17:**Methods for the production of GP-PadeA and its variants. A Trx-fusion tag was used to facilitate purification. An HRV-3C protease site was introduced between the Trx-fusion tag and PadeA. After proteolytic cleavage, GP residues remained on the N-terminus of the PadeA precursor peptide. The identity of the resulting GP-PadeA was confirmed by MS.

After purification by Ni-NTA, the fusion protein was cleaved by digestion with the Hrv-3C protease. As a result of a final purification by HPLC, a PadeA derivative with two additional glycine and proline residues at the N-terminus (GP-PadeA) could be obtained[129] (see Figure 4.18).

**Figure4.18:**Chromatograms of the GP-PadeA purification by preparative HPLC. (A) First purification with water/0.1 % TFA and MeCN/0.1 % TFA. (B) Second purification with water/0.1 % TFA and MeCN/0.1 % TFA and a different gradient.

The base sequence of pET-48b(+)-PadeA plasmid was mutated using the SLIM method for the synthesis of GP-PadeA variants. Subsequently, the newly obtained pET-48b (+) PadeA variant vectors were transformed into *E. coli* BL21 (DE3) cells. The respective thioredoxin PadeA variants fusion protein was expressed in LB medium purified by $Ni^{2+}$-NTA affinity chromatography and cleaved using the Hrv3C protease. The final purification of the precursor peptide variants was achieved by preparative HPLC[129].

## 4.5.3 Recombinant production and purification of active ThcoK kinase

Due to difficulties isolating PadeK, we searched our genome-mining results for a suitable homologue. We found that a kinase from the cluster of *T. composti* KWC4 (ThcoK) could be a good candidate. The high degree of sequence similarity of precursor peptides padeA and thcoA (~ 64 % identity) and associated kinases (~ 41 % identity; ~ 58 % similarity) was expected to facilitate homologous exchange[129].

The ThcoK gene was amplified directly from the chromosomal DNA of *T. composti* KWC4. The ThcoK gene was then cloned into a pET-41a(+) vector, resulting in a pET-41-ThcoK vector. ThcoK was then purified by a Ni-NTA column and size exclusion chromatography (Figure 4.19). It was notable that even though the ThcoK is super soluble, it tends to form aggregates with itself when the temperature is changed drastically. We handled this enzyme consistently at room temperature[129].
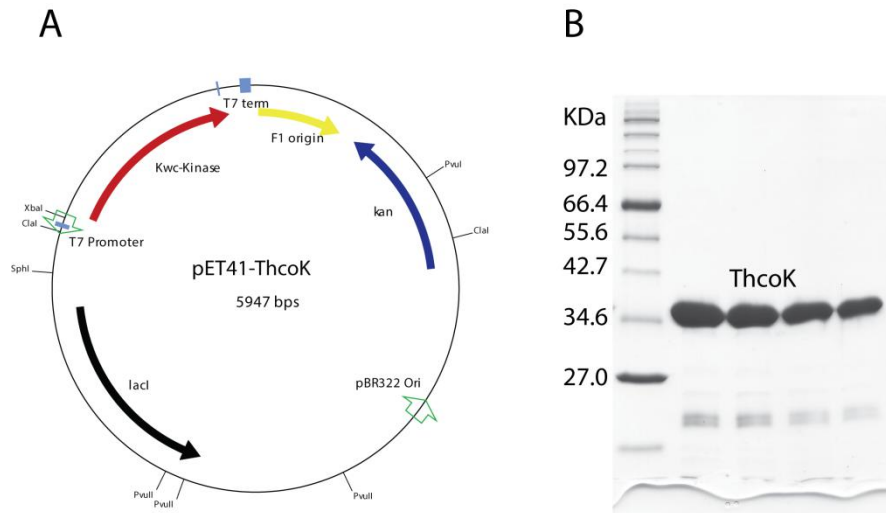
**Figure4.19:**(A) Map of the pET41a-thcoK plasmid used for the expression of His-tagged ThcoK. (B) SDS-PAGE gel of purified ThcoK (lane 2, 35 kDa) and protein marker (lane 1).

## 4.5.4 Exchange of the ThcoK kinase into the paeninodin biosynthesis gene cluster

After we had demonstrated that ThcoK is a more soluble enzyme and easier to handle, we needed to further prove that ThcoK is suitable for homologous exchange. In order to prove that Thcok is able to modify the paeninodin lasso peptide, we exchanged the ThcoK kinase into the paeninodin biosynthesis gene cluster by SLIM mutagenesis. We replaced the PadeK directly with this kinase (ThcoK) in the pET41a-PadeCAKB1B2D vector, resulting a new construct called pET41a-PadeCA(ThcoK)B1B2D and transformed it into an *E. coli* BL21(DE3) cell. New fermentation was performed and, by using the same procedure, extracts from these cultures were analyzed for the presence of phosphorylated paeninodin (Figure 4.20). By incorporating with this kinase, we noticed surprisingly that the modified lasso peptide is now the major product in the extract (≈ 80 %). This demonstrated that the kinase in the lasso peptide gene clusters is changeable[129].
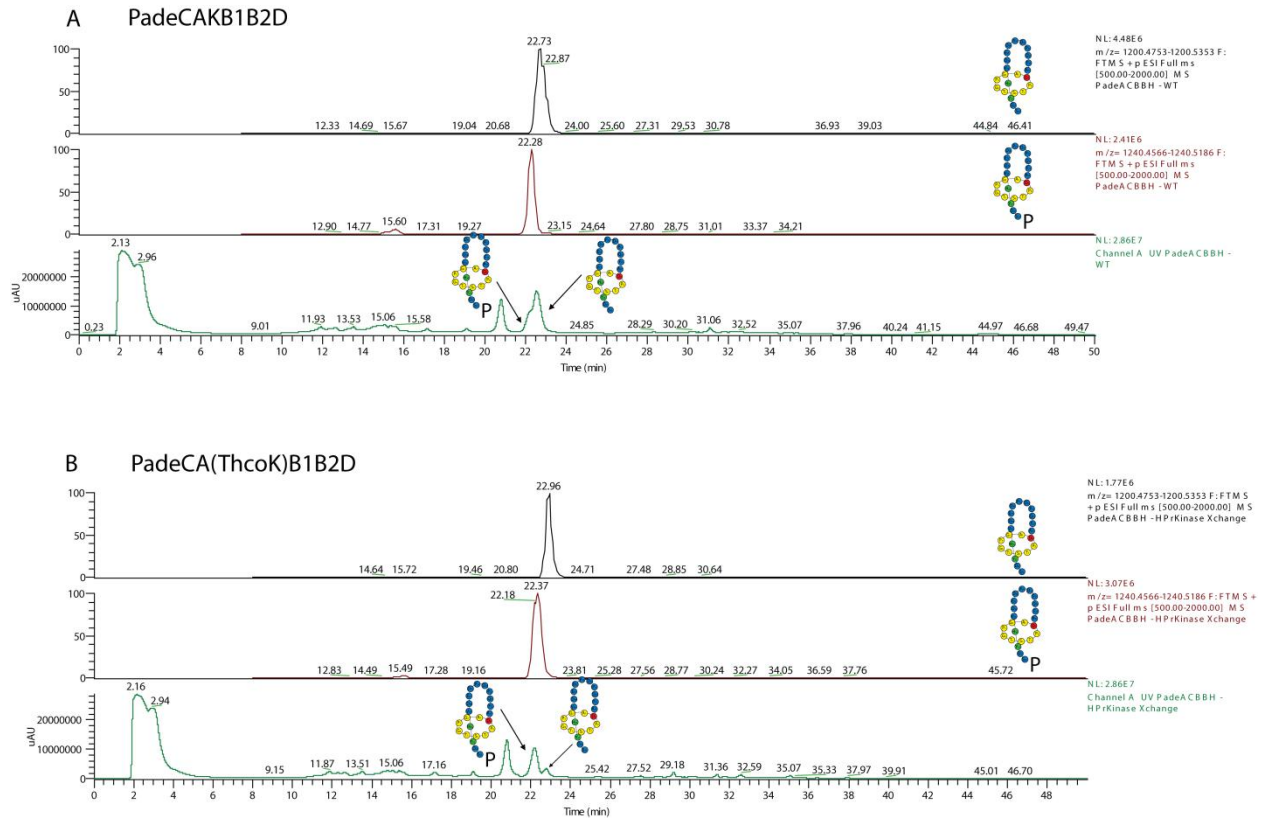
**Figure4.20:** HPLC analysis and extracted ion chromatograms of pellet extracts from pET41a-padeCAKB1B2D and pET41a-padeCA-thcoK-padeB1B2D fermentations.

## 4.5.5 The kinase only modifies the precursor peptide, not the mature lasso peptide

After we had obtained two soluble kinases (PadeK and ThcoK) and two potential substrates (PadeA and paeninodin), we came to the question, how does the modification happen? Due to the unusual knot structure of lasso peptide, the first hypothesis is that lasso peptide is firstly synthesized by the mature enzymes and then modified by the kinase to yield the final products. We, initially, investigated whether this enzyme had any activity toward paedenin lasso peptide. Firstly, PadeK was employed in an *in vitro* assay using paedenin as a substrate. A standard reaction (50 µL) *in vitro* generally contained Tris ·HCl (50 mM, pH 8.0), MgCl$_2$ (5 mM), ATP (1 mM), substrate (50 µM) and enzyme (5 µM). Surprisingly, no activity was observed. This led us to think of the second hypothesis, that the kinase initially modified the precursor to form modified precursor, then the mature enzymes transferred the modified precursor into the final lasso peptide. To prove this hypothesis, another *in vitro* assay using GP-PadeA was performed. This time, trace amounts of the modified precursor predicted were detected[129] (Figure 4.21).
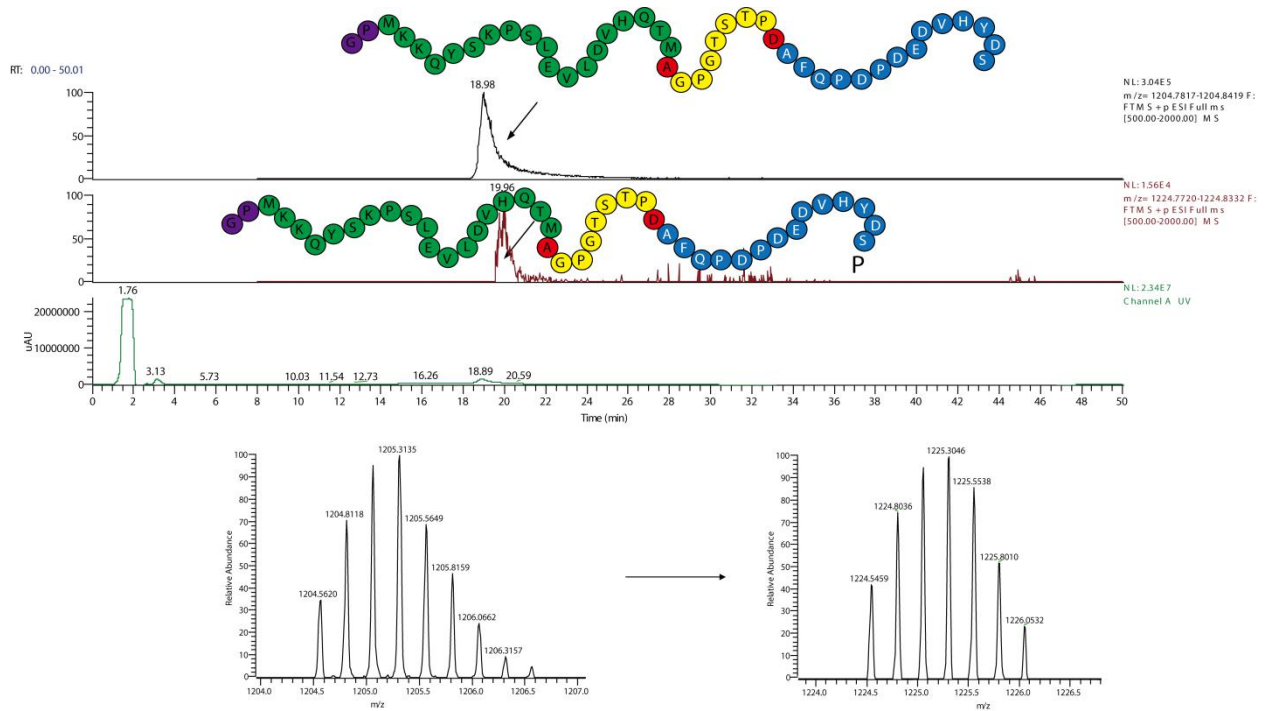
**Figure4.21:**Extracted ion chromatograms and mass spectra of GP-PadeA and GP-PadeA-OPO$_3$$^{2-}$ from the assay with PadeK.

We then directly performed mass spectrometric analysis of the products. MS2 fragmentation and further MS3 fragmentation of the y-11 fragment confirmed that the kinase could actually modify the precursor peptide (Figure 4.22). However, due to the low solubility of the kinase, the activity of the enzyme is very low, which hinders our further studies of the kinase. This could also explain why the modified lasso peptide is only 20 % of the unmodified lasso peptide in the cell extract.
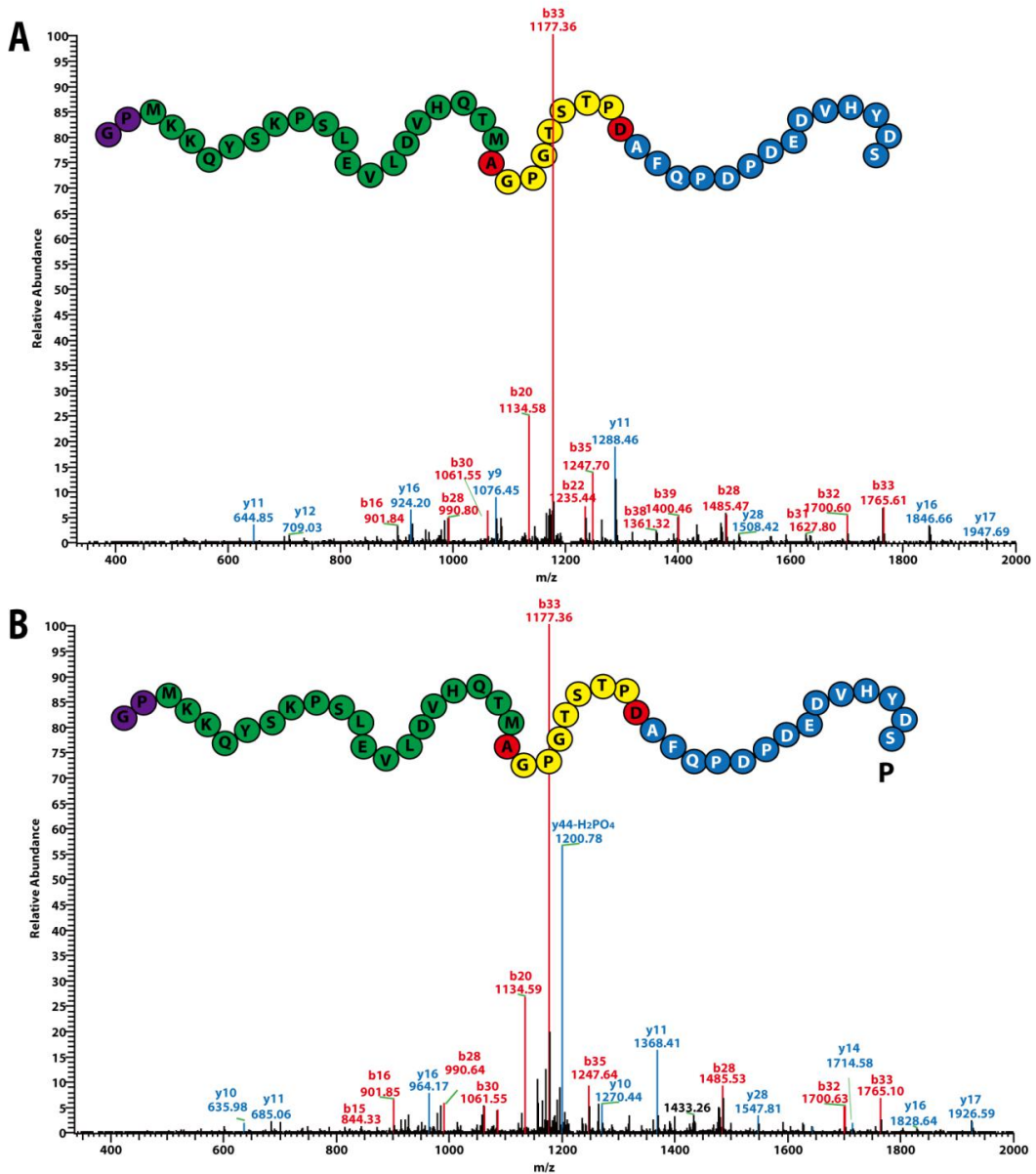
**Figure4.22:**MS$^2$ spectra of GP-PadeA and GP-PadeA-OPO$_3$$^{2-}$. The latter resulted from the assay of PadeK and GP-PadeA.

As shown before, we demonstrated that the kinase in the lasso peptide gene clusters is changeable. Since ThcoK kinase could also effectively modify the peadenin lasso peptide, we then focused on this kinase and repeated the assays. Therefore, another *in vitro* assay using GP-PadeA and ThcoK was performed in the standard assay. As expected, the ThcoK could effectively transform the precursor into modified precursor with a yield of about 80 % (Figure 4.23). Further MS analysis confirmed the products are indeed modified precursor with a phosphate group at the last serine, as it appeared to be in the lasso peptide. We once again used peadenin lasso peptide as a substrate and repeated the assay with new ThcoK kinase and again we did not observe any activity[129] (Figure 4.23).
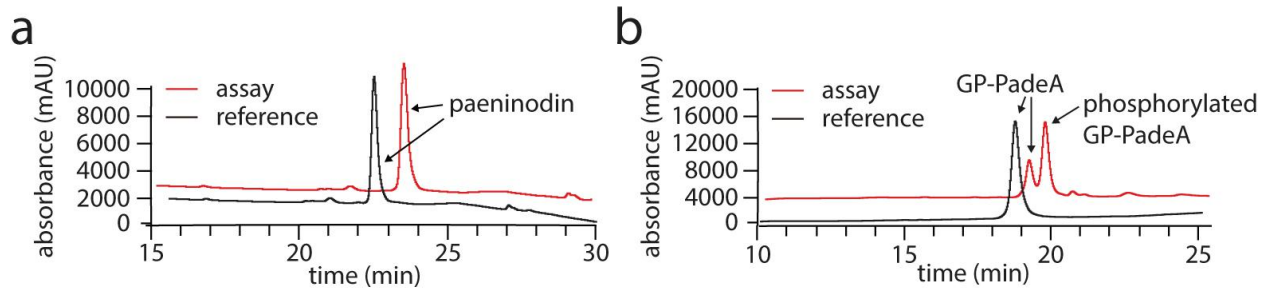
Figure4.23: Phosphorylation assay using ThcoK and (a) paeninodin or (b) GP-PadeA.

In conclusion, we clearly showed by the *in vitro* studies that the kinase could only modify the precursor peptide of the lasso peptide. The kinase phosphorylated the C-terminal Ser of the precursor peptide selectively prior to processing by the B1, B2 and C homologues, thus yielding a phosphorylated lasso peptide (Figure 4.24).
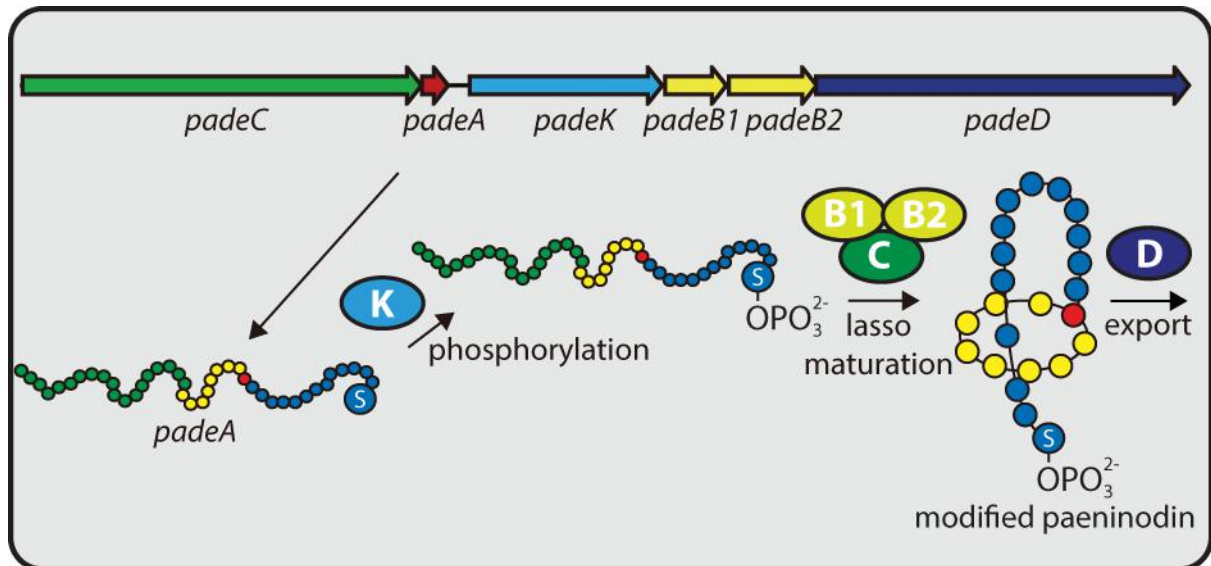


**Figure4.24:**Proposed schematic of paeninodin biosynthesis. Precursor peptide PadeA is phosphorylated at its C-terminal Ser by PadeK, and is then processed into the mature, phosphorylated lasso peptide by PadeB1, PadeB2 and PadeC. Cellular export of the lasso peptide is accomplished by ABC transporter PadeD.

## 4.6 Biochemical Characterization of PadeK and ThcoK Kinase.

### 4.6.1 Bioinformatic analysis of the kinases

In order to characterize the PadeK and ThcoK kinases, we firstly performed a bioinformatic analysis on them. Results showed that parts of the lasso peptide precursor kinase shows 41 % identities with an Hpr kinase (HprK) from *Lactobacillus casei* and an ATP-dependent phosphoenolpyruvate carboxykinase (PCK) from *Thermus thermophilus* Hb8[135-138] (Figure 4.25). These parts are likely to be the active center of the kinase. Past research has showed that PCK and HprK have similar ATP-binding domains and active sites. PCK and HprK generally have similar and conserved catalytic residues of HKDD at the active center. It was shown that

H140K161D178D179 are the key residues in the catalytic process of the crystal structure complex of Hpr kinase (HprK) from *L. casei* with its substrate HPr. Asp-179 is the catalytic base supported by His-140 to deprotonate the Ser-46 hydroxyl group, which is a prerequisite for its phosphorylation. Lys-161 is likely to play a part in the catalysis by neutralizing the negative charge that develops on the phosphate group during the transfer reaction and lowering the energy of the transition state, and Asp-178 is likely to play a role in forming the metal-binding site. All these residues of the HprK active site have their equivalent in PCK and play a similar role in the catalytic process[135-138].
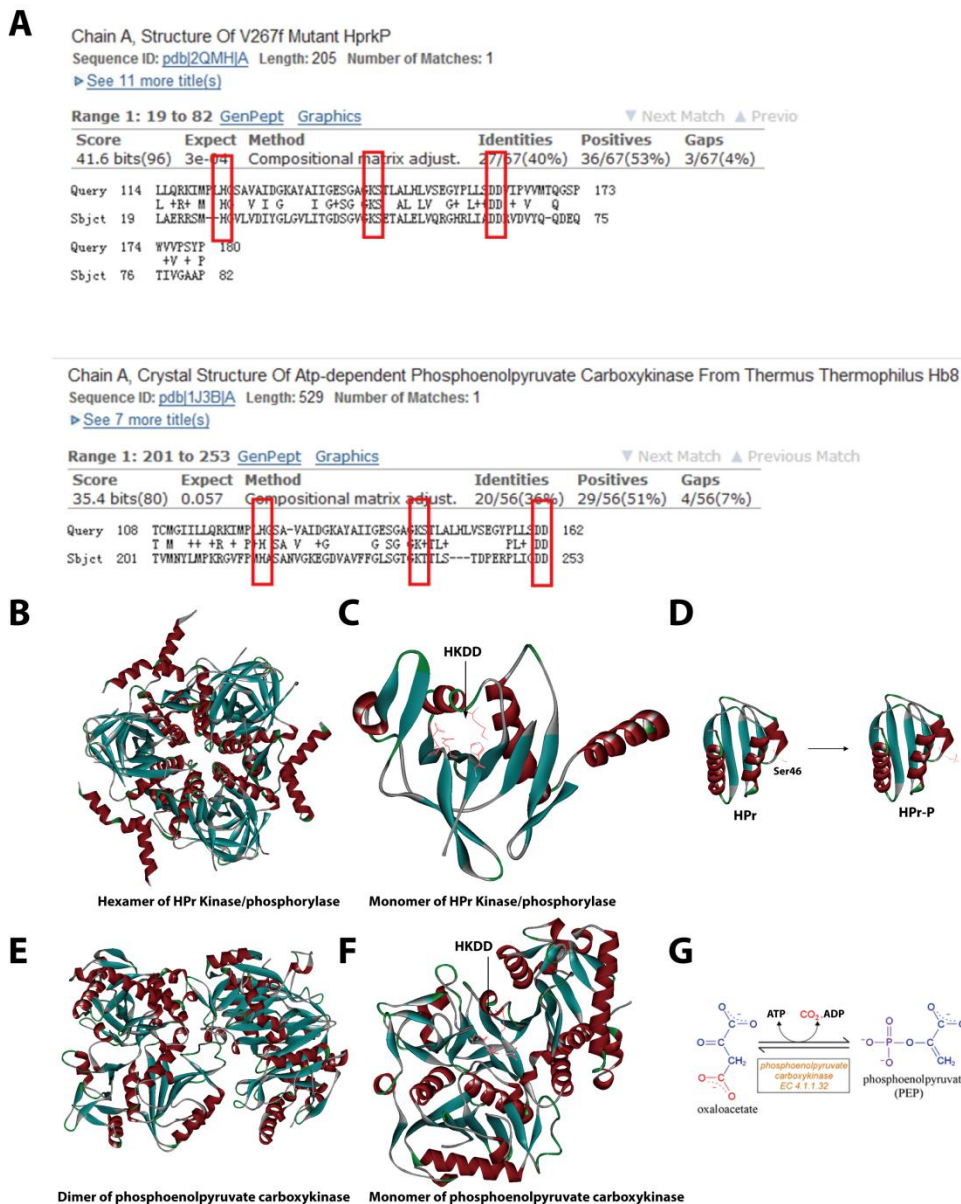


**Figure4.25:** Comparison of ThcoK with kinases of known structures. (A) Alignment of the active site of ThcoK with other kinases found in the PDB. (B) Hexameric and (C) monomeric structures of the HPr kinase/phosphorylase from Lactobacillus casei (PDB: 2QMH). (D) Reaction catalyzed by the HPr kinase/phosphorylase of L. casei. (E) Dimeric and (F) monomeric structures of the phosphoenolpyruvate carboxykinase from Thermus thermophilus Hb8 (PDB: 2PC9). (G) The reaction catalyzed by the phosphoenolpyruvate carboxykinase of T.thermophilus Hb8.

We then used MEME to generate conserved domain motifs for all the kinases identified. All the kinases shared two main motifs. In addition, a third motif was identified in the kinases from Firmicutes but not Proteobacteria (Figure 4.26).
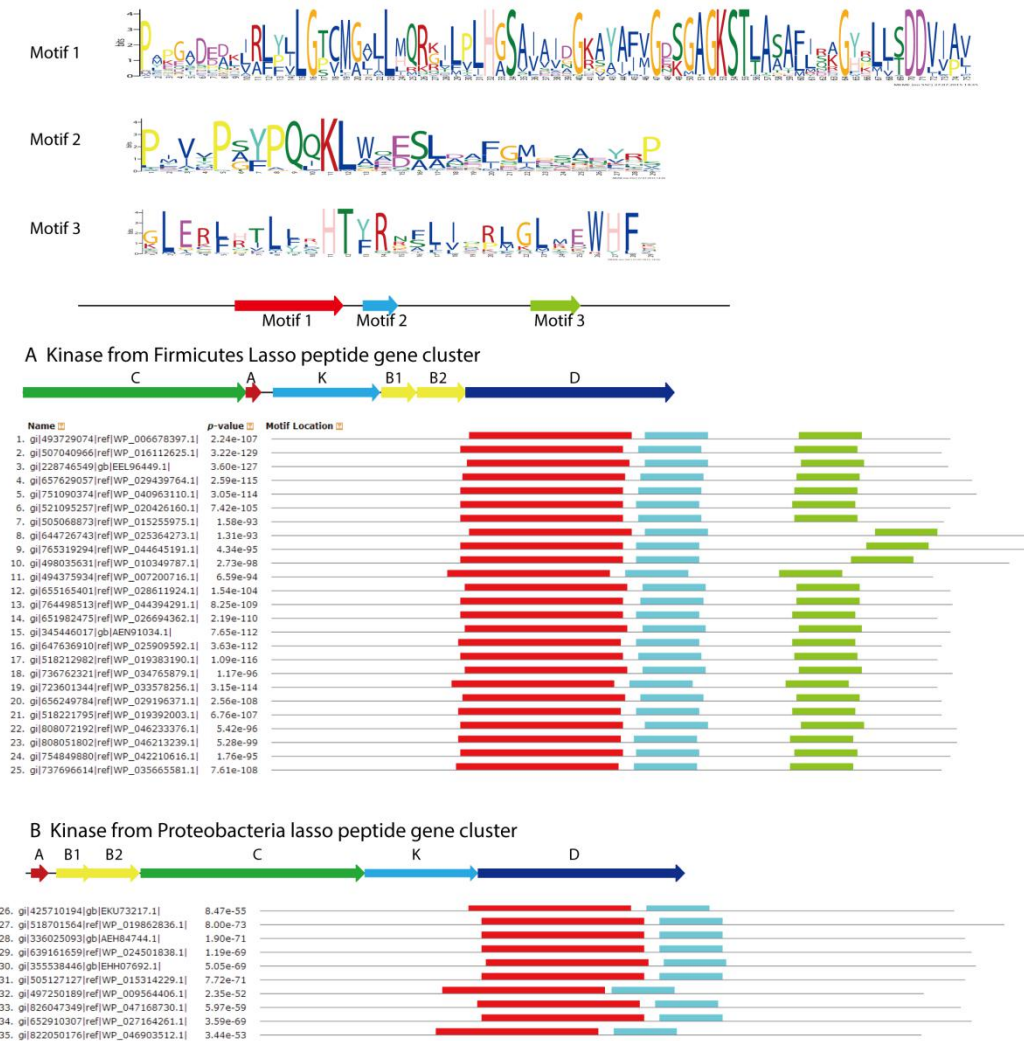


**Figure4.26:**Common motifs in the kinases found in lasso peptide biosynthetic gene clusters identified by comparing 25 kinase sequences from Firmicutes and 10 kinase sequences from Proteobacterial clusters using the MEME algorithm. The locations of the different motifs in each enzyme are shown to approximate scale.

As expected, the first big motifs containing the conserved catalytic residues of the HKDD are highly conserved in all lasso peptide precursor kinases (Figure 4.27), indicating it has a similar catalytic mechanism with HprK and PCK[129].

87

**Figure4.27:** Multiple sequence alignment of the catalytic active site proposed in the first conserved motif of kinases from lasso peptide biosynthetic gene clusters. The essential His-Lys-Asp-Asp motif is highlighted by red boxes.
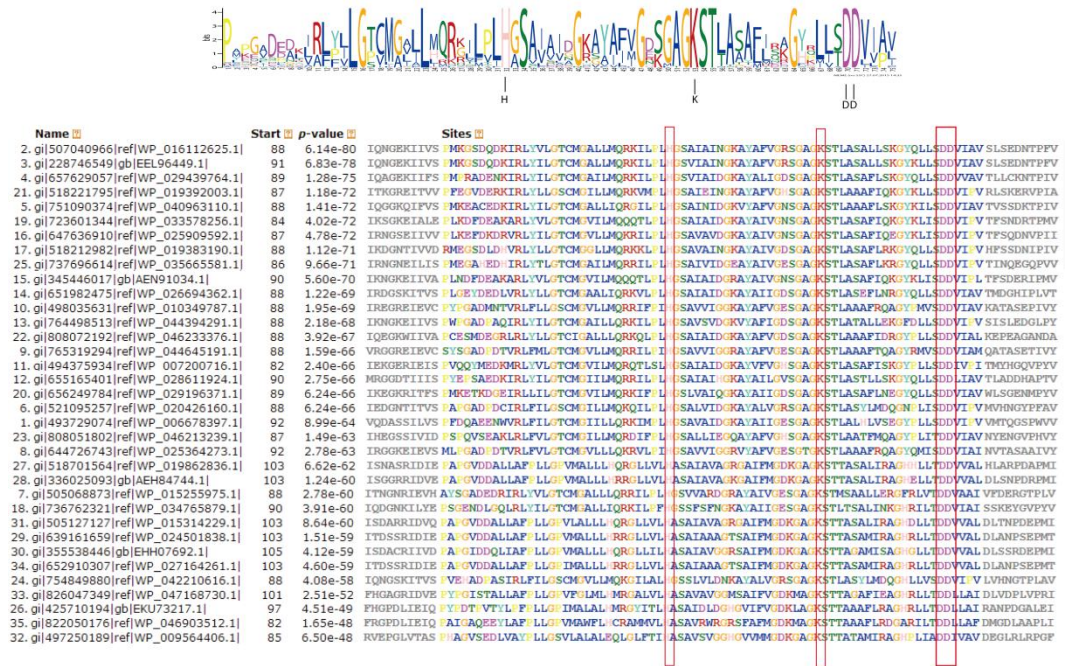
## 4.6.2 Kinetic Parameters for ThcoK

After the bioinformatic analysis, we turned to study the kinetic properties of the kinase. Since PadeK kinase is insoluble and hard to handle, we only investigated the Thcok kinase. We used the GP-PadeA as the substrate and performed the measurement in a standard reaction assay with a different substrate concentration. HPLC-MS was used for quantitative determination of the modified and unmodified precursor peptide. Firstly, a time course analysis of the phosphorylation of GP-PadeA was conducted. After incubation for 4 h with ThcoK, formation of the final modified precursor had almost come to a maximum value. The highest conversion ratio is about 74 %. We carried out standard assays to further calculate Kcat and Km associated with precursor phosphorylation while varying the concentration of GP-PadeA at a constant concentration of ThcoK at 37 °C for 20 min. For the phosphorylation of GP-PadeA, an apparent $K_m$ of 279.3 ± 19.7 µM and a $K_{cat}$ of 0.41 ± 0.02 min$^{-1}$ were determined[129] (Figure 4.28).
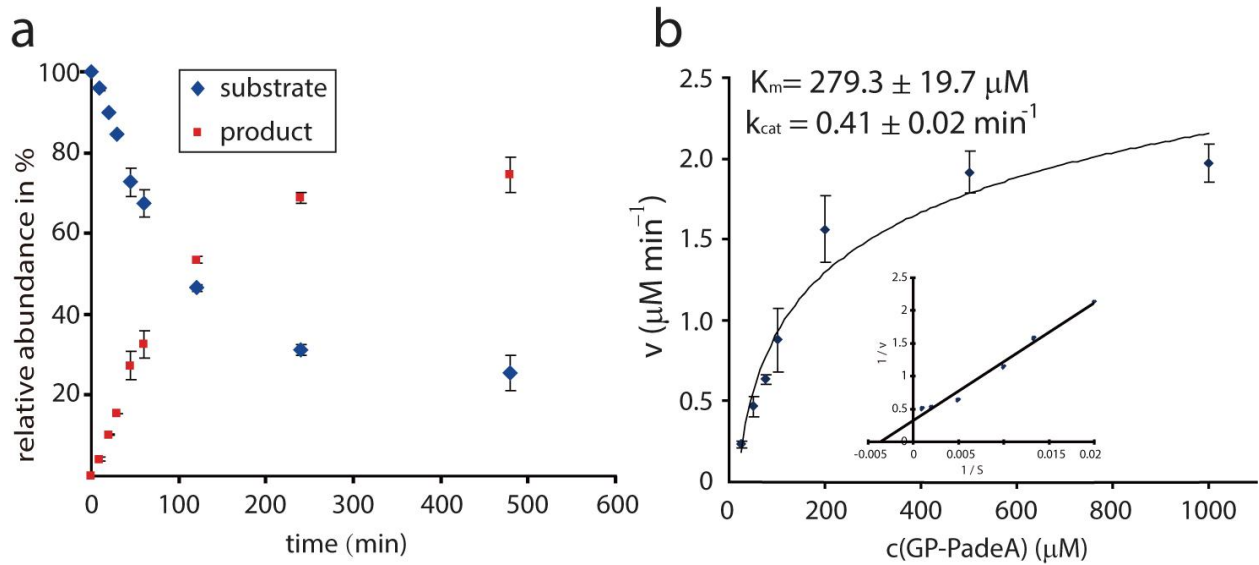
**Figure4.28:**Exemplary time dependency ThcoK-catalyzed phosphorylation of GP-PadeA under assay conditions. (i) Michaelis-Menten plot used to determine the $K_m$ and $k_{cat}$ values.

### 4.6.3 Catalytic mechanism of the Kinase

We hypothesized by bioinformatic analysis that His-Lys-Asp-Asp is the essential catalytic motif for both kinases. To confirm this hypothesis, *in vivo* studies were performed by SLIM mutagenesis. We chose these four key residues in both PadeK and ThcoK and carried out alanine scans (Figure4.29). Expression cultures were grown under normal conditions and their extracts were analyzed by high-resolution LC-FT-MS. No phosphorylated lasso peptide was observed for any variant, although the overall yields of lasso peptide remained unaffected. This proved that the lasso peptide precursor kinase has a similar mechanism and each of the four residues is crucial to catalysis[129] (Figure 4.29).

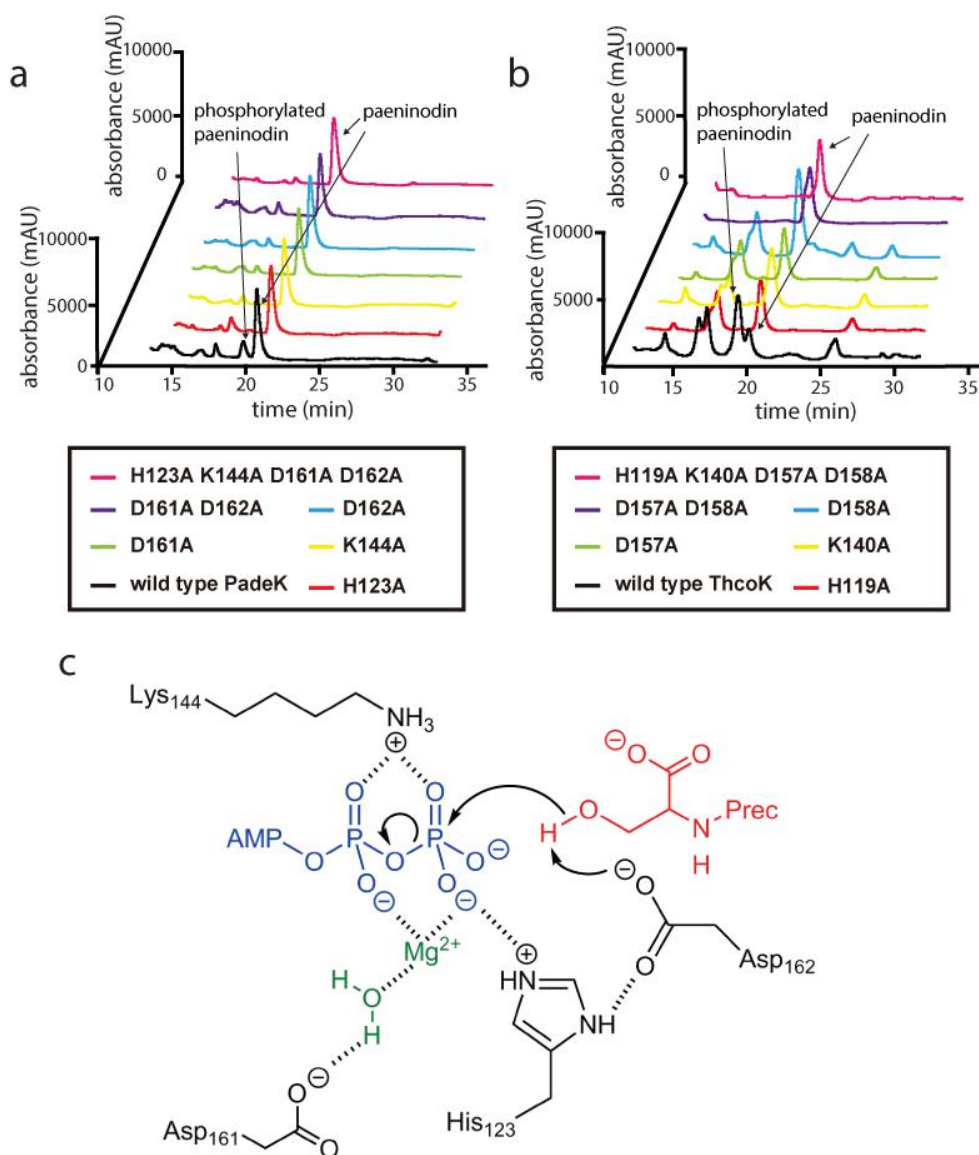**Figure4.29:** Mutational analysis study of the catalytic essential residues in the kinases (a) PadeK and (b) ThcoK. (c) Proposed mechanism for the phosphorylation of the C-terminal Ser of a precursor peptide by PadeK.

## 4.6.4 Substrate specificity of ThcoK

Most natural products produced by posttranslationally modified peptides (RiPPs) are synthesized initially as a long precursor peptide. These precursors typically contain N-terminal leader peptides and C-terminal core peptides[40]. The leader peptide which is appended to the N-terminus of the core peptide is usually important for recognition by many of the posttranslational modification enzymes. Since the kinase here firstly modified the precursor peptide, we initially wanted to investigate whether this modification is a leader peptide dependent or not. A leaderless peptide (PadeA(2-23)) and a shorter truncation peptide (PadeA (13-23)) were generated b SLIM mutagenesis. Both peptides were treated with ThcoK kinase and, surprisingly, both could be effectively phosphorylated[129]. This indicates that neither the leader peptide nor the first half of the core peptide is essential for the modification. With this information, we

generated several PadeA variants and synthesized several derivatives of the truncation peptide of PadeA to investigate the substrate specificity of this unusual kinase further.
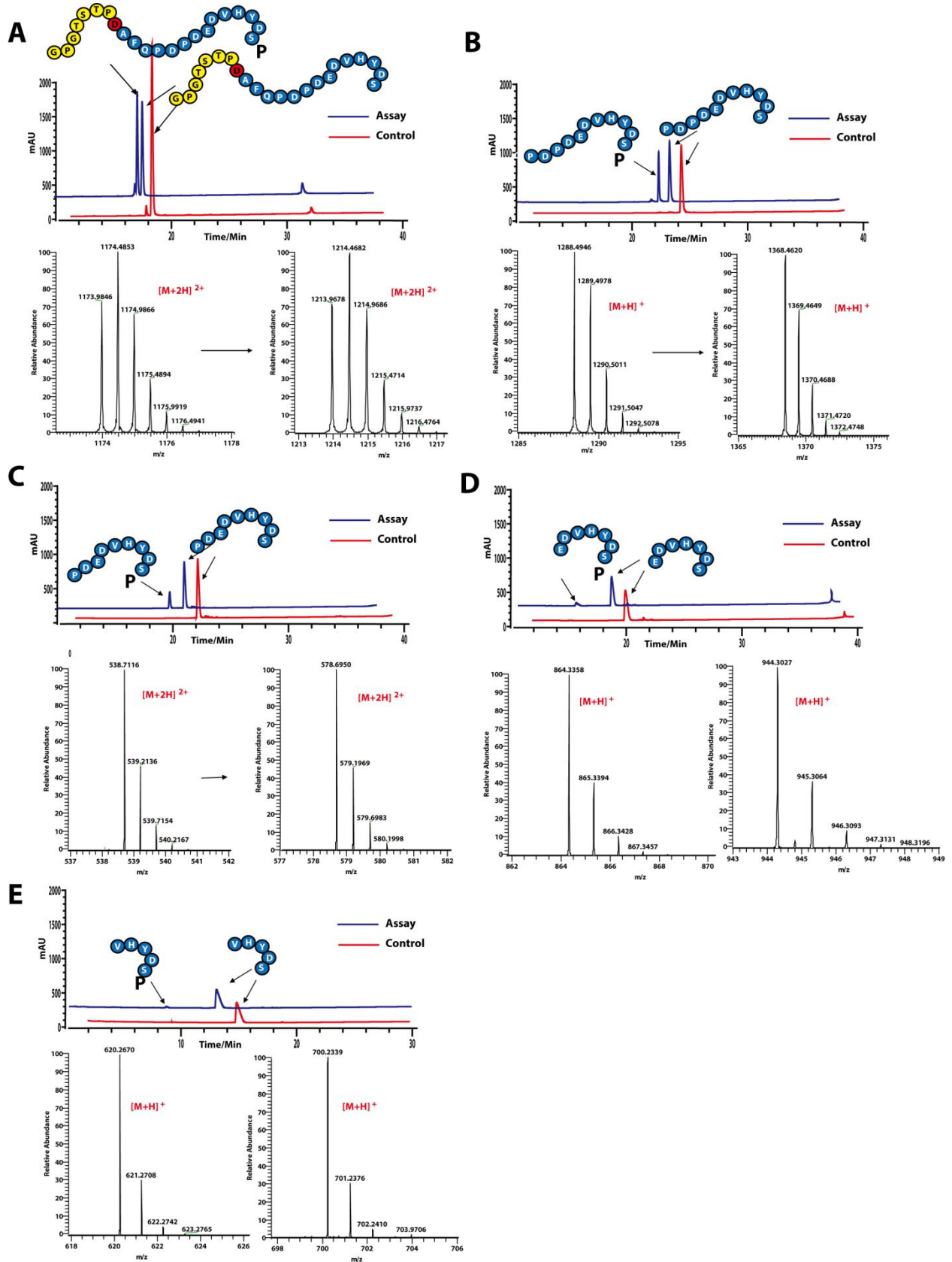
**Figure4.30:**HPLC analysis and observed masses after *in vitro* assays. ThcoK was assayed with the substrates (A) PadeA(2-23), (B) PadeA(13-23), (C) PadeA(15-23), (B) PadeA(17-23) and (E) PadeA(19-23).

Obviously, with shorter length peptides, the conversion rate decreases gradually (Figure 4.30). This indicates that all parts of the precursor peptide contribute to the activity in such a way that the binding affinity between ligand and the kinase is affected.

We then generated additional variants of full-length GP-PadeA, wherein the three residues before Ser23 or the seven residues before the HYDS sequence were exchanged with the same number of alanines. As shown in Figure 4.31, even though long stretches of the amino acid sequence were altered in these variants, the corresponding decreases in phosphorylation efficiency were rather small[129].
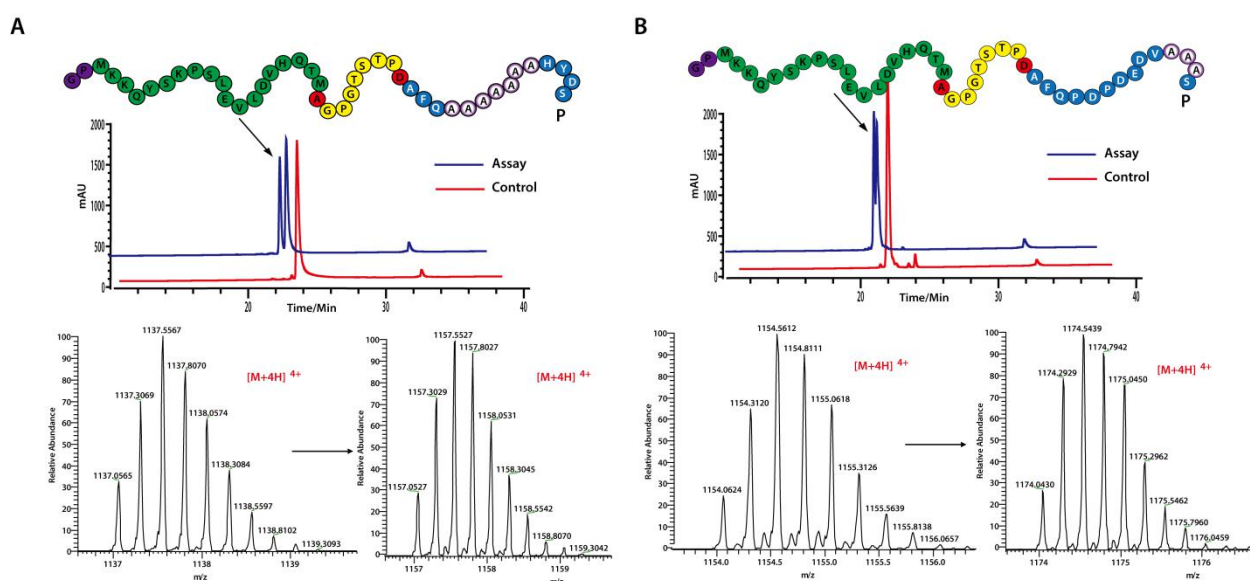


**Figure4.31:**HPLC analysis and observed masses after *in vitro* assays. ThcoK was assayed with the substrates (A) GP-PadeA 7xA-HYDS and (B) GP-PadeA 3xA-S.

We have already shown that mature lasso peptide was not able to be modified by ThcoK. However, we were curious whether the branched-cyclic peptide, whose C-terminus is not shielded by the macrolactam ring, could be phosphorylated. To test this, pure paeninodin was first heated at 65 °C for 4 h to generate a branched-cyclic peptide and then the mixture was incubated with ThcoK under assay conditions. To our surprise, a significant amount of phosphorylated, branched-cyclic peptide was observed (Figure 4.32). Because the lasso fold itself appears to hinder access of the kinase to the C-terminal Ser, the only feasible route to production of phosphorylated paeninodin involves modification of the precursor peptide prior to processing[129].
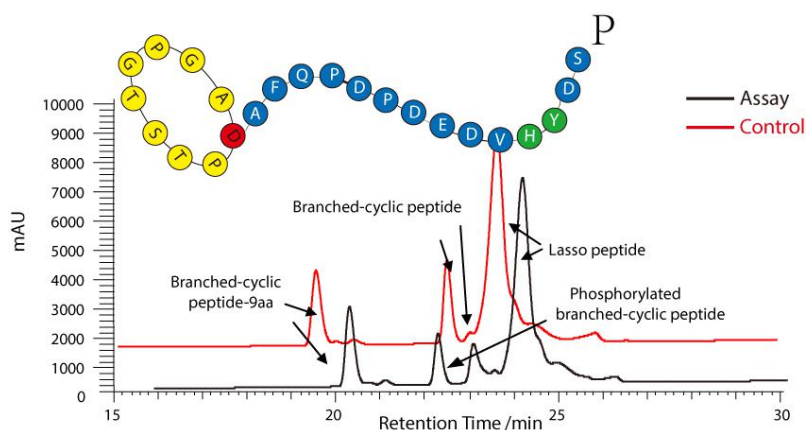
**Figure4.32:** HPLC analysis of an *in vitro* assay with ThcoK and a paeninodin sample after incubation at 65 °C for 4 h. A measurement of a sample of thermally treated paeninodin is shown as reference.

We then tried to investigate the C-terminal to study the substrate specificity of the kinase. As shown in Figure 4.33, no modification was observed for a S23A substitution in GP-PadeA, emphasizing the absolutely essential role of the C-terminal Ser for enzyme activity. The PadeA(13-23) peptide consisting of the last 11 residues of the precursor peptide can be quickly and easily accessed by chemical synthesis. We, therefore, selected it as a scaffold for more intricate exchanges of the C-terminal Ser. As for the full-length precursor peptide, S23A exchange abolishes enzymatic modification completely (4.33). This is also true of S23T and S23Y substitutions, both of which preserve the side chain hydroxyl moiety. To further assess the position of phosphate transfer, we analyzed 11-residue peptides that were methylated at the C-terminal carboxylic acid or the Ser23 side chain. Surprisingly, phosphorylation was observed in only trace amounts or not at all, respectively[129].

**Figure4.33:** HPLC analysis of *in vitro* assays. ThcoK was assayed with the substrates (A) GP-PadeA-S23A, (B) PadeA(13-23)-S23A, (C) PadeA(13-23)-S23Y, (D) PadeA(13-23)-S23T, (E) PadeA(13-23)-COO-CH₃ and (F) PadeA(13-23)-S23-OCH₃.

Finally, we tested a series of Ser shift variants in which the C-terminal Ser was exchanged with Alaand a new Ser residue was introduced at position 14, 17, 20 or 22. None of these variants were phosphorylated, confirming the importance of the C-terminal Ser for kinase activity[129] (Figure 4.34).
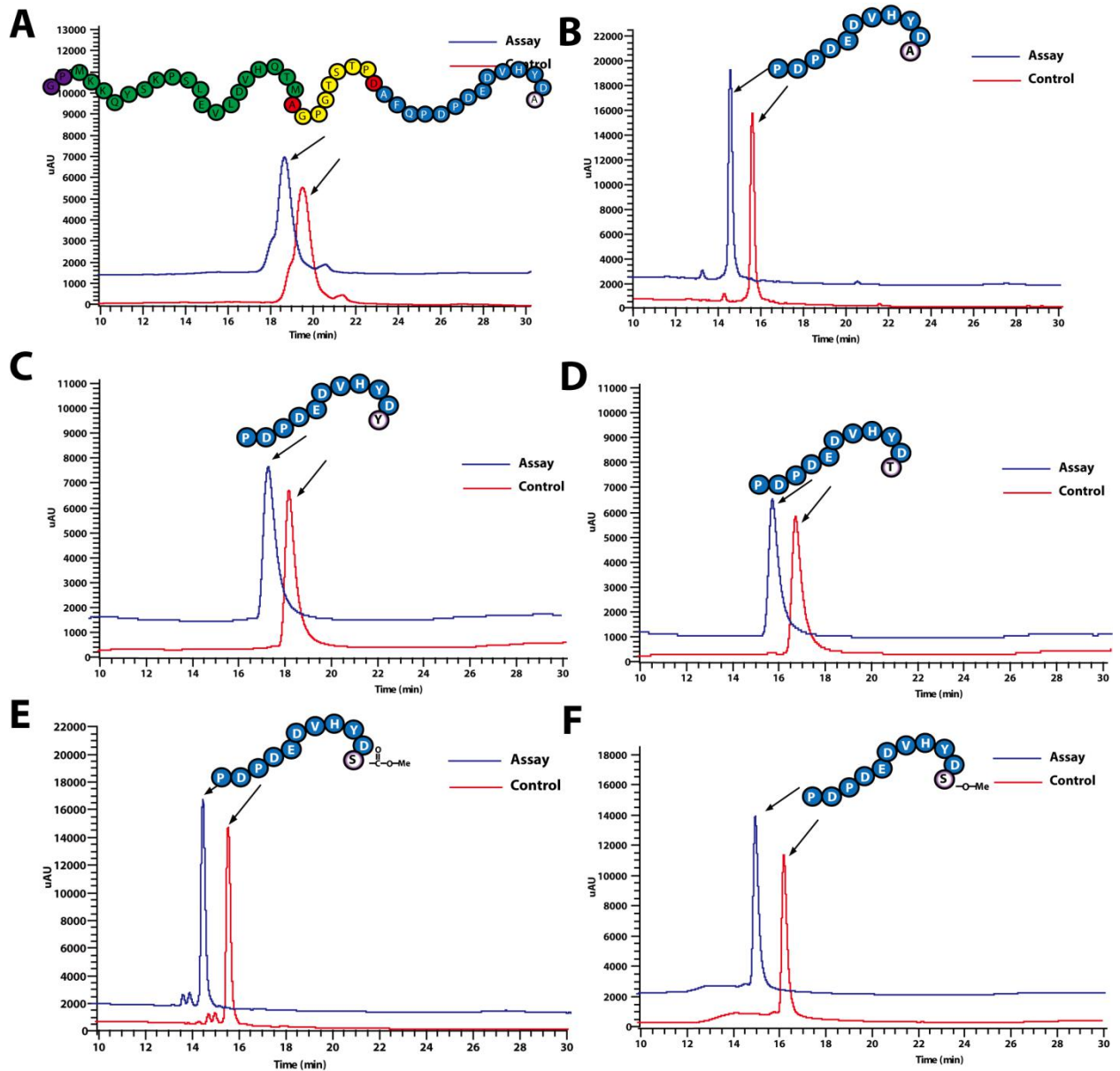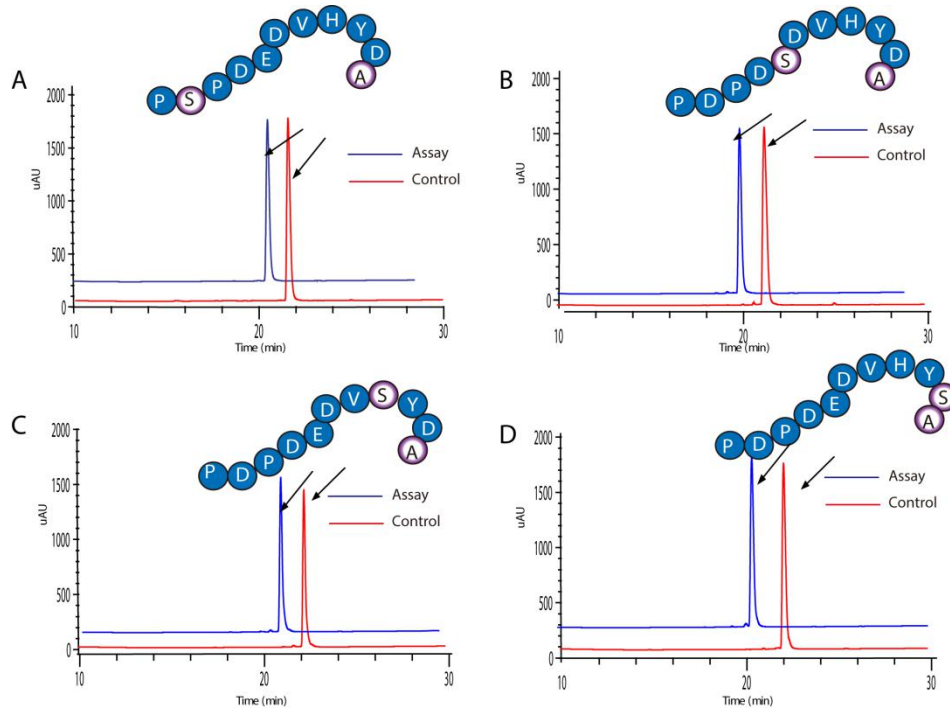
**Figure4.34:** HPLC analysis of *in vitro* assays. ThcoK was assayed with the substrates (A) GP-PadeA-D14S-S23A, (B) PadeA(13-23)-E17S-S23A, (C) PadeA(13-23)- H20S-S23A (D) PadeA(13-23)-D22S-S23A.

Collectively, these results suggest that the kinase specifically recognized the serine at the C-terminus and showed surprising enzyme promiscuous activity towards a wide scope of substrates. Table 4.4 summarizes all the substrates we used and the relative activity[129].

**Table 4.4** Results of the substrate specificity assays with ThcoK.

| Substrate | Sequence | Conversion Ratio |
|---|---|---|
| Paeninodin (lasso fold) | AGPGTSTPDAFQPDPDEDVHYDS | 0 |
| GP-PadeA | GPMKKQYSKPSLEVLDVHQTMAGPGTSTPDAFQPDPDEDVHYDS | 74.4±4.3 |
| PadeA(2-23) | GPGTSTPDAFQPDPDEDVHYDS | 50.4±2.7 |
| PadeA(13-23) | PDPDEDVHYDS | 25.5±10.0 |
| PadeA(15-23) | PDEDVHYDS | 18.1±3.6 |
| PadeA(17-23) | EDVHYDS | 8.2±2.0 |
| PadeA(19-23) | VHYDS | 3.4±0.2 |
| GP-PadeA 7xA-HYDS | GPMKKQYSKPSLEVLDVHQTMAGPGTSTPDAFQAAAAAAAHYDS | 41.2±7.8 |
| GP-PadeA 3xA-S | GPMKKQYSKPSLEVLDVHQTMAGPGTSTPDAFQPDPDEDVAAAS | 39.7±2.1 |
| GP-PadeA S23A | GPMKKQYSKPSLEVLDVHQTMAGPGTSTPDAFQPDPDEDVHYDA | 0 |
| Unthreaded paeninodin | AGPGTSTPDAFQPDPDEDVHYDS | 54.4±9.9 |
| PadeA(13-23)-S23A | PDPDEDVHYDA | 0 |
| PadeA(13-23)-S23Y | PDPDEDVHYDY | 0 |
| PadeA(13-23)-S23T | PDPDEDVHYDT | 0 |

| | | |
|---|---:|---:|
| PadeA(13-23)-COO-CH₃ | **PDPDEDVHYDS-COO-CH₃** | <1 |
| PadeA(13-23)-S23-OCH₃ | **PDPDEDVHYD(S-OCH₃)** | 0 |
| PadeA(13-23)-D14S-S23A | **PSPDEDVHYDA** | 0 |
| PadeA(13-23)-E17S-S23A | **PDPDSDVHYDA** | 0 |
| PadeA(13-23)-H20S-S23A | **PDPDEDVSYDA** | 0 |
| PadeA(13-23)-D22S-S23A | **PDPDEDVHYSA** | 0 |

## 4.6.5 Phosphorylation occurs at the C-terminal Ser23 side chain

The kinase specifically recognized the serine at the C-terminus, but we cannot judge the modification position based on an *in vitro* assay. Three assays were employed to prove the modification position to definitively prove the site of phosphorylation: MS2 analysis, NMR analysis and a chemical modification assay.

### 4.6.5.1 MS2 analysis of the *in vitro* assays

Firstly, when we used PadeA(13-23) as the substrate in the *in vitro* assays, the MS2 fragmentation demonstrated clearly that the phosphate group was attached at the C-terminal serine (Figure 4.35).
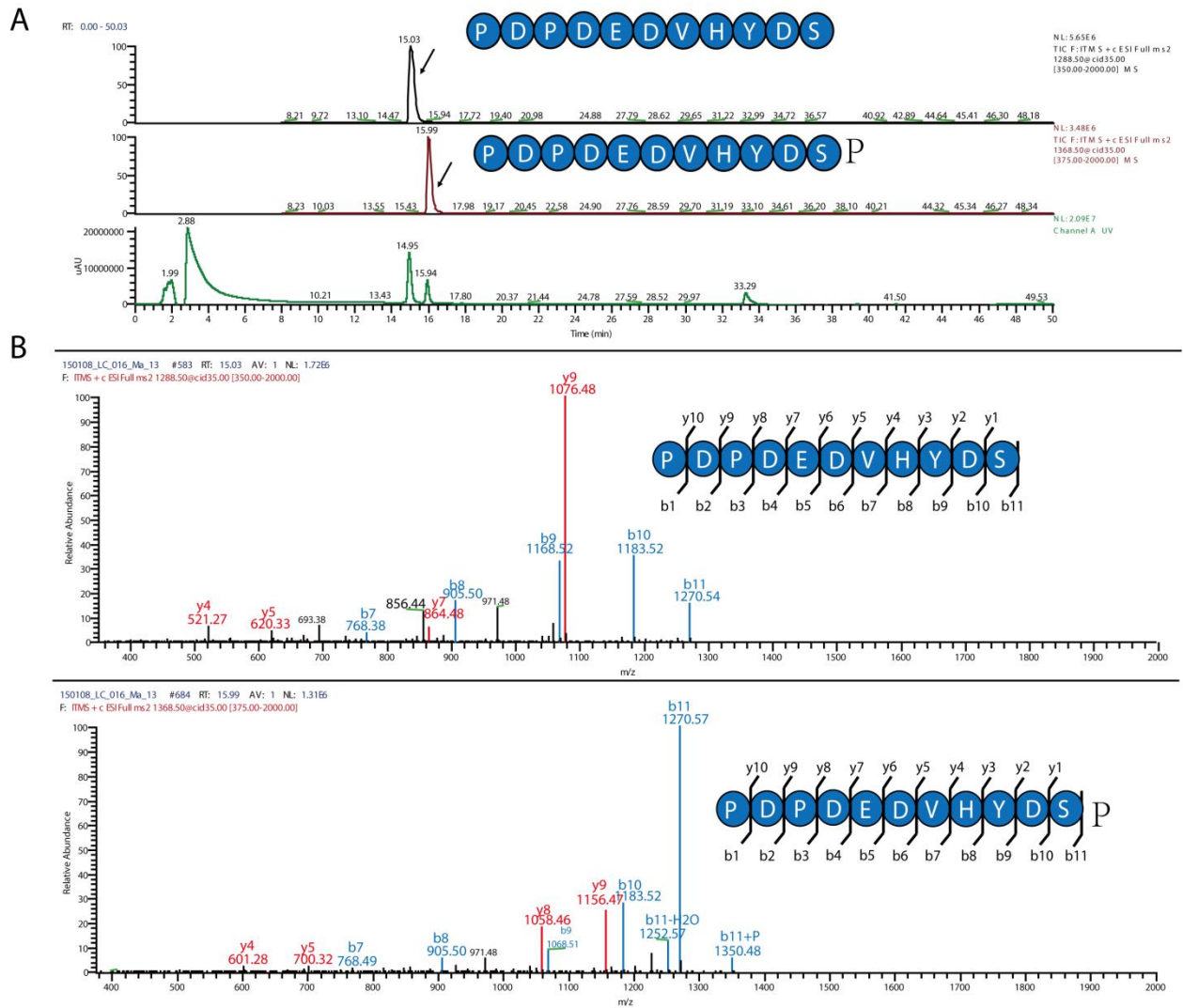
**Figure4.35.** High-resolution LC-FT-MS analysis and extracted ion chromatograms of the *in vitro* assay with ThcoK and PadeA(13-23) as substrate. (A) Extracted ion chromatograms and LC-MS analysis of the *in vitro* assays. (B) MS$^2$ fragmentation of substrate and modified peptide.

When we used PadeA(13-23)-COO-CH$_3$ as the substrate, trace amounts of the predicted modified peptides were detected. Compared with the spectrum obtained from PadeA(13-23) (Figure 4.36), the MS2 fragmentation of both modified peptide and unmodified peptide gives a clearly y and b series of fragment ions. The b series fragment observed plus a phosphate group at the modification peptide demonstrates that the phosphate group is attached at the side chain.
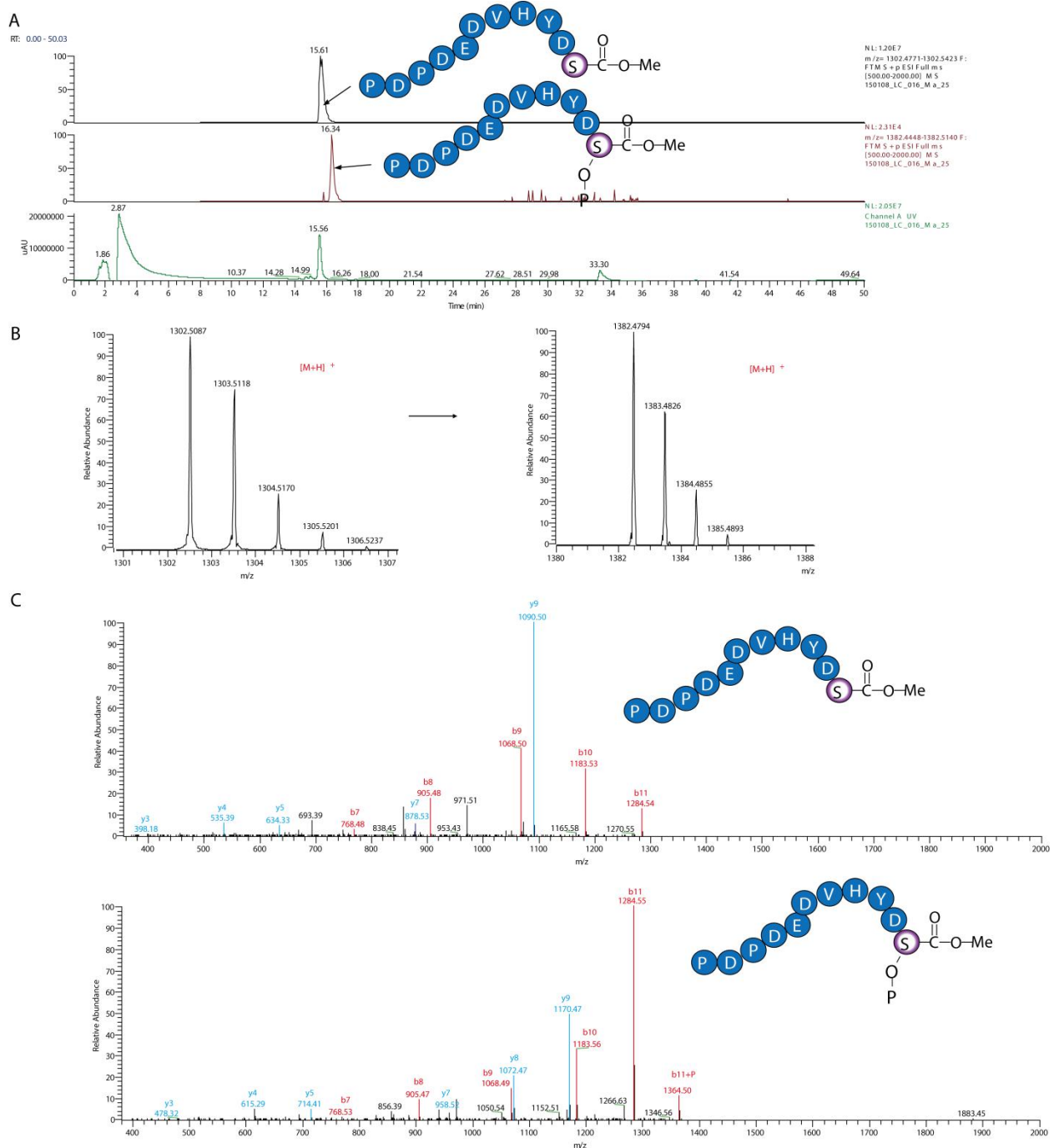
**Figure4.36.** High-resolution LC-FT-MS analysis and extracted ion chromatograms of the *in vitro* assay with ThcoK and PadeA(13-23)-S23-OCH$_3$ as a substrate. (A) Extracted ion chromatograms and LC-MS analysis of the *in vitro* assays. (B) MS$^2$ fragmentation of substrate and modified peptide.

## 4.6.5.2 Esterification assays

Secondly, we wanted to find a chemical modification method[132]. Thus, an optimized modification assay was performed on a large scale on PadeA(13-23) to obtain pure, phosphorylated peptide. Additionally, a version of the 11-residue peptide already carrying a phosphorylated Ser side chain (PadeA(13-23)-Ser23-OPO$_3^{2-}$) was chemically synthesized as a reference.

We, firstly, tried esterification assays using these peptides.[53] Each 10 μg peptide was redissolved in 100 μl 2M methanolic HCl and esterification was allowed to proceed at 25 °C for 4 h. Solvent was then removed by lyophilization and the resulting sample was redissolved in 100 μl of 20 % acetonitrile. LC-MS analysis clearly suggested that the entire carboxyl group in the peptides was esterified. All the peptides were converted into new peptides with six esters, indicating that five carboxyl groups at the side chain and one carboxyl group at the C-terminus were esterified (Figure 4.37). On the other hand, the peptides we synthesized by *in vitro* assays and the one we synthesized from Biomatik have the same behavior in the assays. In both cases, we could find the peptide with six esters plus one phosphate. Phosphate methyl esters were not observed under these conditions. Further MS2 fragmentation demonstrated clearly that the phosphate group was attached at the side chain of the last Serine at the C-terminus of our peptide synthesized *in vitro*[129] (Figure 4.38).
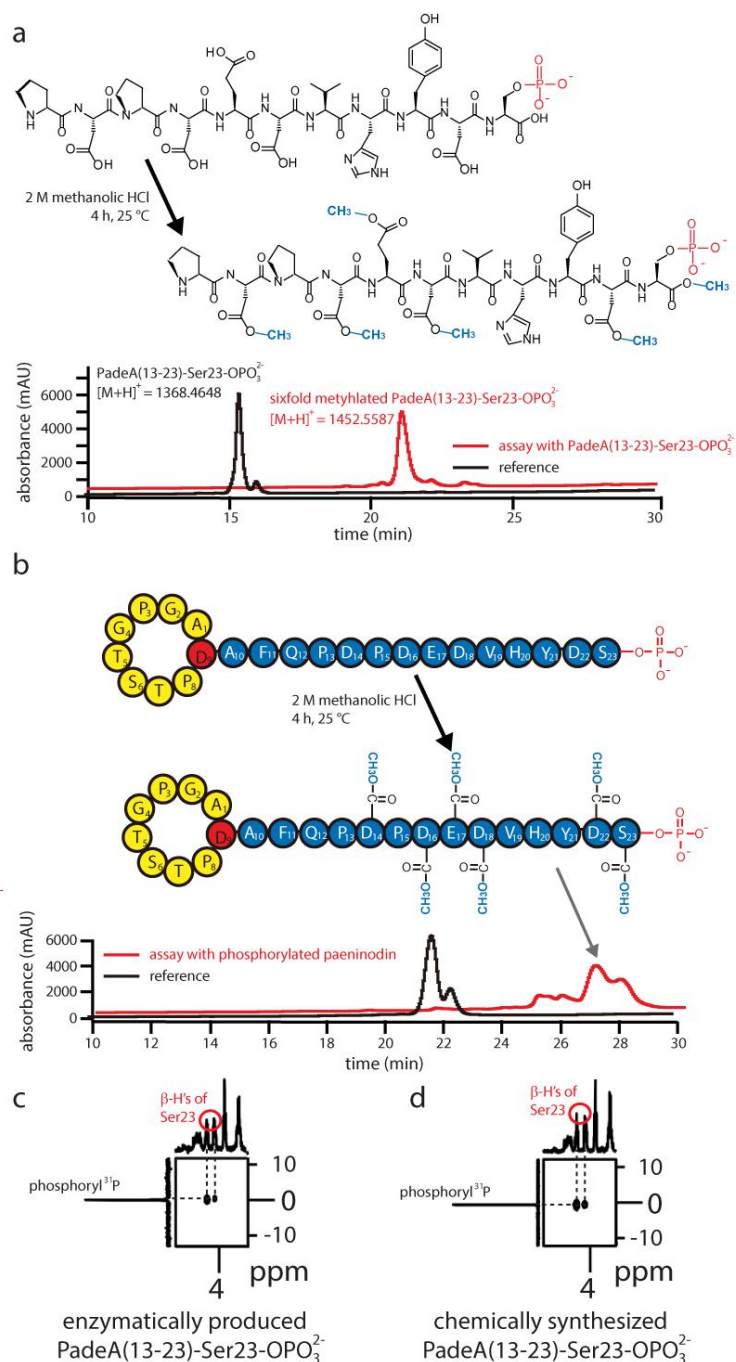
**Figure4.37:**Results of the esterification assays of (a) the enzymatically modified 11-residue peptide and (b) the isolated phosphorylated paenonodin lasso peptide. Excerpts of the $^1$H-$^{31}$P HMBC spectra of (c) enzymatically produced and (d) synthetically derived PadeA(13-23)-Ser23-OPO$_3$$^{2-}$ that show the phosphoryl $^{31}$P–Ser23 b-H cross peaks.

After we calculated the modification position from the *in vitro* assays, we repeated the esterification assays with paeninodin lasso peptide and phosphorylated paeninodin lasso peptide. Both modified peadenin and unmodified peadenin were converted into peptides with six esters, indicating the phosphate group is also attached at the side chain of the Serine at the C-terminus[129] (Figure 4.37).

Further MS2 fragmentation experiments also demonstrated our hypothesis clearly (Figure 4.39).
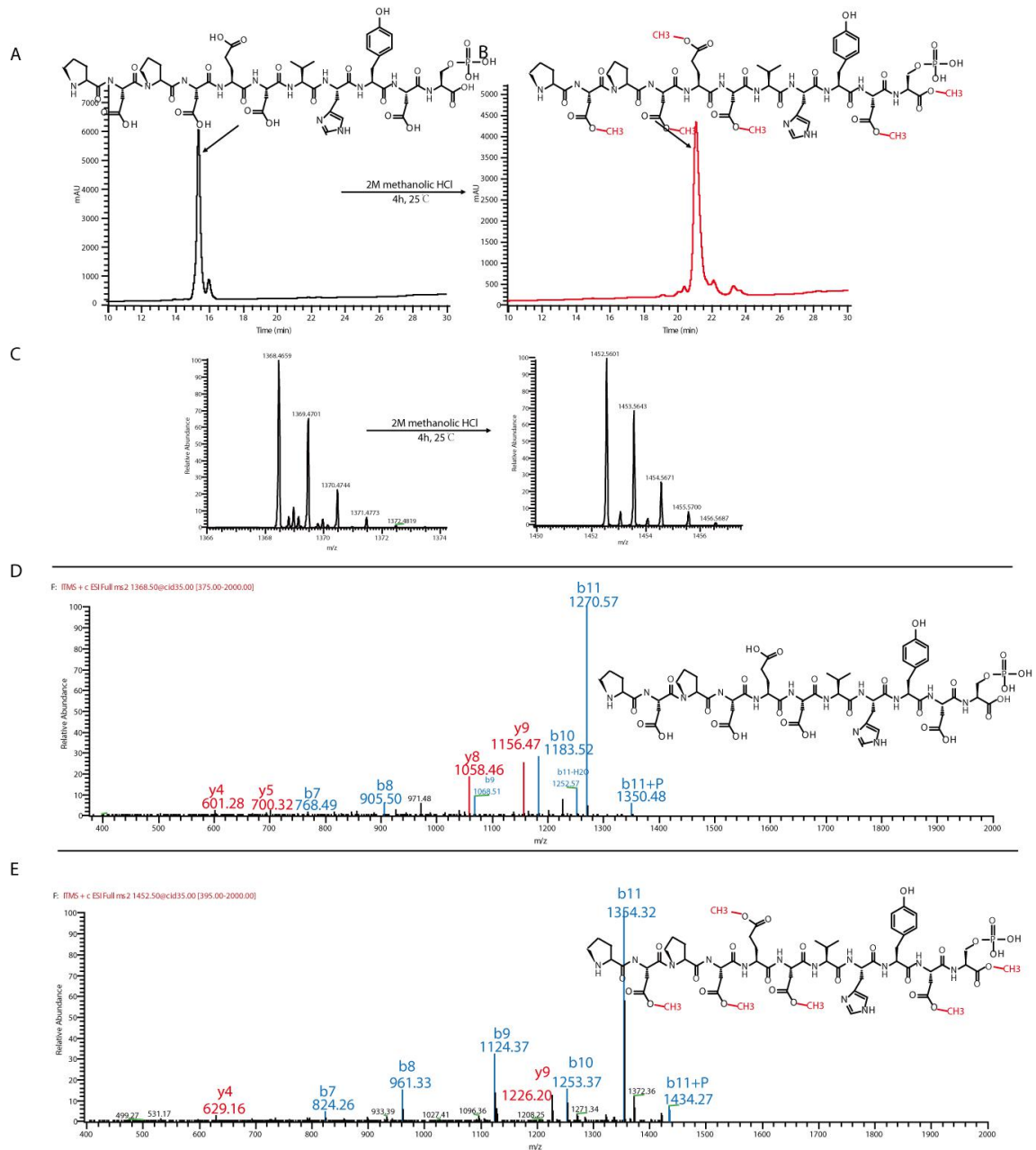
**Figure4.38:** HPLC analysis of the esterification assays with enzymatically generated PadeA(13-23)-Ser23-OPO$_3^{2-}$ (A) HPLC chromatogram of a PadeA(13-23)-Ser23-OPO$_3^{2-}$ reference. (B) HPLC chromatogram of PadeA(13-23)-Ser23-OPO$_3^{2-}$ after the esterification assay. (C) High-resolution MS signals observed before and after the assay reaction. MS2 fragmentation spectra of (D) PadeA(13-23)-Ser23-OPO$_3^{2-}$ and (E) fully esterified PadeA(13-23)-Ser23-OPO$_3^{2-}$ .The same results were obtained when using chemically synthesized PadeA(13-23)-Ser23-OPO$_3^{2-}$ instead of the enzymatically generated one. The same masses minus the phosphate group were observed for the esterification assay with unphosphorylated PadeA(13-23), showing that it was also methylated at six different positions.
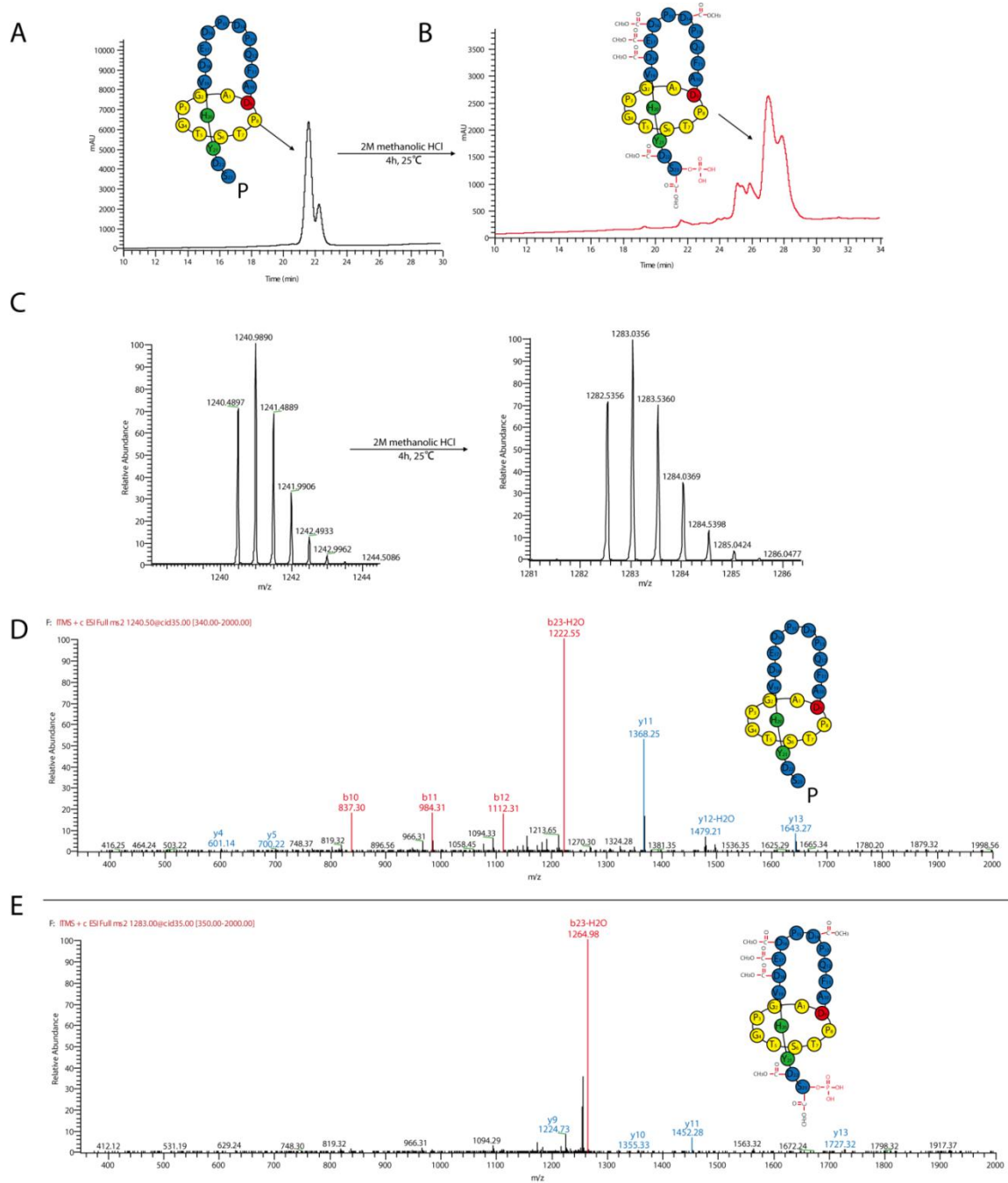
**Figure4.39:**HPLC analysis of the esterification assays with paeninodin-$OPO_3^{2-}$. (A) HPLC chromatogram of a paeninodin-$OPO_3^{2-}$ reference. (B) HPLC chromatogram of paeninodin-$OPO_3^{2-}$ after the esterification reaction. (C) High-resolution MS signals observed before and after the assay reaction. MS2 fragmentation spectra of (D) the paeninodin-$OPO_3^{2-}$ reference and (E) fully esterified paeninodin-$OPO_3^{2-}$. The same number of six methylations was observed in the esterification assay with unmodified paeninodin.

### 4.6.5.3 NMR analysis

NMR studies were performed with the synthetically and enzymatically phosphorylated 11-residue peptides as an additional control.

Nearly identical NMR spectra were obtained for both peptides and, after signal assignment, a long-range correlation between the β-hydrogen of Ser23 and phosphor could be observed in 1H-31P HMBC spectra (Figures 4.40and 4.41). Hence, it could be determined unambiguously

that ThcoK exclusively transfers a phosphate group from ATP to the side chain of the C-terminal Ser, as proposed earlier[129].
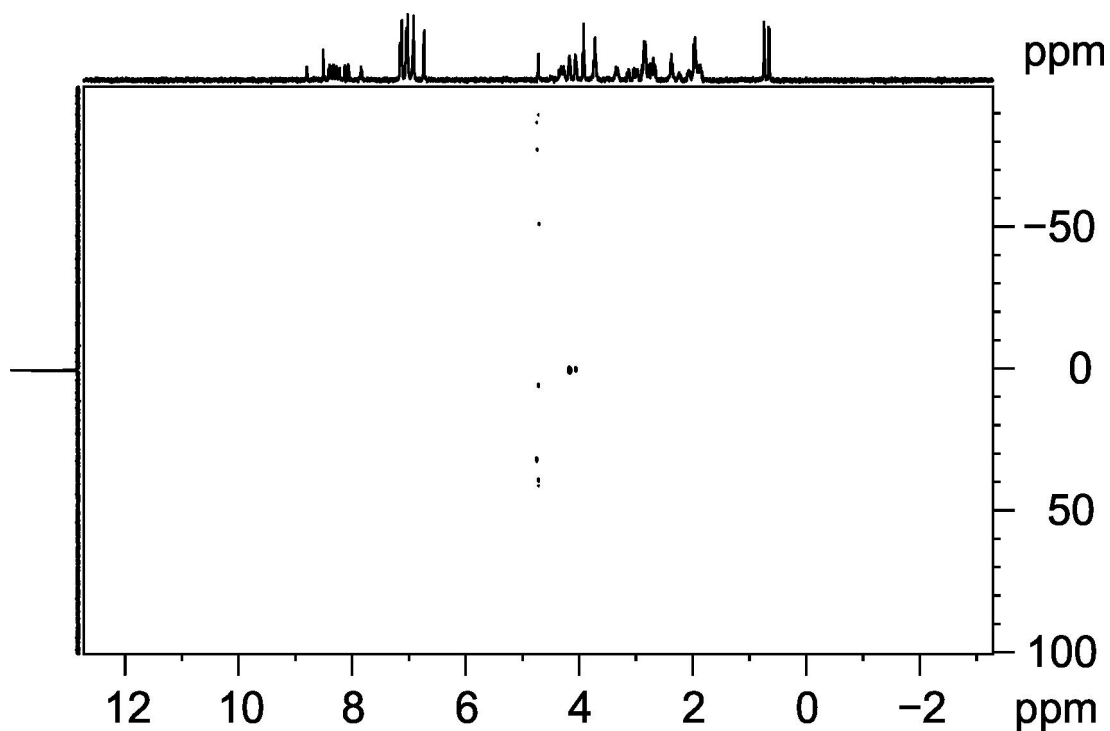


**Figure4.40:**1H-31P HMBC spectrum of enzymatically produced PadeA(13-23)-Ser-OPO32- in H2O/D2O (9:1). The spectrum was recorded on a Bruker Avance III 500 Mhz spectrometer at 298 K.
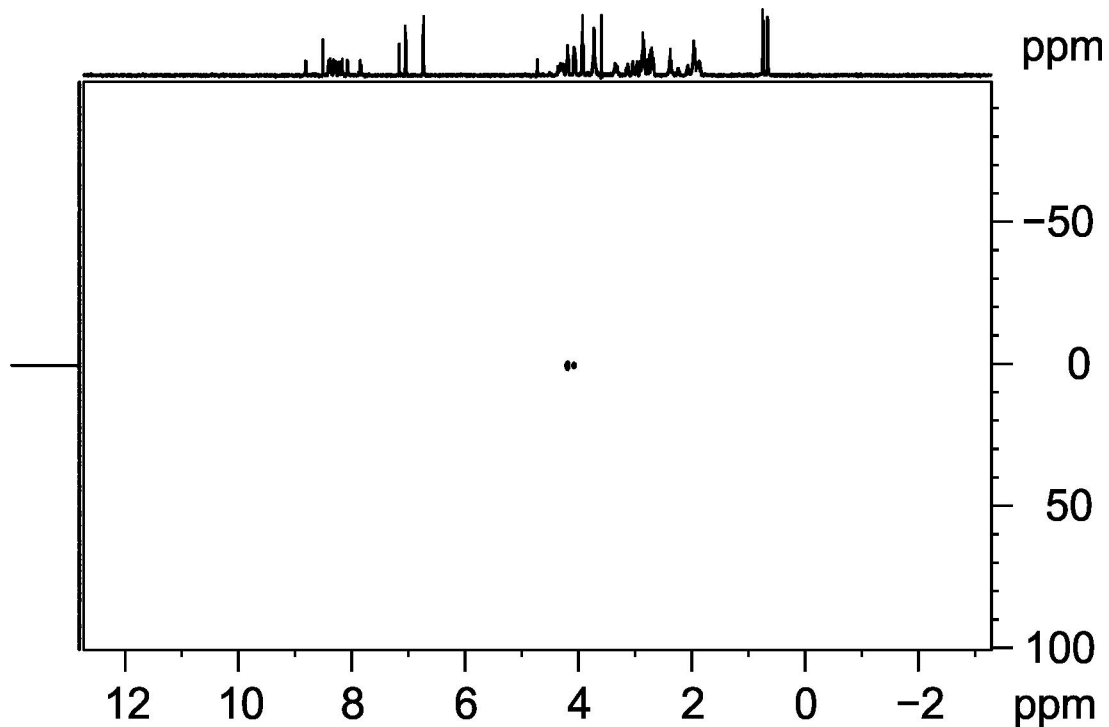


**Figure4.41:**1H-31P HMBC spectrum of chemically synthesized PadeA(13-23)-Ser-OPO32- in H2O/D2O (9:1). The spectrum was recorded on a Bruker Avance III 500 Mhz spectrometer at 298 K.

In conclusion, we clearly demonstrated by using three different methods that the kinase in the lasso peptide gene cluster could modify the side chain of the serine at the C-terminusspecifically. This is a rare example of a kinase that modified the C-terminus of a long peptide.

## 4.7 Bioinformatic Analysis of Putative Lasso Peptide Biosynthetic Gene Clusters Featuring Kinase-Encoding Genes

The majority of all known lasso peptides originate from either Proteobacterial or Actinobacterial sources[97]. Paeninodin is the first lasso peptide isolated from Firmicutes and the first phosphorylated lasso peptide reported so far. This example proved that there are a group of novel and cryptic lasso peptide gene clusters which have been overlooked in the past. The discovery of this first lasso peptide tailoring enzyme opens up a new lasso peptide biosynthesis world.

We then performed a phylogenetic analysis of C protein and its homologies. Phylogenetic trees derived from the C gene homologues have an interesting branching (Figure 4.42). These trees could segregate into five clades, which is in general accordance with the differences of their biosynthetic gene cluster. Hence, there is one group of Actinobacterial clusters, one of Firmicute clusters and three of Proteobacterial clusters. The latter is divided into clusters with a simple ABC organization adjacent to a gene encoding a lasso peptide isopeptidase, clusters containing an additional gene encoding an ABC transporter (D protein) and clusters featuring genes coding for both a D protein and a putative kinase[129].
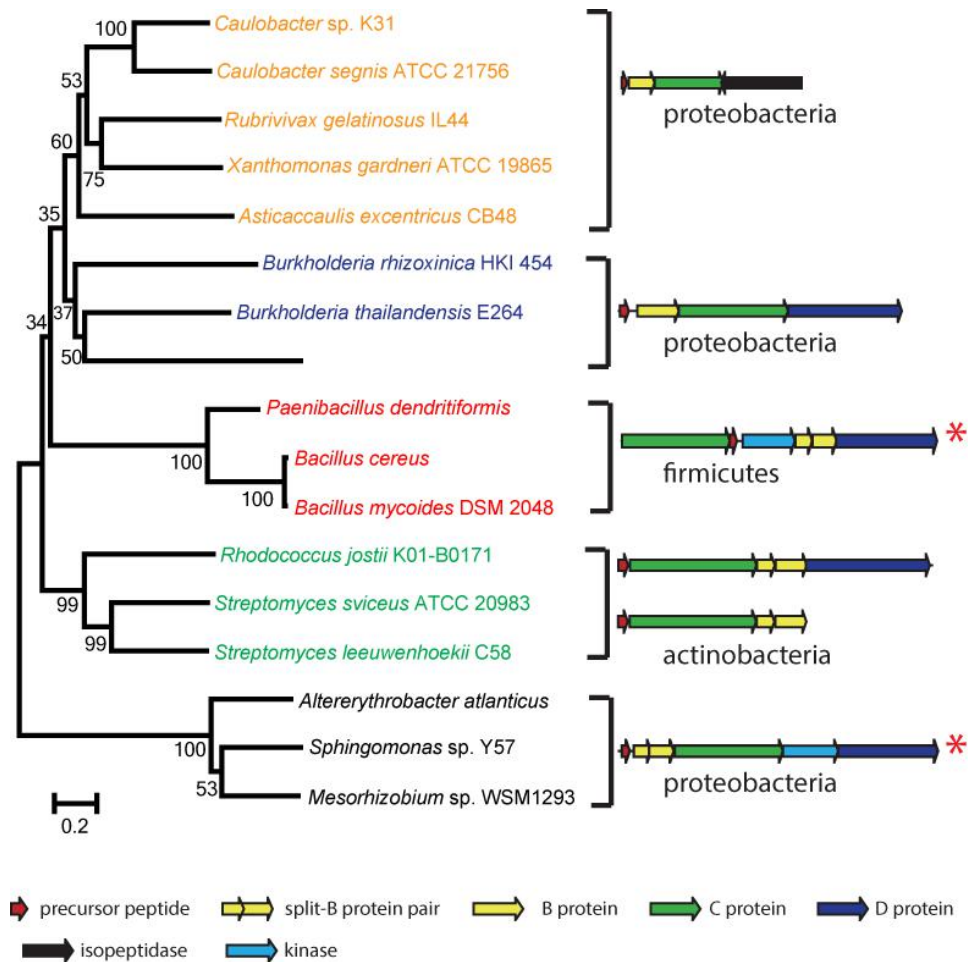
**Figure4.42:**Phylogenetic tree of the C proteins of different lasso peptide biosynthetic gene clusters. The red asterisks mark the clusters containing kinase.

It is highly likely that the lasso peptides have similar functions with similar organization of the biosynthetic machinery, thus, it would be smarter to classify the lasso peptides by the organization of their biosynthetic gene cluster. Therefore, we hypothesize that most lasso peptides could be generally divided into five clades. It appears that the five lasso peptide clades evolved by divergence from a common ancestor. This is reasonable when considering the known and presumed biological functions of these peptides. The clusters containing isopeptidase from Proteobacteria, which lack an ABC transporter, are able to catabolize their lasso peptides, indicating that the peptides play a role in signaling or function in scavenging through a catch-and-release mechanism. On the other hand, lasso peptide gene clusters from Actinobacteria and Proteobacteria, which feature an ABC transporter, often produce lasso peptides with antibacterial properties, indicating that the molecules are synthesized and then exported for competition. That the organization of these gene clusters is different, indicates that they arose independently. The newly identified clusters from Firmicutes featuring kinase and the potential clusters featuring kinase from Proteobacteria represent two new families of lasso peptide gene clusters. Although both clusters contain a kinase, there is little similarity between their precursor peptides other than the C-terminal serine. Thus, it seems that Firmicutes and

105

Proteobacteria evolved a similar strategy independently to produce phosphorylated lasso peptides. The diversity of the biosynthetic machinery and functions of the molecules make them an intriguing class of bacterial natural products.

# 5. Discussion

In this study, we describe the first example of modified lasso peptide and the first tailoring enzyme involved in the lasso peptide biosynthesis pathway. The biosynthetic pathway of modified lasso peptide has been delineated through *in vivo* and *in vitro* studies. This unusual kinase recognized the last serine of the precursor peptide specifically and transferred a phosphate group onto the C-terminus of the peptide, resulting in a modified precursor peptide, and then folded into a modified lasso peptide. This first lasso peptide isolated from Firmicutes revealed several cryptic, novel lasso peptide gene clusters overlooked in the past. The kinase studied here and several potential transferases open up new avenues for engineering and diversifying the chemical nature of lasso peptides.

## 5.1 Isolation and Characterization of the Paeninodin and Phosphorylated Paeninodin Lasso Peptide

### 5.1.1 *Paenibacillus*as a new source for novel natural products

*Paenibacillus* is a genus originally included within the genus *Bacillus* and then reclassified as *paenibacillus*. As the name implies,*paene* in Latin means almost, so bacteria belonging to this genus are almost *Bacilli*. Similar to *Bacillus*, this genus has been detected in all kinds of environments. It has been shown to be important for agriculture and horticulture. Exemplary are *P. larvae*, which is known to cause American foulbrood in honeybees[139], pattern-forming strains, such as P. dendritiformis studied in this thesis, which are already known to develop complex colonies with complex architectures[140], and *P. polymyxa*, which is capable of fixing nitrogen and is used, therefore, in agriculture and horticulture[141].

This genus has been shown as a new source for novel natural products. *P. larvae*, for example, a pathogen that causes American foulbrood, which is the most serious infectious disease of honey bees, has been shown to produce diverse natural products[142]. These secondary metabolites, such as nonribosomal peptides and peptide–polyketide hybrids, are highly likely to be involved in virulence and pathogenicity towards the bee larvae. These natural products included Sevadicin-A nonribosomal tripeptide, bacillibactin-siderophores, paenilarvins-iturin-like lipopeptides and paenilamicins-complex peptide-polyketide hybrids (Figure 5.1).
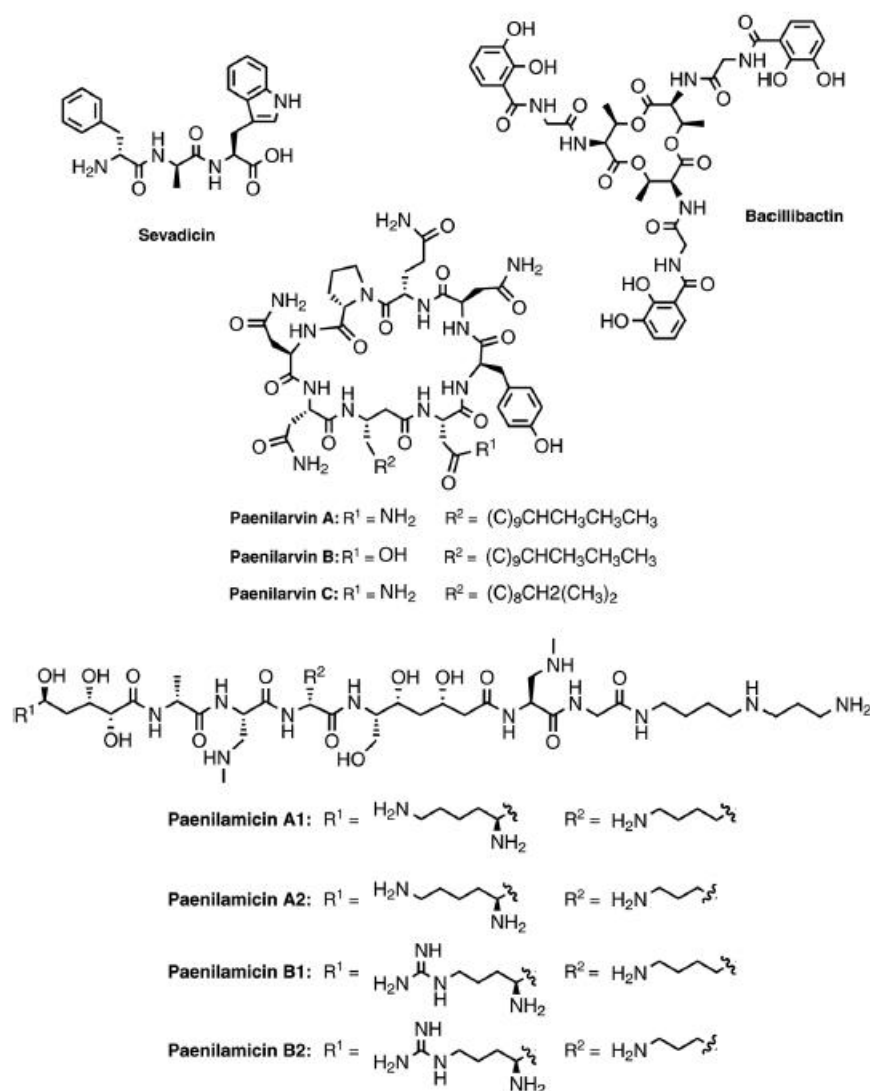
**Figure5.1:**Known compounds isolated from *P. larvae*(Figure from ref [142]).

Genomes sequence projects in recent years have revealed the presence of an impressive number of secondary metabolite clusters. There are more cryptic gene clusters in *P. larvae*, besides the four gene clusters responsible for the biosynthesis of known compounds isolated, of which the natural products are not yet known.

Beside the *P. larvae*, another 56 *Paenibacillus* strains have been sequenced and all of them contain diverse secondary metabolite gene clusters. These *Paenibacillus* seem to represent a new and rich source of secondary metabolites that need to be explored in the future.

*P. dendritiformis* is a species of pattern-forming bacteria, first discovered in the early 1990s by Ben-Jacob's group. It is a social microorganism that forms colonies with complex and dynamic architectures[143]. The complete genome sequence of this strain was sequenced in 2012. The latter provides an excellent opportunity that facilitates the exploitation of natural product discovery in this strain. No secondary metabolites have been isolated from *P. dendritiformis* so far*,* but a lethal protein was produced in response to competition between sibling colonies[143].

Genome analysis of the *P. dendritiformis* chromosome by antismash identified 12 secondary metabolite gene clusters, suggesting the possibility of an extensive secondary metabolism (Table 5.1). Most of them encode for peptides of nonribosomal origin, though PKS and hybrid PKS-NRPS systems are present as well. Moreover, RiPPs gene clusters have also been detected.

**Table5.1**. Potential natural products in biosynthetic gene clusters from *P. dendritiformis*.

| No. | Cluster location | Predicted products |
|-----|------------------|--------------------|
| 1 | WP_006674824.1-WP_006674859.1 | Nrps-t1pks |
| 2 | WP_006675255.1-WP_006675330.1 | Nrps-transatpks-t1pks |
| 3 | WP_040730674.1-WP_006676104.1 | Bacteriocin |
| 4 | WP_006677185.1- WP_006677207.1 | Bacteriocin-nrps |
| 5 | WP_006678389.1- WP_006678408.1 | Lassopeptide |
| 6 | WP_050979467.1-WP_006678468.1 | NRPS |
| 7 | WP_006679151.1-WP_006679160.1 | Nrps-arylpolyene |
| 8 | WP_006679448.1- WP_006679450.1 | NRPS |
| 9 | WP_006679469.1-WP_006679497.1 | NRPS |
| 10 | WP_006679486.1- WP_006679504.1 | NRPS |
| 11 | EHQ59413.1 | NRPS |
| 12 | WP_006679973.1- WP_006679985.1 | Ectoine |

Notably, the gene cluster No. 5 is the paeninodin lasso peptide gene cluster studied in this thesis. The C and D proteins were identified in the prediction by antismash.

## 5.1.2 Discovery and isolation of the paeninodin and phosphorylated paeninodin lasso peptide

The isolation and characterization of the paeninodin and phosphorylated paeninodin lasso peptide are reported in this work. The paeninodin lasso peptide is also known to be the first secondary metabolite isolated from the producing strain *P. dendritiformis*. Moreover, modified paeninodin is the first phosphorylated lasso peptide isolated so far. The fermentation extracts of *P. dendritiformis* C454 failed to yield any lasso peptide, which proved that these lasso peptide gene clusters are silent in lab conditions. Thus, a heterologous expression approach was applied in this study.

Natural products gene clusters from Gram-negative species can generally typically be expressed in greater yield using *E. coli* as a heterologous host. A strong promoter combined with optimum fermentation conditions can normally trigger the production of the target compound. However, expression of natural products gene clusters from Gram-positive species in *E. coli* could be a challenging task. However, the advantage of using *E. coli* as a

109

heterologous host is that it is well understood and easy to handle. Paeninodin lasso peptides were produced by IPTG induction by heterologous expression in *E. coli* that was further purified by HPLC. This proved that *E. coli* could also be a general expression host for Gram-positive species.

The peadenin isolated contains 23 amino acids, which is the largest lasso peptide isolated so far[97]. Tandem MS confirmed its primary structure, which features a nine-residue macrolactam ring between Ala1 and Asp9. This lasso peptide bears an N-terminal Ala residue, a feature only recently observed for a series of lasso peptides from *Caulobacter* sp. K31[116]. This could also explain why these lasso gene clusters were overlooked in Link's paper on "Precursor-centric genome-mining approach for lasso peptide discovery"[117]. They only considered that lasso peptides start with a glycine or cysteine. Actually, much to our surprise, all the lasso peptides from Firmicutes prefer an N-terminal Alanine, indicating that these gene clusters are different from all lasso peptide gene clusters reported.

Even though paeninodin is the first isolated lasso peptide from Firmicutes, we are more interested in the function of the kinase involved in the biosynthetic pathway. Lasso peptide modification in the literature is notably rare. There are only two examples. One is the C-terminal methylation of lassomycin, which has a methyltransferase involved in the lasso peptide gene cluster[104]. The other is the Trp hydroxylation of some RES-701 family lasso peptides whose biosynthetic pathway has not yet been identified.

The kinases identified in lasso peptide biosynthetic gene clusters from Firmicutes are often annotated as homologues of HPr kinase (e.g. from *Lactococcuscasei*). Hpr kinase is the key protein in the regulation of carbon metabolism in *Bacillus subtilis* and many other Gram-positive bacteria[136, 137]. At the beginning, we hypothesized that this kinase may play a role in regulation of paeninodin lasso peptide production. However, a small peak with similar retention time to that of paeninodin in the high-resolution LC-FT-MS analysis of cell extracts caught our attention. Closer inspection showed that it belonged to a compound with the molecular weight of the lasso peptide plus a phosphate group. Based on MS2 results, we proposed that the compound results from phosphorylation of the C-terminal Ser of paeninodin by the kinase encoded within the cluster. Deletion of the kinase gene from the expression plasmid yielded an extract lacking the modified compound, while non-phosphorylated paeninodin was still produced at the same level as before. Co-expression of the knock-out plasmid with a pACYC vector carrying padeK restored production of the phosphorylated compound. Hence, a link between the expression of the putative kinase PadeK and the occurrence of the modified lasso peptide was established and led us to identify a novel lasso peptide tailoring enzyme for the first time.

## 5.1.3 Characterization of the paeninodin and phosphorylated paeninodin lasso peptide

Both the paeninodin and phosphorylated paeninodin lasso peptide could be purified by two-step preparative HPLC. Notably, the production of both compounds ispretty low compared with other lasso peptides isolated from Proteobacteria[112, 129]. (The pure paeninodin lasso peptide was obtained with a yield of ~ 0.1 mg/L; phosphorylated lasso peptide was obtained with a yield of ~ 0.05 mg/L.) Considering the low production of paeninodin lasso peptide, *E. coli* might not be the most ideal host for expression of these lasso peptide gene clusters from Firmicutes. A Gram-positive related host might be a better choice. However, in our case, large-scale fermentation is already enough for isolation of an adequate amount, so we continued to use *E. coli* as the first choice.

We then characterized the paeninodin lasso peptide using the pure samples. We first tried to substantiate that paeninodin exhibits the lasso topology predicted. A assay established previously in which pure lasso peptide is heat-denatured and then digested with carboxypeptidase Y was performed. The results showed that the isolated compound is only resistant to carboxypeptidase Y prior to thermal treatment, indicating that the peptide has adopted a lasso topology (the threaded topology of the lasso peptide occludes access of carboxypeptidase Y to the C-terminal tail; however, following thermally-induced unthreading, the C-terminus becomes accessible to carboxypeptidase Y)[129].

We then used a previously established IM-MS method to further prove the topology of paeninodin. The results corroborate the proposed lasso fold of the isolated compound and, additionally, indicate that paeninodin can adopt multiple, distinct and interchangeable conformations in its native state. We noticed that the paeninodin lasso peptide (AGPGTSTPDAFQPDPDEDVHYDS) contains a lot of proline, aspartic acid and glutamic acid in the potential loop region of the lasso peptide. These residues might contribute to the multiple conformations of the lasso peptide.

Mutagenesis studies were performed to study the biosynthesis of paeninodin and propose the plug residues of the lasso peptide. Sequence analysis of the C-terminal region of paeninodin suggests that His20 and Tyr21 are the most suitable plug residues. Moreover, it could be shown that the paeninodin biosynthetic machinery operates according to the same set of principles identified for Proteobacterial systems[129].

Efforts were then taken to determine the three-dimensional structure of paeninodin. As paeninodin can adopt multiple, distinct and interchangeable conformations in its native state, both the sparse-matrix crystal screening and NMR methods failed. Although structure elucidation was not possible, a series of analytical assays shows that paeninodin exhibits the

111

typical characteristics of a lasso peptide, in accordance with numerous previous publications. These experimental data provide a solid body of evidence for the true lasso nature of paeninodin. However, elucidation of its three-dimensional structure is necessary to prove this beyond a doubt[129].

## 5.2 Biochemical and Genetic Model for Phosphorylated Paeninodin Biosynthesis

No experimentally supported biosynthetic pathway for lasso peptide modification has been reported to date. This raised the question: how does the modification happen? Due to the unique structure of lasso peptide, there would be two disparate biosynthetic scenarios for lasso peptide modification (Figure 5.2). The first scenario could be that the lasso peptide is, firstly, synthesized by the mature enzymes and then modified by the tailoring enzymes to yield the final products. A second scenario could be that the kinase, sulfotransferase and other potential tailing enzymes encoded by the lasso peptide gene cluster, firstly, modify the precursor to form modified precursor, then the mature enzymes transfer the modified precursor into the final lasso peptide.

### 5.2.1 A biosynthetic pathway for phosphorylated paeninodin assembly

We heterologously expressed and isolated the PadeK kinase and performed *in vitro* studies to investigate how paeninodin becomes phosphorylated. Two potential substrates (the padeA precursor peptide and paeninodin) were also prepared. *In vitro* sudies clearly demonstrated that the PadeK kinase could only modify the precursor peptide. Thus, the biosynthetic pathway of modified lasso peptide was delineated. It turned out that the second scenario is right, which is that the kinase acts on the precursor peptide prior to processing by the lasso peptide biosynthetic enzymes.
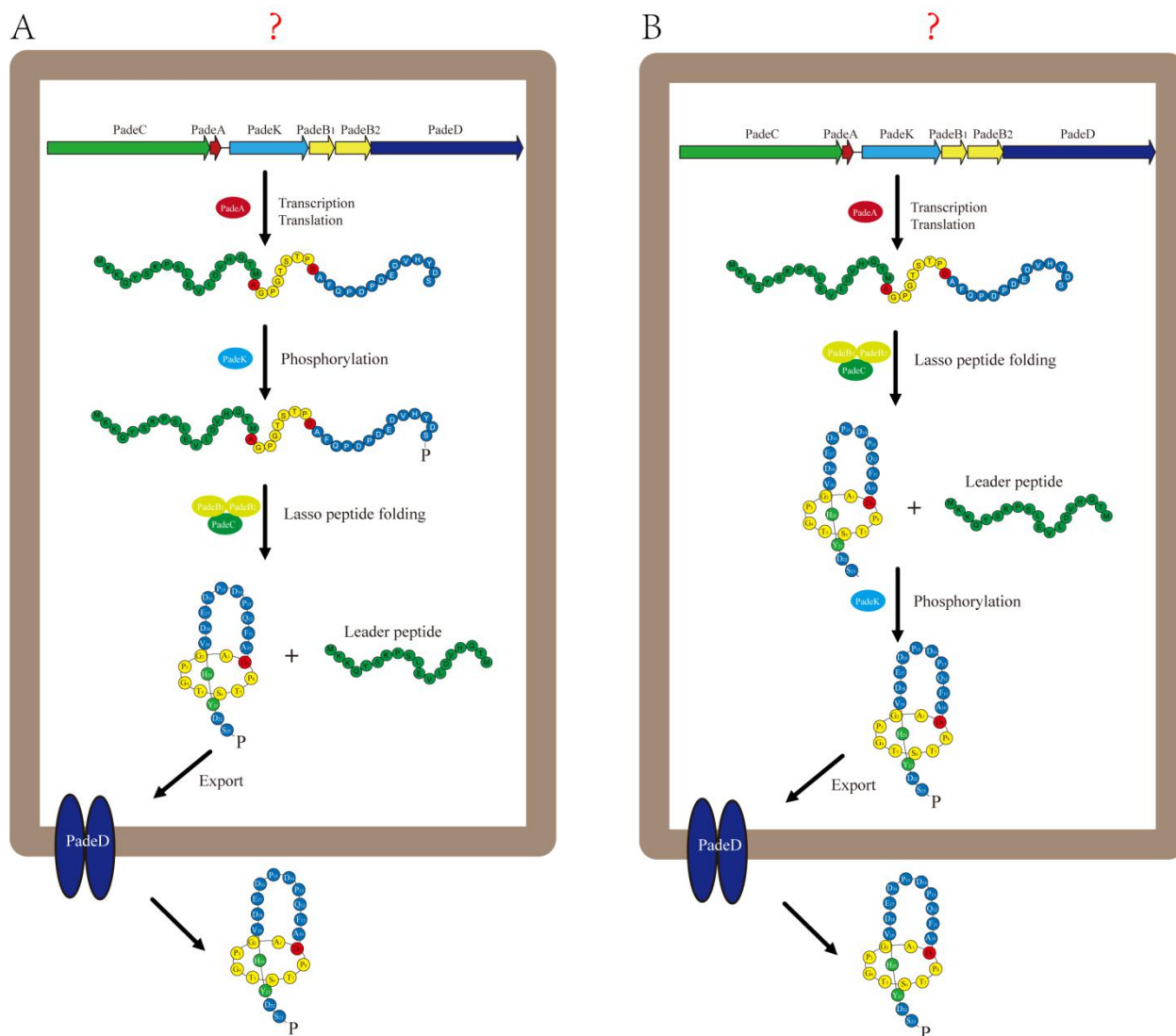
**Figure5.2:**Two potential biosynthetic pathway for modified lasso peptide.

We searched our genome mining results for a suitable homologue due to difficulties isolating PadeKas as well as its weak activity. We identified a kinase from the cluster of *T. composti* KWC4 (ThcoK) that could be readily isolated in high yields and be used to facilitate homologous exchange. When pure ThcoK was subjected to *in vitro* assays with the lasso peptide and unmodified precursor peptideGP-PadeA, phosphorylation of the precursor peptide (~ 80 % conversion), but not the lasso peptide was observed.

We replaced PadeK with ThcoK in our heterologous expression system and the phosphorylated lasso peptide was found to be the main product in the presence of ThcoK. These *in vivo* and *in vitro* experiments for both PadeK and ThcoK kinase clearly provided us a rare glimpse of how lasso peptide further increases its chemical diversity and provide a way for the generation of novel lasso peptide analogs[129].

## 5.2.2 Putative regulation of the phosphorylated paeninodin biosynthesis

A natural products biosynthetic gene cluster normally contains regulatory genes. This is also true for this newly identified paeninodin lasso peptide gene cluster. A putative transcriptional regulator was identified at the downstream of the whole paeninodin lasso gene cluster (Figure 5.3).



**Figure5.3:**Putative regulator involved in paeninodin biosynthesis.

This transcriptional regulator contains a XER-family (xenobiotic response element) HTH(Helix-turn-helix) domain[144] (Figure 5.4). This kind of regulator is believed to be involved in several late growth processes. We hypothesized that this regulator takes part in the biosynthesis of paeninodin in the native strain and could be a reason why no lasso peptide was detected in lab conditions.
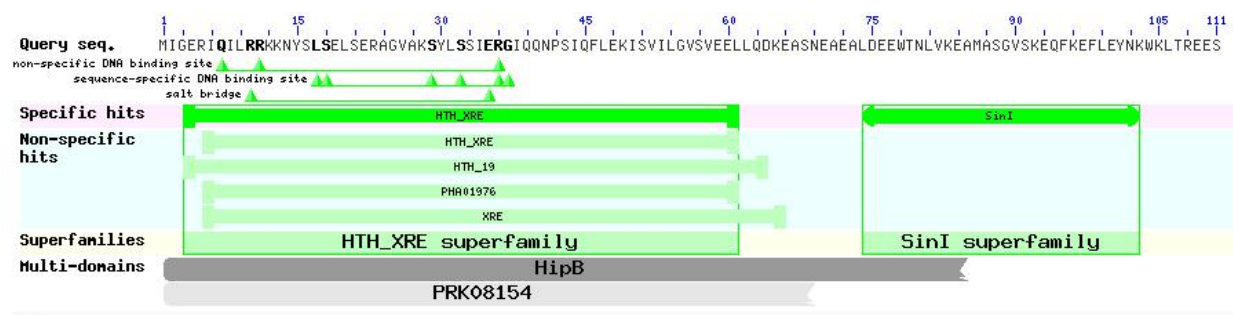


**Figure5.4:**Conserved domain of the potential regulator.

## 5.2.3 Putative function of the phosphorylated paeninodin

The natural function of phosphorylated paeninodin is still elusive. Past research has showed that lasso peptide can generally be divided into two clades based on whether the pathway features an ABC-transporter or an isopeptidase. The gene clusters for recently newly discovered astexins, caulosegnins and xanthomonin have a TonB transporter and an isopeptidase to specifically deconstruct the lasso peptide in the cell[113]. The natural function of this class of lasso peptides is believed to be a type of scavenging molecule that is involved in species competition. On the other hand, the gene cluster for MccJ25, capistruin and lariatin has an ABC transporter serving as an exporter and immunity factor of the lasso peptides[97]. All these molecules are antimicrobial peptides. An ABC transporter could alwaysbe observed in the gene cluster for peadenin lasso peptide and its homologues. However, notably, the PadeD shows very low identities (30 %) to MccjD. Considering these gene clusters featured unusual tailoring enzymes, they are distinct from both clades mentioned above. The lasso peptide from these gene clusters is highly likely to function as an antimicrobial peptide, since the ABC transporter is

present. The phosphate group attached at the C-terminus of the lasso peptide could provide a means of diversifying the lasso peptide surfaces. In particular, the phosphate group with its large hydrated shell could increase the negative charge of lasso peptide significantly, thus, increasing its binding affinity to the target receptor. On the other hand, nature widely choosesthe phosphate group to modify protein. The phosphate group has special properties that can be exploited to regulate critical biological functions when attached to proteins. The phosphate group generally serves as a switch to promote inducible protein-protein interactions, which allows signal transduction networks to transmit transient signals in response to extracellular stimuli. Thus, a second guess for the function of the phosphorylated peadenin would be that it serves as a signal molecule that corresponds to the environmental transformation. It could also be possible that this modification is a self-protective mechanism. The phosphate group could be removed during the transporting process. The last hypothesis is that the phosphorylated precursor peptide is only an intermediate of other tailoring enzymes. The uncharacterized nucleotidyltransferase, for example, might use it as a substrate and will perform another modification at the same position. More future effort needs to be placed on the natural function of these fascinating molecules.

## 5.3 Characterization of the Lasso Peptide Precursor Kinase

### 5.3.1 Characterization of the precursor kinases PadeK and ThcoK

After we figured out how the paeninodin lasso peptide was modified, we then turned our attention to the characterization of the precursor kinases. After demonstrating the much higher *in vivo* and *in vitro* phosphorylation activity of ThcoK compared with PadeK towards PadeA, full characterization of ThcoK was performed. Kinetic parameters for the modification of GP-PadeA were obtained by varying the substrate concentration under constant enzyme concentration. In this way, the Michaelis-Menten parameters $K_m = 279 \pm 19.7$ $\mu$M and $k_{cat} = 0.41 \pm 0.02$ min$^{-1}$ for ThcoK were determined[129].

We also successfully identified the catalytic mechanism of these kinases by informatics analysis combined with mutagenesis studies. Because ThcoK and other kinases identified in lasso peptide biosynthetic gene clusters from Firmicutes are often annotated as homologues of HPr kinase (e.g. from *Lactococcus casei*) and phosphoenolpyruvate carboxykinase (e.g. from *Thermus thermophilus* Hb8), we then checked all the conserved regions of kinases from these families. Intriguingly, all of these kinases appear to share conserved active site motifs consisting of one His, one Lys and two Asp residues (HKDD). The catalytic mechanism for *L. casei* HPr kinase has been studied previously. The roles of these residues in catalysis (His140-Lys161-Asp178-Asp179) can be derived from the co-crystal structure of the kinase with its substrate.

115

The conjugate base of Asp179, stabilized by His140, deprotonates the Ser46 side chain of the HPr substrate, which then abstracts the $\gamma$-PO4$^{3-}$ group of ATP through nucleophilic attack. Lys161 and Asp178 facilitate the binding of ATP by counteracting negative charges and coordinating a metal ion, respectively[135, 137, 138].

We hypothesize that a similar mode of catalysis is feasible for PadeK, ThcoK and their homologues. Indeed, this was confirmed by mutagenesis. We generated single or quadruple Ala variants of these residues, as well as double Asp-to-Ala variants of both PadeK and ThcoK in our lasso peptide production plasmids. No phosphorylated lasso peptide was observed for any variant, although the overall yields of lasso peptide remained unaffected. This suggests strongly that the kinase variants were rendered inactive and that each of the four residues is crucial to catalysis[129].

We then investigated the substrate specificity of ThcoK. Several questions needed to be addressed for these studies: Firstly, the minimal length of a peptide necessary for modification, secondly, the motifs recognized by the kinase and, thirdly, the precise site of phosphate addition. To address these questions, a wide variety of peptides were treated with ThcoK under our established assay conditions. These results collectively suggest that the kinase specifically recognized the serine at the C-terminus and showed surprising enzyme-promiscuous activity towards a wide scope of substrates[129].

## 5.3.2 Lasso peptide precursor kinase as a useful biocatalyst

Phosphorylation is generally one of the most prominent types of posttranslational modification of protein, because of its versatility and ready reversibility. It commonly forms a phosphate ester at the side chain of serine, Thr and tyrosine residues, although phosphorylation of six other amino acids is chemically possible (arginine, cysteine, lysine, histidine, glutamine and asparagine), and in many cases is known to occur. Here, we show a rare example that this family of kinases recognized the serine at the C-terminus specifically.
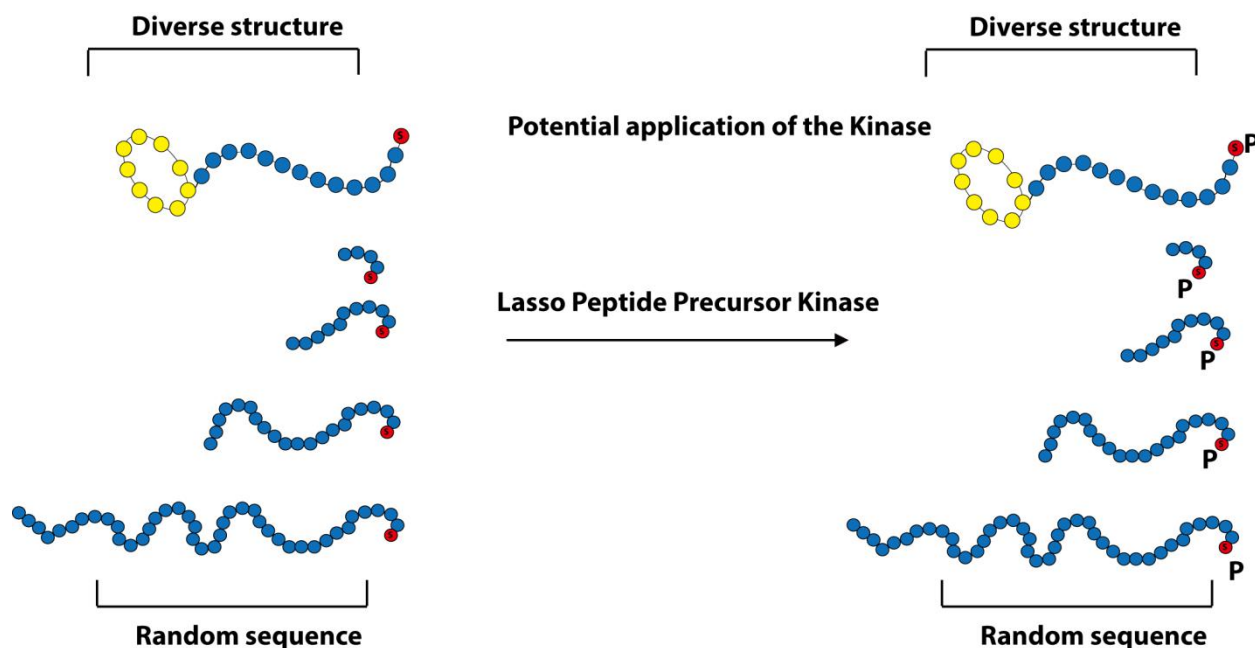
**Figure5.5:**Potential application of the lasso peptide precursor kinase.

It turned out that, in spite of the strict requirement for a C-terminal Ser with free hydroxyl and carboxyl groups, the enzyme shows a remarkable degree of promiscuity toward its substrates, to the extent that their primary structures may otherwise be completely flexible. To the best of our knowledge, no other kinase has been shown to phosphorylate a C-terminal Ser, whereas those which phosphorylate an internal Ser residue are quite common in nature. This unique combination of relaxed substrate specificity and highly selective modification of a C-terminal Ser makes these kinases interesting candidates for applications in synthetic biology, and could complement the chemical toolbox available for *in vitro* peptide synthesis and modification nicely (Figure 5.5). Future studies could aim at evolving one of these kinases for robustness, speed and efficiency for these purposes[129].

## 5.4 Perspective and Outlook

Microorganisms often produce diverse secondary metabolites that are not directly involved in the normal growth, development or reproduction of the producing organism, but take part in important physiological and ecological roles, such as defense strategies. These secondary metabolites have been widely used by humans as medicines, flavorings and recreational drugs[1, 16].

The discovery and isolation of natural products in the past has often relied on bioassay-guided methods. This strategy has indeed led to the discovery of many bioactive compounds used in clinical therapy, such as artemisinin and vancomycin. In recent years, with the concurrent advances in microbial genome sequencing projects, fundamental understanding of natural product biosynthesis and analytical technologies, we are entering a Golden Age of natural products drug discovery[145]. In many cases, a single strain has revealed a multitude of gene

117

clusters associated with the biosynthesis of secondary metabolites. In combination with bioinformatics analysis, this has paved the way for the mining of genomes for diverse unknown bioactive compounds.

The genes generally responsible for the biosynthesis of natural products are often clustered together in a genome. This is also true for the *P. dendritiformis* strain studied in this thesis. The 12 different natural products gene clusters identified put *P. dendritiformis* forward as a rich source for the isolation and the structural characterization of novel natural products. However it is notable that most of the gene clusters identified remain "cryptic" with respect to their cognate metabolites.

Several approaches have been developed during recent years to activate the "cryptic" gene clusters, leading to a successful characterization of new natural products[146]. The first method is called "altering of chemical and physical conditions." This includes methods such as using Histone deacetylases (HDACs) as elicitors or with cocultivation in the presence of other species. A second strategy is called "genetic modification approaches." The latter includes method such as ribosome engineering and alteration of the transcriptional machinery, manipulation of pathway-specific and global regulators, and using quorum sensing as a tool for the regulation of secondary metabolism. The third method is called "heterologous expression." Series of useful, engineered heterologous expression hosts are available for this approach, such as *E. coli* and yeast.

In this thesis, the genome mining and heterologous expression strategy were successfully applied resulting in the discovery of paeninodin and phosphorylated paeninodin. This is the first natural product isolated from *P. dendritiformis.* Further investigation on other gene clusters identified could lead to discovery of more compounds from this strain.

Paeninodin is also the first lasso peptide isolated from Firmicutes. Our genome-mining data revealed that the lasso peptide gene clusters in Firmicutes represent a completely new family of lasso peptide. Further investigation into the isolation and bioactivity of these novel lasso peptides will open up new avenues for lasso peptides.
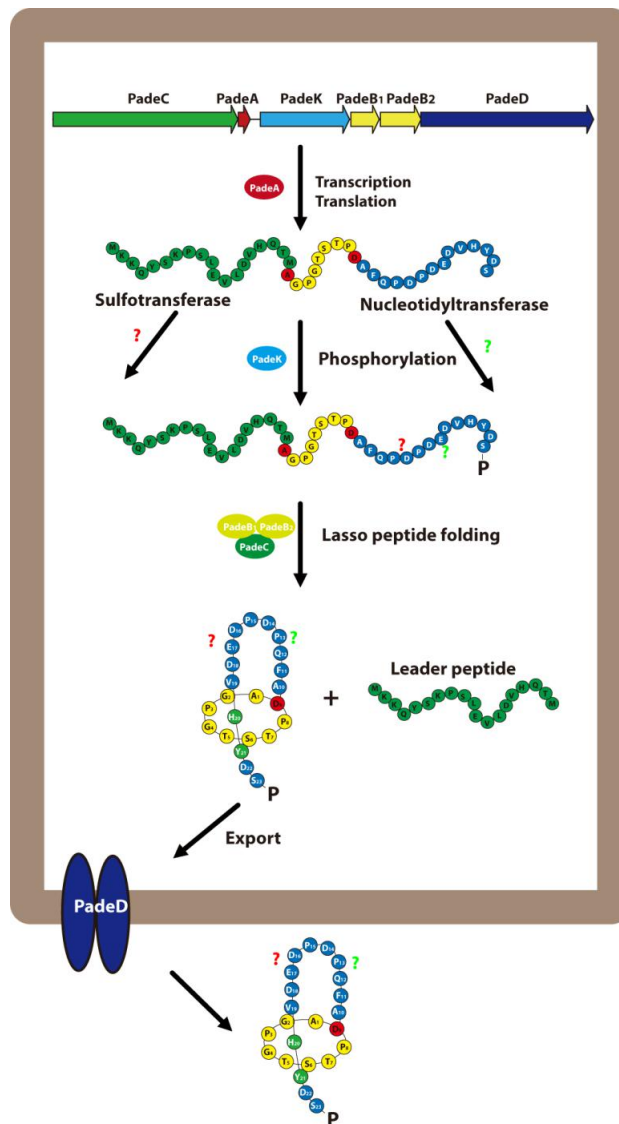
**Figure5.6:**Potential tailoring enzymes involved in lasso peptide modification.

In addition, the identification of various architectures of lasso peptide clusters with potential tailoring enzymes are a "gold mine" that will facilitate diversifying the lasso peptide structure and, hence, improve the physiological and biochemical properties of these fascinating molecules in the future (Figure 5.6). Moreover, these potential tailoring enzymes could also enrich chemists' enzyme toolkits and would facilitate drug design strategies in the future.

119

# 6. REFERENCES

1. Cragg, G. M., Newman, D. J., and Snader, K. M. (1997) Natural Products in Drug Discovery and Development, *Journal of Natural Products 60*, 52-60.
2. Newman, D. J., Cragg, G. M., and Snader, K. M. (2003) Natural Products as Sources of New Drugs over the Period 1981−2002, *Journal of Natural Products 66*, 1022-1037.
3. Koehn, F. E., and Carter, G. T. (2005) The evolving role of natural products in drug discovery, *Nat Rev Drug Discov 4*, 206-220.
4. Harvey, A. L. (2008) Natural products in drug discovery, *Drug Discovery Today 13*, 894-901.
5. South East Asian Quinine Artesunate Malaria Trial, g. Artesunate versus quinine for treatment of severe falciparum malaria: a randomised trial, *The Lancet 366*, 717-725.
6. Newman, D. J., and Cragg, G. M. (2007) Natural Products as Sources of New Drugs over the Last 25 Years, *Journal of Natural Products 70*, 461-477.
7. Vane, J. R., and Botting, R. M. (2003) The mechanism of action of aspirin, *Thrombosis Research 110*, 255-258.
8. Winkle, W. V., and Herwick, R. P. (1945) Penicillin-a review, *Journal of the American Pharmaceutical Association 34*, 97-109.
9. Hertweck, C. (2015) Natural Products as Source of Therapeutics against Parasitic Diseases, *Angewandte Chemie International Edition 54*, 14622-14624.
10. Peláez, F. (2006) The historical delivery of antibiotics from microbial natural products—Can history repeat?, *Biochemical Pharmacology 71*, 981-990.
11. Knight, V., Sanglier, J.-J., DiTullio, D., Braccili, S., Bonner, P., Waters, J., Hughes, D., and Zhang, L. (2003) Diversifying microbial natural products for drug discovery, *Applied Microbiology and Biotechnology 62*, 446-458.
12. Demain, L. A. Pharmaceutically active secondary metabolites of microorganisms, *Applied Microbiology and Biotechnology 52*, 455-463.
13. Williams, D. H., Stone, M. J., Hauck, P. R., and Rahman, S. K. (1989) Why Are Secondary Metabolites (Natural Products) Biosynthesized?, *Journal of Natural Products 52*, 1189-1208.
14. Martínez, J. L. (2008) Antibiotics and Antibiotic Resistance Genes in Natural Environments, *Science 321*, 365-367.
15. Demain, A. L. (1974) HOW DO ANTIBIOTIC-PRODUCING MICROORGANISMS AVOID SUICIDE?*, *Annals of the New York Academy of Sciences 235*, 601-612.
16. Arias , C. A., and Murray , B. E. (2009) Antibiotic-Resistant Bugs in the 21st Century — A Clinical Super-Challenge, *New England Journal of Medicine 360*, 439-443.
17. Finking, R., and Marahiel, M. A. (2004) Biosynthesis of Nonribosomal Peptides, *Annual Review of Microbiology 58*, 453-488.
18. Schwarzer, D., Finking, R., and Marahiel, M. A. (2003) Nonribosomal peptides: from genes to products, *Natural Product Reports 20*, 275-287.
19. Walsh, C. T. (2004) Polyketide and Nonribosomal Peptide Antibiotics: Modularity and Versatility, *Science 303*, 1805-1810.
20. Marahiel, M. A. (2009) Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis, *Journal of Peptide Science 15*, 799-807.
21. Watanabe, K., Hotta, K., Praseuth, A. P., Koketsu, K., Migita, A., Boddy, C. N., Wang, C. C. C., Oguri, H., and Oikawa, H. (2006) Total biosynthesis of antitumor nonribosomal peptides in Escherichia coli, *Nat Chem Biol 2*, 423-428.
22. Steller, S., Vollenbroich, D., Leenders, F., Stein, T., Conrad, B., Hofemeister, J., Jacques, P., Thonart, P., and Vater, J. (1999) Structural and functional organization of the fengycin synthetase multienzyme system from Bacillus subtilis b213 and A1/3, *Chemistry & Biology 6*, 31-41.
23. Velkov, T., and Lawen, A. (2003) Non-ribosomal peptide synthetases as technological platforms for the synthesis of highly modified peptide bioeffectors – Cyclosporin synthetase as a complex example, In *Biotechnology Annual Review*, pp 151-197, Elsevier.

24. Du, L., Sánchez, C., Chen, M., Edwards, D. J., and Shen, B. (2000) The biosynthetic gene cluster for the antitumor drug bleomycin from Streptomyces verticillus ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase, *Chemistry & Biology 7*, 623-642.

25. Sieber, S. A., and Marahiel, M. A. (2005) Molecular Mechanisms Underlying Nonribosomal Peptide Synthesis: Approaches to New Antibiotics, *Chemical Reviews 105*, 715-738.

26. Fischbach, M. A., and Walsh, C. T. (2006) Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms, *Chemical Reviews 106*, 3468-3496.

27. Walsh, C. T., Chen, H., Keating, T. A., Hubbard, B. K., Losey, H. C., Luo, L., Marshall, C. G., Miller, D. A., and Patel, H. M. (2001) Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines, *Current Opinion in Chemical Biology 5*, 525-534.

28. Ling, L. L., Schneider, T., Peoples, A. J., Spoering, A. L., Engels, I., Conlon, B. P., Mueller, A., Schaberle, T. F., Hughes, D. E., Epstein, S., Jones, M., Lazarides, L., Steadman, V. A., Cohen, D. R., Felix, C. R., Fetterman, K. A., Millett, W. P., Nitti, A. G., Zullo, A. M., Chen, C., and Lewis, K. (2015) A new antibiotic kills pathogens without detectable resistance, *Nature 517*, 455-459.

29. von Nussbaum, F., and Süssmuth, R. D. (2015) Multiple Attack on Bacteria by the New Antibiotic Teixobactin, *Angewandte Chemie International Edition 54*, 6684-6686.

30. Belin, P., Moutiez, M., Lautru, S., Seguin, J., Pernodet, J.-L., and Gondry, M. (2012) The nonribosomal synthesis of diketopiperazines in tRNA-dependent cyclodipeptide synthase pathways, *Natural Product Reports 29*, 961-979.

31. Gondry, M., Sauguet, L., Belin, P., Thai, R., Amouroux, R., Tellier, C., Tuphile, K., Jacquet, M., Braud, S., Courcon, M., Masson, C., Dubois, S., Lautru, S., Lecoq, A., Hashimoto, S.-i., Genet, R., and Pernodet, J.-L. (2009) Cyclodipeptide synthases are a family of tRNA-dependent peptide bond-forming enzymes, *Nat Chem Biol 5*, 414-420.

32. Ortiz-Castro, R., Díaz-Pérez, C., Martínez-Trujillo, M., del Río, R. E., Campos-García, J., and López-Bucio, J. (2011) Transkingdom signaling based on bacterial cyclodipeptides with auxin activity in plants, *Proceedings of the National Academy of Sciences 108*, 7253-7258.

33. Sauguet, L., Moutiez, M., Li, Y., Belin, P., Seguin, J., Le Du, M.-H., Thai, R., Masson, C., Fonvielle, M., Pernodet, J.-L., Charbonnier, J.-B., and Gondry, M. (2011) Cyclodipeptide synthases, a family of class-I aminoacyl-tRNA synthetase-like enzymes involved in non-ribosomal peptide synthesis, *Nucleic Acids Research*. 39(10):4475-89

34. Lautru, S., Gondry, M., Genet, R., and Pernodet, J.-L. (2002) The Albonoursin Gene Cluster of S. noursei: Biosynthesis of Diketopiperazine Metabolites Independent of Nonribosomal Peptide Synthetases, *Chemistry & Biology 9*, 1355-1364.

35. Giessen, T. W., von Tesmar, A. M., and Marahiel, M. A. (2013) A tRNA-Dependent Two-Enzyme Pathway for the Generation of Singly and Doubly Methylated Ditryptophan 2,5-Diketopiperazines, *Biochemistry 52*, 4274-4283.

36. Giessen, Tobias W., von Tesmar, Alexander M., and Marahiel, Mohamed A. (2013) Insights into the Generation of Structural Diversity in a tRNA-Dependent Pathway for Highly Modified Bioactive Cyclic Dipeptides, *Chemistry & Biology 20*, 828-838.

37. Giessen, T., and Marahiel, M. (2014) The tRNA-Dependent Biosynthesis of Modified Cyclic Dipeptides, *International Journal of Molecular Sciences 15*, 14610.

38. Kolter, R., and Moreno, F. (1992) Genetics of Ribosomally Synthesized Peptide Antibiotics, *Annual Review of Microbiology 46*, 141-161.

39. Papagianni, M. (2003) Ribosomally synthesized peptides with antimicrobial properties: biosynthesis, structure, function, and applications, *Biotechnology Advances 21*, 465-499.

40. Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., Camarero, J. A., Campopiano, D. J., Challis, G. L., Clardy, J., Cotter, P. D., Craik, D. J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P. C., Entian, K.-D., Fischbach, M. A., Garavelli, J. S., Goransson, U., Gruber, C. W., Haft, D. H., Hemscheidt, T. K., Hertweck, C., Hill, C.,

121

Horswill, A. R., Jaspars, M., Kelly, W. L., Klinman, J. P., Kuipers, O. P., Link, A. J., Liu, W., Marahiel, M. A., Mitchell, D. A., Moll, G. N., Moore, B. S., Muller, R., Nair, S. K., Nes, I. F., Norris, G. E., Olivera, B. M., Onaka, H., Patchett, M. L., Piel, J., Reaney, M. J. T., Rebuffat, S., Ross, R. P., Sahl, H.-G., Schmidt, E. W., Selsted, M. E., Severinov, K., Shen, B., Sivonen, K., Smith, L., Stein, T., Sussmuth, R. D., Tagg, J. R., Tang, G.-L., Truman, A. W., Vederas, J. C., Walsh, C. T., Walton, J. D., Wenzel, S. C., Willey, J. M., and van der Donk, W. A. (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature, *Natural Product Reports 30*, 108-160.

41. Velásquez, J. E., and van der Donk, W. A. (2011) Genome mining for ribosomally synthesized natural products, *Current Opinion in Chemical Biology 15*, 11-21.

42. Kersten, R. D., Yang, Y.-L., Xu, Y., Cimermancic, P., Nam, S.-J., Fenical, W., Fischbach, M. A., Moore, B. S., and Dorrestein, P. C. (2011) A mass spectrometry–guided genome mining approach for natural product peptidogenomics, *Nat Chem Biol 7*, 794-802.

43. Begley, M., Cotter, P. D., Hill, C., and Ross, R. P. (2009) Identification of a Novel Two-Peptide Lantibiotic, Lichenicidin, following Rational Genome Mining for LanM Proteins, *Applied and Environmental Microbiology 75*, 5451-5460.

44. Oman, T. J., and van der Donk, W. A. (2010) Follow the leader: the use of leader peptides to guide natural product biosynthesis, *Nat Chem Biol 6*, 9-18.

45. Crone, W. J. K., Leeper, F. J., and Truman, A. W. (2012) Identification and characterisation of the gene cluster for the anti-MRSA antibiotic bottromycin: expanding the biosynthetic diversity of ribosomal peptides, *Chemical Science 3*, 3516-3521.

46. Gomez-Escribano, J. P., Song, L., Bibb, M. J., and Challis, G. L. (2012) Posttranslational [small beta]-methylation and macrolactamidination in the biosynthesis of the bottromycin complex of ribosomal peptide antibiotics, *Chemical Science 3*, 3522-3525.

47. Huo, L., Rachid, S., Stadler, M., Wenzel, Silke C., and Müller, R. (2012) Synthetic Biotechnology to Study and Engineer Ribosomal Bottromycin Biosynthesis, *Chemistry & Biology 19*, 1278-1287.

48. Yang, X., and van der Donk, W. A. (2013) Ribosomally Synthesized and Post-Translationally Modified Peptide Natural Products: New Insights into the Role of Leader and Core Peptides during Biosynthesis, *Chemistry – A European Journal 19*, 7662-7677.

49. Willey, J. M., and van der Donk, W. A. (2007) Lantibiotics: Peptides of Diverse Structure and Function, *Annual Review of Microbiology 61*, 477-501.

50. Knerr, P. J., and Donk, W. A. v. d. (2012) Discovery, Biosynthesis, and Engineering of Lantipeptides, *Annual Review of Biochemistry 81*, 479-505.

51. Chatterjee, C., Paul, M., Xie, L., and van der Donk, W. A. (2005) Biosynthesis and Mode of Action of Lantibiotics, *Chemical Reviews 105*, 633-684.

52. Yu, Y., Zhang, Q., and van der Donk, W. A. (2013) Insights into the evolution of lanthipeptide biosynthesis, *Protein Science 22*, 1478-1489.

53. Delves-Broughton, J., Blackburn, P., Evans, R. J., and Hugenholtz, J. Applications of the bacteriocin, nisin, *Antonie van Leeuwenhoek 69*, 193-202.

54. de Ruyter, P. G., Kuipers, O. P., and de Vos, W. M. (1996) Controlled gene expression systems for Lactococcus lactis with the food-grade inducer nisin, *Applied and Environmental Microbiology 62*, 3662-3667.

55. Stevens, K. A., Sheldon, B. W., Klapes, N. A., and Klaenhammer, T. R. (1991) Nisin treatment for inactivation of Salmonella species and other gram-negative bacteria, *Applied and Environmental Microbiology 57*, 3613-3615.

56. Navarro, J., Chabot, J., Sherrill, K., Aneja, R., Zahler, S. A., and Racker, E. (1985) Interaction of duramycin with artificial and natural membranes, *Biochemistry 24*, 4645-4650.

57. Stone, D. K., Xie, X. S., and Racker, E. (1984) Inhibition of clathrin-coated vesicle acidification by duramycin, *Journal of Biological Chemistry 259*, 2701-2703.

58. Märki, F., Hänni, E., Fredenhagen, A., and van Oostrum, J. (1991) Mode of action of the lanthionine-containing peptide antibiotics duramycin, duramycin B and C, and

cinnamycin as indirect inhibitors of phospholipase A2, *Biochemical Pharmacology 42*, 2027-2035.

59. Jones, A. M., and Helm, J. M. (2012) Emerging Treatments in Cystic Fibrosis, *Drugs 69*, 1903-1910.

60. Mohr, K. I., Volz, C., Jansen, R., Wray, V., Hoffmann, J., Bernecker, S., Wink, J., Gerth, K., Stadler, M., and Müller, R. (2015) Pinensins: The First Antifungal Lantibiotics, *Angewandte Chemie International Edition 54*, 11254-11258.

61. Bassler, B. L., and Losick, R. (2006) Bacterially Speaking, *Cell 125*, 237-246.

62. Kodani, S., Lodato, M. A., Durrant, M. C., Picart, F., and Willey, J. M. (2005) SapT, a lanthionine-containing peptide involved in aerial hyphae formation in the streptomycetes, *Molecular Microbiology 58*, 1368-1380.

63. Waisvisz, J. M., van der Hoeven, M. G., van Peppen, J., and Zwennis, W. C. M. (1957) Bottromycin. I. A New Sulfur-containing Antibiotic, *Journal of the American Chemical Society 79*, 4520-4521.

64. Waisvisz, J. M., van der Hoeven, M. G., and Nijenhuis, B. t. (1957) The Structure of the Sulfur-containing Moiety of Bottromycin, *Journal of the American Chemical Society 79*, 4524-4527.

65. Hou, Y., Tianero, M. D. B., Kwan, J. C., Wyche, T. P., Michel, C. R., Ellis, G. A., Vazquez-Rivera, E., Braun, D. R., Rose, W. E., Schmidt, E. W., and Bugni, T. S. (2012) Structure and Biosynthesis of the Antibiotic Bottromycin D, *Organic Letters 14*, 5050-5053.

66. Hughes, R. A., and Moody, C. J. (2007) From Amino Acids to Heteroaromatics—Thiopeptide Antibiotics, Nature's Heterocyclic Peptides, *Angewandte Chemie International Edition 46*, 7930-7954.

67. Bagley, M. C., Dale, J. W., Merritt, E. A., and Xiong, X. (2005) Thiopeptide Antibiotics, *Chemical Reviews 105*, 685-714.

68. Liao, R., Duan, L., Lei, C., Pan, H., Ding, Y., Zhang, Q., Chen, D., Shen, B., Yu, Y., and Liu, W. (2009) Thiopeptide Biosynthesis Featuring Ribosomally Synthesized Precursor Peptides and Conserved Posttranslational Modifications, *Chemistry & Biology 16*, 141-147.

69. Morris, R. P., Leeds, J. A., Naegeli, H. U., Oberer, L., Memmert, K., Weber, E., LaMarche, M. J., Parker, C. N., Burrer, N., Esterow, S., Hein, A. E., Schmitt, E. K., and Krastel, P. (2009) Ribosomally Synthesized Thiopeptide Antibiotics Targeting Elongation Factor Tu, *Journal of the American Chemical Society 131*, 5946-5955.

70. Arndt, H.-D., Schoof, S., and Lu, J.-Y. (2009) Thiopeptide Antibiotic Biosynthesis, *Angewandte Chemie International Edition 48*, 6770-6773.

71. Li, C., and Kelly, W. L. (2010) Recent advances in thiopeptide antibiotic biosynthesis, *Natural Product Reports 27*, 153-164.

72. Bowers, A. A., Walsh, C. T., and Acker, M. G. (2010) Genetic Interception and Structural Characterization of Thiopeptide Cyclization Precursors from Bacillus cereus, *Journal of the American Chemical Society 132*, 12182-12184.

73. Leikoski, N., Liu, L., Jokela, J., Wahlsten, M., Gugger, M., Calteau, A., Permi, P., Kerfeld, Cheryl A., Sivonen, K., and Fewer, David P. (2013) Genome Mining Expands the Chemical Diversity of the Cyanobactin Family to Include Highly Modified Linear Peptides, *Chemistry & Biology 20*, 1033-1043.

74. Donia, Mohamed S., and Schmidt, Eric W. (2011) Linking Chemistry and Genetics in the Growing Cyanobactin Natural Products Family, *Chemistry & Biology 18*, 508-519.

75. McIntosh, J. A., and Schmidt, E. W. (2010) Marine Molecular Machines: Heterocyclization in Cyanobactin Biosynthesis, *ChemBioChem 11*, 1413-1421.

76. Leikoski, N., Fewer, D. P., and Sivonen, K. (2009) Widespread Occurrence and Lateral Transfer of the Cyanobactin Biosynthesis Gene Cluster in Cyanobacteria, *Applied and Environmental Microbiology 75*, 853-857.

77. Sivonen, K., Leikoski, N., Fewer, D. P., and Jokela, J. (2010) Cyanobactins—ribosomal cyclic peptides produced by cyanobacteria, *Applied Microbiology and Biotechnology 86*, 1213-1225.

78. Donia, M. S., Ravel, J., and Schmidt, E. W. (2008) A global assembly line for cyanobactins, *Nat Chem Biol 4*, 341-343.
79. Williams, A. B., and Jacobs, R. S. (1993) A marine natural product, patellamide D, reverses multidrug resistance in a human leukemic cell line, *Cancer Letters 71*, 97-102.
80. Long, P. F., Dunlap, W. C., Battershill, C. N., and Jaspars, M. (2005) Shotgun Cloning and Heterologous Expression of the Patellamide Gene Cluster as a Strategy to Achieving Sustained Metabolite Production, *ChemBioChem 6*, 1760-1765.
81. Schmidt, E. W., Nelson, J. T., Rasko, D. A., Sudek, S., Eisen, J. A., Haygood, M. G., and Ravel, J. (2005) Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella, *Proceedings of the National Academy of Sciences of the United States of America 102*, 7315-7320.
82. Hamada, T., Sugawara, T., Matsunaga, S., and Fusetani, N. (1994) Polytheonamides, unprecedented highly cytotoxic polypeptides from the marine sponge Theonella swinhoei 2. Structure elucidation, *Tetrahedron Letters 35*, 609-612.
83. Hamada, T., Sugawara, T., Matsunaga, S., and Fusetani, N. (1994) Polytheonamides, unprecedented highly cytotoxic polypeptides, from the marine sponge theonella swinhoei : 1. Isolation and component amino acids, *Tetrahedron Letters 35*, 719-720.
84. Hamada, T., Matsunaga, S., Fujiwara, M., Fujita, K., Hirota, H., Schmucki, R., Güntert, P., and Fusetani, N. (2010) Solution Structure of Polytheonamide B, a Highly Cytotoxic Nonribosomal Polypeptide from Marine Sponge, *Journal of the American Chemical Society 132*, 12941-12945.
85. Hamada, T., Matsunaga, S., Yano, G., and Fusetani, N. (2005) Polytheonamides A and B, Highly Cytotoxic, Linear Polypeptides with Unprecedented Structural Features, from the Marine Sponge, Theonella swinhoei, *Journal of the American Chemical Society 127*, 110-118.
86. Freeman, M. F., Gurgui, C., Helf, M. J., Morinaka, B. I., Uria, A. R., Oldham, N. J., Sahl, H.-G., Matsunaga, S., and Piel, J. (2012) Metagenome Mining Reveals Polytheonamides as Posttranslationally Modified Ribosomal Peptides, *Science 338*, 387-390.
87. Flühe, L., and Marahiel, M. A. (2013) Radical S-adenosylmethionine enzyme catalyzed thioether bond formation in sactipeptide biosynthesis, *Current Opinion in Chemical Biology 17*, 605-612.
88. Wieckowski, B. M., Hegemann, J. D., Mielcarek, A., Boss, L., Burghaus, O., and Marahiel, M. A. (2015) The PqqD homologous domain of the radical SAM enzyme ThnB is required for thioether bond formation during thurincin H maturation, *FEBS Letters 589*, 1802-1806.
89. Flühe, L., Burghaus, O., Wieckowski, B. M., Giessen, T. W., Linne, U., and Marahiel, M. A. (2013) Two [4Fe-4S] Clusters Containing Radical SAM Enzyme SkfB Catalyze Thioether Bond Formation during the Maturation of the Sporulation Killing Factor, *Journal of the American Chemical Society 135*, 959-962.
90. Flühe, L., Knappe, T. A., Gattner, M. J., Schäfer, A., Burghaus, O., Linne, U., and Marahiel, M. A. (2012) The radical SAM enzyme AlbA catalyzes thioether bond formation in subtilosin A, *Nat Chem Biol 8*, 350-357.
91. BABASAKI, K., TAKAO, T., SHIMONISHI, Y., and KURAHASHI, K. (1985) Subtilosin A, a New Antibiotic Peptide Produced by Bacillus subtilis 168: Isolation, Structural Analysis, and Biogenesis, *Journal of Biochemistry 98*, 585-603.
92. Shelburne, C. E., An, F. Y., Dholpe, V., Ramamoorthy, A., Lopatin, D. E., and Lantz, M. S. (2007) The spectrum of antimicrobial activity of the bacteriocin subtilosin A, *Journal of Antimicrobial Chemotherapy 59*, 297-300.
93. Kawulka, K. E., Sprules, T., Diaper, C. M., Whittal, R. M., McKay, R. T., Mercier, P., Zuber, P., and Vederas, J. C. (2004) Structure of Subtilosin A, a Cyclic Antimicrobial Peptide from Bacillus subtilis with Unusual Sulfur to α-Carbon Cross-Links: Formation and Reduction of α-Thio-α-Amino Acid Derivatives, *Biochemistry 43*, 3385-3395.
94. Kawulka, K., Sprules, T., McKay, R. T., Mercier, P., Diaper, C. M., Zuber, P., and Vederas, J. C. (2003) Structure of Subtilosin A, an Antimicrobial Peptide from Bacillus subtilis with

Unusual Posttranslational Modifications Linking Cysteine Sulfurs to α-Carbons of Phenylalanine and Threonine, *Journal of the American Chemical Society 125*, 4726-4727.

95. Stein, T. (2005) Bacillus subtilis antibiotics: structures, syntheses and specific functions, *Molecular Microbiology 56*, 845-857.

96. Allenby, N. E. E., Watts, C. A., Homuth, G., Prágai, Z., Wipat, A., Ward, A. C., and Harwood, C. R. (2006) Phosphate Starvation Induces the Sporulation Killing Factor of Bacillus subtilis, *Journal of Bacteriology 188*, 5299-5303.

97. Hegemann, J. D., Zimmermann, M., Xie, X., and Marahiel, M. A. (2015) Lasso Peptides: An Intriguing Class of Bacterial Natural Products, *Accounts of Chemical Research 48*, 1909-1919.

98. Maksimov, M. O., Pan, S. J., and James Link, A. (2012) Lasso peptides: structure, function, biosynthesis, and engineering, *Natural Product Reports 29*, 996-1006.

99. Maksimov, M. O., and Link, A. J. (2013) Prospecting genomes for lasso peptides, *Journal of Industrial Microbiology & Biotechnology 41*, 333-344.

100. Knappe, T. A., Linne, U., Xie, X., and Marahiel, M. A. (2010) The glucagon receptor antagonist BI-32169 constitutes a new class of lasso peptides, *FEBS Letters 584*, 785-789.

101. Ducasse, R., Yan, K.-P., Goulard, C., Blond, A., Li, Y., Lescop, E., Guittet, E., Rebuffat, S., and Zirah, S. (2012) Sequence Determinants Governing the Topology and Biological Activity of a Lasso Peptide, Microcin J25, *ChemBioChem 13*, 371-380.

102. Pavlova, O., Mukhopadhyay, J., Sineva, E., Ebright, R. H., and Severinov, K. (2008) Systematic Structure-Activity Analysis of Microcin J25, *Journal of Biological Chemistry 283*, 25589-25595.

103. Solbiati, J. O., Ciaccio, M., Farías, R. N., González-Pastor, J. E., Moreno, F., and Salomón, R. A. (1999) Sequence Analysis of the Four Plasmid Genes Required To Produce the Circular Peptide Antibiotic Microcin J25, *Journal of Bacteriology 181*, 2659-2662.

104. Gavrish, E., Sit, Clarissa S., Cao, S., Kandror, O., Spoering, A., Peoples, A., Ling, L., Fetterman, A., Hughes, D., Bissell, A., Torrey, H., Akopian, T., Mueller, A., Epstein, S., Goldberg, A., Clardy, J., and Lewis, K. (2014) Lassomycin, a Ribosomally Synthesized Cyclic Peptide, Kills Mycobacterium tuberculosis by Targeting the ATP-Dependent Protease ClpC1P1P2, *Chemistry & Biology 21*, 509-518.

105. Metelev, M., Tietz, Jonathan I., Melby, Joel O., Blair, Patricia M., Zhu, L., Livnat, I., Severinov, K., and Mitchell, Douglas A. (2015) Structure, Bioactivity, and Resistance Mechanism of Streptomonomicin, an Unusual Lasso Peptide from an Understudied Halophilic Actinomycete, *Chemistry & Biology 22*, 241-250.

106. Inokoshi, J., Matsuhama, M., Miyake, M., Ikeda, H., and Tomoda, H. (2012) Molecular cloning of the gene cluster for lariatin biosynthesis of Rhodococcus jostii K01-B0171, *Applied Microbiology and Biotechnology 95*, 451-460.

107. Elsayed, S. S., Trusch, F., Deng, H., Raab, A., Prokes, I., Busarakam, K., Asenjo, J. A., Andrews, B. A., van West, P., Bull, A. T., Goodfellow, M., Yi, Y., Ebel, R., Jaspars, M., and Rateb, M. E. (2015) Chaxapeptin, a Lasso Peptide from Extremotolerant Streptomyces leeuwenhoekii Strain C58 from the Hyperarid Atacama Desert, *The Journal of Organic Chemistry 80*, 10252-10260.

108. Li, Y., Ducasse, R., Zirah, S., Blond, A., Goulard, C., Lescop, E., Giraud, C., Hartke, A., Guittet, E., Pernodet, J.-L., and Rebuffat, S. (2015) Characterization of Sviceucin from Streptomyces Provides Insight into Enzyme Exchangeability and Disulfide Bond Formation in Lasso Peptides, *ACS Chemical Biology 10*, 2641-2649.

109. Xie, X., and Marahiel, M. A. (2012) NMR as an Effective Tool for the Structure Determination of Lasso Peptides, *ChemBioChem 13*, 621-625.

110. Ogawa, T., Ochiai, K., Tanaka, T., Tsukuda, E., Chiba, S., Yano, K., Yamasaki, M., Yoshida, M., and Matsuda, Y. (1995) RES-701-2,-3 and-4, novel and selective endothelin type B receptor antagonists produced by Streptomyces sp. I. Taxonomy of producing strains, fermentation, isolation, and biochemical properties, *The Journal of antibiotics 48*, 1213-1220.

125

111. Hegemann, J. D., Zimmermann, M., Xie, X., and Marahiel, M. A. (2013) Caulosegnins I–III: A Highly Diverse Group of Lasso Peptides Derived from a Single Biosynthetic Gene Cluster, *Journal of the American Chemical Society 135*, 210-222.

112. Hegemann, J. D., Zimmermann, M., Zhu, S., Klug, D., and Marahiel, M. A. (2013) Lasso peptides from proteobacteria: Genome mining employing heterologous expression and mass spectrometry, *Peptide Science 100*, 527-542.

113. Maksimov, M. O., and Link, A. J. (2013) Discovery and Characterization of an Isopeptidase That Linearizes Lasso Peptides, *Journal of the American Chemical Society 135*, 12038-12047.

114. Zimmermann, M., Hegemann, Julian D., Xie, X., and Marahiel, Mohamed A. (2013) The Astexin-1 Lasso Peptides: Biosynthesis, Stability, and Structural Studies, *Chemistry & Biology 20*, 558-569.

115. Hegemann, J. D., Zimmermann, M., Zhu, S., Steuber, H., Harms, K., Xie, X., and Marahiel, M. A. (2014) Xanthomonins I–III: A New Class of Lasso Peptides with a Seven-Residue Macrolactam Ring, *Angewandte Chemie International Edition 53*, 2230-2234.

116. Zimmermann, M., Hegemann, J. D., Xie, X., and Marahiel, M. A. (2014) Characterization of caulonodin lasso peptides revealed unprecedented N-terminal residues and a precursor motif essential for peptide maturation, *Chemical Science 5*, 4032-4043.

117. Maksimov, M. O., Pelczer, I., and Link, A. J. (2012) Precursor-centric genome-mining approach for lasso peptide discovery, *Proceedings of the National Academy of Sciences 109*, 15223-15228.

118. Maksimov, M. O., Koos, J. D., Zong, C., Lisko, B., and Link, A. J. (2015) Elucidating the Specificity Determinants of the AtxE2 Lasso Peptide Isopeptidase, *Journal of Biological Chemistry 290*, 30806-30812.

119. Knappe, T. A., Manzenrieder, F., Mas-Moruno, C., Linne, U., Sasse, F., Kessler, H., Xie, X., and Marahiel, M. A. (2011) Introducing Lasso Peptides as Molecular Scaffolds for Drug Design: Engineering of an Integrin Antagonist, *Angewandte Chemie International Edition 50*, 8714-8717.

120. Hegemann, J. D., De Simone, M., Zimmermann, M., Knappe, T. A., Xie, X., Di Leva, F. S., Marinelli, L., Novellino, E., Zahler, S., Kessler, H., and Marahiel, M. A. (2014) Rational Improvement of the Affinity and Selectivity of Integrin Binding of Grafted Lasso Peptides, *Journal of Medicinal Chemistry 57*, 5829-5834.

121. Clarke, D. J., and Campopiano, D. J. (2007) Maturation of McjA precursor peptide into active microcin MccJ25, *Organic & Biomolecular Chemistry 5*, 2564-2566.

122. Yan, K.-P., Li, Y., Zirah, S., Goulard, C., Knappe, T. A., Marahiel, M. A., and Rebuffat, S. (2012) Dissecting the Maturation Steps of the Lasso Peptide Microcin J25 in vitro, *ChemBioChem 13*, 1046-1052.

123. Cheung, W. L., Pan, S. J., and Link, A. J. (2010) Much of the Microcin J25 Leader Peptide is Dispensable, *Journal of the American Chemical Society 132*, 2514-2515.

124. Choudhury, H. G., Tong, Z., Mathavan, I., Li, Y., Iwata, S., Zirah, S., Rebuffat, S., van Veen, H. W., and Beis, K. (2014) Structure of an antibacterial peptide ATP-binding cassette transporter in a novel outward occluded state, *Proceedings of the National Academy of Sciences 111*, 9145-9150.

125. Rosengren, K. J., Clark, R. J., Daly, N. L., Göransson, U., Jones, A., and Craik, D. J. (2003) Microcin J25 Has a Threaded Sidechain-to-Backbone Ring Structure and Not a Head-to-Tail Cyclized Backbone, *Journal of the American Chemical Society 125*, 12464-12474.

126. Pan, S. J., Rajniak, J., Maksimov, M. O., and Link, A. J. (2012) The role of a conserved threonine residue in the leader peptide of lasso peptide precursors, *Chemical Communications 48*, 1880-1882.

127. Knappe, T. A., Linne, U., Zirah, S., Rebuffat, S., Xie, X., and Marahiel, M. A. (2008) Isolation and Structural Characterization of Capistruin, a Lasso Peptide Predicted from the Genome Sequence of Burkholderia thailandensis E264, *Journal of the American Chemical Society 130*, 11446-11454.

128. Burkhart, B. J., Hudson, G. A., Dunbar, K. L., and Mitchell, D. A. (2015) A prevalent peptide-binding domain guides ribosomal natural product biosynthesis, *Nat Chem Biol* *11*, 564-570.

129. Zhu, S., Hegemann, J. D., Fage, C. D., Zimmermann, M., Xie, X., Linne, U., and Marahiel, M. A. (2016) Insights into the Unique Phosphorylation of the Lasso Peptide Paeninodin, *Journal of Biological Chemistry* .doi: 10.1074/jbc.M116.72210

130. Bailey, T. L., Williams, N., Misleh, C., and Li, W. W. (2006) MEME: discovering and analyzing DNA and protein sequence motifs, *Nucleic Acids Research* *34*, W369-W373.

131. Chiu, J., March, P. E., Lee, R., and Tillett, D. (2004) Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h, *Nucleic Acids Research* *32*, e174.

132. Ficarro, S. B., McCleland, M. L., Stukenberg, P. T., Burke, D. J., Ross, M. M., Shabanowitz, J., Hunt, D. F., and White, F. M. (2002) Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae, *Nat Biotech* *20*, 301-305.

133. Jeanne Dit Fouque, K., Afonso, C., Zirah, S., Hegemann, J. D., Zimmermann, M., Marahiel, M. A., Rebuffat, S., and Lavanant, H. (2015) Ion Mobility–Mass Spectrometry of Lasso Peptides: Signature of a Rotaxane Topology, *Analytical Chemistry* *87*, 1166-1172.

134. Knappe, T. A., Linne, U., Robbel, L., and Marahiel, M. A. (2009) Insights into the Biosynthesis and Stability of the Lasso Peptide Capistruin, *Chemistry & Biology* *16*, 1290-1298.

135. Fieulaine, S., Morera, S., Poncet, S., Mijakovic, I., Galinier, A., Janin, J., Deutscher, J., and Nessler, S. (2002) X-ray structure of a bifunctional protein kinase in complex with its protein substrate HPr, *Proceedings of the National Academy of Sciences* *99*, 13437-13441.

136. Nessler, S. (2005) The bacterial HPr kinase/phosphorylase: A new type of Ser/Thr kinase as antimicrobial target, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* *1754*, 126-131.

137. Nessler, S., Fieulaine, S., Poncet, S., Galinier, A., Deutscher, J., and Janin, J. (2003) HPr Kinase/Phosphorylase, the Sensor Enzyme of Catabolite Repression in Gram-Positive Bacteria: Structural Aspects of the Enzyme and the Complex with Its Protein Substrate, *Journal of Bacteriology* *185*, 4003-4010.

138. Poncet, S., Mijakovic, I., Nessler, S., Gueguen-Chaignon, V., Chaptal, V., Galinier, A., Boël, G., Mazé, A., and Deutscher, J. (2004) HPr kinase/phosphorylase, a Walker motif A-containing bifunctional sensor enzyme controlling catabolite repression in Gram-positive bacteria, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* *1697*, 123-135.

139. Alippi, A. M., López, A. C., and Aguilar, O. M. (2002) Differentiation of Paenibacillus larvae subsp. larvae, the Cause of American Foulbrood of Honeybees, by Using PCR and Restriction Fragment Analysis of Genes Encoding 16S rRNA, *Applied and Environmental Microbiology* *68*, 3655-3660.

140. Sirota-Madi, A., Olender, T., Helman, Y., Brainis, I., Finkelshtein, A., Roth, D., Hagai, E., Leshkowitz, D., Brodsky, L., Galatenko, V., Nikolaev, V., Gutnick, D. L., Lancet, D., and Ben-Jacob, E. (2012) Genome Sequence of the Pattern-Forming Social Bacterium Paenibacillus dendritiformis C454 Chiral Morphotype, *Journal of Bacteriology* *194*, 2127-2128.

141. Timmusk, S., Grantcharova, N., and Wagner, E. G. H. (2005) Paenibacillus polymyxa Invades Plant Roots and Forms Biofilms, *Applied and Environmental Microbiology* *71*, 7292-7300.

142. Muller, S., Garcia-Gonzalez, E., Genersch, E., and Sussmuth, R. D. (2015) Involvement of secondary metabolites in the pathogenesis of the American foulbrood of honey bees caused by Paenibacillus larvae, *Natural Product Reports* *32*, 765-778.

143. Be'er, A., Zhang, H. P., Florin, E.-L., Payne, S. M., Ben-Jacob, E., and Swinney, H. L. (2009) Deadly competition between sibling bacterial colonies, *Proceedings of the National Academy of Sciences* *106*, 428-433.

127

144. Durante-Rodríguez, G., Valderrama, J. A., Mancheño, J. M., Rivas, G., Alfonso, C., Arias-Palomo, E., Llorca, O., García, J. L., Díaz, E., and Carmona, M. (2010) Biochemical Characterization of the Transcriptional Regulator BzdR from Azoarcus sp. CIB, *Journal of Biological Chemistry 285*, 35694-35705.

145. Shen, B. (2015) A New Golden Age of Natural Products Drug Discovery, *Cell 163*, 1297-1300.

146. Zarins-Tutt, J. S., Barberi, T. T., Gao, H., Mearns-Spragg, A., Zhang, L., Newman, D. J., and Goss, R. J. M. (2016) Prospecting for new bacterial metabolites: a glossary of approaches for inducing, activating and upregulating the biosynthesis of bacterial cryptic or silent natural products, *Natural Product Reports 33*, 54-72.

# 7. Appendix

## DNA sequences of the studied proteins

In the following, the DNA sequences of the proteins recombinantly produced in this work are shown.

### PadeK

```
ATGACCGAACGAGCAGCGGTAAGGACTGACCATTATAAAGCTTTTGGCTTCCGTATTGAGAGCGACTTTGTATTACC
GGAACTGCCTCCGGCTGGCGAGCGCGAGCCTCTGGACAACATTACGGTACGGCGGACAGACTTGCAGCCTCTCTGGA
ACAGCTCCATTCATTTCTATGGAAACTTCGCTATTCTGGATCACGGCCGTACGGTTATGTTTCGGGTACCTGGGGCT
GCCATATATGCGGTACAGGATGCCTCCTCCATCCTGGTCTCTCCGTTCGATCAGGCGGAAGAGAACTGGGTTCGGCT
TTTTATTCTGGGGACTTGTATGGGCATTATCTTGCTGCAGCGCAAAATTATGCCTTTGCACGGCAGCGCCGTAGCTA
TTGACGGGAAGGCCTATGCCATTATCGGCGAATCCGGGGCGGGCAAATCGACCTTGGCCTTGCACCTGGTGTCCGAG
GGATATCCGCTGCTTAGCGATGATGTTATTCCCGTCGTGATGACTCAAGGCTCTCCTTGGGTTGTTCCTTCTTACCC
ACAGCAAAATTATGGCTGGACACACTGAAGCACATGGGAATGGACAATGCCAATTACACCCCGCTTTATGAGCGGA
ATACGAAATTTGCCGTGCCCGTGGGGTCCAACTTTCATGATGAACCTCTTCCGTTGGCAAGTATCTTTGAGTTGGTT
CCATGGGATGCCGCAACTCATATTGCGCCGATCCAGGGGATGGAGCGCTTTCGCGTCCTCTTTCACCATACGTACCG
TAATTTTTTGGTGCAGCCGCTGGGATTGATGGAATGGCACTTCAAGACCCTGTCATCCTTCGTTCACCAGATCGGGA
TGTACCGTTTACACCGTCCCATGGTTGGATTCAGCACGCTTGATCTGACATCGCACATTTTAAACATTACCCGGCAA
GGAGAGAATGACCAATGA
```

### ThocK

```
GTGACCAGGACAAATACCGGCTATCGGTACCGGGCCTTCGGGCTCCGGATTGACAGCGACATCCCGCTGC
CCGAACTCGGAGACGGCCACCCGGCCTGACGGCGACGCCGATCTGACCGTCGTCCGCTGCGGCGAGGCGGA
GCCGGAATGGGCGGAAGGCGGAGGCGGCGGCAGACTGTATGCGGCGGAAGGCATCGTCTCGTTCCGGGTT
CCGCAGACGGCCGCTTTCCGGATCACGAACGGCAACCGGATCGAAGTGCATGCTTACAGCGGAGCAGACG
AGGACCGGATCCGCCTGTACGTGCTCGGCACGTGCATGGGCGCGCTCCTGCTGCAGCGGCGCATCCTGCC
GCTTCACGGCAGCGTCGTGGCGCGCGACGGGCGCGCATACGCCATCGTCGGCGAATCGGGCGCGGGCAAG
TCGACGATGTCCGCGGCGCTGCTGGAACGCGGATTCCGTCTGGTGACGGACGACGTCGCCGCCATCGTCT
TCGATGAGCGCGGCACGCCGCTGGTCATGCCGGCCTACCCGCAGCAGAAGCTCTGGCAGGACAGCCTCGA
CCGGCTCCAGATTGCCGGCAGCGGGCTGAGGCCGCTCTTCGAGAGGGAGACGAAATACGCGGTCCCGGCG
GACGGCGCGTTCTGGCCGGAACCGGTGCCGCTTGTACATATCTATGAGCTCGTCCATTCCGACGGGCAGA
CGCCGGAGCTTCAGCCGATTGCCAAGCTGGAGCGGTGCTATACCTTGTATCGCCACACGTTCCGGCGGTC
GCTGATCGTGCCGTCCGGGCTTTCCGCCTGGCATTTCGAGACGGCGGTGAAGCTGGCGGAGAAGACCGGC
ATGTATCGGCTGATGCGGCCGGCGAAGGTGTTCGCGGCGCGGGAATCCGCGCGTCTCATAGAGACGCATG
CCGATGGGGAGGTGAGCCGATGA
```

## Sequences of the proteins studied

In the following, the amino acid sequences of the proteins recombinantly produced in this work are shown.

### PadeK

MTERAAVRTDHYKAFGFRIESDFVLPELPPAGEREPLDNITVRRTDLQPLWNSSIHFYGN

FAILDHGRTVMFRVPGAAIYAVQDASSILVSPFDQAEENWVRLFILGTCMGIILLQRKIM

PLHGSAVAIDGKAYAIIGESGAGKSTLALHLVSEGYPLLSDDVIPVVMTQGSPWVVPSYP

QQKLWLDTLKHMGMDNANYTPLYERNTKFAVPVGSNFHDEPLPLASIFELVPWDAATHIA

i

```
PIQGMERFRVLFHHTYRNFLVQPLGLMEWHFKTLSSFVHQIGMYRLHRPMVGFSTLDLTS

HILNITRQGENDQ
```

**ThcoK**

```
VTRTNTGYRYRAFGLRIDSDIPLPELGDGTRPDGDADLTVVRCGEAEPEWAEGGGGGRLY

AAEGIVSFRVPQTAAFRITNGNRIEVHAYSGADEDRIRLYVLGTCMGALLLQRRILPLHG

SVVARDGRAYAIVGESGAGKSTMSAALLERGFRLVTDDVAAIVFDERGTPLVMPAYPQQK

LWQDSLDRLQIAGSGLRPLFERETKYAVPADGAFWPEPVPLVHIYELVHSDGQTPELQPI

AKLERCYTLYRHTFRRSLIVPSGLSAWHFETAVKLAEKTGMYRLMRPAKVFAARESARLI

ETHADGEVSR
```

**Table 7.1.** $^1$H chemical shifts of PadeA(13-23) in $H_2O/D_2O$ (9:1) at 298 K.

| amino acid | NH | H | H | others |
|---|---|---|---|---|
| Pro13 | / | 4.968 | | |
| Asp14 | 9.046 | 5.164 | 3.113 2.902 | |
| Pro15 | / | 4.655 | | |
| Asp16 | 8.588 | 4.814 | 3.102 3.003 | |
| Glu17 | 8.318 | 4.515 | 2.173 2.141 | $CH_2$: 2.622, 2.319 |
| Asp18 | 8.521 | 4.851 | 3.083 2.952 | |
| Val19 | 8.085 | 4.162 | 2.112 | $CH_3$: 0.985; 0.901 |
| His20 | 8.614 | 4.852 | 3.368 3.263 | $H_2$: 8.758 $H_4$: 7.399 |
| Tyr21 | 8.436 | 4.741 | 3.233 3.082 | $H_{2,6}$: 7.292; $H_{3,5}$: 6.981 |
| Asp22 | 8.688 | 4.902 | 3.085 2.933 | |
| Ser23 | 8.191 | 4.584 | 4.107 4.037 | |

**Table 7.2 .** $^1$H chemical shifts of chemically synthesized PadeA(13-23)-Ser-OPO$_3^{2-}$ in $H_2O/D_2O$ (9:1) at 298 K.*

| amino acid | NH | H | H | others |
|---|---|---|---|---|
| Pro13 | / | 4.968 | | |
| Asp14 | 9.062 | 5.180 | 3.140 2.935 | |
| Pro15 | / | 4.592 | 2.453 | |
| Asp16 | 8.566 | 4.829 | 3.105 3.034 | |

| | | | | |
|---|---|---|---|---|
| Glu17 | 8.322 | 4.528 | 2.169 | CH$_2$: 2.622, 2.318 |
| | | | 2.134 | |
| Asp18 | 8.534 | 4.868 | 3.094 | |
| | | | 2.988 | |
| Val19 | 8.093 | 4.173 | 2.106 | CH$_3$: 0.985; 0.901 |
| His20 | 8.621 | 4.873 | 3.382 | H$_2$: 8.756 |
| | | | 3.275 | H$_4$: 7.406 |
| Tyr21 | 8.478 | 4.738 | 3.216 | H$_{2,6}$: 7.291; H$_{3,5}$: 6.983 |
| | | | 3.092 | |
| Asp22 | 8.657 | 4.909 | 3.094 | |
| | | | 2.939 | |
| Ser23 | 8.424 | 4.760 | 4.439 | |
| | | | 4.315 | |

*chemically synthesized PadeA(13-23)-Ser-OPO$_3^{2-}$; $\delta(^{31}P)$ = 0.8 ppm; long-range $^1$H-$^{31}$P correlation (HMBC): cross peaks between H$\beta$(Ser23) at 4.439 and 4.315 ppm and R-OPO$_3^{2-}$ at 0.8 ppm were observed.

**Table 7.3.** $^1$H chemical shifts of enzymatically generated PadeA(13-23)-Ser-OPO$_3^{2-}$ in H$_2$O/D$_2$O (9:1) at 298 K.*

| amino acid | NH | H | H | others |
|---|---|---|---|---|
| Pro13 | / | 4.968 | | |
| Asp14 | 9.051 | 5.166 | 3.119 | |
| | | | 2.915 | |
| Pro15 | / | 4.592 | 2.453 | |
| Asp16 | 8.575 | 4.819 | 3.098 | |
| | | | 3.015 | |
| Glu17 | 8.315 | 4.519 | 2.170 | CH$_2$: 2.625, 2.320 |
| | | | 2.137 | |
| Asp18 | 8.529 | 4.857 | 3.080 | |
| | | | 2.964 | |
| Val19 | 8.091 | 4.168 | 2.104 | CH$_3$: 0.985; 0.899 |
| His20 | 8.619 | 4.867 | 3.376 | H$_2$: 8.757 |
| | | | 3.272 | H$_4$: 7.406 |
| Tyr21 | 8.480 | 4.729 | 3.223 | H$_{2,6}$: 7.294; H$_{3,5}$: 6.985 |
| | | | 3.082 | |
| Asp22 | 8.659 | 4.900 | 3.083 | |
| | | | 2.935 | |
| Ser23 | 8.383 | 4.719 | 4.417 | |
| | | | 4.303 | |

*enzymatically generated PadeA(13-23)-Ser-OPO$_3^{2-}$; $\delta(^{31}P)$ = 0.7 ppm; long-range $^1$H-$^{31}$P correlation (HMBC): cross peaks between H$\beta$(Ser23) at 4.417 and 4.303 ppm and R-OPO$_3^{2-}$ at 0.7 ppm were observed.
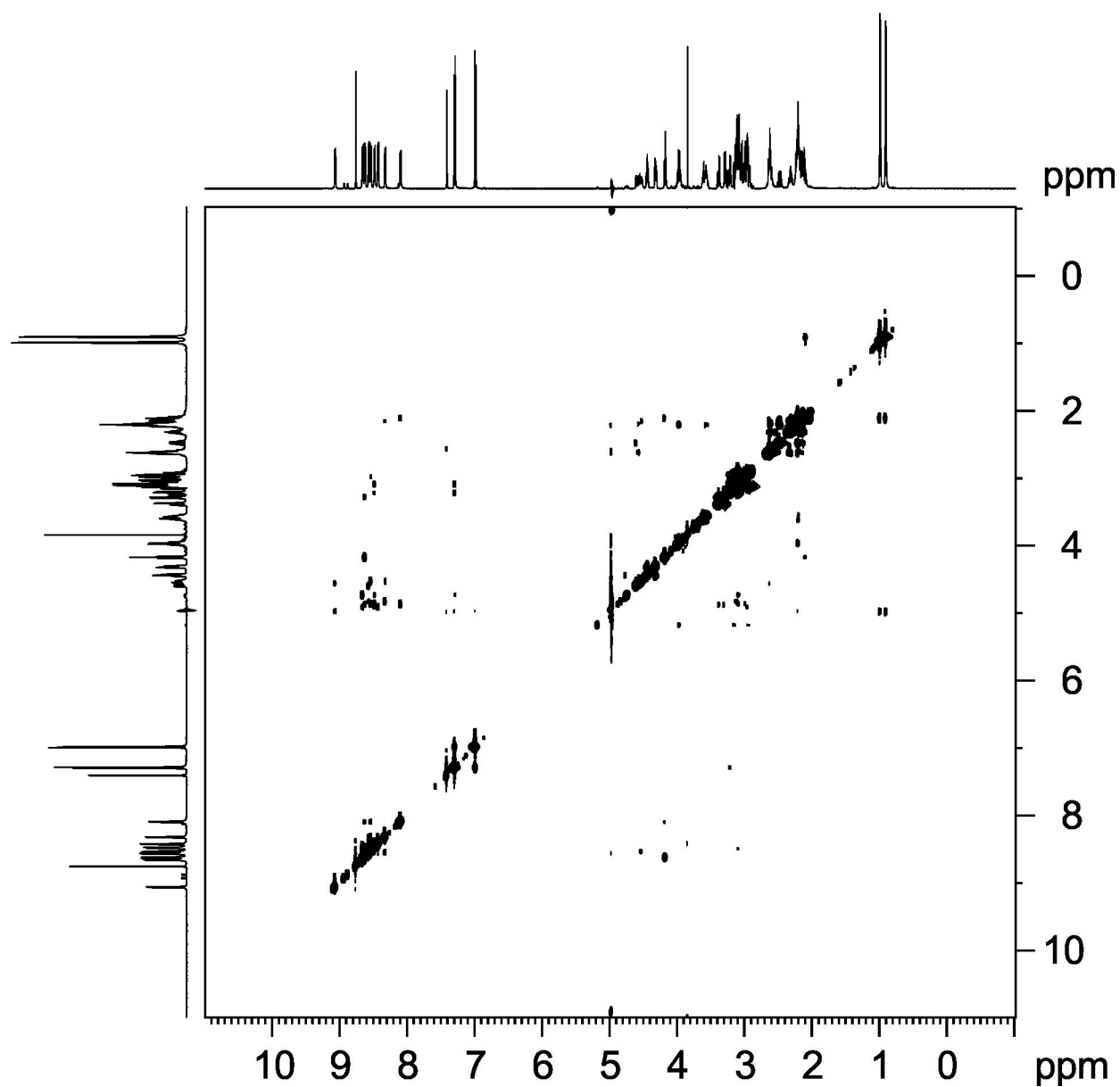
**Figure 7.1** NOESY spectrum of enzymatically produced PadeA(13-23)-Ser-OPO$_3^{2-}$ in H$_2$O/D$_2$O (9:1). The spectrumwas recorded on a Bruker Avance 600 Mhz spectrometer at 298 K.
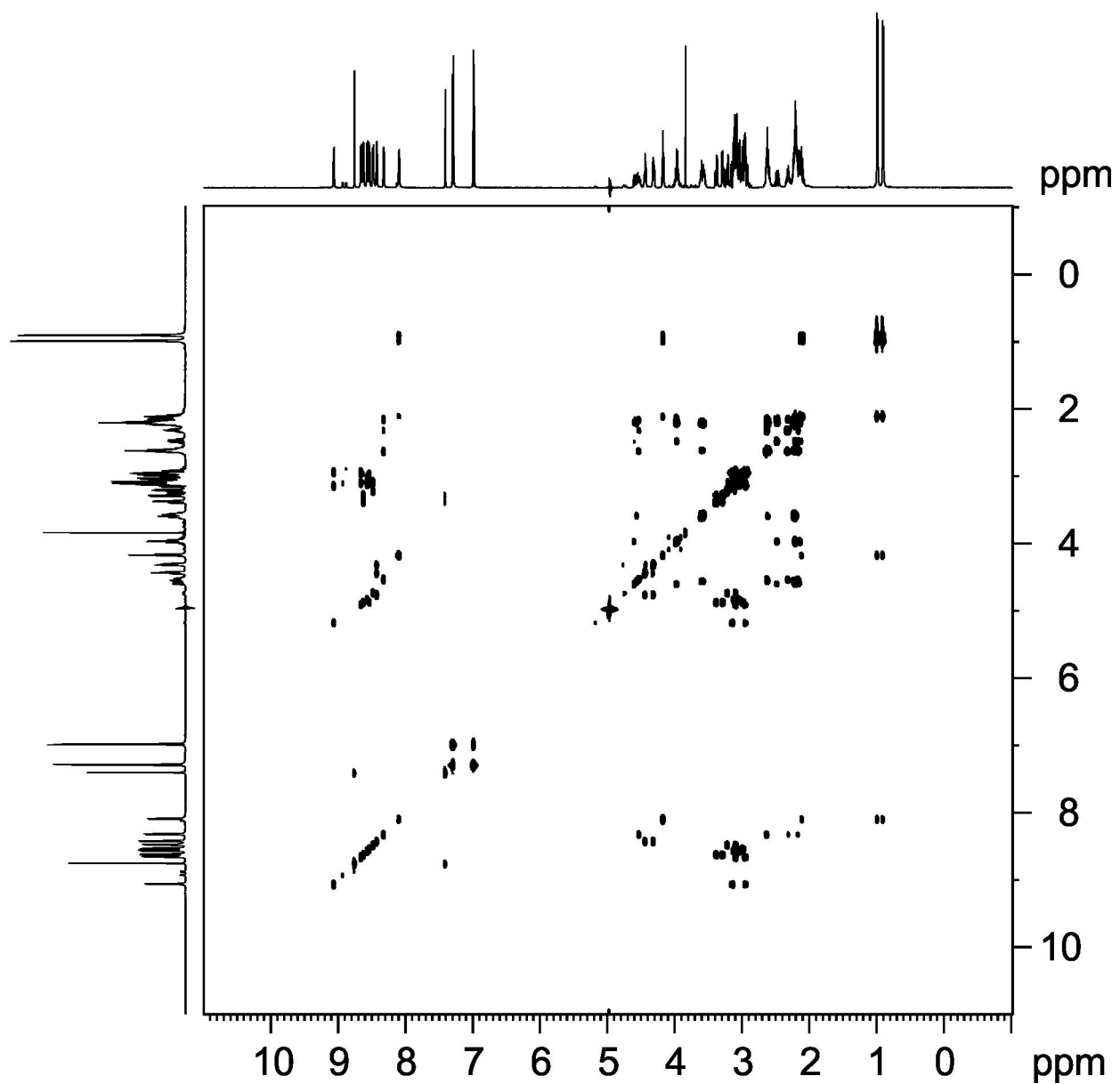
**Figure 7.2** TOCSY spectrum of enzymatically produced PadeA(13-23)-Ser-OPO$_3^{2-}$ in H$_2$O/D$_2$O (9:1). The spectrum was recorded on a Bruker Avance 600 Mhz spectrometer at 298 K.

# Acknowledgements

First and foremost, I would like to thank Prof. Dr. M. A. Marahiel for letting me do the coolest lasso peptide project in his research group and for providing great scientific guidance and constant support during my Ph.D thesis. To be honest, without his supports I would have already quit my studies after my first year in Marburg. Even though sometime he is critical of me, I know that he is always on my side and wants me to success. I appreciate all he has done for me and sorry for the trouble I made during the past 3 years. I am also grateful for the opportunity he gave me to attend the "lasso peptide meeting" every year. These meetings have truly been outstanding experiences that expandedmy scientific horizons.

I also want to take this chance to thank my former supervisor Prof. Dr. Guojun Zheng. Most of my basic skills for doing microbiology experiment were trained in his lab. Besides, he keeps supporting me during my Ph.D thesis. I am very lucky to have these two supervisors during my academic life and will appreciate for my whole life.

I would like to extend my thanks to Prof. Dr. Peter Graumann, Prof. Dr. Armin Geyer and Prof. Dr. Wolfgang Buckel for being a part of my thesis committee and thanks for their interest in my work and their time.

I also would like to thank Dr. Uwe Linne for his excellent support with mass spectroscopy and for his expert advices in LC-MS.

A special thank you goes to everyone in the Lasso Gang: Dr. Julian Hegemann, Dr. Christopher Fage and Dr. Marcel Zimmermann. Without their help, I can not finish the project. Thanks for the fruitful discussions and the friendly research atmosphere they made for me.

I also want to thank all the technical assistants in our lab：Antje Schäfer, Christiane Bomm and Gabriele Schimpff-Weiland. They are gratefully acknowledged for the excellent technical assistance through the years.

I thank all the current and former colleagues from AK66. I miss the great time we worked together.

In Marburg, I also learnt that a good teacher could significantly change a student's life and could become a dreamer supporter like Prof. Marahiel and Prof. Zheng. I am lucky that most of the teachers I met from primary school until now are nice. I want to thank these nice teachers here and wish them the best.

I also would like to thank the Chinese government for providing scholarship to me.

Finally, I thank all my friends, my family and my wife for their long time sacrifice and supporting. With your supporting, I will keep pursuing my dreams

Die vorliegende Dissertation wurde in der Zeit vom September 2012 bis März2016 am Fachbereich Chemie der Philipps-Universität Marburg unter Leitung von Herrn Prof. Dr. Mohamed A. Marahiel angefertigt.

## Erklärung

nach § 10, Abs. 1 der Promotionsordnung der Mathematisch-Naturwissenschaftlichen Fachbereiche und des Medizinischen Fachbereichs für seine mathematisch-naturwissenschaftlichen Fächer der Philipps-Universität Marburg.

Ich erkläre, dass eine Promotion noch an keiner anderen Hochschule als der Philipps-Universität Marburg, Fachbereich Chemie, versucht wurde.

Ich versichere, dass ich meine vorgelegte Dissertation

**Discovery and Insights into the Unique Tailoring of the Paeninodin Lasso Peptide from *Paenibacillus dendritiformis* C454**

selbst und ohne fremde Hilfe verfasst, nicht andere als die in ihr angegebenen Quellen oder Hilfsmittel benutzt, alle vollständig oder sinngemäß übernommenen Zitate als solche gekennzeichnet sowie die Dissertation in der vorliegenden oder einer ähnlichen Form noch bei keiner anderen in- oder ausländischen Hochschule anlässlich eines Promotionsgesuchs oder zu anderen Prüfungszwecken eingereicht habe.

Gehrden, den:

_____

Shaozhou Zhu