



Skolkovo Institute of Science and Technology

Skolkovo Institute of Science and Technology

PHAZOLICIN — A NOVEL AZOLE-MODIFIED PEPTIDE ANTIBIOTIC:
STRUCTURE, MECHANISMS OF ACTION, TRANSPORT, AND BIOSYNTHESIS

Doctoral Thesis

by

DMITRII TRAVIN

DOCTORAL PROGRAM IN LIFE SCIENCES

Supervisor

Professor Konstantin Severinov

Moscow - 2022

© Dmitrii Travin 2022

I hereby declare that the work presented in this thesis was carried out by myself at Skolkovo Institute of Science and Technology, Moscow, except where due acknowledgement is made, and has not been submitted for any other degree.

Candidate (Dmitrii Travin)

Supervisor (Prof. Konstantin Severinov)

Abstract

People widely use antibiotics in healthcare, agriculture, and for preservation of food for almost a century. However, the spread of antimicrobial resistance across the populations of clinically and industrially important microorganisms reduces the arsenal of compounds we can rely on in the fight against undesirable bacteria. Furthermore, the rates of new compound discovery using classical screening approaches dropped dramatically in the past decades primarily due to the high levels of rediscovery of known molecules. The search for novel antibiotics starting from the analysis of accumulated genomic data (“genome mining”) is a viable alternative, which makes use of the improvements in DNA sequencing technologies.

In this work, using genome mining, we identified a biosynthetic gene cluster (BGC) of a putative new antibiotic in the genome of a symbiotic nitrogen-fixing bacterium *Rhizobium* sp. Pop5. Further, we purified and characterized the compound, which biosynthesis is guided by the BGC. The molecule, which we named phazolicin (PHZ) belongs to the growing class of ribosomally synthesized posttranslationally-modified peptide (RiPP) natural products. PHZ is a linear azol(in)e-containing peptide (LAP) exhibiting narrow-spectrum activity against the strains of rhizobia closely related to its producer. PHZ inhibits bacterial translation by the obstruction of the ribosomal nascent peptide exit tunnel and demonstrates the mode of interaction with the ribosome, which is distinct from that of previously described translation-targeting antibiotics. PHZ uses two different non-specific peptide import systems to enter susceptible cells. Such dual mode of

uptake dramatically decreases the levels of spontaneous PHZ resistance acquisition. PHZ production solely defines the ability of *Rhizobium* sp. Pop5 to eliminate closely related strains from the co-culture. However, our attempts to investigate the role of PHZ production in soil and upon the nodulation of plants did not reveal any competitive advantage of PHZ-producing strains. Finally, we perform a systematic search for BGCs of novel putative linear azol(in)e-containing peptides in the publicly available bacterial genomes and describe several previously overlooked groups, which are interesting leads for further experimental validation.

The work present is a comprehensive study of a single compound starting with its prediction by the methods of bioinformatics and finishing with the experiments aiming to validate the role of its production in the complex environment mimicking natural conditions. The application of diverse methodologies of biochemistry, genetic engineering, microbiology, and structural biology enabled us to characterize PHZ from both chemical and biological points of view.

Keywords: phazolicin, RiPPs, LAPs, azol(in)e-modified peptides, translation inhibitors, *Rhizobium*, natural products, peptide antibiotics.

Publications

- [1] **D. Y. Travin**, Z. L. Watson, M. Metelev, F. R. Ward, I. A. Osterman, I. M. Khven, N. F. Khabibullina, M. Serebryakova, P. Mergaert, Y. S. Polikanov, J. H. D. Cate and K. Severinov, “Structure of ribosome-bound azole-modified peptide phazolicin rationalizes its species-specific mode of bacterial translation inhibition” *Nat. Commun.*, 10:4563, Oct. 2019
- [2] Q. Nicoud, Q. Barrière, N. Busset, S. Dendene, **D. Travin**, M. Bourge, R. Le Bars, C. Boulogne, M. Lacroël, S. Jenei, A. Kereszt, E. Kondorosi, E. G. Biondi, T. Timchenko, B. Alunni, and P. Mergaert “Sinorhizobium meliloti functions required for resistance to antimicrobial NCR peptides and bacteroid differentiation”, *mBio*, 12:e00895-21, Jul. 2021
- [3] **D. Y. Travin**, D. Bikmetov and K. Severinov, “Translation-Targeting RiPPs and Where to Find Them,” *Front. Genet.*, 11:226, Mar. 2020
- [4] **D. Y. Travin**, K. Severinov, and S. Dubiley, “Natural Trojan horse inhibitors of aminoacyl-tRNA synthetases,” *RSC Chem. Biol.*, vol. 2, no. 2, pp. 468–485, Mar. 2021

Conference presentations

- [1] **D. Y. Travin**, Z. Watson, M. Metelev, F. Ward, I. A. Osterman, I. Khven, M. Serebryakova, Y. Polikanov, J. Cate, and K. Severinov, “Phazolicin – a founding member of a new group of thiazole/oxazole modified peptides inhibiting prokaryotic translation” at *1st International RiPPs Conference* (Granada, Spain, April 2019)
- [2] **D. Y. Travin**, D. Bikmetov, K. Severinov “Genome mining for novel bioactive peptides” at *12th International Multiconference “Bioinformatics of Genome Regulation and Structure/Systems Biology”* (Novosibirsk, Russia, July 2020)

Acknowledgments

This Ph.D. Thesis would never see the light without the work, advice, and support of a large number of great people.

I would like to express my gratitude to my scientific advisor Prof. Konstantin Severinov. Working under your supervision for almost ten years allowed me not only to perform the research in the wonderful lab, but also to learn how to think scientifically, how to raise new questions, and how to call into question the facts, which everyone takes for granted. Thank you for giving me the freedom of choice of the problems to investigate and introducing to the international scientific community.

I am immensely grateful to Dr. Svetlana Dubiley. Without your help, valuable advice, support, and experience in everyday life in the laboratory, I would not be able to fulfill many tasks. Thank you for your attention and care, and for your ability to give hope in the most difficult situations.

I would like to thank all the members of Konstantin Severinov laboratory at Skoltech, who helped me a lot during the years of work there and, especially, Dr. Marina Serebryakova, Dr. Dmitry Ghilarov, Dr. Mikhail Metelev, Dr. Alexey Kulikovskiy, Aleksander Popov, Dmitry Sutormin, Dmitry Bikmetov, Anastasia Rusanova, Karina Karneeva, Konstantin Gilep, Alexey Livenskiy, and Yulia Gordeeva.

I would like to thank Dr. Peter Mergaert (I2BC, Gif-sur-Yvette) for the great opportunity to join the Plant-Bacteria Interactions Lab as a visiting scientist, invaluable help, support, and care. I also thank the members of Peter Mergaert & Denis Faure's Lab

who were involved in the studies of phazolicin conducted in France and helped me a lot with plant competition experiments, fluorescent microscopy, and genetic manipulations with rhizobia. These are Dr. Emanuele Biondi, Dr. Tatiana Timchenko, Dr. Nicolas Busset, Dr. Quentin Barrière, Dr. Quentin Nicoud, Gaëlle Lextrait, Romain Jouan, Sara Dendene, and Justine Del Bianco.

I would like to thank our collaborators involved in the studies of phazolicin in Russia and abroad for their professionalism and dedication:

- Prof. Yury Polikanov, Dr. Nelli Khabibullina (University of Illinois at Chicago),
- Prof. Jamie Cate, Dr. Zoe Watson, Dr. Fred Ward (University of California, Berkeley),
- Dr. Solange Moréra and Dr. Armelle Vigouroux (I2BC),
- Prof. Petr Sergiev and Dr. Ilya Osterman (Skoltech & MSU),
- Dr. Konstantinos Beis and Satomi Inaba-Inoue (Imperial College of London),
- Dr. Victor Zgoda (Institute of Biomedical Chemistry).
- Prof. Maria Esperanza Martinez Romero (National Autonomous University of Mexico)

Finally, I am grateful to my family and close friends Anna Zhurakovskaya, Iaroslav Popov, Margarita Ezhova, and Marina Kalinina for believing in me and supporting me throughout these years.

Table of Contents

Abstract.....	- 3 -
Publications.....	- 5 -
Conference presentations.....	- 5 -
Acknowledgments.....	- 6 -
Table of Contents.....	- 8 -
List of Symbols, Abbreviations.....	- 11 -
List of Figures.....	- 13 -
List of Tables.....	- 15 -
Chapter 1. Introduction.....	- 16 -
Chapter 2. Review of the literature.....	- 20 -
2.1 RiPPs.....	- 20 -
2.1.1 Natural bioactive peptides.....	- 20 -
2.1.2 Common features of RiPP biosynthesis.....	- 23 -
2.1.3 RiPP posttranslational modifications.....	- 26 -
2.1.4 Azol(in)e-modified RiPPs, posttranslational azol(in)e installation.....	- 29 -
2.1.5 Linear azol(in)e-containing peptides.....	- 36 -
2.1.6 Trifolitoxin.....	- 41 -
2.2 Rhizobia: brief characteristics.....	- 44 -
Chapter 3. Thesis Objectives.....	- 50 -
Chapter 4. Materials and methods.....	- 51 -
4.1 Bacterial strains and growth conditions.....	- 51 -
4.2 PHZ production and purification.....	- 51 -
4.3 Mass-spectrometry.....	- 52 -
4.3.1 MALDI-ToF-MS and MS-MS analysis.....	- 52 -
4.3.2 High-resolution MS and ESI-MS-MS analysis.....	- 53 -
4.4 Testing of PHZ activity <i>in vivo</i>	- 54 -
4.5 <i>In vitro</i> translation inhibition assays.....	- 54 -
4.6 Broth microdilution assays and MIC determination.....	- 55 -
4.6.1 Initial screening of strains for the sensitivity to PHZ.....	- 55 -
4.6.2 Testing of transporter-deficient Sm1021 derivatives.....	- 55 -
4.7 Molecular cloning procedures.....	- 56 -
4.8 Sm1021 transposon library construction.....	- 56 -
4.9 Selection of PHZ-resistant mutants.....	- 57 -
4.10 Sm 1021 whole-genome sequencing, transposon insertions identification.....	- 57 -
4.11 Construction of <i>Rhizobium</i> sp. Pop5 derivatives (Ω_{phzD} and Ω_{NI}).....	- 58 -
4.12 Construction of Sm1021 double mutants.....	- 58 -
4.13 Competition experiments.....	- 59 -
4.13.1 Co-cultivation in laboratory medium.....	- 59 -
4.13.2 Co-cultivation in soil.....	- 60 -

4.13.3 Testing the effect of PHZ addition into the soil.....	- 61 -
4.13.4 Nodulation experiments	- 61 -
4.13.5 Analysis of nodules.....	- 62 -
4.14 Colony-forming unit (CFU) assay	- 63 -
4.15 Cryo-electron microscopy (cryo-EM) structure determination.	- 64 -
4.15.1 Ribosome preparation	- 64 -
4.15.2 Cryo-EM sample preparation.....	- 65 -
4.15.3 Data collection	- 65 -
4.15.4 Cryo-EM data analysis.....	- 65 -
4.15.5 Model building.....	- 66 -
4.16 Cloning, expression and purification of YejA Sm	- 67 -
4.17 Crystallization and structure determination of YejA Sm	- 67 -
4.18 Genome mining for novel LAP BGCs	- 68 -
Chapter 5. Phazolicin: biosynthetic gene cluster, structure, and mechanism of action	- 72 -
5.1 Genome-mining guided discovery of putative new LAPs BGCs.....	- 72 -
5.2 Phazolicin purification and characterization	- 75 -
5.3 Phazolicin is a narrow-spectrum antibiotic affecting rhizobia	- 80 -
5.4 Assignment of the <i>phz</i> BGC with its product	- 82 -
5.5 Phazolicin inhibits bacterial translation.....	- 87 -
5.6 Phazolicin targets the nascent peptide exit tunnel of the bacterial ribosome	- 90 -
5.7 Amino acid sequence of the uL4 protein loop determines the specificity of PHZ action	- 96 -
.....	- 96 -
Chapter summary.....	- 101 -
Chapter 6. Phazolicin import into susceptible cells	- 102 -
6.1 Two transporter systems are inactivated in PHZ resistant <i>Sinorhizobium meliloti</i> mutants	- 102 -
6.2 BacA and YejABEF transporters contribute to the uptake of PHZ.....	- 105 -
6.3 BacA and YejABEF homologs are capable of PHZ internalization	- 109 -
6.4 BacA and YejABEF are involved in the internalization of the PHZ-unrelated thiazole-containing antibiotic bleomycin	- 112 -
6.5 Structure of <i>S. meliloti</i> YejA	- 113 -
Discussion.....	- 117 -
Chapter summary.....	- 119 -
Chapter 7. Phazolicin: homologs and the ecological role.....	- 120 -
7.1 <i>phz</i> -like BGC are found across the genomes of proteobacteria.....	- 120 -
7.2 <i>Mesorhizobium loti</i> strain 582 produces PHZ-likeazole-modified peptides	- 126 -
7.3 PHZ-production underlies <i>Rhizobium</i> sp. Pop5 competitiveness in culture	- 129 -
7.4 The effects of PHZ production and addition are not detected in soil	- 130 -
7.5 PHZ production and the competition for root nodulation	- 134 -
Discussion.....	- 143 -
Chapter summary.....	- 146 -
Chapter 8. Systematic search for novel LAP BGCs	- 147 -

8.1 Construction of networks of <i>ycaO</i> -containing BGCs.....	- 147 -
8.2 Analysis of the network of <i>ycaO</i> -containing BGCs	- 149 -
8.2.1 Group 1	- 152 -
8.2.2 Group 2	- 153 -
8.2.3 Group 3 (Lactazolicins)	- 154 -
8.2.4 Group 4 (<i>mcb</i> -like BGCs from pseudomonads)	- 155 -
8.2.5 Group 5 (Flavazolicins)	- 157 -
Chapter summary.....	- 160 -
Chapter 9. Conclusion.....	- 161 -
Bibliography	- 163 -
Appendix.....	- 178 -
Supplementary Tables	- 178 -

List of Symbols, Abbreviations

ABC – ATP-binding cassette
ATP – adenosine triphosphate
BGC – biosynthetic gene cluster
BLM – bleomycin
BNF – biological nitrogen fixation
BSA – bovine serum albumin
CFU – colony-forming unit
Cryo-EM – cryogenic electron microscopy
DAP – diaminopimelic acid
DMSO – dimethyl sulfoxide
DUF – domain of unknown function
ERY – erythromycin
FMN – flavin mononucleotide
FSC – Fourier shell correlation
gDNA – genomic DNA
HCA – heterocycloanthracin
HIV – human immunodeficiency virus
HPLC – high-performance liquid chromatography
IMAC – immobilized metal affinity chromatography
IPTG – isopropyl β -D-1-thiogalactopyranoside
KLB – klebsazolicin
LAP – linear azol(in)e-containing peptide
MALDI – matrix-assisted laser desorption/ionization
McB – microcin B17
MCS – multiple cloning site
MIC – minimal inhibitory concentration
MOA – mechanism of action
NCRs – nodule-specific cysteine-rich peptides
NGS – next generation sequencing
NP – natural product

NPET – nascent polypeptide exit tunnel
NRP – non-ribosomal peptide
NRPS – non-ribosomal peptide synthetase
ORF – open reading frame
PBP – periplasmic solute-binding protein
PDB – protein data bank
PHZ – phazolicin
PQQ – pyrroloquinoline quinone
PRP – pentapeptide repeat protein
PTM – posttranslational modification
PTC – peptidyl transferase center
RFP – red fluorescent protein
RiPP – ribosomally synthesized posttranslationally-modified peptide
RMSD – root-mean-square deviation
RRE – RiPP precursor peptide recognition element
RT – room temperature
SBP – substrate-binding protein
SRP – signal recognition particle
SSN – sequence similarity network
sfGFP – superfolder green fluorescent protein
SLiPT – SbmA-like peptide transporters
TFA – trifluoroacetic acid
TFX – trifolitoxin
ToF-MS – time-of-flight mass spectrometry
TOMM – thiazole/oxazole-modified microcins

List of Figures

- Figure 2.1.1** Two major pathways of synthesis of peptide natural products with nonproteinogenic amino acids.
- Figure 2.1.2** Examples of primary PTMs of RiPPs.
- Figure 2.1.3** Examples of RiPP tailoring modifications.
- Figure 2.1.4** Azol(in)e cycles found in RiPPs.
- Figure 2.1.5** Representatives of azol(in)e-modified RiPPs.
- Figure 2.1.6** Mechanisms of azole formation in NRP and RiPP biosynthetic pathways.
- Figure 2.1.7** LAPs.
- Figure 2.1.8** Trifolitoxin.
- Figure 2.2.1** Establishment of legume-rhizobia symbiosis.
- Figure 2.2.2** Chemical structures of representative rhizobial specialized metabolites.
- Figure 5.1.1** Comparison of the *phz* BGC with previously described *kfp*-like BGCs.
- Figure 5.2.1** Purification of PHZ-related compounds.
- Figure 5.2.2** PHZ MS/MS analysis and structure.
- Figure 5.2.3** High-resolution MS spectra for PHZ, A-PHZ, and TA-PHZ.
- Figure 5.4.1** Construction and phenotype verification of *Rhizobium* sp. Pop5 Ω *phzD*.
- Figure 5.4.2** PHZ production in *E. coli* Rosetta (DE3) heterologous host.
- Figure 5.5.1** PHZ inhibits bacterial translation *in vivo* and *in vitro*.
- Figure 5.5.2** *In vitro* translation inhibition activity of PHZ and its naturally occurring longer forms (TA-PHZ and A-PHZ).
- Figure 5.6.1** Effects chloramphenicol (CHL), klebsazolicin (KLB), and of phazolicin (PHZ) on *in vitro* translation of superfolder green fluorescent protein (sfGFP).
- Figure 5.6.2** Fourier shell correlation curve for the cryo-EM structure of *E. coli* ribosome with PHZ bound.
- Figure 5.6.3** The structure of the bacterial ribosome in complex with PHZ.

Figure 5.7.1 Amino acid sequence of the uL4 ribosomal protein loop determines the species-specific mode of PHZ binding to the ribosome.

Figure 6.1.1 BacA and YejABEF transporters independently contribute to the PHZ internalization into the cells of *Sinorhizobium meliloti*.

Figure 6.2.1 Genetic basis of the PHZ resistance of the isolates derived from Sm1021 Ω yejA.

Figure 6.3.1 Resistance to PHZ can be reverted by the episomal expression of BacA, YejABEF or their orthologs from various Alpha- and Gammaproteobacteria.

Figure 6.5.1 Crystal structure of YejASm solute-binding protein.

Figure 7.1.1 Sampling sites of the environmentally isolated strains with *phz*-like biosynthetic gene clusters in the genomes.

Figure 7.1.2 *phz*-like biosynthetic gene clusters.

Figure 7.2.1 PHZ-like peptides produced by *Mesorhizobium loti* strain 582.

Figure 7.3.1 PHZ production mediates the competitiveness of *Rhizobium* sp. Pop5 in co-culture.

Figure 7.4.1 The effects of PHZ on the growth of the PHZ-sensitive *Rhizobium leguminosarum* bv. *phaseoli* 4292 in soil.

Figure 7.5.1 *Rhizobium* sp. Pop5 forms functional nodules on the roots of *Phaseolus vulgaris*.

Figure 7.5.2 Selection of strains for nodulation competition experiment.

Figure 7.5.3 Plant nodulation competition experiments.

Figure 8.1.1 The pipeline used for the identification of YcaO-containing RiPP BGCs.

Figure 8.2.1 A similarity network of *ycaO*-containing BGCs containing the genes of E1-like partner proteins.

Figure 8.2.2 The two groups of new BGCs of putative LAPs.

Figure 8.2.3 BGCs of lactazolicins and *mcb*-like BGCs from pseudomonads.

Figure 8.2.4 Flavazolicines.

List of Tables

Table 2.1 Characteristics of the classes of azol(in)e-containing RiPPs

Table 2.2 Selected characteristics of linear azol(in)e-containing peptides

Table 4.1 Cultivation media used in the study

Table 4.2 List of protein domains annotated in the BGCs of putative LAPs

Table 5.1 Characteristics of the PHZ-related compounds identified in the *Rhizobium* sp. Pop5 cultivation medium extract

Table 5.2 Phazolicin MIC values for various strains of rhizobia

Table 6.1 Phazolicin and bleomycin MIC values for Sm1021 and its derivatives

Table 7.1 Isolation sources for the strains with *phz*-like BGCs in the genome

Chapter 1. Introduction

For thousands of years, people have been using biologically active compounds produced by diverse organisms inhabiting the Earth. The oldest written evidence of the use of medicinal plants for the preparation of drugs comes from Sumerian clay tablets (3000 BC), Egyptian papyri (1500 BC), and holy books such as the Vedas and the Bible [1]. With the rise of biology and chemistry in the 19th-20th centuries, we started attributing certain biological activities to exact molecules produced by the organism rather than just empirically accumulating knowledge about the beneficial features of selected plants, animals, fungi, *etc.* Natural products (NPs), which, in the broadest sense, include any substance of biological origin found in nature, still remain the primary source of novel compounds with diverse biological activities for the growing needs of humanity.

This is well illustrated by the history of the discovery of new antibiotics – an important group of chemicals, the introduction of which in the everyday life dramatically changed the life expectancy and the mortality structure of the human population. The majority of clinically relevant antibiotics are derivatives of NPs produced by bacteria and filamentous fungi [2]. Many of the compounds currently in use were discovered as a result of screening of collections of cultivable strains obtained from various sources for desired antimicrobial activities against selected pathogens. Such an approach was extremely effective throughout the so-called “Golden Era” of antibiotic discovery in 1950-1960 but started to result in a high percentage of the discovery of already known compounds by the

end of the previous century once all “low-hanging fruits”, i.e., active compounds readily produced in considerable amounts by easily cultivated strains, were identified. This marked the beginning of the current crisis in antibiotic discovery, which is getting even more serious given the ever-increasing rates of the development and spread of antimicrobial resistance in clinically important pathogenic bacteria [3] and the abandoning the field of antibiotic research by many big pharmaceutical companies [4]. Research groups worldwide have implemented multiple alternative approaches to the identification of new previously overlooked compounds in the past couple of decades. These include, for instance, sampling of bacteria from previously understudied and exotic environments [5]–[7], the attempts to identify the metabolites of uncultivable bacteria grown in their natural environment [8], and development of high-throughput systems for the screening of bacterial strains without the need of conventional large-scale fermentation [9], [10].

Another method includes the search for the genes predicted to encode the biosynthesis of putative novel compounds in the genomic data. This approach referred to as “genome mining” benefits from the rapid accumulation of publicly available genomic data, which became possible due to the dramatic decrease in the cost of DNA sequencing in the last 15-20 years [11]. Another key factor enabling the effective search of new compounds in the paradigm “from genes to molecules” is the significant improvement in our understanding of the biochemical processes underlying the biosynthesis of various natural products. Once we know, which proteins (and hence, genes encoding them) are required for one or another step in the assembly of known molecules, we can search in the genomes for new variants and combinations of biosynthetic genes, which likely guide the

biosynthesis of the previously unknown metabolites. The genome mining approach works especially effectively for the genomes of prokaryotes, where the genes encoding the proteins related to a single biosynthetic pathway are most often colocalized and share common regulation. Such groups of functionally linked genes are known as biosynthetic gene clusters (BGCs). BGCs frequently code not only for the enzymes required for the biosynthesis of a certain NP but also for other proteins such as transporters, regulators, immunity proteins, *etc.*, which can often provide additional hints about the mechanism of transport and even the mechanism of action of a putative new compound. A variety of computational tools, which enable rapid automated search for putative BGCs in the newly sequenced DNA, have been developed in the last decades, with antibiotics & Secondary Metabolite Analysis Shell (antiSMASH) [12] being currently the most widely used and most actively updated platform. Specialized databases (e.g. MIBiG [13]) are designed to systematize the data about the BGCs with determined function and natural products, whose biosynthesis they guide.

Once the first full genome of the model actinomycete *Streptomyces coelicolor* A3(2) was sequenced in 2002 [14], it became evident, that the number of BGCs it encodes exceeds by several times the number of known specialized metabolites produced by the strain. This means that the majority of BGCs are dormant under the commonly used conditions of bacteria cultivation, an expected result given the high cost of specialized metabolites biosynthesis. Numerous approaches have been developed for the activation of such silent BGCs of interest in their native hosts or for the heterologous production of the compounds encoded after the transfer of the BGCs genes into a heterologous host by the

means of genetic engineering [15]. Another obvious obstacle on the way from the BGC in the genome to the bioactive molecule discovery is the search for the biological activity of the metabolite, as it can be proposed in advance only based on the close homology with the clusters of known compounds or, in rare cases, on the presence of specific self-immunity determinants [16]. Often the compounds are characterized structurally, but the function they can play in the natural microbial community remains unknown. Thus, the identification of a previously unknown BGC in the genomic sequence of a certain organism is only the first step on a long and thorny path towards the discovery of a new biologically active compound.

In this PhD Thesis, I describe the characterization of a new azole-modified peptide antibiotic phazolicin (PHZ) starting with its BGC identification in the genome of *Rhizobium* sp. Pop5 during genome mining. The work includes the determination of the PHZ structure, detailed mechanism of action, identification of the strains susceptible to the action of PHZ, and the mechanism of its import into the susceptible cells. The Thesis also covers the identification of PHZ homologs produced by other strains of rhizobia, and genome-mining-guided discovery of new putative families of azole-modified peptides across sequenced prokaryotic genomes.

Chapter 2. Review of the literature

2.1 RiPPs

2.1.1 *Natural bioactive peptides*

Living organisms use amino acids to create diverse polymers. On the one hand, ribosomally synthesized proteins display an enormous diversity resulting from the combinatorial assembly of a relatively small set of genetically encoded monomers in long spatially structured polymers. Additional complexity and diversity of proteins is achieved by posttranslational modifications (PTMs), which play key roles in many processes, e.g. signal transduction and protein degradation. On the other hand, nature comprises the immense diversity of amino acid polymers of lower molecular weight – peptides. Known functions of bioactive peptides are diverse, they can act as antibiotics, toxins [17], signaling molecules (e.g. mammalian hormones and growth factors [18], or quorum sensing signals of gram-positive bacteria [19]), antivirals [20], antioxidants (e.g. glutathione), and siderophores [21].

Such a great functional diversity requires a large arsenal of chemical moieties, which can be introduced into the peptide structures. Two fundamentally different strategies were developed by nature to overcome the restrictions imposed by the genetic code and the mechanics of the ribosome, which assembles peptides from a limited set of L- α -amino acids. First, specific enzymes can posttranslationally modify ribosomally synthesized

peptides comprised of the proteinogenic amino acids converting them into mature compounds with new properties (**Fig. 2.1.1A**). This approach is implemented in the biosynthesis of a rapidly growing group of natural products known as Ribosomally synthesized Posttranslationally-modified Peptides (RiPPs) [22], [23]. Second, the biosynthetic machinery other than the ribosome can perform the assembly of amino acid polymers. Examples of nonribosomal peptide (NRP) assembly lines include, for instance, the enzymatic pathways towards glutathione [24] and the peptide cross-links of the peptidoglycan of the bacterial cell wall [25]. However, the vast majority of NRPs are assembled by the huge multidomain proteins known as NRP synthetases (NRPSs, **Fig. 2.1.1B**) [26]. Each NRPS module incorporates one amino acid residue into the growing peptide chain. The specificity of this process is determined by the structure of the active site of the adenylation domain, which is selective for a certain amino acid [27]. Accumulation of structural data on the binding of amino acids by adenylation enzymes and genomic sequences of the NRP BGCs of known NRPs enabled the deciphering of the so-called *nonribosomal code*. Now it allows the bioinformatics prediction of the NRPs encoded by newly discovered BGCs based on the amino acid sequence of the NRPSs they encode [28], [29]. The diversity of building blocks used by NRPSs exceeds that of proteinogenic amino acids by more than an order of magnitude and includes various exotic moieties, such as, for instance, halogenated, cyclopropyl-containing, and β -amino acids [26], [30]. NRPSs are predominantly found in the genomes of bacteria and fungi. They are responsible for the assembly of many clinically relevant natural products such as antibiotics

(e.g., vancomycin) [31], anticancer compounds (e.g., bleomycin) [32], and immunosuppressants (e.g., cyclosporin) [33].

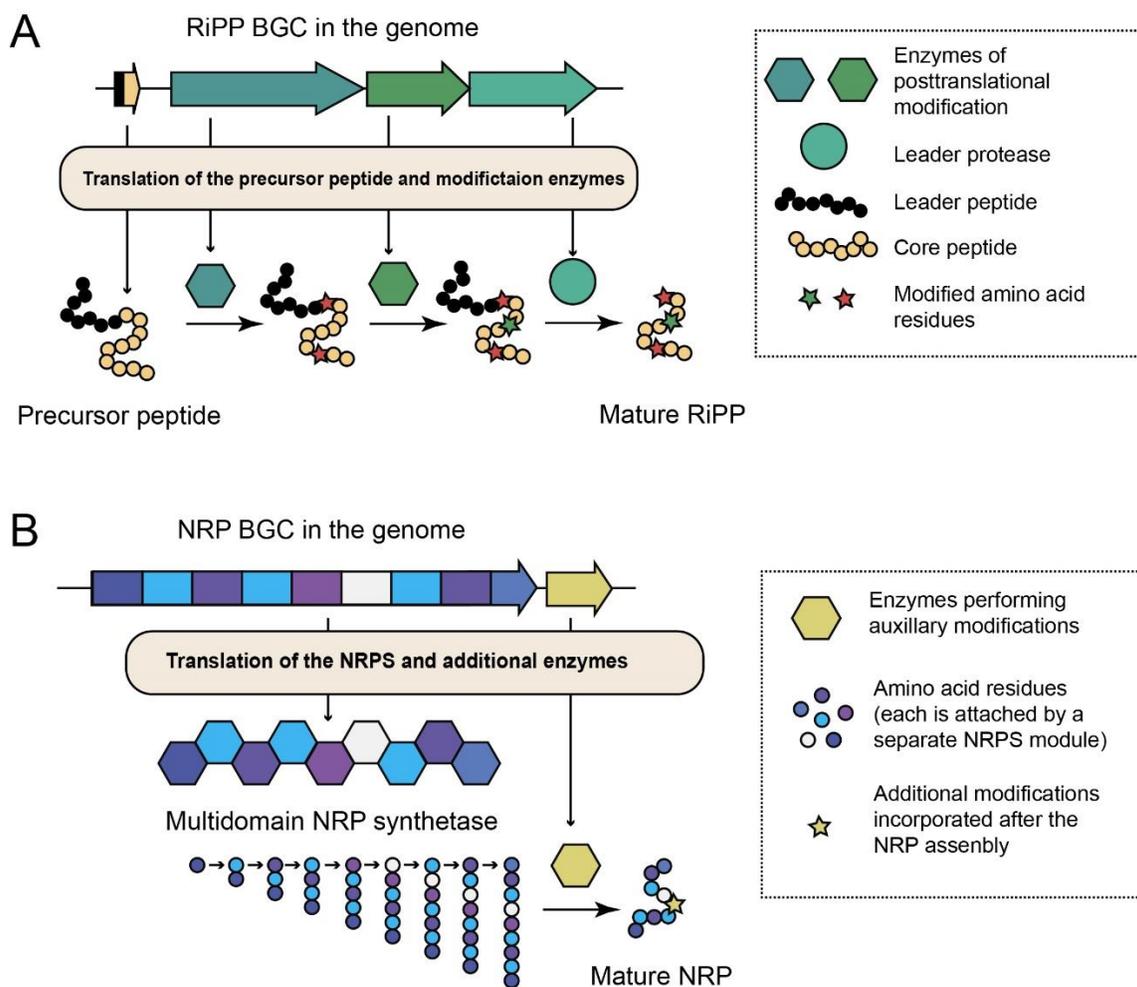


Figure 2.1.1 Two major pathways of synthesis of peptide natural products with nonproteinogenic amino acids. (A) Biosynthesis of RiPPs includes post-translational modification of ribosomally-assembled peptides by specific enzymes. Arrows indicate genes in a biosynthetic gene cluster. (B) The assembly of NRPs is performed by large multidomain NRP synthetases with each residue being incorporated into the growing amino acid chain by a separate specific module.

2.1.2 Common features of RiPP biosynthesis

The abbreviation “RiPPs” was first proposed in a review by Arnison *et al.* in 2013 [23]. This work, key for the emerging field of RiPP research, not only summarized the then available data on the compounds, whose biosynthesis had been already studied by that moment, but also laid the foundation for the currently used nomenclature of genes in RiPP BGCs. In this review, the commonality of the biosynthetic logic was chosen as a key feature, which unites diverse compounds with a wide range of biological activities (see below) produced by various organisms.

A typical RiPP biosynthetic pathway (**Fig. 2.1.1A**) starts with the translation of a peptide, (a *precursor peptide*), encoded as a separate open reading frame (ORF). The precursor is then recognized by specific enzyme(s), which selectively introduce modifications in its structure. In the majority of RiPPs the precursor can be subdivided into two functional parts, an N-terminally located *leader* peptide and a C-terminally located *core*. The leader part is required for the specific recognition of the substrate by the cognate modification machinery, while the core will become the functional mature compound once PTMs are installed. Proteolytic removal of the leader usually follows the modification of the core. In some cases, auxiliary (tailoring) modifications (see below) are installed after that. Enzymes performing such modifications do not normally require the leader part and recognize the modified core instead [34]. A fully modified compound is typically secreted from the producing cell via a dedicated export pump.

As was already mentioned above, in bacteria, the genes coding for the proteins required for a certain biosynthetic pathway are often clustered together in BGCs. The most “complete” RiPP BGC would include the genes of (i) precursor peptide(s), (ii) enzyme(s) involved in the posttranslational modification, (iii) a protease responsible for the leader peptide processing, (iv) an export pump, (v) protein(s) involved in the immunity of the producer to the mature product, if it is toxic, and (iv) regulatory protein(s). In reality, some of the BGC components are often missing. E.g., in many cases, the cleavage of the leader is performed by a housekeeping protease [35], [36], or a specific protease encoded elsewhere in the genome outside of the BGC [37]. If the mature RiPP is not exported, like, for instance, pyrroloquinoline quinone (PQQ), a redox cofactor for certain proteins within the producing cell, an exporter gene is absent from its BGC [38]. There are even examples of BGCs with a precursor-encoding gene located distantly from the cognate maturation proteins operon [39], however, such cases are exceptional. The co-clustering of a short precursor peptide gene with the genes of the cognate biosynthetic machinery is often used in the genomic searches for novel RiPP BGCs, in which short ORFs found close to or between the genes of putative modification enzymes are annotated. Recently this process was implemented in a new mining tool called RiPPER, which performs automated identification and ranking of putative RiPP precursor ORFs in the genomic data [40].

Since each translated precursor peptide molecule results, after the modification, in a single mature RiPP molecule, in several RiPP families, specific features in the architecture of the corresponding BGCs allow for the enhanced production of the

precursors compared to other genes. In some cases, the gene of the precursor peptide, being the first in the operon, is followed by the transcription terminator hairpin, which can be occasionally bypassed by the RNA-polymerase [41]. This leads to significantly higher levels of precursor gene transcripts compared to those of the modification proteins, transporters, *etc.*, located downstream. In some RiPP BGCs, there are multiple copies of the precursor gene, which can code for identical or slightly different amino acid sequences of the core peptides [42], [43] and a cocktail of related mature molecules is produced. Finally, in several RiPP groups, there are genes coding for so-called *cassette* (multicore) precursors, which contain a single leader followed by multiple sequential core sequences (not necessarily identical) separated by recognition sites of dedicated proteases [40], [44]–[47]. A protease performing the cleavage at the C-terminus of the core is required for the maturation in this case in addition to the commonly found leader protease cutting at the N-terminus of the core. All these strategies enable higher production of the molecules derived from the precursor's core.

Since RiPP precursors are genetically encoded, mutations in their genes can lead to amino acid substitutions in mature compounds. Darwinian selection will support the sequence alternations, leading to the biosynthesis of the compounds with enhanced bioactivity, specificity, or stability, which provide a competitive advantage to the producer. The situation is much more complex for NRPs, since multiple simultaneous mutations in the active site residues are required to change the amino acid specificity of a given NRPS module. Genetically encoded precursors also make RiPPs biosynthetic pathways an

exceptionally attractive system for genetic engineering, since the precursor sequences can be easily altered via site-directed mutagenesis in the corresponding genes. Furthermore, precursor gene libraries encompassing thousands of variants can be created for subsequent screening in search of modified peptides with desired properties. For instance, such an approach making use of a library encoding 10^6 lanthipeptides recently led to the selection of a compound effectively preventing the interaction of the HIV p6 protein with its partner in the human cell [48].

2.1.3 RiPP posttranslational modifications

In the previous section, we discussed general features shared by RiPP biosynthetic pathways. Here we briefly review the PTMs found across the families of RiPPs and their contribution to the unique properties of the mature modified peptides. The currently used classification of RiPPs is based on the chemical nature of the modifications installed into the structure of the precursor. To date, it includes 42 classes produced by members of all three domains of life [22]. The majority of known RiPPs are of bacterial origin, however, there is a growing body of evidence on the production of diverse RiPPs by animals, plants, and fungi [49].

The PTMs found in the structure of RiPPs can be roughly divided into two groups: *class-defining* or *primary* PTMs, which significantly change the scaffold and/or the overall three-dimensional structure of the peptide, and *tailoring* or *auxiliary* modifications, which affect selected amino acid residues in the peptide. Primary modifications are required for

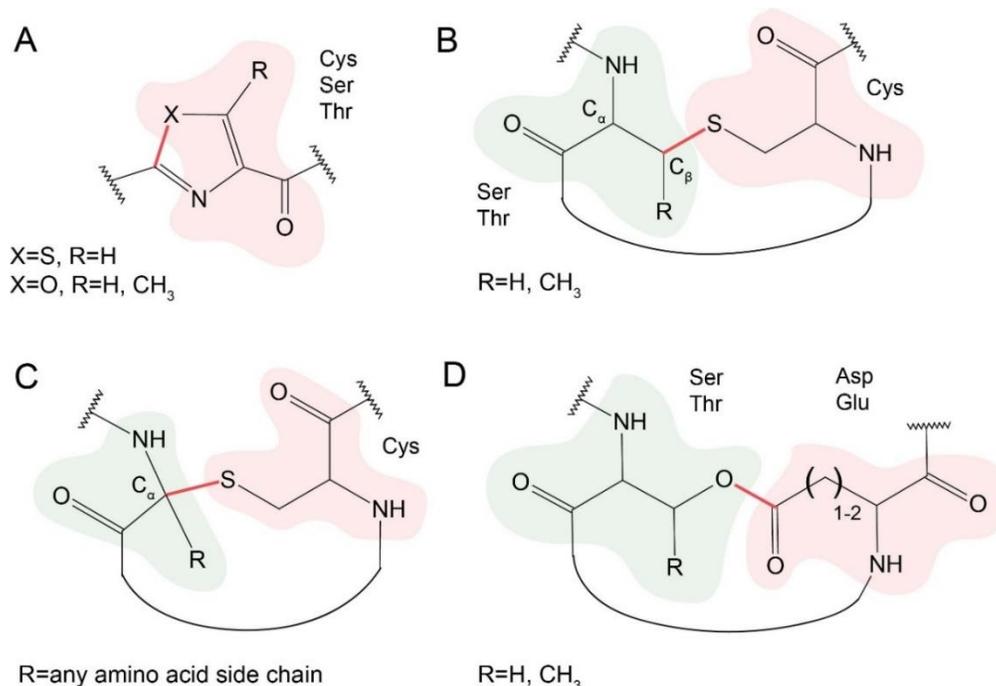


Figure 2.1.2 Examples of primary PTMs of RiPPs. New bonds formed between the amino acid residues are shown in red. The residues involved in the formation of new bonds are shown on the color background and labeled. **(A)** Azole heterocycles found in *thiopeptides*, *linear azol(in)e-modified peptides*, *cyanobactins*, *bottromycins* [50]. **(B)** (Methyl) lanthionine bonds – a class-defining PTM of *lanthipeptides* [51]. **(C)** Sactionine bond typical for *sactipeptides* [52]. **(D)** Macrolactone and macrolactam (not shown) rings formed via the linkages between amino acid side chains are characteristic for *graspptides* [53].

the bioactivity of RiPPs; they shape the molecule making it rigid and often directly mediate the binding with the target. Primary PTMs often include the formation of heterocycles (**Figure 2.1.2A**) or macrocycles (**Figure 2.1.2B**) in the peptide structure. Tailoring PTMs are extremely diverse. In some cases, their installation is a critical step on the way towards a fully-modified compound (see below), while sometimes tailoring PTMs just slightly increase the bioactivity of the resulting RiPP, for instance, via the establishment of additional hydrogen bonds with the molecular target of the modified peptide [34].

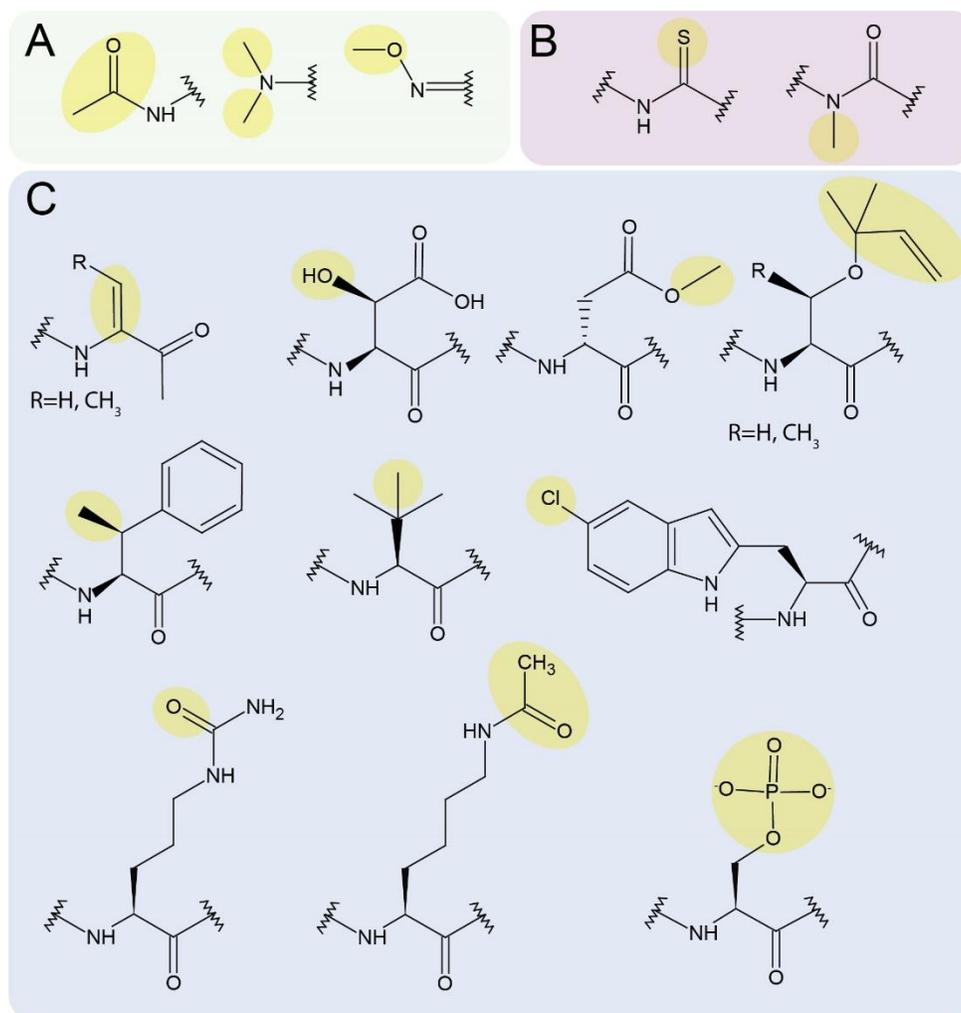


Figure 2.1.3 Examples of RiPP tailoring modifications. Modification installed is shown on the yellow background. Examples of the compounds (compound classes) harboring a given tailoring PTM are provided in brackets. **(A)** Modifications of the peptide N-terminus. From left to right: acetylation (*microviridin J* [54]), demethylation (*plantazolicin* [55]), O-methyl oxime installation (*azolemycin C* [56]) **(B)** Modifications of the peptide backbone: thioamidation (*thioamitides* [57]), N-methylation (*omphalotin* [58]). **(C)** Modifications of amino acid side chains: Ser/Thr dehydration to dehydroalanine and dehydrobutyrine respectively (*lanthipeptides*, *thiopeptides*, *goadsporin*), Asp β -hydroxylation (*cinnamycin* [59]), Asp O-methylation (*bottromycins* [60]), Ser/Thr O-prenylation (*cyanobactins* [61]), Phe β -methylation (*bottromycins* [60]), Val β -methylation (*proteusins* [62], *bottromycins* [60]), Trp halogenation to 5-chloro-Trp (*microbisporicin* [63]), Arg deimination to citrulline (*citrulassin* [64]), Lys acetylation (*albusnodin* [65]), and Ser phosphorylation (*paeninodin* [66]).

2.1.4 Azol(in)e-modified RiPPs, posttranslational azol(in)e installation

Installation of azol(in)e cycles (**Figure 2.1.2A**) is one of the most abundant primary PTMs found in multiple classes of RiPPs. Azoles are a family of aromatic five-membered heterocycles containing a nitrogen atom and at least one other non-carbon atom. Their corresponding reduced analogs are azolines and azolidines. Azol(in)es found across RiPPs include thiazole(in)es containing an atom of sulfur as a second heteroatom in the five-membered ring and (methyl)oxazole(in)es with an atom of oxygen. Azoline cycles are installed into the precursor peptide via the cyclization of the side chains of Cys, Ser, and Thr residues, which results in the formation of thiazolines, oxazolines, and methylthiazolines respectively (**Fig. 2.1.4A and B**). Subsequent azoline oxidation leads to aromatic azoles. Although azoles are more chemically stable compared to azolines, which can undergo spontaneous ring opening under acidic conditions, in certain groups of compounds (e.g. cyanobactins [67] and trifolitoxin [68]) the mature molecules contain azolines.

In rare cases, azolines can be installed on the basis of a previous PTM, which provides a nucleophilic group required for the heterocyclization reaction. Such two-step modification was found in the biosynthesis of polyazole cyclic peptides (e.g., YM-216391): phenyloxazoline is formed out of the β -hydroxyphenylalanine residue synthesized during a previous tailoring PTM (**Fig. 2.1.4C**) [69]. In several rare cases, tailoring modifications can target already installed azole cycles. For instance, in the

biosynthesis of thiopeptide GE2270 a thiazole cycle is additionally decorated by a methoxymethyl group by the sequential action of several enzymes (**Fig. 2.1.4D**) [70].

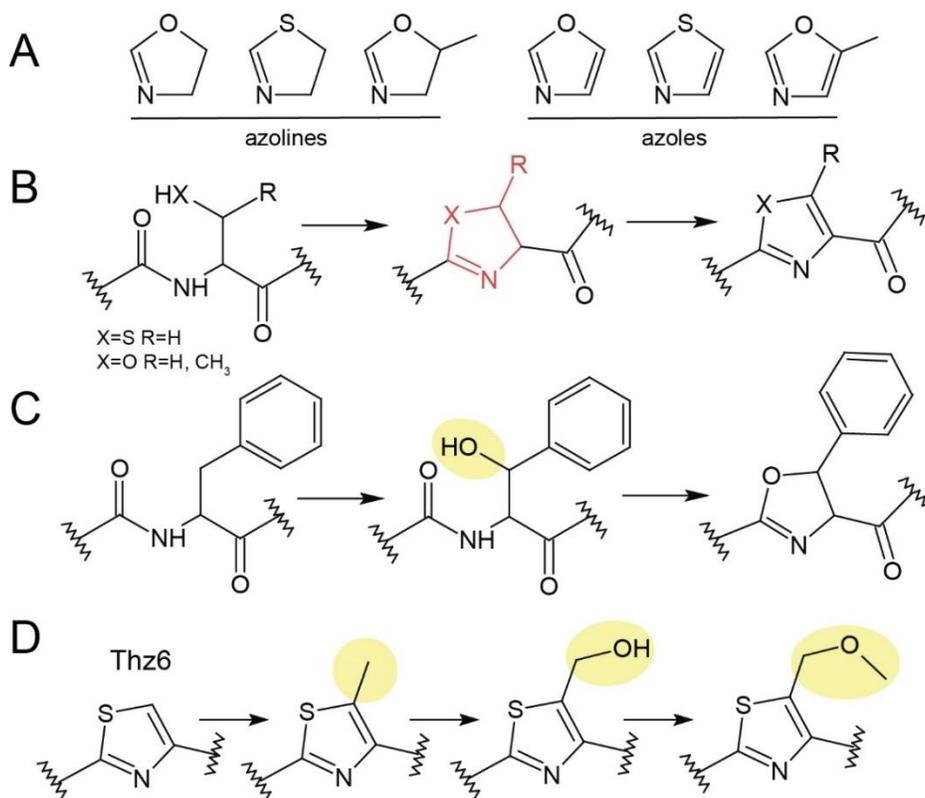


Figure 2.1.4 Azol(in)e cycles found in RiPPs. (A) Azoles and azolines resulting from the cyclization of the proteinogenic amino acids. (B) Azole installation is a two-step process involving the reactions of cyclodehydration (leading to the azoline cycle shown in red) and oxidation. (C) Phenylazoline appears in the course of YM-216391 biosynthesis after the heterocyclization of the hydroxylated Phe residue [69]. (D) Tailoring RiPP PTMs can target azole cycles. The steps leading towards the methoxymethylthiazole in the thiopeptide GE2270 are shown [70].

Azol(in)e cycles are found in several classes of RiPPs [22]. Previously they were united under an umbrella term thiazole/oxazole-modified microcins (TOMMs) [71], however, now this term is not used widely, since representatives of these classes harbor

diverse other core modifications. The data on the PTMs, biological activities, and mechanisms of action of azol(in)e-containing RiPP classes are summarized in **Table 2.1**.

Chemical structures of selected representatives of these classes are shown in **Figure 2.1.5**.

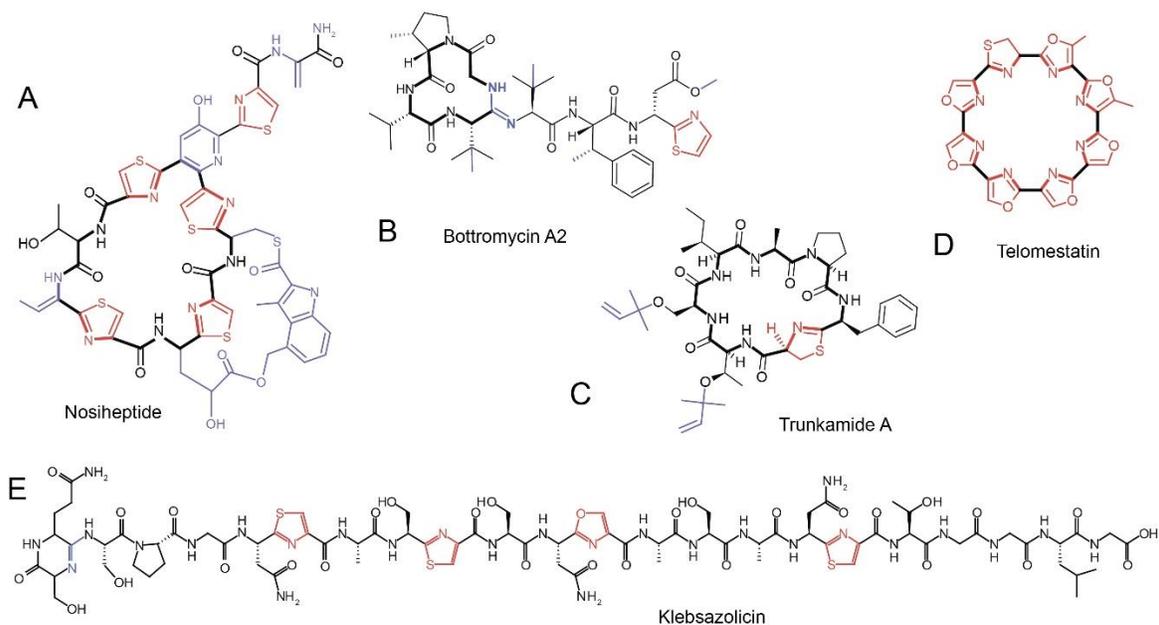


Figure 2.1.5 Representatives of azol(in)e-modified RiPPs. Azol(in)e cycles are shown in red, other modifications are violet, amidine bonds formed by YcaO enzymes in bottromycin and klebsazolicin are blue. The backbones of the peptide macrocycles in A-D are highlighted in bold. **(A)** Thiopeptide nosiheptide [72]. **(B)** Bottromycin A2 [73]. **(C)** Cyanobactin trunkamide [74]. **(D)** Polyazole cyclopeptide telomestatin [75]. **(E)** Linear azol(in)e-modified peptide klebsazolicin [76].

Despite the diversity of chemical structures and additional modifications displayed by known azol(in)e-modified RiPPs, they all share the same mechanism of posttranslational azole installation. In all RiPP biosynthetic pathways studied, this process is catalyzed by *YcaO domain-containing cyclodehydratases*. Here, it is worth noting, that the presence of azole cycles in bioactive peptides is not restricted to RiPPs. There are

Table 2.1 Characteristics of the classes of azol(in)e-containing RiPPs

№	Class	Other core PTMs	Producing organisms	Biological activity, mechanism of action	Refs
1	Thiopeptides	(i) dehydration of amino acids, (ii) macrocyclization via the formation of a six-membered azacycle	Mainly Actinobacteria	Antibacterial and antiplasmodial. Inhibit translation either through the direct interaction with the ribosome GTP-ase associated center (e.g. thiostrepton, nosiheptide, and micrococcin) or prevent the binding of elongation factor EF-Tu with aminoacyl-tRNA (e.g. GE2270A, thiomuracin, and GE37468A).	[77]– [81]
2	Cyanobactins	N-C macrocyclization	Cyanobacteria	Cytotoxic and antitumor (multiple compounds), antibacterial, antiviral, antiplasmodial, inhibitors of proteases, binding of transition metals.	[67], [82], [83]
3	Linear azol(in)e-containing peptides (LAPs)	none	Proteobacteria, Firmicutes, Actinobacteria	Antibacterial (different mechanisms of action (MOAs), see section 2.1.3 for details), cytotoxic and hemolytic.	See Table 2.2
4	Bottromycins	Macrolactamidine cycle formation	<i>Streptomyces</i> spp.	Antibacterial. Inhibit translation via binding with the ribosome A-site. Detailed mechanism is unknown.	[73], [84], [85]
5	Polyazole cyclopeptides	N-C macrocyclization, azoles are catenated	Actinobacteria	Cytotoxic, antitumor. Telomestatin inhibits telomerase via the stabilization of G-quadruplexes.	[75], [86]– [88]
6	Trifolitoxin	5-membered heterocyclic ring formed from a Glu residue	<i>Rhizobium</i> spp.	Antibacterial against rhizobia closely related to the producing strain. Mechanism is unknown. See section 2.1.6 for the details.	[68], [89]

known NRPs containing azol(in)e cycles, e.g., siderophore yersiniabactin [90] or antitumor agent bleomycin [32]. In contrast to RiPPs, in the biosynthetic pathways leading to these molecules, the installation of azoles is performed during the process of peptide chain assembly. Heterocyclization is catalyzed by specific cyclization domains (Cy-domains), which substitute conventional condensation (C-domains) domains in selected NRPS modules (**Fig. 2.1.6A**). Subsequent maturation of the azolines incorporated into an NRP can include either oxidation into azoles by specialized Ox-domains of the same NRPS module or their reduction into azolidines mediated by the activity of a stand-alone NADPH-dependent reductase [26].

The mechanism of azoline installation by YcaO enzymes was studied recently in detail in the group of Prof. D. Mitchell (University of Illinois at Urbana-Champaign, USA). The YcaO enzymes were shown to use a molecule of ATP for direct activation of the peptide backbone [91]. The reaction mechanism is proposed to include a nucleophilic attack of the heteroatom in the side chain of Cys, Ser, or Thr onto the amide carbonyl of the adjacent amino acid residue, which is followed by an ATP-dependent elimination of the carbonyl-derived oxygen (**Fig. 2.1.6B**). Amino acid residues required for the activity of YcaO enzymes include those involved in the binding of Mg^{2+} ions, which in turn coordinate ATP [92], and several residues of the active center, which have been defined based on the recently obtained crystal structures of azoline-forming YcaOs [93].

For RiPP precursor peptide recognition YcaO enzymes usually cooperate with partners belonging to one of the two families: the E1-like proteins or the Ocin/ThiF-like

proteins. These partners harbor conserved N-terminal domains called RiPP precursor peptide recognition elements (RREs) [94], which are responsible for the specific interaction with cognate leader peptides and correct positioning of the cores in the YcaO cyclodehydratase active center. In almost half of known BGCs the genes of YcaO

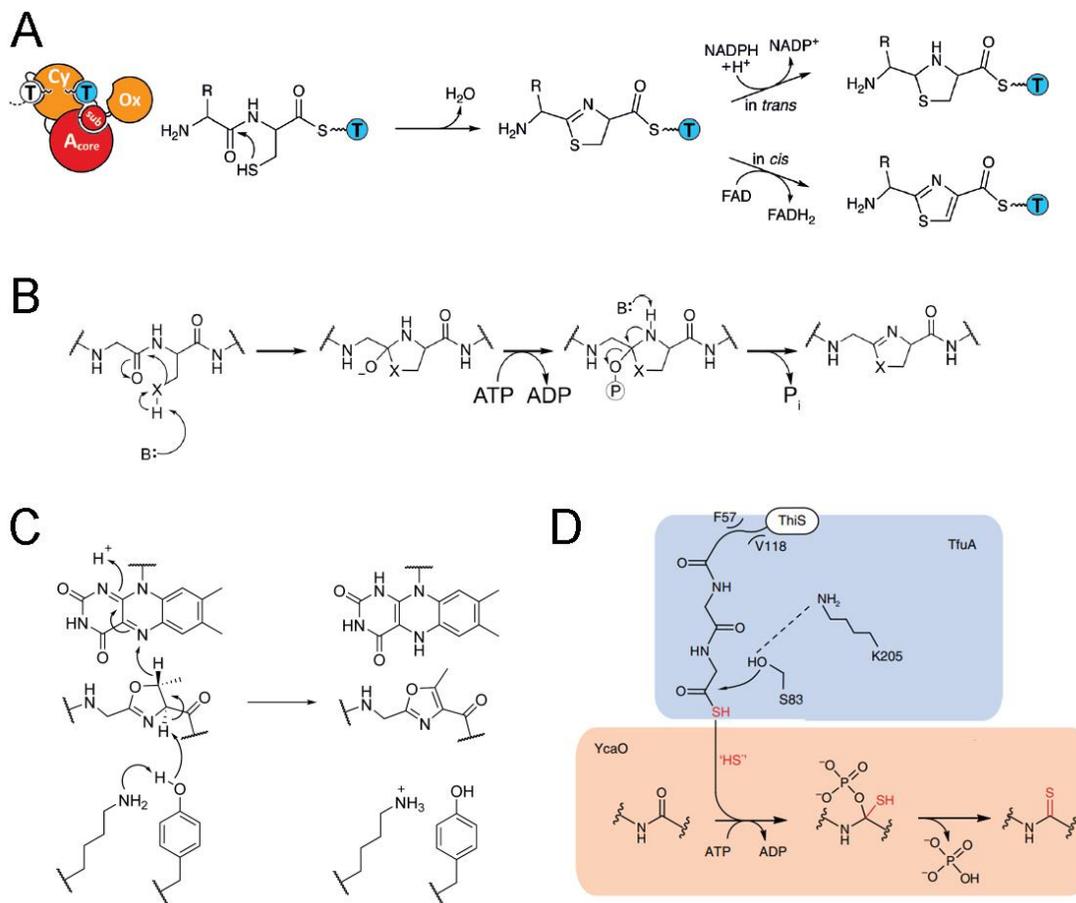


Figure 2.1.6 Mechanisms of azole formation in NRP and RiPP biosynthetic pathways. (A) Installation of azolines into NRPs and their subsequent maturation to azoles and azolidines by specialized NRPS domains. (B) Mechanism of ATP-dependent azoline installation catalyzed by YcaO enzymes. (C) Mechanism of FMN-dependent azoline oxidation. (D) Mechanism of TfuA-assisted thioamide installation in the backbone of the proteins/peptides. ((A) is adapted from [26], (B) and (C) are adapted from [50], (D) is adapted from [95].)

cyclodehydratase and the corresponding E1-like protein are fused and a single protein composed of two functional domains is produced. YcaOs, which do not require the presence of partner proteins (“standalone YcaOs”) are also known, for instance, BmbD, which installs the thiazoline in the biosynthesis of bottromycin [96], however, such cases are an exception rather than a rule.

Oxidation of azolines into azoles is typical for most azol(in)e-modified RiPPs and may be explained by the increased chemical stability of aromatic azoles. Specific FMN-dependent dehydrogenases often encoded next to the genes of YcaO and their partner proteins in RiPP BGCs are responsible for this reaction. These enzymes are dimers, containing a conserved Lys-Tyr amino acid pair in addition to a motif responsible for FMN binding. The proposed mechanism of azoline oxidation is shown in **Figure 2.1.6C**. Azoline dehydrogenases can either recognize the precursor peptide themselves (e.g., ThcOx from the biosynthetic pathway of cyanobactin cyanothecamide [97]) or form a ternary complex with YcaO and its partner. In the latter case, exemplified by the McbBCD complex responsible for the maturation of a LAP microcin B17 [93], the two enzymes responsible for subsequent steps in the installation of azoles (YcaO cyclodehydratase McbD and azoline dehydrogenase McbC) interact with the core part of the peptide bound via its leader to the RRE of the E1-like partner protein McbB.

Recently two other reactions relying on the activity of YcaO domain-containing enzymes were discovered in RiPP biosynthetic pathways. First, the BmbE and PurCD enzymes involved in the maturation of bottromycins were shown to catalyze the formation

of the N-terminal 12-membered macrolactamidine ring characteristic for this class of RiPPs [96], [98] (**Fig. 2.1.5B**). Smaller six-membered N-terminal amidine cycles installed by YcaOs were characterized in the biosynthesis of klebsazolicin [99] (**Fig. 2.1.5E**) and streptamidine [100]. In the reactions of amidine installation, the N-terminal amino group of the peptide is proposed to serve as a nucleophile.

Second, YcaO enzymes install thioamide bonds (**Fig. 2.1.3B**) in several groups of RiPPs (thiopeptins [101], saalfelduracin [79], and thioamitides [57]) as well as in some proteins encoded by archaea (methyl-coenzyme M reductase [102]) and, probably, bacteria (ribosomal protein uL16 [103]). The activity of a partner protein TfuA is required for this reaction in several cases studied: it catalyzes the hydrolysis of thiocarboxylated ThiS, a proteinaceous donor of sulfur (**Fig. 2.1.6D**) [95].

To sum up, azol(in)es can be found in RiPPs belonging to several currently recognized classes. YcaO domain-containing cyclodehydratases are the key enzymes responsible for the formation of azolines. YcaOs are assisted by several partner proteins (E1-like, Ocin/ThiF-like, TfuA), the presence of which may be indicative of the kinds of PTMs the corresponding YcaO is responsible for.

2.1.5 Linear azol(in)e-containing peptides

Linear azol(in)e-containing peptides (LAPs) comprise a group of azol(in)e-modified RiPPs, which do not undergo any core PTMs other than heterocyclizations leading to the formation of azol(in)e cycles in the peptide backbone. Thus, a minimal LAP

BGC comprises only a gene encoding the precursor peptide (gene A) and gene(s) coding for the enzymes involved in the installation of azole cycles. These include a YcaO-cyclodehydratase (the product of the D gene), which in all experimentally studied LAP biosynthetic pathways requires a partner protein for the leader peptide binding (either an E1-like protein, the product of the C gene, or a ThiF-like protein encoded by the F gene), and a dehydrogenase (the product of the B gene), which oxidizes azolines to azoles. Known LAPs display a variety of tailoring PTMs (**Table 2.2**), some of which are unprecedented among the RiPPs, for instance an N-terminal oxime of azolemycin [56].

As the set of chemical characteristics required for the attribution of a compound to LAPs is not particularly restrictive, this group includes peptides, which do not share any obvious sequence similarity. The relationships between the LAPs characterized to date (**Table 2.2**) are best described as a “sea with islands”, where each “island” represents a group of closely related homologs (e.g., streptolysin S with its relatives clostridiolysin S and listeriolysin S, or plantazolicin and coryneazolicin). For now, an understanding of the evolution of LAP BGCs, which resulted in the present-day diversity, is missing, and the relations between these separate groups of LAPs cannot be reconstituted. The phylogenetic tree of azol(in)e-containing RiPPs based on the sequences of YcaO enzymes also supports the heterogeneity of LAPs, since the branches corresponding to their YcaOs are interspersed with well-defined groups of YcaOs from other classes, such as those of thiopeptides, bottromycins, and cyanobactins (**Fig. 2.1.7A**) [104]. Consistently, the biological activities and mechanisms of action of LAPs, when known, are extremely

diverse (**Table 2.2**). For instance, antimicrobial LAPs include the compounds targeting bacterial ribosome (klebsazolicin), DNA gyrase (microcin B17), and, probably, cell membrane (plantazolicin) (see **Table 2.2**). In the case of klebsazolicin (**Figure 2.1.5E**, KLB), for which the exact mode of the interaction with its molecular target is determined structurally, we know that theazole cycles of the peptide form π - π stacking with nucleobases of the 23S ribosomal RNA (**Figure 2.1.7B**) [76]. Their presence is strictly required for KLB activity.

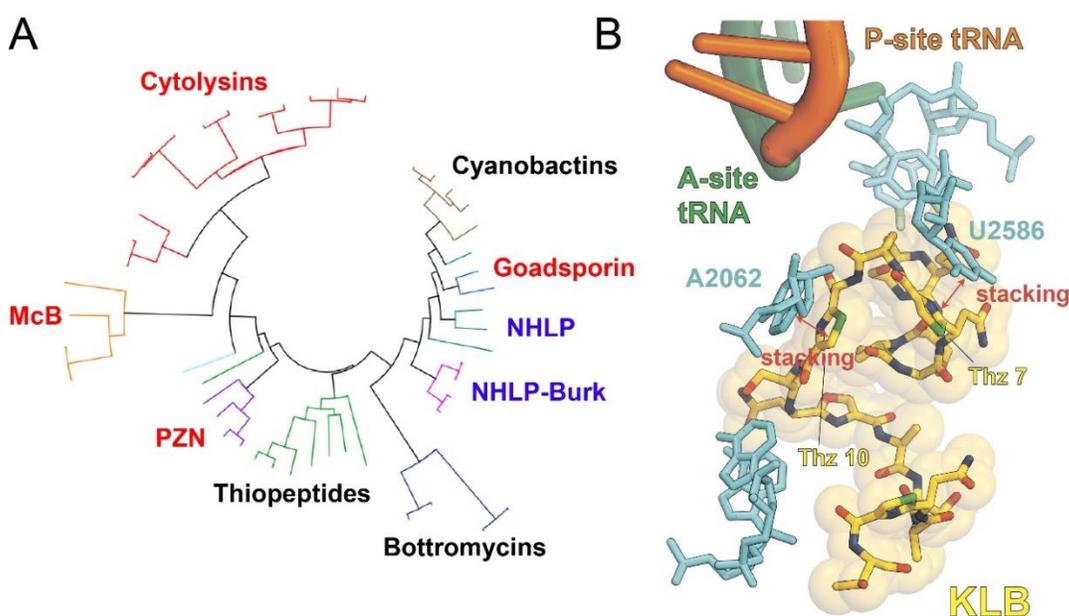


Figure 2.1.7 LAPs. (A) A maximum-likelihood tree of a representative sample of azol(in)e-modified RiPPs based on the YcaO proteins from each cluster. LAPs are labeled red, putative groups of LAPs predicted in [105] are blue, other classes of RiPPs are black. PZN, plantazolicin; McB, microcin B17; NHP, nitrile hydratase leader peptide; NHP-Burk, NHP from *Burkholderia*. Adapted from [104]. (B) KLB bound to the 23S rRNA in the ribosome nascent peptide exit tunnel (PDB ID: 5W4K [76]). Red arrows show the stacking interactions formed by Thz7 and Thz10 of the peptide with nucleobases U2586 and A2062, respectively.

Table 2.2 Selected characteristics of linear azol(in)e-containing peptides

№	LAP	Producing microorganism	Other PTMs	Biological activity, mechanism of action	Refs
1	Microcin B17 (McB17)	<i>Escherichia coli</i> K12	none	Antimicrobial against other <i>E. coli</i> strains; targets DNA-gyrase (type II DNA topoisomerase), leads to SOS-response. Close homologs are produced by several strains of <i>Pseudomonas syringae</i> .	[106]– [111]
2	Klebsazolicin (KLB)	<i>Klebsiella pneumoniae</i> ATCC 11296	N-terminal amidine cycle	Antimicrobial against the strains from genera <i>Escherichia</i> , <i>Yersinia</i> , and <i>Klebsiella</i> . Inhibits translation by the obstruction of the ribosome exit tunnel	[76], [99]
3	Plantazolicin (PZN)	<i>Bacillus amyloliquefaciens</i> FZB42	N-terminal dimethylation	Ultra-narrow antimicrobial activity against <i>Bacillus anthracis</i> , proposed to act through the depolarization of the cell membrane	[112], [113]
4	Coryneazolicin (CZN)	<i>Corynebacterium urealyticum</i> DSM 7109	N-terminal dimethylation	Antimicrobial against Gram-positive bacteria <i>Bacillus subtilis</i> , <i>Staphylococcus aureus</i> , and <i>Micrococcus luteus</i> , cytotoxic, apoptosis-inducing.	[114]
5	Sonorensin	<i>Bacillus sonorensis</i> MT93	N/A (structure of the mature compound remains largely unexplored)	Antimicrobial against <i>Listeria monocytogenes</i> and <i>Vibrio vulnificus</i> , MOA is unknown.	[115], [116]
6	Goadsporin (GS)	<i>Streptomyces</i> sp. TP-A0584	two dehydroalanines	Antimicrobial against some actinobacterial strains, induces morphogenesis and elicits secondary metabolism in other actinobacteria. Proposed to target signal recognition particle (SRP).	[117]– [119]

7	Spongiicolazolicins A and B	<i>Streptomyces sp.</i> CWH03	Dehydroalanines (4 and 3 in A and B respectively)	Unknown	[120]
8	Azolemycins A, B, C, and D	<i>Streptomyces sp.</i> FXJ1.264	N-terminal (methyl) oxime, C-terminal methylation	Unknown	[56]
9	Streptolysin S (SLS)	Group A <i>Streptococcus pyogenes</i> strains (GAS)	N/A (structure is not determined to date)	Cytotoxic and hemolytic virulence factor. Was shown to act through the disruption of band3 – the major anion exchange protein of erythrocytes, which leads to rapid influx of Cl ⁻ ions into the red blood cells and their subsequent lysis (this may be not the only MOA).	[121]– [123]
10	Listeriolysin S (LLS)	<i>Listeria monocytogenes</i> 4b F2365	N/A	Antimicrobial against G ⁺ -bacteria, LLS remains associated with the membrane of the producer and kills sensitive bacteria in a contact-dependent manner inducing their membrane permeabilization.	[124]– [126]
11	Clostridiolysin S (CLS)	<i>Clostridium botulinum</i> , <i>C. sporogenes</i>	N/A	SLS-like hemolytic activity.	[127]

2.1.6 Trifolitoxin

Trifolitoxin (TFX) is the only characterized RiPP produced by rhizobia. The exact chemical structure of TFX was determined only recently [68], although antimicrobial activity of a substance produced by *Rhizobium leguminosarum* bv. *trifolii* T24 was first reported more than 50 years ago [128]. TFX is an 11 amino acid long peptide, which undergoes extensive posttranslational modification: the Cys9 residue is converted into a thiazoline, while the residues Arg6, Gln7, and Gly8 give rise to a blue chromophore moiety that includes a substituted imidazole ring (**Fig. 2.1.8A**) [68]. This set of modifications does not allow the attribution of TFX to any known family of azol(in)e-modified peptides, so it is the only representative of a family of its own (see **Table 2.1**).

The biosynthetic gene cluster responsible for the production of TFX and resistance to the compound was identified through cosmid library construction using the gDNA of the original producer [129]. It comprises the genes of the precursor peptide (*tfxA*), three genes of putative FMN-dependent dehydrogenases (*tfxB*, *tfxC*, *tfxF*), and genes of a YcaO-like enzyme (*tfxE*), a putative phosphotransferase (*tfxG*), and a transporter (*tfxD*) belonging to the multidrug and toxic compound extrusion (MATE) family (**Fig. 2.1.8B**). The exact pathway towards mature TFX remains enigmatic. Based on the compound structure and the set of genes in the BGC, it was proposed that TfxE, acting as a standalone YcaO-like enzyme, converts Cys into the thiazoline, while FMN-dependent oxidoreductases are somehow involved in the chromophore biosynthesis [50], [68]. The biosynthesis of TFX seems to be even more complex, as it was shown that the gene *tfuA* (should not be confused

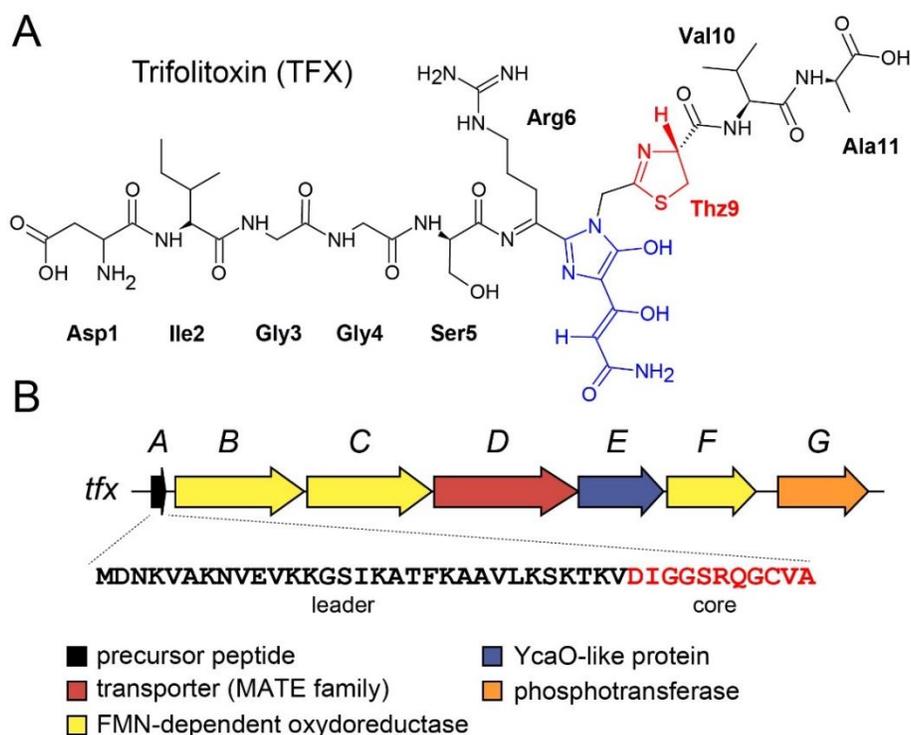


Figure 2.1.8 Trifolitoxin. (A) Chemical structure of TFX. Thiazoline cycle is red, a blue chromophore formed by the residues Arg6, Gln7, and Gly8 is blue. (B) Biosynthetic gene cluster of TFX. Proposed functions of encoded proteins are listed below. The amino acid sequence of the TfxA precursor peptide with highlighted leader (black) and core (red) parts is shown. Thz, thiazoline.

with the YcaO partner protein TfuA described earlier [95]) encoded outside of the *tfx* BGC is required for the biosynthesis of the molecule [130]. The exact function of the TfuA, which is predicted to be a tetratricopeptide repeat motif-containing protein, TFX maturation is unknown.

TFX has potent antimicrobial activity against rhizobia closely related to the TFX producer [89], [128]. In particular the majority of *R. leguminosarum* strains tested appeared to be TFX-sensitive [89]. The molecular mechanism of TFX action as well as the details

of its import into sensitive cells remain unexplored. Remarkably, the substitutions of amino acids other than Arg6, Gln7, Gly8, and Cys9 (those involved in posttranslational modification) do not influence the activity of the compound [131], which may point towards the participation of modified residues in the interaction of the antibiotic with its molecular target or specific importers.

Even for the RiPPs, whose biological activity was determined, little is still known about the roles they play in bacterial ecology and interspecies competition [132]. The TFX-producing strain *R. leguminosarum* bv. *trifolii* T24 is considered “symbiotically ineffective”, i.e. it forms small, white nodules unable to fix atmospheric nitrogen on the roots of red clover *Trifolium dubium* [128]. Interestingly, when mixed inoculums of the T24 strain and a TFX-sensitive strain were used for clover inoculation, T24 was responsible for the formation of the majority of the nodules (88-96% depending on the susceptible strain tested) [128]. Later, the role of TFX production in the increased competitiveness of rhizobia was demonstrated by the conjugative transfer of genes required for TFX production to the symbiotically effective strain *R. leguminosarum* bv. *trifolii* TA1. The resulting TFX-producing strain was very competitive in the presence of the TFX-sensitive partner, while this was not the case for the control conjugant devoid of TFX biosynthetic genes [133]. Subsequently the production of TFX by *R. leguminosarum* bv. *trifolii* T24 was demonstrated in several types of soil [134]. Thus, TFX is a narrow-spectrum antirhizobial RiPP, produced by symbiotically ineffective rhizobia to gain advantage upon the nodulation in the presence of closely related susceptible strains.

2.2 Rhizobia: brief characteristics

Nitrogen-fixing members of the classes Alpha- and Betaproteobacteria capable of forming symbiotic relations with leguminous plants (family Fabaceae) are collectively known as “Rhizobia”. The members of this paraphyletic group are in their free-living state rod-shaped gram-negative motile bacteria, which lead a saprophytic lifestyle consuming a diverse spectrum of carbon sources in soil [135]. Once engaged in symbiosis with plants, rhizobia form *nodules* – specific organs on roots. In nodules, bacteria live intracellularly and differentiate into *bacteroids*, a form, in which they can effectively reduce atmospheric nitrogen in a process of biological nitrogen fixation (BNF). BNF is an energy consuming process (16 molecules of ATP are required per one molecule of N₂ [136]), catalyzed by nitrogenase, an enzymatic complex sensitive to the presence of oxygen. Rhizobia performing BNF in nodules are provided with all required nutrients for bacterial growth, including a source of energy and carbon from the host plant in the form of dicarboxylic acids [137] and protected from the atmospheric oxygen by the plant tissues. In turn, plant host receives ammonium, which is required for the biosynthesis of N-containing macromolecules including proteins and nucleic acids. BNF is a complex process, which requires the biosynthesis of a specific Fe-Mo cofactor and the action of a large number of proteins, which are encoded by *nif* and *fix* gene arrays in rhizobial genomes [138]. The majority of rhizobia are unable to fix nitrogen in free-living state, although several exceptions (e.g. *Azorhizobium caulinodans*) are known [139]. Thanks to their unique

ability of symbiotic nitrogen fixation, rhizobia are widely applied as biofertilizers to improve the production of legume crops.

Nitrogen starvation of host plants is the primary factor, which triggers a chain of events leading to the intracellular infection by free-living soil-resident rhizobia (**Fig. 2.2.1A**). This process involves a complex molecular cross talk between plants and bacteria. Under nitrogen limitation, the former release flavonoids, which cross the membranes and are detected by the regulatory protein NodD of rhizobial cells present in the close proximity to the plant root [140]. NodD bound to flavonoids is a master regulator inducing the expression of *nod* genes [141], which guide the biosynthesis and export of lipochitooligosaccharide compounds known as nodulation (Nod) factors (**Fig. 2.2.1B**) [142]. Nod factors diffusing back to the plant cells bind to specific membrane receptors on their surface and the subsequent activation of a signaling cascade initiates the structural changes in the cells of plant root hairs. These include partial degradation of the cell wall and the rearrangement of the cytoskeleton, which leads to the root hair curling, entrapping of rhizobia in the root hair curl, and subsequent formation of an invagination called an infection thread. Bacteria in the infection thread grow and divide, finally bypassing the epidermis and reaching the cells of the incipient nodule that is initiated in the root cortex simultaneously with the infection process. There, the intracellular infection takes place by the release of rhizobia through endocytosis into the target nodule cells, resulting in bacteroids enclosed in an additional symbiosome membrane. Also as a result of the Nod factor triggered signaling cascade, dedifferentiation of cortex cells gives rise to the

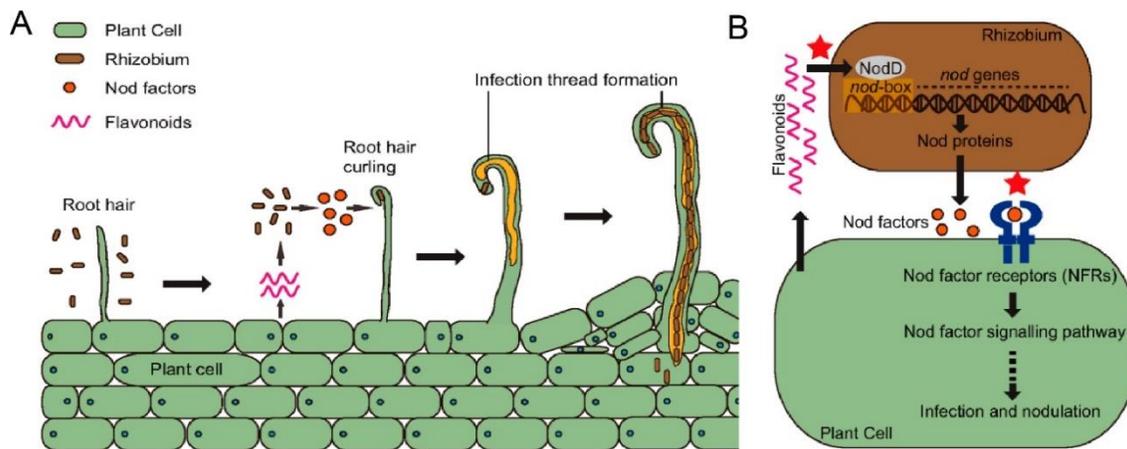


Figure 2.2.1 Establishment of legume-rhizobia symbiosis. (A) Major events in the process of root nodule formation. (B) Schematic representation of the mechanisms involved in the molecular cross talk between rhizobia and the plant. Red stars indicate the examples of receptor-ligand pairs determining the specificity of the symbiosis. Adapted from [143].

primordium of a new nodule, which consist dividing cells and which will allow the growth of the newly formed symbiotic organ. Two types of nodules are distinguished based on the presence of a meristem in the mature organ. In so called *determinate* nodules, the primordium is temporal and will not result in the formation of a meristem. These nodules grow exclusively by expansion of it cells that were initiated in the primordium. On the other hand, in *indeterminate* nodules, a meristem will emerge from the primordium that keeps functioning during the lifetime of the nodule. The cell division activity of this meristem results in continuously growing nodules and leads to the clear stratification of the nodule into functional zones (meristem – cell infection zone – N₂ fixation zone – senescence zone) [135]. The former type is found, for instance, in beans (*Phaseolus vulgaris*), while the latter is typical for alfalfa (*Medicago sativa*).

The majority of rhizobial strains establish highly specific interactions with a restricted number of legume species, although there are examples of strains with a broad host range (e.g. *Sinorhizobium fredii* NGR234 can nodulate legumes from 75 different genera [144]). The specificity of the symbiosis is determined by a range of factors including, for instance, the ability of bacteria to respond to various types of flavonoids and the spectrum of Nod factors and surface determinants (e.g. exopolysaccharides) of rhizobial cells recognized by the host plant in the process of nodulation [145]. The host tightly regulates the growth of rhizobia inside the nodules after the initial infection, which is an additional factor contributing to the specificity of symbiosis. In legumes like *Medicago* species, this control is mediated by the secretion of a cocktail of bioactive peptides known as nodule-specific cysteine-rich peptides (NCRs). NCRs produced by legumes are diverse, they do not display any significant conservation of the amino acid sequence except for 4 or 6 conserved cysteine residues [146]. The mechanisms of NCRs action on bacterial cells also vary: some cationic NCRs effect the cell surface and form pores in the membranes causing cell lysis [147], while others are internalized and bind to different proteins in the cell provoking global transcriptional changes in symbiotic bacteria [148].

Rhizobia feature large genomes (typically 5 to 10 Mb), which is explained by the need to encode multiple genes required for nitrogen fixation and the establishment of symbiosis (*fix*, *nif*, *nod* genes, etc.). Additionally, rhizobial genomes encode an arsenal of metabolic pathways enabling their oligotrophic free-living lifestyle in soil and the use of

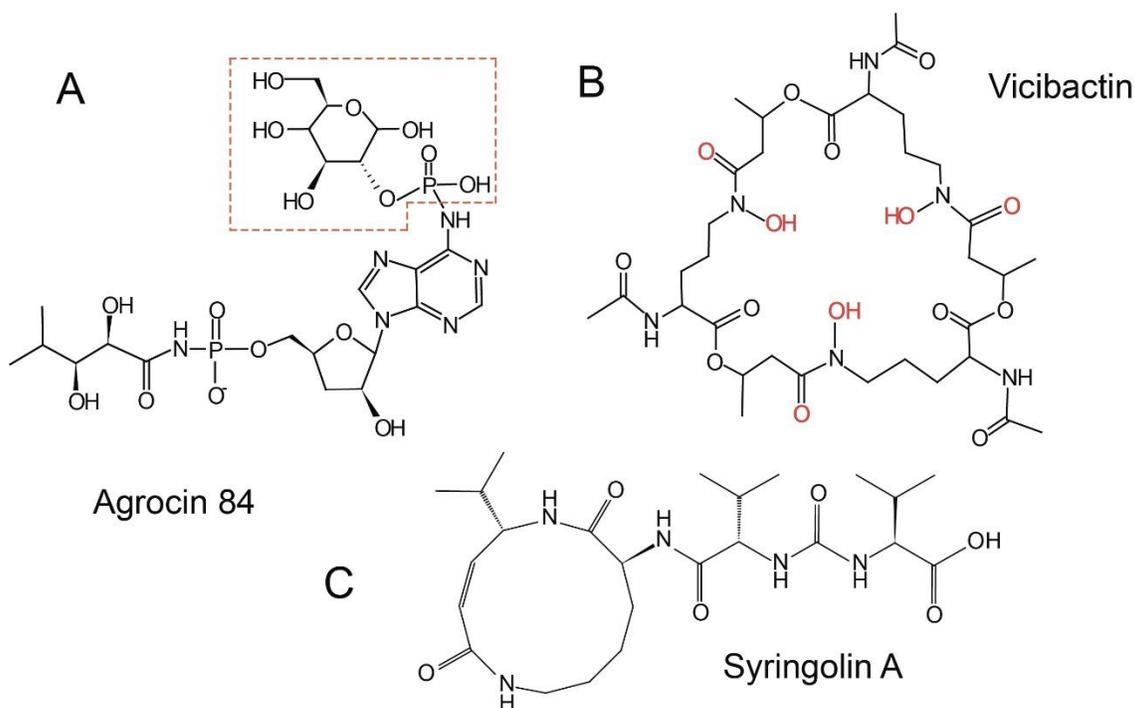


Figure 2.2.2 Chemical structures of representative rhizobial specialized metabolites. (A) Trojan-horse antibiotic agrocin 84 targeting leucyl tRNA synthetase (produced by *Rhizobium rhizogenes* K84) [149]. A moiety mimicking agrocinopine A and required for the efficient uptake into susceptible *Agrobacterium* species is shown in the dashed orange frame [150]. (B) Trihydroxamate siderophore vicibactin produced by *Rhizobium leguminosarum* ATCC14479 [151]. Chemical groups involved in iron chelation are red. (C) PK-NRP hybrid proteasome inhibitor syringolin A produced by *Rhizobium* sp. AP16 [152].

diverse carbon sources [135]. Unlike other groups of bacteria with large genomes such as Actinobacteria or Myxobacteria, rhizobia do not typically produce diverse specialized metabolites. The biosynthetic gene clusters are scarce in their genomes and the majority of those recognized by antiSMASH [12] are responsible for the production of saccharides. In the MIBIG database [13], which contains now data on more than 2000 experimentally studied BGCs, there are only four entries of compounds produced by the members of the

genus *Rhizobium* and three of the genus *Agrobacterium*, compared with 200 entries from *Streptomyces*. Trifolitoxin, belonging to azole-modified RiPPs was already discussed before (see section 2.1.6). **Figure 2.2.2** shows several other examples of specialized metabolites produced by rhizobia.

Chapter 3. Thesis Objectives

The work described in the Thesis was initiated by the identification of a biosynthetic gene cluster of a putative new LAP in the genome of *Rhizobium* sp. strain Pop5 by the methods of genome mining. The workflow of new compounds' characterization in most cases implies that objectives depend on the results obtained in the previous step. For instance, if an antibiotic acts on the membrane of the bacterium, there is no need to identify transporters required for its import.

In our case, the sequence of objectives, which arose in our work on phazolicin, was the following:

1. Identification of the compound, whose biosynthesis is guided by the identified BGC.
2. Determination of its structure.
3. Search for the biological activity of the compound.
4. Identification of its molecular target.
5. Determination of the detailed mechanism of action.
6. Identification of transporters involved in its import into susceptible cells.
7. Identification of the compound's role in the interspecies competition.
8. Systematic genome mining for previously undescribed groups of putative LAPs of potential interest for further experimental validation.

Chapter 4. Materials and methods

4.1 Bacterial strains and growth conditions.

Bacterial strains used in the study are listed in **Supplementary Table 1**. The cultivation media used for the growth of bacteria are listed in **Table 4.1**. Rhizobia were cultivated at 28 °C, *E. coli* strains at 37 °C. Antibiotics were used in the following final concentrations: kanamycin, 50 $\mu\text{g}\cdot\text{mL}^{-1}$ for *E. coli* and 100 $\mu\text{g}\cdot\text{mL}^{-1}$ for rhizobia; chloramphenicol, 34 $\mu\text{g}\cdot\text{mL}^{-1}$; spectinomycin, 25 $\mu\text{g}\cdot\text{mL}^{-1}$; gentamycin 50 $\mu\text{g}\cdot\text{mL}^{-1}$, streptomycin, 500 $\mu\text{g}\cdot\text{mL}^{-1}$; rifampicin 25 $\mu\text{g}\cdot\text{mL}^{-1}$, carbenicillin 100 $\mu\text{g}\cdot\text{mL}^{-1}$ and ampicillin 100 $\mu\text{g}\cdot\text{mL}^{-1}$.

4.2 PHZ production and purification.

Rhizobium sp. Pop5 was cultivated for 24 hours in liquid RM medium (750 mL) with shaking (180 rpm) at 28 °C. Next, the cells were pelleted (30000g, 30 min, 4 °C) and cultivation medium supplied with 0.1% trifluoroacetic acid (TFA) was loaded onto Agilent HF Bond Elut LRC-C18 Cartridge (1 g) pre-equilibrated with 0.1% TFA. The cartridge was then extensively washed with water followed by 50 mL of 10% acetonitrile (MeCN). The peptide fraction was eluted using 10 mL of 30% MeCN, dried out with centrifugal evaporator (GeneVac EZ-2), and dissolved in 500 μL of 100% dimethyl sulfoxide (DMSO). The obtained extract was then subjected to reverse phase HPLC purification on a Luna PREP C18 column (21.2 x 250 mm, 10 μm particle size). The column was pre-equilibrated using 0.1% TFA (Buffer A). The following gradient of 100% MeCN (Buffer B) was used for the separation and elution of the peptides: 0-5 min 0% B, 5-12 min 0-19% B, 12-22 min 19-22% B, 22-27 min 22-70% B, 27-33 70% B. The detection was performed at 254 nm, which is a characteristic wavelength forazole-containing compounds. The

collected fractions were checked with MALDI-ToF MS and those containing the compounds of interest were dried under vacuum, dissolved in 100% DMSO and stored at -20 °C until further use.

Table 4.1 Cultivation media used in the study

Medium	Composition (per 1L of liquid medium)	Agar (%)	Use
LB	5 g NaCl, 10 g tryptone, and 5 g yeast extract	1,5	Growth of <i>E. coli</i>
LB/MC	LB medium + 2.5 mM CaCl ₂ + 2.5 mM MgSO ₄	-	Generalized transduction by <i>S. meliloti</i> Sm1021 phage φM12 (see section 4.11).
TY	5 g tryptone, 3 g yeast extract		Same (see section 4.11)
2xYT	5 g NaCl, 16 g tryptone, and 10 g yeast extract	1,5	Expression of proteins in <i>E. coli</i>
M9	7.25 g Na ₂ HPO ₄ ·2H ₂ O, 3 g KH ₂ PO ₄ , 0.5 g NaCl, 1 g NH ₄ Cl, 2 mM MgSO ₄ , 100 μM CaCl ₂ , 10 μg·mL ⁻¹ thiamine, 1% (v/v) glycerol	1,5	Minimal medium for <i>E. coli</i> , was used in sensitivity tests.
YEB	5 g peptone, 5 g beef extract, 5 g sucrose, 1 g yeast extract, 0.4 g MgSO ₄ ·7H ₂ O, pH 7.5	2	Growth of rhizobia of the genera <i>Mesorhizobium</i> , <i>Sinorhizobium</i> , <i>Agrobacterium</i> , and <i>Azorhizobium</i>
RM	10 g mannitol, 0.5 g K ₂ HPO ₄ , 0.2 g MgSO ₄ , 0.1 g NaCl, 1 g yeast extract, pH 6.8	2	Growth of rhizobia of the genus <i>Rhizobium</i>

4.3 Mass-spectrometry

4.3.1 MALDI-ToF-MS and MS-MS analysis.

MALDI-MS mass spectra were recorded in a positive ion measurement mode using Ultraflextreme MALDI-ToF-ToF-MS instrument (Bruker Daltonics, Germany). The spectra were obtained in reflecto-mode with the accuracy of measuring the monoisotopic

m/z ratio up to 0.1 Da. The fragmentation spectra were obtained using the Lift mode, for which the accuracy of measuring the daughter ions was within 0.2 Da. Sample aliquots were mixed on a steel target with a 30 mg·mL⁻¹ 2,5-dihydroxybenzoic acid in 0.5% trifluoroacetic acid and a 30% MeCN-water solution (Sigma-Aldrich).

4.3.2 High-resolution MS and ESI-MS-MS analysis

One microgram of peptides in a volume of 1-4 µL was loaded onto the Acclaim µ-Precolumn (0.5 mm x 3 mm, 5 µm particle size, Thermo Scientific) at a flow rate of 10 µL/min for 4 min in an isocratic mode of Mobile Phase C (2% MeCN, 0.1% formic acid). Then the peptides were separated with high-performance liquid chromatography (HPLC, Ultimate 3000 Nano LC System, Thermo Scientific, Rockwell, IL, USA) in a 15-cm long C18 column (Acclaim® PepMap™ RSLC inner diameter of 75 µm, Thermo Fisher Scientific, Rockwell, IL, USA). The peptides were eluted with a gradient of buffer B (80% MeCN, 0.1% formic acid) at a flow rate of 0.3 µL/min. The total run time was 60 minutes, which included an initial 4 min of column equilibration with buffer A (0.1% formic acid), then gradient from 5 to 35% of buffer B over 35 min, then 6 min to reach 99% of buffer B, flushing 10 min with 99% of buffer B and 5 min re-equilibration to buffer A.

MS analysis was performed at least in triplicate with a Q Exactive HF mass spectrometer (Q Exactive HF Hybrid Quadrupole-Orbitrap™ Mass spectrometer, Thermo Fisher Scientific, Rockwell, IL, USA). The temperature of the capillary was 240 °C and the voltage at the emitter was 2.1 kV. Mass spectra were acquired at a resolution of 120,000 (MS) in a range of 300–1500 m/z. Tandem mass spectra of fragments were acquired at a resolution of 15,000 (MS/MS) in the range from 100 m/z to 1200 m/z value determined by a charge state of the precursor, but no more than 2000 m/z. The maximum integration time was 50 ms and 110 ms for precursor and fragment ions, respectively. Automatic gain control (AGC) target for precursor and fragment ions were set to 1*10⁶ and 2*10⁵, respectively. An isolation intensity threshold of 50,000 counts was determined for

precursor's selection and up to top 20 precursors were chosen for fragmentation with high-energy collisional dissociation (HCD) at 29 NCE (Normalized collision energy). Precursors with a charge state of +1 and more than +5 were rejected and all measured precursors were dynamically excluded from triggering of a subsequent MS/MS for 20 s.

4.4 Testing of PHZ activity *in vivo*

For the *in vivo* bioactivity test (**Fig. 5.5.1A**), we used the reporter strains BW25113 pDualrep2 and BW25113 Δ *tolC* pDualrep2 as previously described [153]. Briefly, 2 μ L of solutions of tested compounds were applied onto the agar plate that already contained a lawn of the reporter strain. The antibiotics were used as 10 mM solutions (PHZ, KLB, and erythromycin), or 20 μ M solution (levofloxacin). After overnight incubation of the plate at 37 °C, it was scanned by ChemiDoc (Bio-Rad) using “Cy3-blot” settings for RFP and “Cy5-blot” for the Katushka2S protein.

4.5 *In vitro* translation inhibition assays

The inhibition of firefly luciferase synthesis by PHZ (**Fig. 5.5.1B**) was assessed essentially as described previously [154]. Briefly, the *in vitro* transcribed firefly luciferase mRNA was translated in the *E. coli* S30 Extract System for Linear Templates (Promega). Reactions containing 200 ng of mRNA and 0.1 mM of D-luciferin were carried out in 5 μ L aliquots at 37 °C for 15 minutes and the activity of *in vitro* synthesized luciferase was measured by VictorX5 (PerkinElmer) every 30 seconds.

To test the ability of PHZ to act upon the *E. coli* vs. *T. thermophilus* ribosomes (**Fig. 5.5.2**), we used *in vitro* PURExpress translation system (New England Biolabs) reconstituted from purified components and supplied with either the WT 70S ribosomes from *E. coli* or the reconstituted hybrid 70S ribosomes containing purified 30S subunits from *E. coli* and purified 50S subunits from *T. thermophilus*. Translation of superfolder green fluorescent protein (sfGFP) was carried out according to the manufacturer's protocol. The assembled reactions (10 μ L) were supplemented with 100 ng of sfGFP mRNA. PHZ,

KLB, or CHL (chloramphenicol) antibiotics we added to a final concentration of 50 μM when needed. The reactions were placed in a 384-well black-wall plate and the progression of the reactions was monitored over 3 hours by a TECAN microplate scanner.

4.6 Broth microdilution assays and MIC determination

4.6.1 Initial screening of strains for the sensitivity to PHZ

Determination of MICs was performed in 96-well plates in different media for different bacterial strains (**Table 5.2**) using a broth microdilution assay. In each row, the next well contained two times less antibiotic than the previous one. Two different PHZ concentration ranges were tested: 128–0.5 μM for the strains less susceptible to PHZ; and 8 μM to 31.25 nM for the strains more susceptible to PHZ; a well without antibiotic was used as a control. Initially each well contained 100 μL of media. Actively growing bacteria were added to the wells up to the concentration of $\sim 5 \times 10^6$ cells/mL. The plates were incubated for 48 h at 28 °C with shaking (180 rpm). MIC was defined as the concentration of PHZ in the well, in which no bacterial growth was observed, while the two-fold lower concentration of PHZ was not abolishing the growth of bacteria.

4.6.2 Testing of transporter-deficient Sm1021 derivatives

Precultures of the *wt S. meliloti* Sm1021 and mutants were grown in YEB medium with Sm. Overnight grown cultures were diluted to $\text{OD}_{600\text{nm}}$ of 0.2 in fresh YEB medium with Sm and grown until $\text{OD}_{600\text{nm}}$ of 1. The cells were pelleted by centrifugation and resuspended in YEB medium without antibiotics until $\text{OD}_{600\text{nm}}$ of 0.05. The cells were dispatched by 150 μL in a 96-well plate, except for the first column, which contained 300 μL of cultures. PHZ was added to the first column to a final concentration of 20 μM or BLM to a final concentration of 20 $\mu\text{g} \cdot \text{mL}^{-1}$. Two-fold serial dilutions in the subsequent columns were obtained by serial transfer of 150 μL to the next column and mixing by pipetting up and down. No peptide was added to the last column of the 96-well plate,

serving as a control. The 96-well plates were incubated in a SPECTROstar Nano plate incubator (BMG LABTECH). The growth of the cultures in the wells was monitored by measuring the OD_{600nm}; data points were collected every hour for 48 hours. Plates were incubated at 28 °C with double orbital shaking at 200 rpm. Data and growth curves were analyzed using Microsoft Excel. The assay was performed in triplicate for both PHZ and BLM. The obtained MIC values are summarized in **Table 6.1**.

4.7 Molecular cloning procedures

Plasmids used in the study including those constructed are listed in **Supplementary Table 2**. Primers used for DNA amplification are listed in **Supplementary Table 3**.

Molecular cloning was performed either by conventional restriction enzyme digestion and ligation protocol (restriction sites are specified for each gene in the corresponding primer names) or by Gibson Assembly protocol (NEB). For Gibson Assembly pSRK plasmid was PCR-amplified with primers pSRK_GA_F and pSRK_GA_R and treated with DpnI restriction endonuclease (ThermoFisher).

The transformation of *Sinorhizobium meliloti* Sm1021 and *Rhizobium* sp. Pop5 with pSRK and pVO155 derivatives was performed following the triparental mating protocol using pRK600 as a helper plasmid.

4.8 Sm1021 transposon library construction

The strain *S. meliloti* Sm1021 carrying resistance to Sm was used for transposon mutagenesis and was cultured in YEB medium supplemented with Sm at 28 °C. The *E. coli* MFD*pir* strain [155] (Δ *dapA*-derivative, auxotroph for diaminopimelic acid (DAP) synthesis) carrying the plasmid pSAM_Ec [156] was used as a donor strain for the transposon mutagenesis, and cultured in LB supplemented with 300 $\mu\text{g}\cdot\text{mL}^{-1}$ of DAP and kanamycin at 37 °C. The donor strain *E. coli* MFD*pir* pSAM_Ec and the recipient strain *S. meliloti* Sm1021 were grown in 50 mL cultures at 180 rpm until the exponential growth

phase at a final OD_{600nm} of 1. The cultures were washed twice (centrifugation at 1100g for 10 minutes at room temperature) with fresh medium without antibiotics. The pellets were resuspended in fresh medium without antibiotics to obtain a final OD_{600nm} of 50. For conjugation, the donor strain and the recipient strain were mixed at a ratio 1:1. Multiple 100 µL drops of the bacterial mix were spotted on YEB agar plates supplemented with 300 µg·mL⁻¹ of DAP and incubated at 28 °C. After 6 hours of incubation, allowing conjugation of the pSAM_Ec plasmid from the donor *E. coli* strain to the Sm1021 recipient strain and transposition of the transposon into the genome of the target strain, the spots were resuspended in YEB medium, a dilution series was plated on selective medium carrying Sm and Km and subjected to CFU counting to assess the number of individual bacterial mutants obtained by the mutagenesis. In parallel, the remaining bacterial suspension was spread on YEB agar plates supplemented with Sm and Km to obtain the *S. meliloti* Sm1021 transposon mutant population. After 2 days of incubation at 28 °C, the transposon library was resuspended from the agar plates in fresh liquid YEB medium. The suspension was adjusted to 20% glycerol, aliquoted and stored at -80 °C.

4.9 Selection of PHZ-resistant mutants

100 µL of the *S. meliloti* Sm1021 Tn-library prepared as described above with the cell concentration of approximately 1*10⁸ cells·mL⁻¹ were plated on two Petri dishes with YEB medium containing Sm, Km and 20 µM PHZ (>20xMIC). Petri dishes were incubated for 48 hours at 28 °C. Obtained colonies were restreaked on a Petri dish with fresh PHZ-containing YEB medium to confirm the resistance phenotype.

4.10 Sm 1021 whole-genome sequencing, transposon insertions identification

DNA was extracted from 3 mL overnight cultures of *S. meliloti* Sm1021 PHZ-resistant mutants using GeneJET Genomic DNA purification kit (ThermoFisher) according to the manufacturer protocol. NGS libraries were prepared using NEBNext Ultra II DNA Library Prep kit (NEB). DNA sequencing was performed on Illumina MiSeq with the

250+250 bp paired-end protocol. Library preparation and sequencing were performed at Skoltech Sequencing Core Facilities. Raw reads were filtered and trimmed with Trimmomatic [157], genome assembly was performed with SPAdes [158]. Identification of transposon insertion positions was performed with a stand-alone BLAST using the Km^R gene sequence as a bait [159]. NC_003047 genome annotation was used as a reference. Sequencing data were deposited in SRA (BioProject ID PRJNA760523).

4.11 Construction of *Rhizobium* sp. Pop5 derivatives (Ω *phzD* and Ω NI)

To obtain a *Rhizobium* sp. Pop5 mutant with the disruption of the *phzD* gene (YcaO domain-containing cyclodehydratase, locus tag: RCCGEGPOP_21747, protein GenBank accession number: EJZ19165.1), a 566 bp internal fragment of the gene was PCR amplified and cloned into the plasmid pVO155*nptIIgfp* (pVO155 plasmid [160] derivative with constitutively expressed *gfp* gene; does not replicate in *Rhizobium* spp.) between *Sall* and *XbaI* restriction sites. The resulting construct was introduced into *Rhizobium* sp. Pop5 via triparental mating with the helper strain HB101 pRK600 [161]. The cells with the plasmid integrated into the genome were selected on RM medium with Km. As there is no resistance marker in the genome of *Rhizobium* sp. Pop5, we did not perform a counter-selection of the *E. coli* donor, which could be easily distinguished from *Rhizobium* based on the morphology of the colonies growing on the solid RM medium. The construction of the strain *Rhizobium* sp. Pop5 Ω NI (neutral insertion) was performed using the same procedure. In this case a region containing the parts of *phzD* and *phzR* genes was amplified and cloned into pVO155*nptIIgfp*.

4.12 Construction of Sm1021 double mutants

Generalized transduction by *S. meliloti* Sm1021 phage ϕ M12 was used to obtain double mutants lacking both functional BacA and YejABEF importers. *S. meliloti* Sm1021 Δ *bacA* served as a donor, while Sm1021 Ω *yejA* and Sm1021 Ω *yejE* were used as recipient strains. The procedure was performed as described earlier [162]. Briefly, 5 mL of Sm1021

ΔbacA donor strain culture grown overnight at 30 °C in LB/MC medium supplemented with 25 μg·mL⁻¹ of Sp, was inoculated by the phage in cell:phage ratio 1:1. The mixture was incubated overnight with shaking at 30 °C, then sterilized by the addition of 150 μL of chloroform and cleared from the remaining cell debris by centrifugation (7000g, 10 min). The obtained lysate was then used to inoculate 1 mL of the overnight culture of the recipient strains grown in LB/MC up to cell:phage ratio of 2:1. The obtained mixtures were incubated for 30 min at room temperature and pelleted (4500g, 2 min). The pellet was washed with 1 mL of TY medium and resuspended in fresh TY. The suspensions were plated on the TY agar plates supplemented with Sp and Km to select for transductants carrying both resistance markers in the genome. A mixture lacking the lysate served as a negative control. The colonies obtained were screened using PCR with primers specific to *bacA* and *yejA* or *yejE* to confirm the genotype.

4.13 Competition experiments

4.13.1 Co-cultivation in laboratory medium

The strains Sm1021 *wt*, Sm1021 *ΔbacA* *ΩyejA*, Pop5 pIN72, and Pop5 *ΩphzD* pIN72 were used for competition experiments under laboratory cultivation conditions. Pre-cultures of the strains, grown in RM medium with appropriate antibiotics, were washed and diluted in fresh RM medium without antibiotics to OD_{600nm}=0.1 and then grown until OD_{600nm}=1. Single strain cultures and 50%:50% mixtures were prepared from these fresh suspensions to reach an initial OD_{600nm}=0.04 in RM medium without antibiotics. Aliquots were taken from these cultures at t=0 h, t=24 h, t=48 h, and t=72 h. Ten-fold dilution series were prepared from these aliquots, 5 μL of each dilution was spotted on RM plates without antibiotics or with either Sm500 (selective for Sm1021 and derivatives) or Tc10 (selective for pIN72-harboring Pop5 derivatives). CFU counting was performed after 48 hours of incubation at 28 °C.

4.13.2 Co-cultivation in soil

The soil used for cocultivation experiments was obtained from the Ferme de Viltain in Saclay (48°45'14.2"N 2°09'32.3"E) in February 2021. The field selected for the collection of soil was used for the cultivation of beans in 2020. Before use, the soil sample was kept for several weeks for drying and the dry soil was crushed and sifted with a 2 mm sieve.

In all experiments the PHZ-sensitive strain *R. leguminosarum* bv. *phaseoli* 4292 (*Rlp* 4292) was used (Rif^R). For cocultivation experiments the soil. The overnight culture of *Rlp* 4292 was diluted to the OD_{600nm}=0.2 in fresh RM medium and incubated with 200 rpm agitation at 28 °C until the OD_{600nm} was 1. Three samples of the obtained culture (10 mL each) were placed in separate tubes. The cells were pelleted, the supernatant was removed, and each cell pellet was resuspended in 33 mL of sterile water. Each of three obtained suspensions was used to inoculate a pot containing 100 g of dry soil. This corresponds to the addition of about 10⁸ CFU per g of soil. Before the inoculation with Pop5 derivatives, the soil was kept for 7 days at RT to allow the stabilization of the *Rlp* 4292 population at around 10⁷ CFU per g of soil (measured with the recovery/CFU counting procedure described below).

The suspensions of Pop5 *wt* or Pop5 Ω *phzD* were prepared as above for *Rlp* 4292. Either 10⁷ CFUs/g (equal amount to the soil resident *Rlp* 4292) or 10⁹ CFUs/g (hundred fold excess to the soil resident *Rlp* 4292) of each culture were inoculated into 10 g of soil sampled from each of three experimental pots with established populations of *Rlp* 4292. Water without bacteria was added to control samples.

Recovery of viable *Rlp* 4292 cells from soil and subsequent CFU counting was performed according to the following procedure. An aliquot containing 1 g of homogeneous soil from each experimental pot was added to 1 mL of sterile water in a separate tube. Each tube was vortexed intensively and placed for 30 min in a shaker. Then, the clumps of soil were pelleted by centrifugation (1 min, 0.1g), while bacterial cells remained in the supernatant. An aliquot of 100 μ L of the obtained supernatant was

transferred into a new Eppendorf tube and supplemented with 900 μL of RM medium. Serial dilutions of the obtained sample (50 μL per dilution) were spread on Petri dishes with RM medium supplemented with rifampicin and cycloheximide. Cycloheximide at 200 $\mu\text{g}/\text{mL}$ inhibits the growth of fungi on the plates. The counting of the colonies was performed after 2 days of incubation at 28 $^{\circ}\text{C}$.

4.13.3 Testing the effect of PHZ addition into the soil

Three samples of soil (200 mg each) were prepared in Eppendorf tubes from the pot inoculated with *Rlp* 4292. 20 μL of mQ or a PHZ-containing solution were added to the samples. The final concentration of PHZ was either 0 (mQ added), 100 μM , or 200 μM . Each tube was vortexed. The same procedure was performed using 200 μL of water instead of soil. The experiment was performed in triplicate using the soil from independently inoculated pots. The resulting number of samples was 18. After 3 days of incubation at RT, 200 μL of water were added to each sample. Each tube was vortexed and then placed for 30 min in a shaker. The tubes were centrifuged for 1 min at 0.1g. The supernatant from each tube was transferred to a new tube and 10-fold serial dilutions were prepared. Aliquots of the serial dilution samples (50 μL each) were spread on Petri dishes with RM medium supplemented with rifampicin and cycloheximide. We counted the number of colonies on the dishes after 2 days of incubator at 28 $^{\circ}\text{C}$.

4.13.4 Nodulation experiments

Phaseolus vulgaris (variety “Magical”) plants growing in a perlite-sand (2:1) mixture were used in all experiments. This nutrient-free substrate allows for the good development of the root system of the plants, does not contain any source of nitrogen and bacteria capable of *P. vulgaris* nodulation.

Before planting, the seeds of *P. vulgaris* were sterilized and germinated according to the following procedure. Approximately 200 seeds were placed in a flask, covered with concentrated sulfuric acid, and incubated for 7 minutes. Sulfuric acid was removed, the

seeds were washed with an excess of sterile water (300 mL) five times. Then, the seeds were incubated for 20 min in bleach (0.36%), followed by rinsing with sterile water for five times. After that, the seeds were placed on the Petri dishes with Kalys agar (0.7%) and placed in an incubator (28 °C) for 2 days.

The germinated seeds were planted in individual square pots (2 L each) filled with perlite-sand mixture. The planted seeds were watered with nitrogen-free fertilizer (PlantPro, N-P-K: 0-15-40; 1 g/L) every other day for one week. After the appearance of first leaves, the plants were watered every two days, alternating with water and with fertilizer. The greenhouse was maintained at a temperature of approximately 23 °C.

The inoculation of seedlings was performed on the 10th day after planting. The bacterial cultures ($OD_{600nm}=0.05$ or 0.1, see section 7.5 for the discussion of the initial density used) of *Rlp* 4292 and Pop5 derivatives were used for the inoculation. The precultures grown overnight in the RM medium with relevant antibiotics added (Rif50 for *Rlp* 4292, Kan100 for Pop5 pSRK and Pop5 $\Omega phzD$) were pelleted (12 min, RT, 4500g), washed twice with fresh medium without antibiotics, and resuspended in water until the desired cell density. Aliquots containing 10 mL of the obtained cell suspensions were inoculated next to the stem of each plant. The mixed suspension of the two tested strains (1:1) used for the inoculation of experimental groups was prepared immediately before the inoculation and used for all the plants in one experimental group (10 mL per plant). In each competition experiment, we had groups of plants inoculated with single strains and a noninoculated group, as controls. The pots with plants from every experimental or control group were placed on a separate tray, which prevented subsequent mixing of the inoculums upon watering of the plants.

4.13.5 Analysis of nodules

Three-week old nodules were examined by microscopy for the presence of bacteria inside the plant cells. The preparation of nodules and microscopy was performed essentially as described before [163]. In brief, nodule sections (60 μ m) were prepared with

a Leica VT1200S vibratome and incubated in staining solution (5 μ M SYTO9, 30 μ M PI, 0.01% Calcofluor White M2R in 50 mM Tris-HCl buffer pH 7.0). Sections were rinsed in water and placed in 50 mM Tris-HCl buffer, pH 7.0, for microscopic observations. Images were taken using a Leica confocal laser scanning microscope TCS SP2.

Nodules for the analysis were collected 3 weeks after the day of inoculation. 20 nodules were picked randomly from different parts of the root system for each plant in the experimental samples. Each individual nodule was surface-sterilized by soaking for 1 min in 100 μ L of 70% EtOH, EtOH was diluted by the addition of 1 mL of sterile mQ. After an additional round of wash with 1 mL of sterile mQ, each nodule was crushed in 500 μ L of bacteroid extraction buffer (125 mM KCl, 50 mM sodium succinate, 50 mM TES, 0.1% BSA, pH 7.0) until no large clumps of plant tissues were remaining. Aliquots of the obtained nodule homogenate (10 μ L each) were spotted on square Petri dishes with RM agar supplemented with selective antibiotics (Rif50 or Kan100). One plate was used for the simultaneous analysis of bacteria recovered from 50 nodules (**Fig. 7.5.3A**).

4.14 Colony-forming unit (CFU) assay

CFU counting was used to assess the sensitivity of strains to the action of PHZ as a complementary method to the broth microdilution assay, as it allows identifying the occurrence of resistant clones, which appear as colonies growing in the undiluted to hundred-fold diluted samples. For CFU counting the overnight cultures of selected Sm1021 derivatives were diluted with fresh YEB medium with relevant antibiotics added to the $OD_{600nm}=0.2$ and allowed to grow to the $OD_{600nm}=0.6$ at 28 °C with shaking. Then the cultures were adjusted to $OD_{600nm}=0.2$ and ten-fold dilution series of the obtained cell suspensions were prepared, 5 μ L of each dilution was spotted on YEB plates supplemented with either Sm (negative control) or Sm and PHZ. Each experiment was performed in triplicate using independent starting cultures inoculated with single colonies of the corresponding strains. For the strains carrying the pSRK plasmids with a panel of genes under the control of the *lac* promotor, 1 mM IPTG was added to the YEB medium to induce

the expression. Plates were incubated for 3 days at 28 °C, after which the number of CFUs was counted.

4.15 Cryo-electron microscopy (cryo-EM) structure determination.

4.15.1 Ribosome preparation

E. coli strain MRE600 overnight cultures were diluted 1:100 into 3 L of LB. After growth to mid-log phase ($OD_{600nm}=0.6$), cells were cooled on ice for 3 min, pelleted, and washed with buffer A (20 mM Tris-HCl pH 7.5, 100 mM NH_4Cl , 10 mM $MgCl_2$, 0.5 mM EDTA, 2 mM DTT). Cells were resuspended in 100 mL buffer AS (buffer A with 150 mM sucrose) and lysed with two passes through an Emulsiflex-C3 homogenizer at 18000 psi. Lysate was clarified by centrifugation for 30 min at 18000 rpm in a JA-20 rotor (Beckman-Coulter). The supernatant was layered over a sucrose cushion of 24 mL buffer B (buffer A with 500 mM NH_4Cl) with 0.5 M sucrose and 17 mL buffer C (20 mM Tris-HCl pH 7.5, 60 mM NH_4Cl , 6 mM $MgCl_2$, 0.5 mM EDTA, 2 mM DTT) with 0.7 M sucrose in ti-45 tubes (Beckman-Coulter). Ribosomes were pelleted in a Ti-45 rotor at 27000 rpm for 16 hours at 4 °C. Crude ribosome pellets were resuspended in disassociation buffer (buffer C with 1 mM $MgCl_2$) and clarified via centrifugation at 15000g for 10 min. The cleared supernatant was layered over 15-40% sucrose gradients in disassociation buffer and spun 28000 rpm for 16 hours at 4 °C in a SW-32 rotor (Beckman-Coulter). Gradients were fractionated with an ISCO fractionation system and 30S + 50S peaks were combined. The combined subunits were concentrated in 15 mL Millipore 100k molecular weight cutoff spin filter and buffer exchanged with reassociation buffer (buffer C with 10 mM $MgCl_2$). Ribosomes were incubated 45 min at 37 °C, layered onto 15-40% sucrose gradients in reassociation buffer, and spun 27000 rpm for 15 hours at 4°C in an SW-32 rotor. Gradients were fractionated as before, 70S peaks were collected, and ribosomes were concentrated and washed with buffer C in an Amicon stirred cell filtration system using 100k cutoff filters. Ribosomes were stored in aliquots at -80 °C.

4.15.2 Cryo-EM sample preparation

Roughly 4-6 μM PHZ was incubated with 100 nM ribosomes in RC buffer (50 mM HEPES pH 7.5, 150 mM KOAc, 12 mM MgOAc, 7 mM β -mercaptoethanol) for 30 min at 37°C. Complexes were deposited in 4- μL aliquots on 300 mesh Quantifoil UltraAuFoil R1.2/1.3 grids that were topped with a layer of amorphous continuous carbon and glow-discharged for 15 seconds with a Pelco SC-6 sputter coater. After incubation period of about a minute, the excess sample was washed off on drops of RC-LS buffer (RC with 25 mM KOAc). Samples were blotted and plunge-frozen in liquid ethane using an FEI Vitrobot with blot force 6, humidity 100%, and 20 °C or 4 °C depending on the freezing session.

4.15.3 Data collection

Images were collected in two separate sessions on an FEI Titan Krios Microscope operated at 300 keV with a GIF energy filter and GATAN Summit K2 direct electron detector in super-resolution mode. Images were collected at 215,000x magnification for a pixel size of 0.56 Å (0.28 Å super-resolution). Automated movie collection was performed with SerialEM [164] over the defocus range -0.6 to -2.0 μm , and Focus software [165] was used to monitor data collection. The total dose was 29.97 e-/Å² for session 1 and 29.83 e-/Å² for session 2, each over 30 frames.

4.15.4 Cryo-EM data analysis

Cryo-EM data processing was done using RELION 3.0 software [166] (**Supplementary Table 4**). The two datasets were processed separately until the final round of map refinement, for which the best particles from both sets were combined. Movies were motion-corrected and dose-weighted using RELION's motion correction algorithm. Contrast transfer function (CTF) parameters were estimated using CTFFind4 [167] and micrographs with poor CTF fit, as determined by visual inspection, were sorted

out. Particles were picked automatically with the Laplacian-of-Gaussian method and subjected to several rounds of careful 2D class-based cleaning before an initial round of 3D auto-refine. CTF refinement with per-particle defocus and beam tilt correction was performed, and particles were further sorted through 3D classification without alignment. Another round of 3D auto-refine was performed on the best 3D classes with focused refinement on the 50S subunit. This was followed by one round of Bayesian polishing [166] and another round of CTF refinement, which were each followed by another round of 3D auto-refine with a mask around the 50S subunit. Particles from the two sessions were then pooled for the final round of 3D auto-refine. A charge density map was calculated in Chimera [168] as previously described [169]. Post-processing was also performed in RELION to generate a Fourier shell correlation (FSC) curve. All refinement procedures were run on 2x binned particles because of the small pixel size.

4.15.5 Model building

Assembly #2 from the PDB entry 4YBB [170] was used as a starting ribosome model. Only the 50S subunit was subjected to real-space refinement in PHENIX [171], because charge density corresponding to the 30S subunit was relatively poor after focused refinement on the 50S, where PHZ binds. Ions and water molecules were removed from the starting model prior to real-space refinement. The atomic model for the PHZ was built as follows: first, baton building in COOT [172] was used to trace the main chain; next, the resulting polyalanine chain was mutated to the unmodified PHZ sequence after visual search for the characteristic features in the map; next, oxazole and thiazole modifications were built in Avogadro version 1.2.0 [173] based on the density. Finally, CIF restraints for the oxazole and thiazole moieties were created based on the structural data from [174]. USCF Chimera was used for EM density map shown in **Fig. 5.6.2**.

4.16 Cloning, expression and purification of YejASm

YejASm signal peptide prediction was performed using SignalP 5.0 [175]. *yejA*Sm lacking the fragment encoding the first 30 amino acids (signal peptide) was PCR-amplified from Sm1021 genomic DNA and cloned into a pET29b(+) vector (Novagen) generating a C-terminal 6His-tag. The resulting plasmid pET29-*yejA*Sm-CHis6 was electroporated into *E. coli* Rosetta 2 (DE3) pLysS. Two liters of 2xYT medium supplemented with Km and Cm were inoculated with 20 mL of an overnight culture of the obtained strain. Cells were grown at 37 °C, 180 rpm to an OD_{600nm}=0.7, then induced with 0.5 mM IPTG and incubated at 28 °C for another 5 hours. The cultures were cooled down on ice, pelleted (4500g, 20 min, 4 °C), and frozen in liquid nitrogen.

Cells were resuspended in 80 mL of Lysis Buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 10% glycerol, 20 mM imidazole) supplemented with homemade purified DNase and protease inhibitors cocktail (Sigma-Aldrich) and disrupted by sonication. After centrifugation (30000g, 25 min, 4 °C), the supernatant was loaded onto a 5 mL HisTrap HP column (Cytiva). Protein elution was performed with 50 mM Tris-HCl pH 8.0, 300 mM imidazole and 300 mM NaCl. Protein fractions were loaded onto a gel filtration column (HiLoad 26/60 Superdex 200 prep grade, Cytiva) equilibrated with 50 mM Tris-HCl pH 8.0 and 150 mM NaCl. The fractions with the highest protein concentration were pooled, concentrated, and stored at -80 °C.

4.17 Crystallization and structure determination of YejASm

Crystallization conditions for YejASm at 14 mg·mL⁻¹ were screened using QIAGEN kits (Valencia, CA) with a Mosquito nanodrop robot (SPT Labtech). YejASm crystals were manually optimized in the condition specified in **Supplementary Table 5**. Crystals were transferred to a cryo-protectant solution (mother liquor supplemented with 25% PEG 400) and flash-frozen in liquid nitrogen. Diffraction data were collected at 100 K on PROXIMA 2 beamline at SOLEIL synchrotron (Saint-Aubin, France). Data processing was performed using the XDS package [176] (**Supplementary Table 5**). Because of the diffraction

anisotropy, the DEBYE and STARANISO programs developed by Global phasing Ltd were applied to the data scaled with AIMLESS using the STARANISO server (<http://staraniso.globalphasing.org>). These programs perform an anisotropic cut-off of merge intensity data on the basis of an analysis of local $I/\sigma(I)$, compute Bayesian estimates of structures amplitudes, taking into account their anisotropic fall-off, and apply an anisotropic correction to the data. The structure was solved by molecular replacement with PHASER [177] using the coordinates of the separate N- and C-terminal lobes of the Cu(I)-methanobactin complex-binding protein MbnE from *Methylocystis parvus* OBBP (PDB ID: 5ICQ, [178]) as search models. Inspection of the resulting model using COOT [179] showed strong electron density maps at the lobes interface, which were attributed to peptides likely coming from protein degradation during overexpression. The backbones of the short peptide ligands (2 and 5 amino acid peptides) were modeled at two different places of the interface based on the electron density. Electron density for peptide side chains was more ambiguous and no electron density was present linking the two short bound peptides, indicating that a population of different peptides might be present in the ligand-binding site of YejASm molecules within the crystal. Refinement of the structure was performed with BUSTER-2.10 [180] employing TLS (translation/libration/screw) groups restraints. Refinement details are shown in **Supplementary Table 5**. Molecular graphics images were generated using PyMOL (<http://www.pymol.org>). The YejASm structure factors and coordinates have been deposited to the Protein Data Bank with PDB ID 7Z8E.

4.18 Genome mining for novel LAP BGCs

We downloaded 146 381 bacterial genomes from RefSeq [181] database on March 27th, 2019. To obtain all YcaO domain-containing proteins we searched the database with profile HMMs (TIGR03549, TIGR03604, and PF02624) using hmmer package (hmmer.org). We clustered resulting hits with mmseqs2 [182] (90% identity; 90%

coverage) to remove duplicates and redundant highly similar sequences from organisms, which genome sequences are overrepresented in the database.

The genomic regions of 12.5 kbp to each side of the identified unique YcaO protein-coding genes were annotated with RODEO [183] using Pfam 32.0 and TIGRFAMs 15.0 databases. For further analysis, we selected genomic regions according to the following rules. First, we collected regions that encode proteins containing E1-like (PF00881, TIGR03603, TIGR04424) or ThiF-like (PF00899, TIGR02354, TIGR02356, TIGR03693, TIGR03736, TIGR03882) domains. Since the initial hmmer search was performed with relatively permissive parameters, for some of the hits RODEO proposed a profile HMM other than those for YcaOs (TIGR03549, TIGR03604, or PF02624) as the most probable. We removed such genomic regions from further analysis. In order to exclude thiopeptides, studied comprehensively in several other works, we removed genomic regions containing genes of lantipeptide dehydratase (PF14028, PF04738, TIGR03897, PF05147). Putative precursor peptides were predicted with RiPPER [40]. For each BGC the best predicted precursor peptide was selected as the one bearing the highest number of cyclizable residues (Ser, Thr, Cys residues) within the C-terminal half.

Using a custom script we converted RODEO output to genbank files imitating antiSMASH [184] output (<http://github.com/bikdm12/RODEO2antiSMASH>). The script adds a feature “cluster” with information about the class of the product. The coordinates of this feature are boundaries of the group of genes located on the same strand no further than 100 bps from each other and containing a gene of an YcaO protein. Genes that may be related to the biosynthesis of azol(in)e-containing RiPPs (for the list of domains see **Table 4.2**) were marked as biosynthetic. These files were then used to build a sequence similarity network with BiG-SCAPE [185], which was subsequently visualized using Cytoscape [186].

Table 4.2 List of protein domains annotated in the BGCs of putative LAPs.

Name	Accession number	Description
Methyltransf_11	PF08241.12	Methyltransferase domain
Methyltransf_2	PF00891.18	O-methyltransferase domain
Methyltransf_25	PF13649.6	Methyltransferase domain
TfuA	PF07812.12	TfuA-like protein
ThiF	PF00899.21	ThiF family
YcaO	PF02624.16	YcaO cyclodehydratase, ATP-ad Mg ²⁺ -binding
TIGR00027	TIGR00027	mthyl_TIGR00027: methyltransferase, TIGR00027 family
TIGR00702	TIGR00702	TIGR00702: YcaO-type kinase domain
TIGR02356	TIGR02356	adenyl_thiF: thiazole biosynthesis adenylyltransferase ThiF
TIGR03534	TIGR03534	RF_mod_PrmC: protein-(glutamine-N5) methyltransferase, release factor-specific
TIGR03549	TIGR03549	TIGR03549: YcaO domain protein
TIGR03601	TIGR03601	B_an_ocin: bacteriocin, heterocycloanthracin/sonorensin family
TIGR03603	TIGR03603	cyclo_dehy_ocin: thiazole/oxazole-forming peptide maturase, SagC family component
TIGR03604	TIGR03604	TOMM_cyclo_SagD: thiazole/oxazole-forming peptide maturase, SagD family component
TIGR03605	TIGR03605	antibiot_sagB: SagB-type dehydrogenase domain
TIGR03693	TIGR03693	ocin_ThiF_like: putative thiazole-containing bacteriocin maturation protein
TIGR03793	TIGR03793	TOMM_pelo: NHLP leader peptide domain
TIGR03798	TIGR03798	ocin_TIGR03798: nif11-like leader peptide domain
TIGR03882	TIGR03882	cyclo_dehyd_2: bacteriocin biosynthesis cyclodehydratase domain
TIGR03891	TIGR03891	thiopep_ocin: thiopeptide-type bacteriocin biosynthesis domain
TIGR03892	TIGR03892	thiopep_precurs: thiazolylpeptide-type bacteriocin precursor
TIGR03893	TIGR03893	lant_SP_1948: type 2 lantibiotic, SP_1948 family
TIGR03949	TIGR03949	bact_Iib_cerein: class Iib bacteriocin, lactobin A/cerein 7B family
TIGR03965	TIGR03965	mycofact_glyco: mycofactocin system glycosyltransferase
TIGR03971	TIGR03971	SDR_subfam_1: SDR family mycofactocin-dependent oxidoreductase

TIGR03975	TIGR03975	rSAM_ocin_1: ribosomal peptide maturation radical SAM protein 1
TIGR03997	TIGR03997	mycofact_OYE_2: mycofactocin system FadH/OYE family oxidoreductase 2
TIGR04078	TIGR04078	rSAM_yydG: peptide modification radical SAM enzyme, YydG family
TIGR04107	TIGR04107	rSAM_HutW: putative heme utilization radical SAM enzyme HutW
TIGR04187	TIGR04187	GRASP_SAV_5884: ATP-grasp ribosomal peptide maturase, SAV_5884 family
TIGR04188	TIGR04188	methyltr_grsp: methyltransferase, ATP-grasp peptide maturase system
TIGR04223	TIGR04223	quorum_AgrD: cyclic lactone autoinducer peptide
TIGR04282	TIGR04282	glyco_like_cofC: transferase 1, rSAM/selenodomain-associated
TIGR04424	TIGR04424	metallo_McbB: McbB family protein
TIGR04470	TIGR04470	rSAM_mob_pairB: radical SAM mobile pair protein B
TIGR04479	TIGR04479	bcpD_PhpK_rSAM: radical SAM P-methyltransferase, PhpK family
TIGR04511	TIGR04511	SagB_rel_DH_2: putative peptide maturation dehydrogenase
TIGR04538	TIGR04538	P450_cycloAA_1: cytochrome P450, cyclodipeptide synthase-associated
TIGR04545	TIGR04545	rSAM_ahbD_hemeb: heme b synthase

Chapter 5. Phazolicin: biosynthetic gene cluster, structure, and mechanism of action

5.1 Genome-mining guided discovery of putative new LAPs BGCs

A ribosome-targeting LAP klebsazolicin (KLB) and its BGC (*klpACBDE*) were previously characterized in our laboratory (see section 2.1.5) [76], [99]. To find biosynthetic gene clusters resembling that of KLB we performed a BLASTP search using the sequence of KlpD (WP_077257196), a YcaO domain-containing cyclodehydratase encoded in the *klp* BGC (**Fig. 5.1A**), as a bait. This protein was selected, as it is involved in the modification of all azol(in)e-containing RiPPs catalyzing the first step in the formation of heterocycles (see section 2.1.2). Using the top hits of the BLASTP search output, we identified a number of clusters with the same gene composition as in the *klp* BGC. Manual annotation of small ORFs adjacent to genes involved in the posttranslational modification (homologs of *klpC*, *klpB*, and *klpD*) allowed us to identify the genes of putative precursor peptides, located nearby. Automated genome annotation tools frequently miss short ORFs of RiPPs precursor peptides. In the case of LAPs, manual identification is relatively reliable, as core peptides are enriched in amino acid residues capable of posttranslational conversion into azoles, i.e. Cys, Ser, and Thr.

Although all biosynthetic gene clusters found contained the same set of core genes, including those of the precursor, three proteins required for the installation of azole cycles, and an ABC-transporter, it became evident that they can be split into two groups based on the amino acid sequence of the precursor peptide encoded. The first group included the BGCs of close KLB homologs, some of which were described previously (**Fig. 5.1B**). Their precursor peptides are short (approx. 40-47 amino acids long) and contain an XQSP motif involved in the installation of the N-terminal amidine cycle present in klebsazolicin [99]. The second group of clusters found in our BLASTP search included three sequences identified in the genomes of Alphaproteobacteria: *Rhizobium* sp. PDO1-076, *Rhizobium* sp. Pop5, and *Phyllobacterium myrsinacearum* DSM 5893. Precursor peptides they encode are longer than KlpA-like precursors, have a different distribution of Cys, Ser, and Thr residues in the core part, lack the motif for amidine installation at the leader-core junction, and contain 3-4 conserved positively charged residues in the core part (**Fig. 5.1C**). Moreover, in these clusters, the gene of the transmembrane export pump homologous to *klpE* was oriented in the reverse direction to the rest of the genes in the cluster. Two of the clusters found in the genomes of *Rhizobium* strains also contained a gene of a MucR-family transcription regulator located downstream of the D gene and oriented in the reverse direction. Given these features distinguishing the newly identified clusters from *klp*-like ones, we hypothesized that the modified peptides they encode may have a distinct mode of action and proceeded to characterize their products.

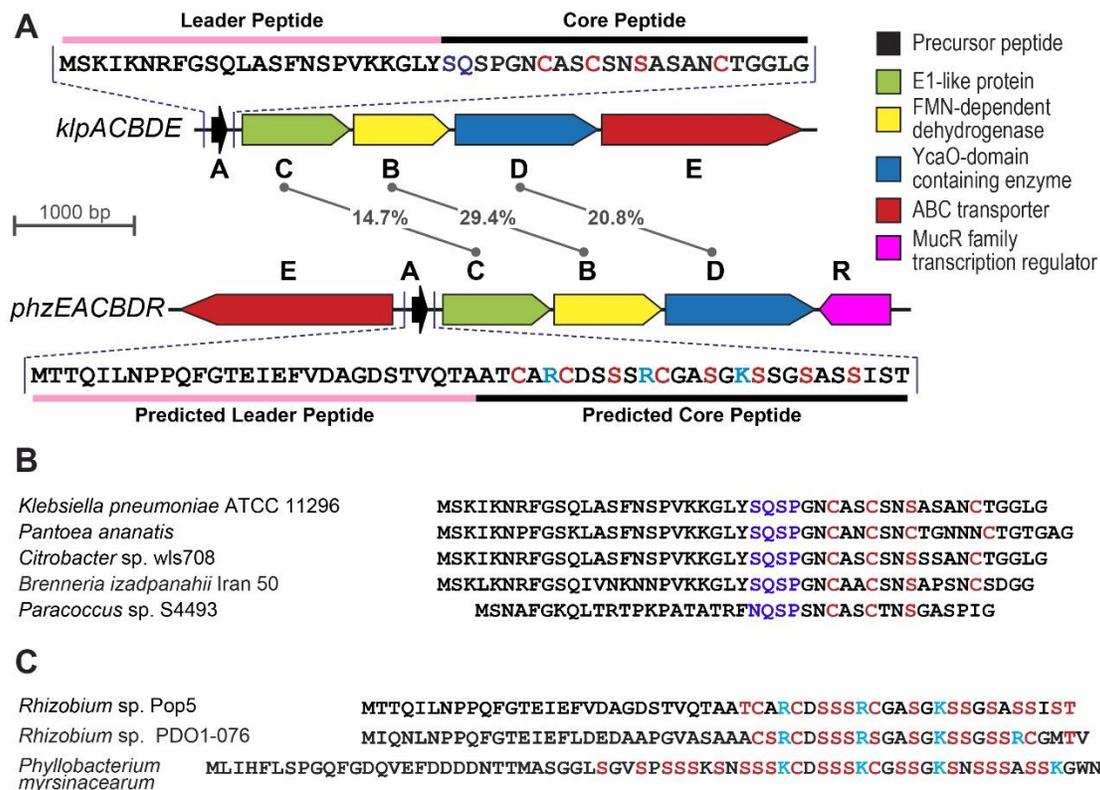


Figure 5.1.1 Comparison of the *phz* BGC with previously described *klp*-like BGCs

(A) Gene composition of the biosynthetic gene cluster of KLB (*klpACBDE*) and a newly identified LAP BGC from the genome of *Rhizobium* sp. Pop5 (*phzEACBDR*). Functions of the encoded proteins are listed on the right. Numbers in the middle indicate the level of sequence identity (in %) between the amino acid sequences of proteins forming the azole-installing modification complexes (C, B, and D proteins). Amino acid sequences of the precursor peptides are shown next to each BGC, core peptides are shown in black, leader peptides in pink. Residues converted into azoles in mature modified peptides are shown in red, positively charged residues in the core part of precursor from *Rhizobium* sp. Pop5 are shown in blue. (B) A representative set of KLB-like precursor peptides. Note the conserved XQSP motif (shown in blue) involved in the installation of the N-terminal amidine ring in KLB. The residues converted into azoles in the KLB core part and corresponding residues in the precursors of the homologs are shown in red. (C) Amino acid sequences of the precursor peptides encoded in the BGCs found in the genomes of Alphaproteobacteria. Residues of Ser, Thr, and Cys in the C-terminal parts of the precursors are shown in red, positively charged residues – in light blue.

5.2 Phazolicin purification and characterization

The strain *Rhizobium* sp. Pop5 was kindly provided to us by Prof. Maria Esperanza Martinez Romero (National Autonomous University of Mexico), whose group isolated it from the nodules of wild beans (*Phaseolus vulgaris*) collected in the tropical forest Los Tuxtlas (South Mexico) and sequenced its genome. Based on the presence of the ABC export pump in the BGC we proposed that the modified peptide is secreted into the medium upon production.

The HPLC analysis of extracts prepared from the RM medium after the 24 h long cultivation of *Rhizobium* sp. Pop5 revealed several peaks absorbing at 254 nm (**Fig. 5.2.1A**). This wavelength was previously used for detection of other compounds containing azoles by monitoring their elution from HPLC columns. MALDI-ToF-MS of HPLC fractions allowed us to identify three major and five minor compounds, which could be the products of the PhzA precursor peptide posttranslational modification followed by the leader peptide removal, and, for some of them, partial proteolysis of the C-terminal part of the core. **Table 5.1** summarizes the features of the identified peptides. The major compound with monoisotopic $m/z = 2363.9$ $[M+H]^+$ (**Fig. 5.2.1B**) was named phazolicin (PHZ), following the nomenclature proposed for LAPs meaning “an azole-modified peptide from *Phaseolus*”. Its mass corresponds to that of the 27 C-terminal amino acid residues of the PhzA precursor cleaved off between Ala28 and Ala 29 with 8 azole cycles installed (each azole installation leads to the loss of 20 Da).

Table 5.1 Characteristics of the PHZ-related compounds identified in the *Rhizobium* sp. Pop5 cultivation medium extract

HPLC peak (Fig. 5.2.1A)	Amino acid sequence*	Mass (av.)	Number of cycles	Structural features
3	AT C AR C DS S SR C GA S GK S SG S AS S IS T	2364.5	8	Phazolicin (PHZ)
4	AAT C AR C DS S SR C GA S GK S SG S AS S IS T	2435.5	8	Alternative leader peptide cleavage site
4	TAAT C AR C DS S SR C GA S GK S SG S AS S IS T	2536.6	8	Alternative leader peptide cleavage site
1	AT C AR C DS S SR C GA S GK S SG S AS S IS T	2384.5	7	Smaller number of cycles in the product
1	AT C AR C DS S SR C GA S GK S SG S AS S IS T	2404.5	6	Smaller number of cycles in the product
1	AT C AR C DS S SR C GA S GK S SG S AS S IS T	2424.5	5	Smaller number of cycles in the product
2	AAT C AR C DS S SR C GA S GK S SG S AS S IS T	2455.5	7	Alternative leader peptide cleavage site + smaller number of cycles
2	TAAT C AR C DS S SR C GA S GK S SG S AS S IS T	2556.6	7	Alternative leader peptide cleavage site + smaller number of cycles
1,2	AT C AR C DS S SR C GA S GK S SG S AS S	2083.4	7	C-terminal degradation of the peptide + alternative cleavage
1,2	AAT C AR C DS S SR C GA S GK S SG S AS S	2154.4	7	C-terminal degradation of the peptide + alternative cleavage

*residues converted into azole cycles are shown in bold red

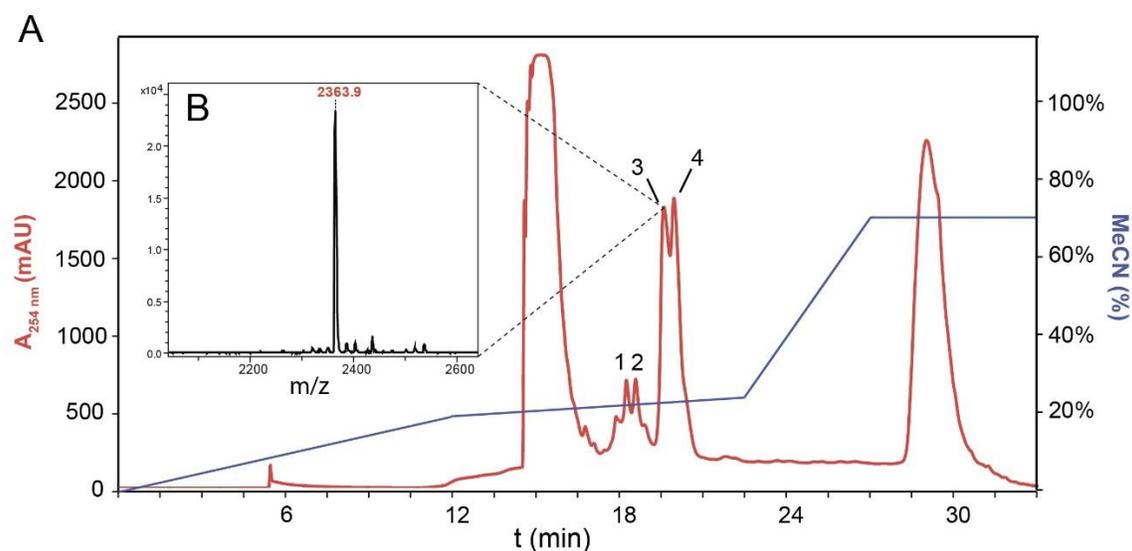


Figure 5.2.1 Purification of PHZ-related compounds (A) HPLC profile of the RM medium extract after *Rhizobium* sp. Pop5 cultivation. The absorbance curve ($A_{254 \text{ nm}}$, red trace) and MeCN-gradient profile (blue trace) are shown. HPLC peaks (1-4) were collected separately and analyzed using the MALDI MS. (B) MS spectrum of the HPLC peak 3. The mass peak with $m/z = 2363.9$ $[M+H]^+$ corresponding to PHZ is labeled.

ESI and MALDI-ToF-MS/MS fragmentation spectra (**Figs. 5.2.2A** and **5.2.2B**) were obtained for PHZ and its minor forms present in the extract. This allowed the location of theazole cycles installed in the core and confirmed that the amino acid sequence of the produced compound matches that of the peptide encoded by the *phzA* gene. In PHZ every third amino acid residue is converted into anazole. This includes all three cysteines transformed into thiazoles and five serines forming oxazoles. Interestingly, there are six more residues (five serines and one threonine), which could potentially form azoles but are left intact by the peptide modification complex. Such selectivity of theazole-installing machinery, when certain residues are omitted, is not unique for PHZ. In other LAPs, e.g., microcin B17 [187], klebsazolicin [76], and goadsporin [118] nucleophilic side chains of several Ser and Thr residues for some reason do not undergo the conversion into azoles. The mechanism of the modification enzyme's selectivity, however, remains unknown.

The diversity of the PHZ-related compounds identified in the cultivation medium extract likely results from the combination of three factors. First, there were forms with an alternative site of the leader peptide removal, which contained one or two additional residues at the N-terminus (**Table 5.1**). These included two major compounds otherwise identical to PHZ but extended N-terminally by one Ala or Thr-Ala pair of residues. Together with PHZ, these two forms further referred to as A-PHZ ($MH^+ = 2434.9$) and TA-PHZ ($MH^+ = 2536.0$) constituted the major portion of all the PHZ-related compounds produced. A gene coding for a specific leader protease is absent from the *phz* BGC, which means that PHZ biosynthesis relies on the activity of cellular protease(s) encoded

elsewhere in the genome for the removal of the leader. In the biosynthesis of other LAPs including microcin B17 and klebsazolicin, a conserved nonspecific cellular metalloprotease TldDE was shown to be responsible for the leader peptide cleavage [36]. The participation of a nonspecific protease in the process of PhzA leader removal is consistent with the presence of A-PHZ and TA-PHZ forms in the extract.

Second, we observed several minor forms lacking up to three C-terminally locatedazole cycles (**Table 5.1**) originating most likely from the peptides, whose modification was aborted at some point. The presence of these forms is consistent with the azole installation from the N- to C-terminus in the course of the PhzA precursor modification. Finally, we detected several minor forms lacking up to three C-terminal amino acids (**Table 5.1**), which presumably result from proteolysis by carboxypeptidases in the cytoplasm prior to the mature compound export. The proteolytic degradation of PHZ by aminopeptidases may include only the removal of the N-terminal Ala1 (Ala29 of the precursor), since the carbonyl of Thr2 is involved in the formation of the first thiazole cycle, which protects the peptide bond from the cleavage.

High-resolution MS¹ measurements for the three major detected compounds including PHZ, A-PHZ, and TA-PHZ were performed to confirm the proposed structures. The obtained values of 2362.8679, 2433.9052, and 2534.9524 are all within 1.5 ppm of the corresponding masses calculated based on empirical formulas (**Fig. 5.2.3**).

¹ High-resolution spectra and ESI-MS-MS spectra (see above) were obtained by Dr. Viktor Zgoda (Institute of Biomedical Chemistry, Moscow, Russia).

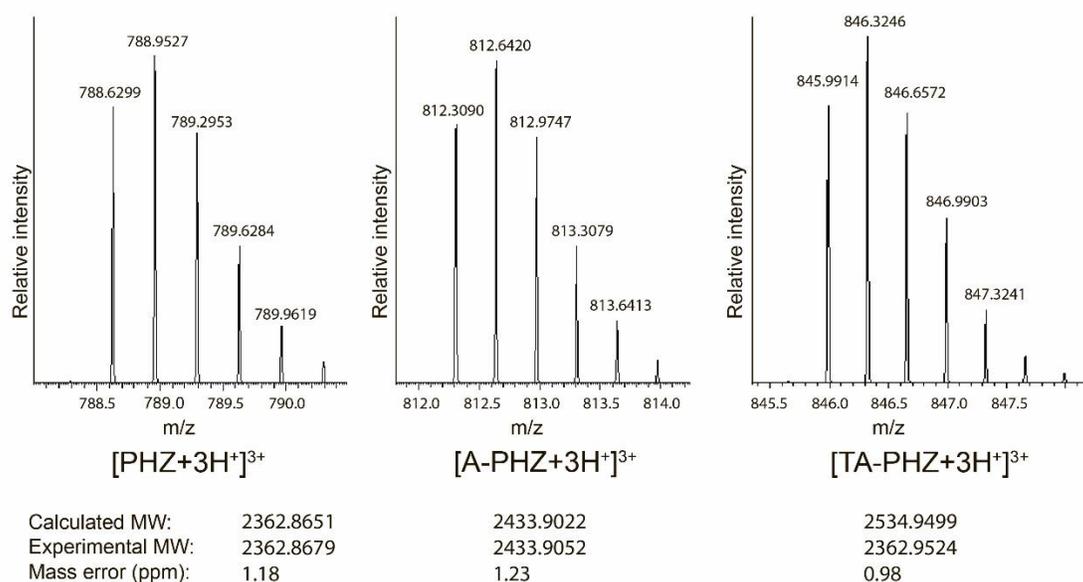


Figure 5.2.3 High-resolution MS spectra for PHZ, A-PHZ, and TA-PHZ The m/z values are indicated for each peak. The calculated mass, experimentally obtained mass, and mass error in ppm are shown for each compound below the spectra.

5.3 Phazolicin is a narrow-spectrum antibiotic affecting rhizobia

Once we identified PHZ and related compounds from the *Rhizobium* sp. Pop5 cultivation medium as LAPs originating from the PhzA precursor, we sought to determine whether PHZ and its forms have biological activity. First, we spotted PHZ in high concentration (5 mM) on top of a lawn of *E. coli* BW25113 grown on solid rich medium, which led to the appearance of a very small inhibition zone. Multiple previously characterized RiPPs exhibit narrow-spectrum antimicrobial activity against strains closely related to their producers mediating the competition for the niche between evolutionary close bacterial lineages [76], [188], [189]. With this in mind, we tested PHZ against a panel of strains of the order *Rhizobiales* (class Alphaproteobacteria) belonging to the genera

Rhizobium, *Mesorhizobium*, *Sinorhizobium*, *Azorhizobium*, and *Agrobacterium*. **Table 5.2** summarizes the data on the minimal inhibitory concentrations (MICs) of PHZ determined in the broth microdilution assay against the selected set of rhizobia. The MIC values were determined either in the *Rhizobium* medium (RM) or the YEB medium depending on which rich medium is optimal for the growth of a strain under study. As can be seen from **Table 5.2**, PHZ is highly active against the members of genera *Rhizobium*, *Azorhizobium*, and *Sinorhizobium* with MICs ranging between 0.125 and 4 μM for the tested representatives. Bacteria from the genus *Mesorhizobium* were resistant to the action of PHZ, while against both *Agrobacterium* strains tested PHZ displayed high MICs (64 mM).

Table 5.2 Phazolicin MIC values for various strains of rhizobia

Microorganism	Medium	MIC, μM	MIC, $\mu\text{g}\cdot\text{mL}^{-1}$
<i>Rhizobium leguminosarum</i> bv. <i>viciae</i> 3841	RM	1	2.4
<i>Rhizobium leguminosarum</i> bv. <i>phaseoli</i> 4292	RM	0.125	0.3
<i>Rhizobium leguminosarum</i> bv. <i>phaseoli</i> RCR 3622	RM	2	4.7
<i>Rhizobium etli</i> DSM 11541	RM	0.25	0.6
<i>Rhizobium tibeticum</i> DSM 21102	RM	1	2.4
<i>Agrobacterium tumefaciens</i> C58C1	YEB	64	151.3
<i>Agrobacterium rhizogenes</i> ARQUA1	YEB	64	151.3
<i>Mesorhizobium loti</i> MAFF303099	YEB	>128	>302.6
<i>Mesorhizobium thianshanense</i> HAMBI 3372	YEB	>128	>302.6
<i>Azorhizobium caulinodans</i> ORS571	YEB	2	4.7
<i>Sinorhizobium fredii</i> HH103	YEB	0.25	0.6
<i>Sinorhizobium medicae</i> WSM419	YEB	4	9.5
<i>Sinorhizobium meliloti</i> Sm1021	YEB	0.3	0.7

We also tested a small set of plant pathogenic, plant-associated, and soil microorganisms for their sensitivity to PHZ. The set included Gammaproteobacteria (*Erwinia amylovora*, *Pantoea ananatis* PA4, *Pseudomonas putida* KT2440, and *P. fluorescens* Pf-5), Firmicutes (*Bacillus subtilis* 168, *B. cereus* ATCC 4342), and Actinobacteria (*Arthrobacter* sp. ATCC21022, *Microtetraspora glauca* NRRL B-3735). However, no activity was observed. Thus, like other previously characterized LAPs, PHZ displays narrow-spectrum antimicrobial activity with the lowest MICs against the strains phylogenetically close to the producing strain *Rhizobium* sp. Pop5. Specific antirhizobial activity of PHZ is of particular interest, given that only several other compounds produced by rhizobia and acting exclusively against closely related strains are known (e.g., trifolitoxin (see section 2.1.6, [89]) and agrocin 84 [190]).

5.4 Assignment of the *phz* BGC with its product

The amino acid sequence of the PhzA precursor matching that of PHZ azole-modified peptide identified in the cultivation medium was the only link between PHZ and the *phz* BGC in the *Rhizobium* sp. Pop5 genome. We aimed to determine if the participation of other genes in the cluster is required for the production of the compound. Using plasmid insertion mutagenesis in the native PHZ-producing strain *Rhizobium* sp. Pop5 we constructed a derivative with inactivation of *phzD* (Fig. 5.4.1A and 5.4.1B), gene coding for the YcaO cyclodehydratase, a key enzyme required for the installation of azole cycles (see 2.1.4). This strain referred further to as *Rhizobium* sp. Pop5. Ω *phzD* (symbol Ω before

the name of a gene shows that the inactivation of the gene was achieved via plasmid insertion) was tested for the production of PHZ by MALDI-ToF-MS of the whole cells. No mass-peaks corresponding to PHZ or its longer forms A-PHZ and TA-PHZ were detected, while in the initial strain the corresponding peaks were readily identified (**Fig. 5.4.1C**). The expression of a plasmid-borne copy of the *phzD* gene under the control of the *lac* promoter (pSRK broad-host-range vector [191]) restored the PHZ-producing phenotype, while no production was detected in the Ω *phzD* cells harboring the empty vector or in the absence of IPTG in the medium (**Fig. 5.4.1C**). These results demonstrate that *phzD* gene is essential for the production of the mature PHZ in the native host. We tested the Ω *phzD* strain for the presence of antagonistic activity in a spot test over a lawn of the PHZ-sensitive strain *Rhizobium leguminosarum* bv. *phaseoli* 4292; wt *Rhizobium* sp. Pop5 served as a control. As can be seen in **Figure 5.4.1D**, no inhibition zone formation was observed around the colony of the Ω *phzD* derivative, while a clear zone was surrounding a colony of the original PHZ-producing strain. This experiment demonstrates that PHZ production is the only mechanism mediating the antagonistic activity of the Pop5 strain against PHZ-sensitive bacteria.

Due to the possible polar effect on the *phzD* gene located downstream, it was impossible to test, if the *phzC* and *phzB* gene products participate in the modification of the PhzA precursor peptide using insertional gene inactivation successfully performed for *phzD*. Thus, we set to reconstruct this process heterologously in the cells of *E. coli*. To achieve this goal a pET Duet-based vector with *phzA* cloned into one multiple cloning site

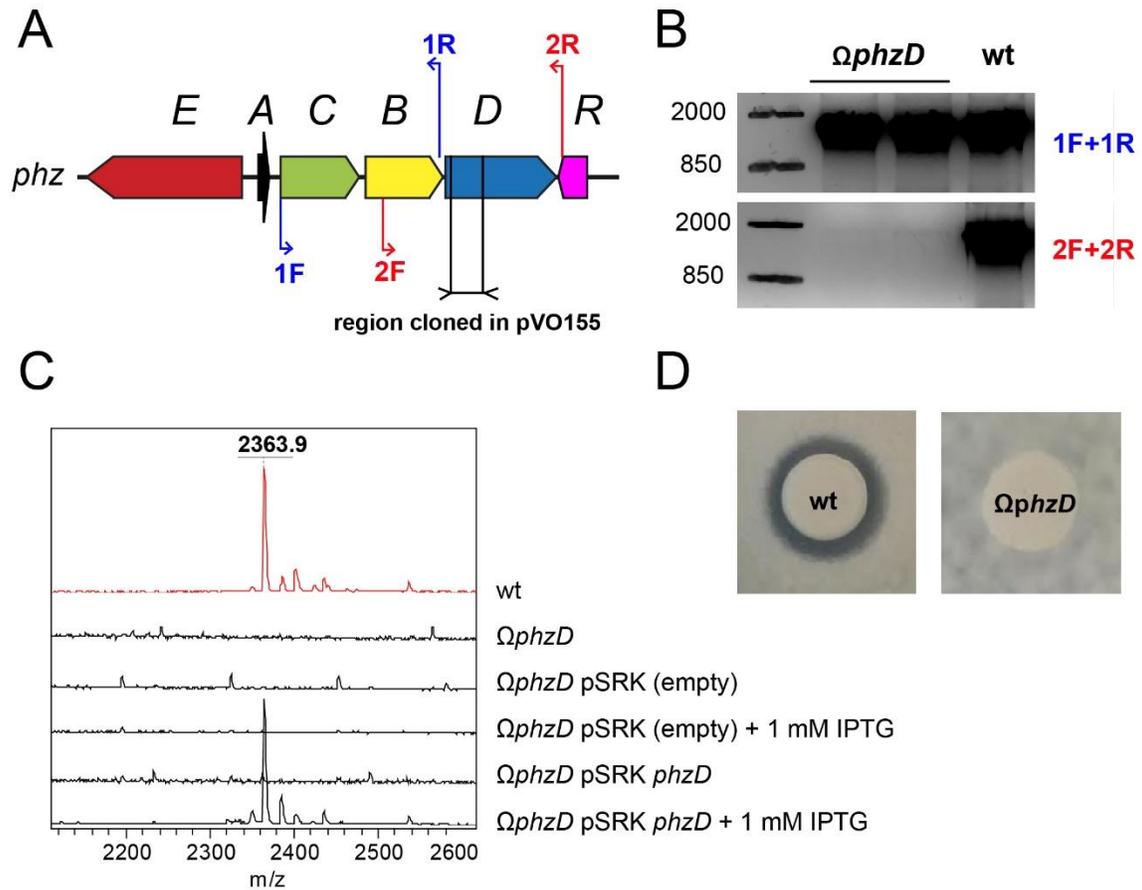


Figure 5.4.1 Construction and phenotype verification of *Rhizobium* sp. Pop5 Ω *phzD*
(A) PHZ BGC (*phzEACBDR*) in the genome of *Rhizobium* sp. Pop5. A region cloned into pVO155 for subsequent plasmid insertion mutagenesis is shown. The oligonucleotide primer pairs for genomic insertion verification are shown as pairs of blue and red arrows.
(B) DNA gel electrophoresis of the PCR products, amplified from the genomic DNA purified from two clones of Ω *phzD* mutant and wt *Rhizobium* sp. Pop5.
(C) Mass-spectra of whole cells show the absence of the prominent mass-peak of mature PHZ (2363.9 [M+H]⁺) in Ω *phzD* mutant compared to wt *Rhizobium* sp. Pop5. The induction of *phzD* gene expression from the pSRK plasmid by IPTG leads to the restoration of PHZ production by the Ω *phzD* mutant.
(D) *Rhizobium leguminosarum* 4292 growth inhibition zone visible around wt *Rhizobium* sp. Pop5 colony is not observed for Ω *phzD* derivative.

(MCS) and *phzCBD* genes into the other MCS was constructed. The DNA sequence coding for the N-terminal 8xHis-tag was fused with the *phzA* gene to allow subsequent purification of the modified peptides. We were unable to enrich the modified peptides using IMAC from the lysates of *E. coli* Rosetta (DE3) cells expressing the obtained vector. Moreover, we observed that its expression was toxic for the cells. Using MALDI-ToF-MS of whole cells grown on the solid medium in the presence of IPTG inducer we detected a pair of peaks corresponding to mature PHZ and A-PHZ, which were absent in the mass spectra of the cells grown in the absence of IPTG and in the empty plasmid control (**Figure 5.4.2**). Thus, the presence of the genes coding for the PhzA precursor and the three components of the modification enzymatic complex (PhzC, PhzB, and PhzD) is sufficient for the production of the fully modified PHZ. Interestingly, we observed the cleavage of the leader peptide in *E. coli* at positions matching those in *Rhizobium* sp. Pop5. This may indicate that the leader peptide removal is performed either by a protease conserved across Proteobacteria, or by nonrelated peptidase(s), which cleave off the N-terminal part of the precursor until they encounter the core part protected from the proteolytic degradation by installedazole cycles.

The *phzE* gene is predicted to encode an ABC exporter, which is a common feature across BGCs of known LAPs. Such transmembrane pumps not only enable the export of the mature product but also provide self-resistance to the antibiotic produced. Some LAP BGCs encode additional specific immunity proteins (e.g., gyrase rescue protein McbG in the BGC of microcin B17 [192]), however, in many cases, the active efflux of the toxic

compound by the transmembrane pump is sufficient to keep its intracellular level below the toxic concentration. To test, if the PhzE exporter provides immunity to PHZ, we cloned *phzE* gene into the pSRK vector under the control of the *lac* promoter and transformed the obtained vector into *Sinorhizobium meliloti* Sm1021, a strain sensitive to PHZ (see above). The resulting strain grew well in the medium supplemented with 128 μ M PHZ in the presence of IPTG. Thus, the MIC for this strain is at least two orders of magnitude higher than that for the parental *wt* Sm1021, which supports the role of PhzE as a major determinant of self-immunity to PHZ in the producing strain.

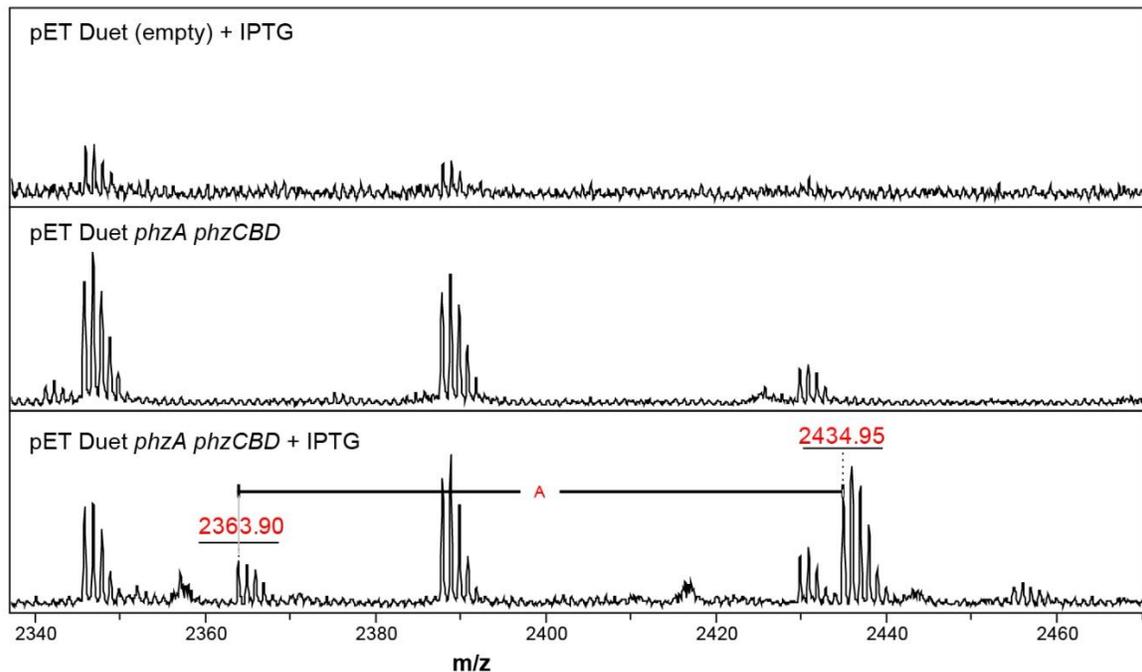


Figure 5.4.2 PHZ production in *E. coli* Rosetta (DE3) heterologous host. MALDI-ToF mass spectra of the whole cells transformed with denoted vectors. For *E. coli* harboring pET Duet *phzA phzCBD* vector the spectra for the cells growing on the medium with and without IPTG are shown. The mass peaks corresponding to PHZ (2363.90 MH^+) and A-PHZ (2434.95 MH^+) are labeled.

To sum up, experiments with the native producer and the *E. coli* heterologous host showed that PHZ is a product of posttranslational modification of the PhzA precursor peptide by the proteins encoded by *phzCBD* genes. PhzE provides self-immunity enabling specific active export of the mature compound. The protease(s) responsible for the leader cleavage remain unknown, most likely these are nonspecific cellular proteases conserved across Proteobacteria.

5.5 Phazolicin inhibits bacterial translation²

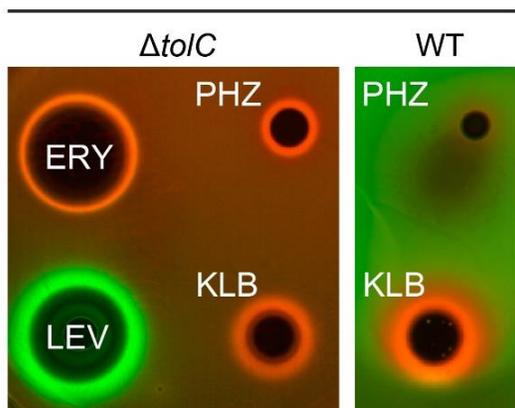
As was already reviewed above (see section 2.1.5), known LAPs can have diverse mechanisms of action. To identify the target of PHZ we tested its activity using a previously developed *E. coli*-based *in vivo* reporter system, which allows identifying the compounds affecting two major intracellular processes – translation and DNA replication [153]. Both processes are the targets of previously described LAPs klebsazolicin [76] and microcin B17 [107] respectively. In this *in vivo* system, a plasmid pDualRep2 encodes two fluorescent proteins (RFP and Katushka2S). The expression of the *rfp* gene is driven by the SOS-response inducible *sulA* promoter, which is activated once SOS-inducing agents (e.g., fluoroquinolones affecting the DNA gyrase) are present in the medium in sublethal concentrations. The presence of translation inhibitors stalling the ribosome (e.g., macrolides or tetracycline) leads to the expression of Katushka2S gene [153]. Thus, when

² *In vivo* and *in vitro* experiments with purified PHZ described in this chapter were performed by Dr. Ilya Osterman (Lomonosov Moscow State University / Skoltech, Moscow, Russia).

a drop of solution with a compound with an unknown mechanism of action is spotted over a lawn of a pDualRep2-harboring strain, the fluorescence surrounding the growth inhibition zone indicates which of the two mechanisms (if any) underlies the compound activity. A 10 mM PHZ solution was tested against the two *E. coli* reporter strains harboring pDualRep2 (BW25113 and BW25113 $\Delta tolC$). The latter strain lacks the gene coding for TolC – the major outer membrane multidrug efflux protein expelling chemically diverse compounds including multiple antibiotics from the cell [193]. As can be seen in **Figure 5.5.1A**, similarly to previously studied ribosome-targeting LAP klebsazolicin, PHZ induces production of Katushka2S fluorescent protein, which indicates that the antibiotic leads to ribosome stalling. Interestingly, unlike KLB, PHZ caused larger inhibition zones once spotted over BW25113 $\Delta tolC$ compared to the *wt*, which may indicate that the compound is a subject to the TolC-mediated export in *E. coli*.

To prove that protein synthesis is the target of PHZ, we tested the effect of PHZ on translation of luciferase mRNA in S30 *E. coli* lysate. Luciferase activity was decreased with the increase in the concentration of PHZ in the reaction (**Figure 5.5.1B**). The concentration of PHZ matching previously identified MICs for sensitive rhizobia (around 1 μ M) decreased the level of the reporter activity approximately three-fold (**Figure 5.5.1B**, blue curve). In a separate *in vitro* translation experiment, we compared the activity of PHZ and its naturally occurring forms with an alternative leader cleavage site (A-PHZ and TA-PHZ). The mixture of the two longer forms was purified as separate HPLC fraction. As can be seen in **Figure 5.5.2** protein synthesis inhibition by A/TA-PHZ mixture was

A *E. coli* BW25113 pDualRep2



B

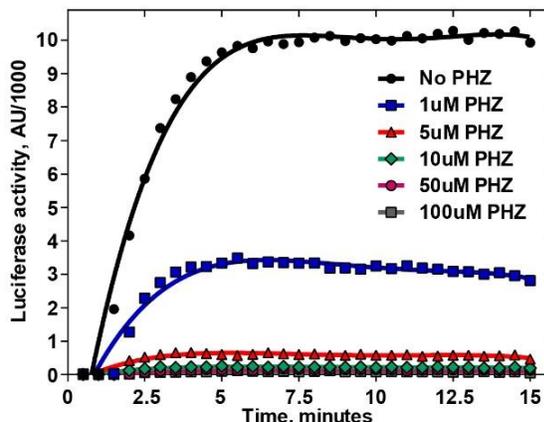
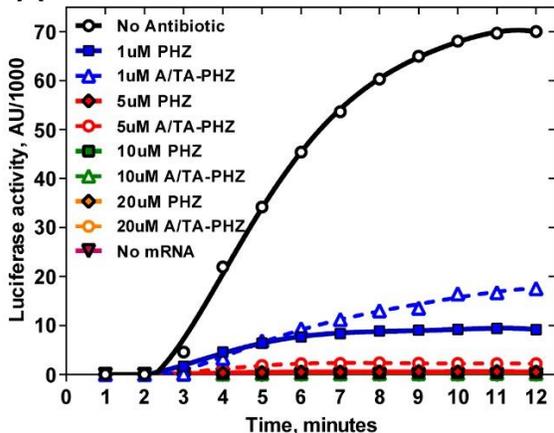


Figure 5.5.1 PHZ inhibits bacterial translation *in vivo* and *in vitro*. (A) Testing of PHZ activity against the reporter strains *E. coli* BW25113 pDualRep2 and *E. coli* BW25113 $\Delta tolC$ pDualRep2. Red (pseudocolor is shown, Katushka2S signal) halo around the growth inhibition zones is visible for PHZ as well as ERY (erythromycin) and KLB (klebsazolicin) serving as positive controls. Green (pseudocolor, RFP signal) halo surrounds the LEV (levofloxacin) inhibition zone. (B) Kinetic curves showing the concentration-dependent inhibition of *in vitro* translation of the luciferase mRNA in S30 *E. coli* lysate upon the addition of various concentrations of PHZ. AU – arbitrary units.

A



B

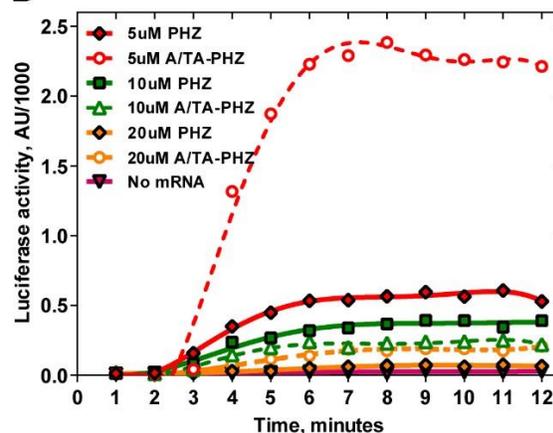


Figure 5.5.2 *In vitro* translation inhibition activity of PHZ and its naturally occurring longer forms A-PHZ and TA-PHZ). Graphs corresponding to PHZ are shown as solid lines. Dashed lines correspond to the mixture of A/TA-PHZ, (HPLC peak 4, Fig. 5.2.1A). (B) is a close-up view of (A) in the upper range concentrations (5-20 μ M).

comparable to that of PHZ. The result suggests that the N-terminus of the peptide does not contribute significantly to the binding with the antibiotic target, as it can be extended by at least two residues without a change in inhibitory activity.

5.6 Phazolicin targets the nascent peptide exit tunnel of the bacterial ribosome

Bacterial translation is a complex process, which involves the coordinated action of tens of macromolecules including the ribosome, a set of translation factors, tRNAs, and aminoacyl-tRNA synthetases. Known RiPPs can target different steps of translation including those, which do not involve the ribosome. For instance, microcin C is a specific inhibitor of aspartyl-tRNA synthetase [194], while thiopeptide GE2270A prevents the binding of EF-Tu to the aminoacylated tRNAs [195]. By analogy with KLB, whose BGC served as a starting point for identification of PHZ, we expected that the ribosome itself is the target of PHZ. To get insights into the mechanism of PHZ action we set to crystallize the ribosome of *Thermus thermophilus* in a complex with PHZ, mRNA, and tRNAs³. This approach was previously applied to identify the ribosome-binding sites of multiple antibiotics, including that of KLB [76]. However, our multiple crystallization attempts were unsuccessful and we could not observe electron density that could be attributed to PHZ anywhere on the *Th* ribosome. The *in vitro* translation inhibition assays performed with either *E. coli* (*Eco*) 70S ribosome or with a hybrid ribosome, which contained the 30S

³ Ribosome crystallization trials and *in vitro* translation with the *Th* ribosome were performed by Dr. Nelli Khabibullina and Prof. Yury Polikanov (University of Illinois at Chicago, Chicago, USA).

subunit of *E. coli* and 50S subunit from *T. thermophilus* demonstrated the absence of inhibition by PHZ when the hybrid ribosome was used (**Figure 5.6.1**). This result not only showed the unexpected species-specificity of PHZ action on such a highly conserved target as the ribosome but also allowed us to localize the PHZ binding site at the large ribosomal subunit.

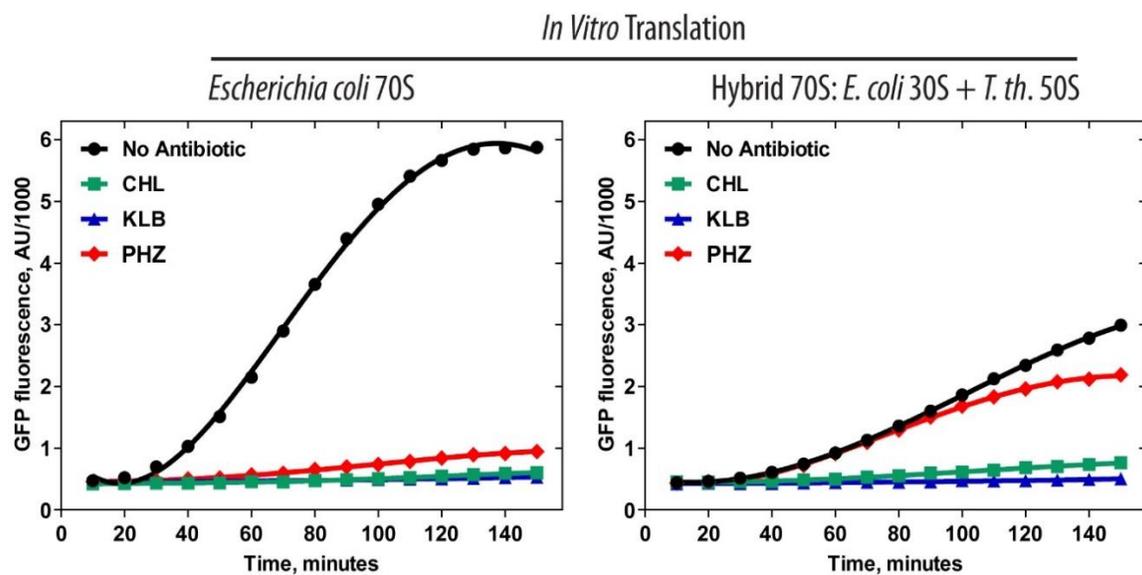


Figure 5.6.1 Effects of chloramphenicol (CHL), klebsazolicin (KLB), and phazolicin (PHZ) on *in vitro* translation of superfolder green fluorescent protein (sfGFP). PURExpress system supplied with either the *E. coli* 70S WT ribosomes or hybrid ribosomes (*E. coli* 30S + *T. thermophilus* 50S) were used. CHL and KLB served as positive controls. All antibiotics were added to the final concentration of 50 μ M. Note, that the overall rate of protein synthesis by the hybrid ribosome is approximately two-fold lower than that of the WT 70S (shown by the black curves). AU – arbitrary units.

Since we were unable to obtain the crystal structure of the *Tth* ribosome with bound PHZ, we tried to get insights into the exact mechanism of PHZ binding to the *Eco* ribosome using an alternative structural approach – cryogenic electron microscopy (cryo-EM). We obtained the charge density map⁴, characterized by the overall resolution of 2.87 Å according to the “gold-standard” Fourier shell correlation (FSC) method (**Fig. 5.6.2A**). It revealed PHZ bound to the ribosome in the nascent peptide exit channel (NPET). The PHZ binding site had high resolution and excellent quality of the map (**Fig. 5.6.2B**). This enabled us to fit PHZ atomic model (residues 2–23) and confirm the positions of seven out of eight azoles in the modified peptide, which were proposed previously based on the MS-MS

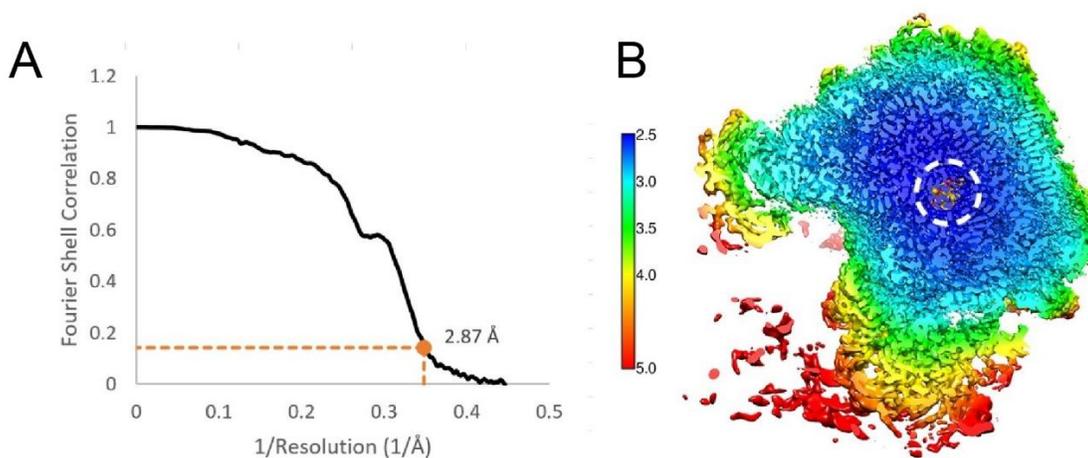


Figure 5.6.2 Fourier shell correlation curve for the cryo-EM structure of *E. coli* ribosome with PHZ bound. (A) Based on the “gold-standard FSC” cutoff value 0.143, the overall resolution for the large ribosomal subunit in the 70S ribosome is 2.87 Å. **(B)** Slice through the 50S subunit showing local resolution limits. The region of PHZ binding (dashed white circle) has the highest local resolution. The color scale bar on the left is in Ångstroms. 30S ribosomal subunit is not shown and was not included in the final model.

⁴ Cryo-EM data for the *Eco* ribosome with PHZ was collected and analyzed by Dr. Zoe Watson, Dr. Fred Ward, and Prof. Jamie Cate (University of California, Berkeley, USA). PHZ atomic model fitting and visualizations were made by Prof. Yury Polikanov (University of Illinois at Chicago, Chicago, USA).

analysis (**Fig. 5.6.3A**). While tandem mass spectrometry results indicated the presence of another oxazole cycle at position 24, in our model the last built residue Ser23 was left unmodified since Ser24 was not visible in the map due to its poor quality in that region (**Fig. 5.6.3B**).

In the structure, the linear modified PHZ peptide forms a compact globule binding in the upper part of the NPET of the 50S ribosomal subunit where it extensively interacts with the 23S rRNA and ribosomal proteins uL22 and uL4 (**Fig. 5.6.3C and D**). There are two major types of interactions between the PHZ and the nucleotides of the 23S rRNA. First, three azoles (Thz3, Oxz15, and Oxz18) are involved in π - π stacking with nucleotides A751, C2611, and U2609, respectively (**Fig. 5.6.3E and F**). Second, positively charged side chains of Arg5 and Arg11, as well as hydrophilic side chains of Asp7 and Ser8 form additional stabilizing hydrogen bonds with nucleobases and backbone phosphate groups of the 23S RNA (**Fig. 5.6.3E and F**). A notable feature of the PHZ molecule in its ribosome-bound state is the formation of complex intramolecular interactions that include both displaced face-to-face and edge-to-face π - π stacking of Thz12, Oxz21, Thz6, and Oxz18, along with the nucleobase U2609 from the 23S rRNA (**Fig. 5.6.3F**).

The binding site of PHZ in the ribosome NPET partially overlaps with that of klebsazolicin (KLB) – a previously characterized ribosome targeting LAP (see **2.1.5**). Both linear modified peptides form a globule once bound to the ribosome, and, in both cases, some of the azole cycles stack with rRNA nucleobases. However, the interactions PHZ and KLB make with 23S rRNA differ. KLB lacks the intramolecular π - π stacking system and

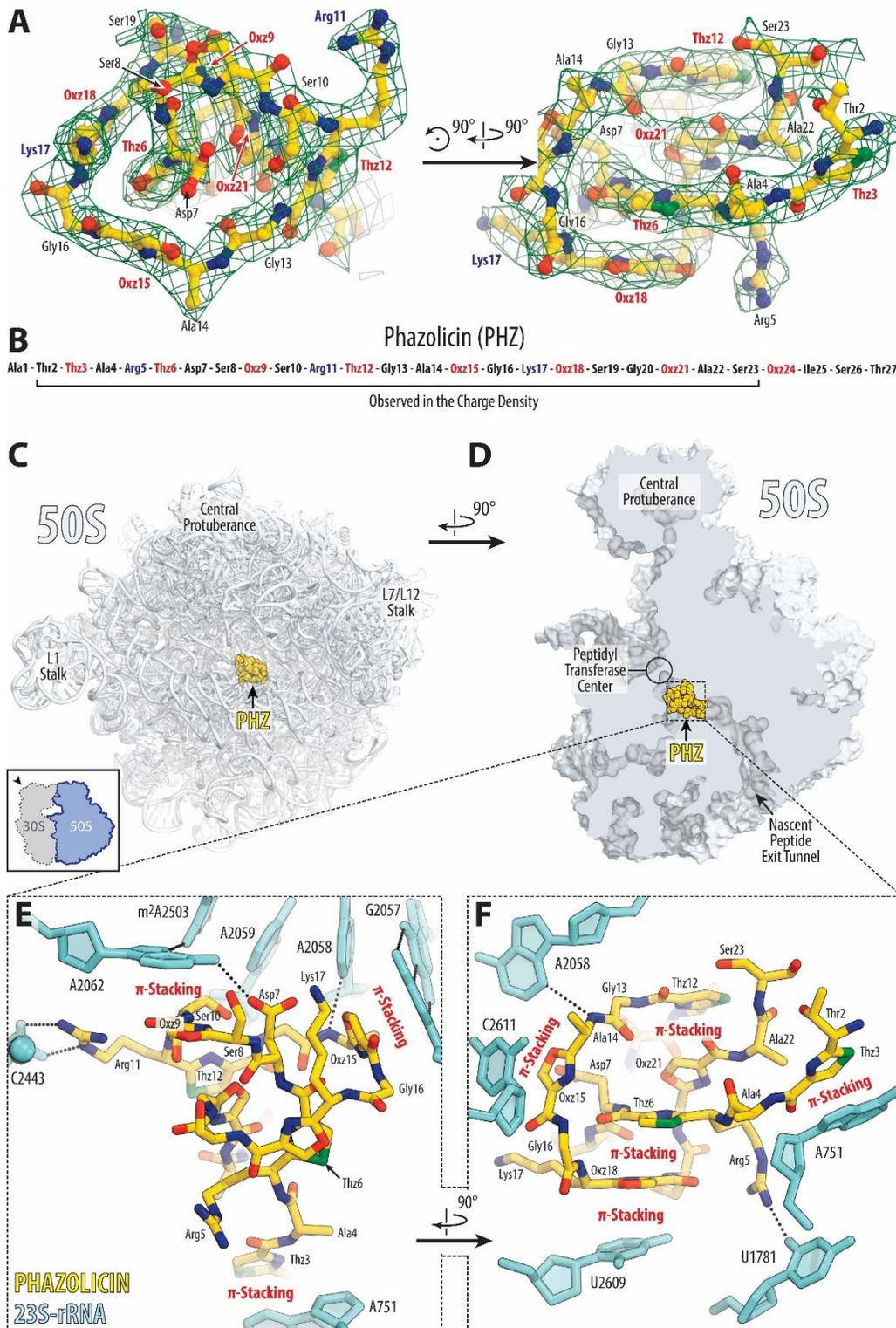


Figure 5.6.3 The structure of the bacterial ribosome in complex with PHZ. (A) Cryo-EM map of PHZ in complex with the *E. coli* 70S ribosome (green mesh) contoured at 2.5σ . The fitted model of the compound is displayed in its respective charge density viewed from two different perspectives. Carbon atoms are yellow, nitrogen atoms are blue, oxygen atoms are red, and sulfur atoms are green. (B) Schematic diagram showing the part of the ribosome-bound PHZ visible in the cryo-EM map. (C), (D) PHZ-binding site (yellow) on the *E. coli* large ribosomal subunit viewed from two different perspectives. In (C), the 50S subunit is viewed from the inter-subunit interface (30S subunit is removed for clarity) as indicated by the inset. The view in (D) is from the cytoplasm onto the A site. (E), (F) Close-up views of the PHZ-binding site in the ribosome exit tunnel with the contacts between PHZ and the 23S rRNA indicated. *E. coli* numbering of the nucleotides is used.

its N-terminal amidine ring (strictly required for the bioactivity of the compound) establishes the interactions through hydrogen bonds with nucleobases U2584 and U2585 of the peptidyl transferase center (PTC) [76]. The unmodified N-terminus of PHZ is oriented towards the outer part of the NPET and does not play an important role in the binding to the ribosome. This rationalizes the previous observation, that naturally occurring longer PHZ forms (A-PHZ and TA-PHZ) have inhibitory activities comparable to that of the major compound (see section 5.5).

A comparison of the structures of ribosome-bound PHZ and KLB reveals that PHZ binds further away from ribosome PTC than KLB. Superimposition of our structure of the ribosome-bound PHZ (**Fig. 5.6.4A**) or the previously published KLB (**Fig. 5.6.4B**) with the cryo-EM structure of the 70S *Eco* ribosome containing the ErmBL nascent peptide chain connected to the P-site tRNA [196] showed that there is more space between the ribosome-bound compound and the PTC in the case of PHZ compared to KLB. This allows suggesting that PHZ should leave space for a few more amino acids to be incorporated into the nascent polypeptide chain before it encounters the NPET-bound PHZ molecule, which

stalls the translation. The upper part of the ribosome NPET serves as a binding site for several classes of antimicrobials including macrolides (e.g. erythromycin) [197] and streptogramins B (e.g. quinupristin) [198]. Compared to these antibiotics of smaller molecular weight, which only partially occlude the NPET (**Fig. 5.6.4E**), both KLB and PHZ nearly completely obstruct the ribosome exit tunnel (**Fig. 5.6.4D and F**).

5.7 Amino acid sequence of the uL4 protein loop determines the specificity of PHZ action

Our cryo-EM structure of the ribosome-bound PHZ showed that its binding site is located next to the NPET constriction site – the narrowest part of the tunnel formed by the loops of the ribosomal proteins uL4 and uL22, which protrude into the core of the large ribosomal subunit formed by the 23S rRNA. The N-terminal and C-terminal parts of PHZ are located between the loops of these two proteins and occupy more space compared to the KLB C-terminal part (**Fig. 5.7.1A and B**). The amino acid sequences of the uL22 and uL4 proteins forming the constriction vary between species. We superimposed our *Eco* 70S ribosome structure with bound PHZ with the structure of the *Tth* 70S ribosome. The comparison revealed that residue His69^{Tth} of the *Tth* protein uL4 clashes sterically with the PHZ molecule (**Fig. 5.7.1C**). The corresponding residue in the uL4 of *Eco* ribosome is represented by a smaller Gly64^{Eco} (**Fig. 5.7.1E**). Another clash is observed for Arg90^{Tth} of the uL22 protein, while there is no clash in the structure with the *Eco* ribosome featuring Lys90^{Eco} at this position (**Fig. 5.7.1C**). We hypothesized that these differences in the

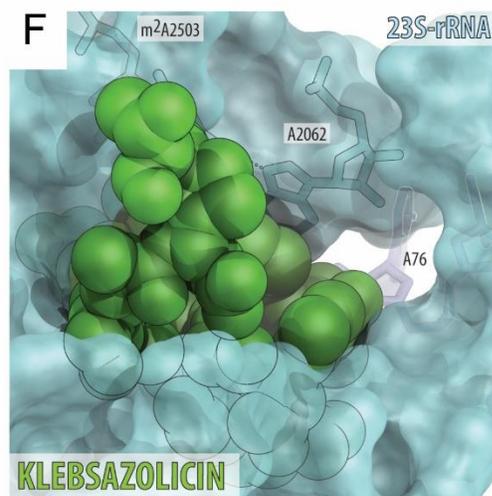
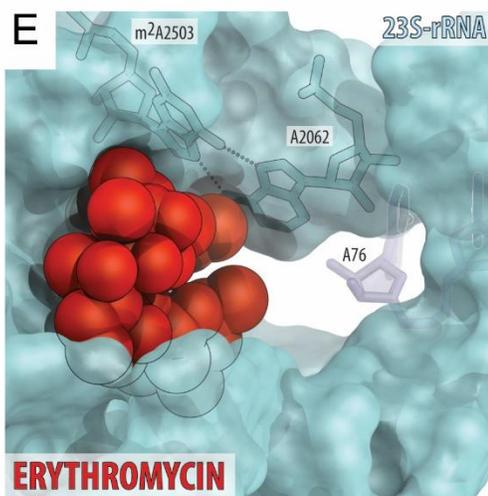
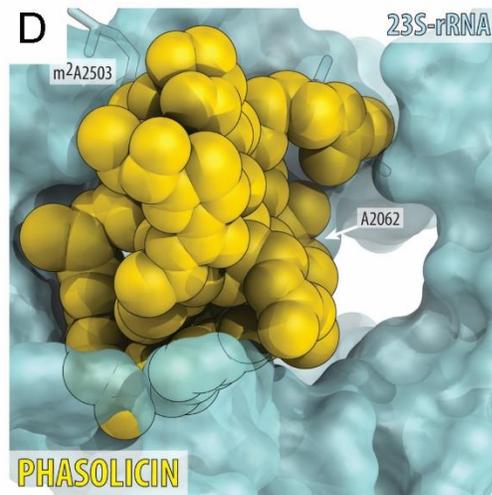
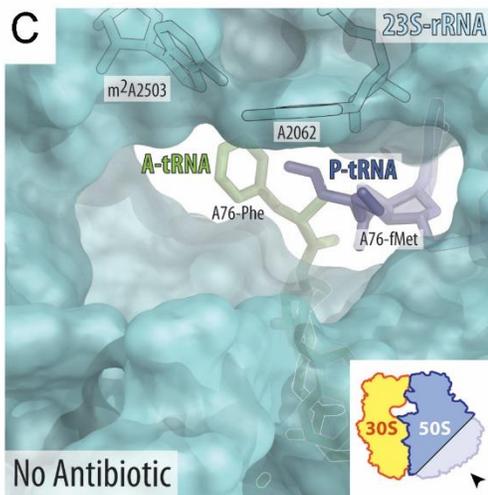
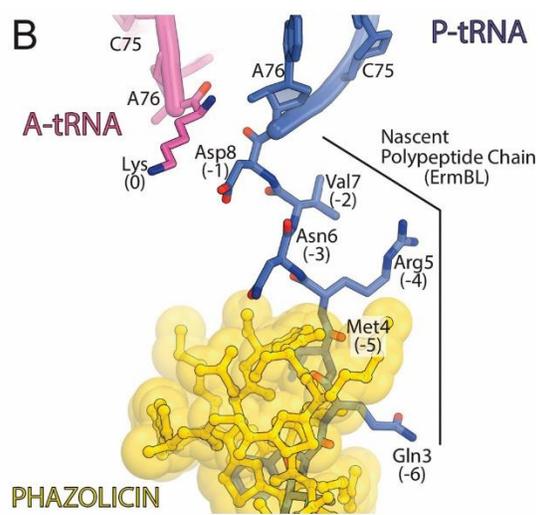
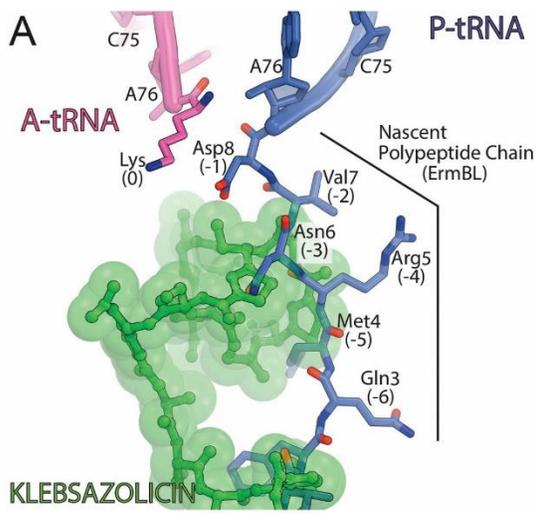


Figure 5.6.4 Comparison of the PHZ-bound ribosome structure with those of other NPET targeting antibiotics. (A), (B) *In silico* modeling of the nascent polypeptide chain in the NPET in the presence of klebsazolicin (A, PDB: 5W4K8) or phazolicin (B) using the cryo-EM structure of the 70S ribosome with ErmBL peptide bound to the P-site tRNA (PDB: 5JTE31). (C)-(F) Occlusion of the nascent peptide exit tunnel by antibiotics. (C) NPET of the drug-free 70S ribosome (PDB: 4Y4P1). The view is from the wide-open part of the tunnel onto the PTC (see the inset). (D, E, F) Occlusion of the nascent peptide exit tunnel by ERY (E), PHZ (D), and KLB (F). Structures of ERY and KLB are from PDB entries 6ND62 and 5W4K3, respectively. Unlike ERY and similar to KLB, PHZ almost completely occludes the lumen of the peptide exit tunnel.

uL4 and uL22 loops can influence the binding of PHZ to the ribosome, and hence explain the previous observations on the species-specificity of the PHZ action, specifically, its inability to inhibit the *Tth* ribosome. To test this hypothesis, we constructed pSRK-based vectors harboring *S. meliloti* genes *rplD* or *rplV* coding for the proteins uL4 and uL22, respectively, as well as their mutant variants, whose products have single amino acid substitutions of residues forming the NPET constriction (*rplD*^{K65A}, *rplD*^{G68H}, and *rplV*^{K90R}). These substitutions aimed to mimic the sequence of the corresponding loop region found in the PHZ-resistant *Tth* ribosome. As can be seen in **Figure 5.7.1F** episomal overexpression of the uL4^{G68H} variant in *S. meliloti* Sm1021 confers partial resistance to PHZ and allows the growth on the medium supplemented with 8 μM of the antibiotic. No growth was observed at these conditions for a strain, in which the *rplD* *wt* variant was expressed. No such effects were observed for the *rplV*^{K90R} gene variant. Thus, the difference in the size of the side chains between Lys90 and Arg90 is not that significant compared to that for His69 and Gly64. Moreover, unlike His69^{Tth}, which is located in a confined pocket and cannot easily re-adjust its position in response to PHZ binding, residues Arg90^{Tth} and Lys90^{Eco} have more space around and should be able to readjust their

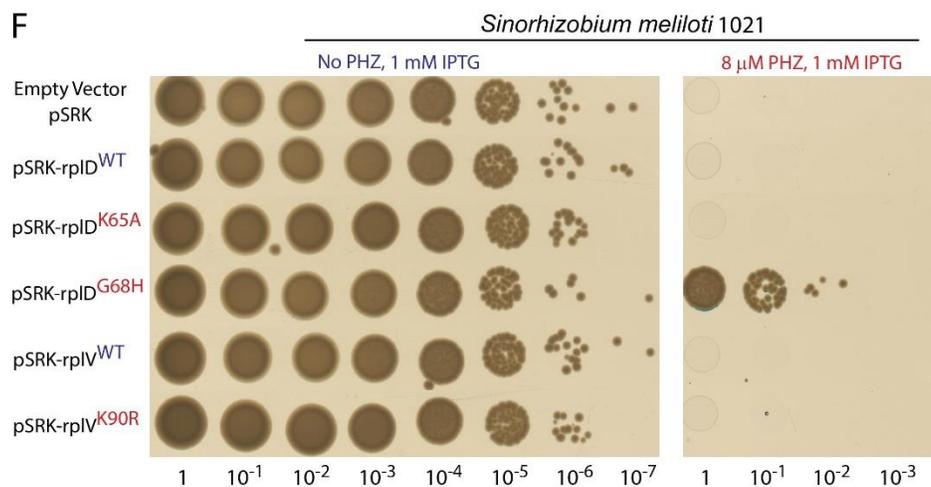
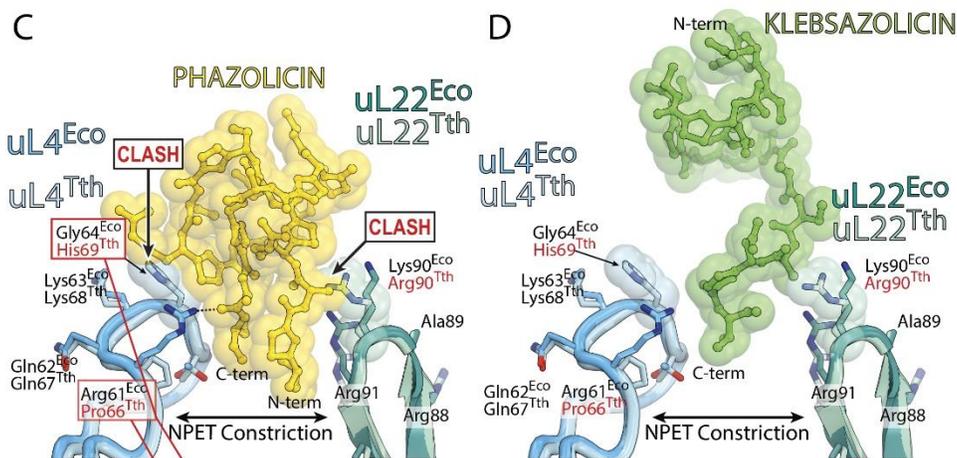
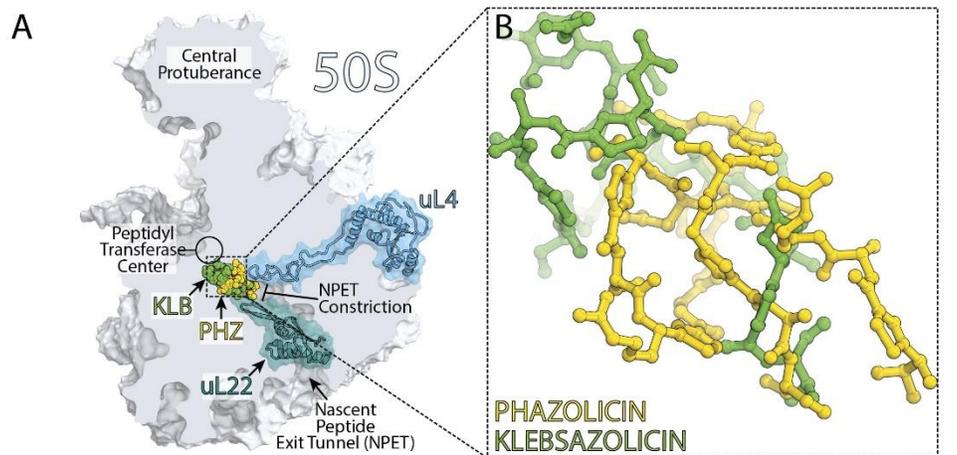


Figure 5.7.1 Amino acid sequence of the uL4 ribosomal protein loop determines the species-specific mode of PHZ binding to the ribosome (A) Comparison of PHZ (yellow) and KLB (green) binding sites in the NPET. The 50S subunit is shown in light blue. Ribosomal proteins uL4 and uL22 are highlighted in blue and teal, respectively. (B) Superposition of the ribosome structures with bound KLB and PHZ based on the alignment of the 23S rRNAs. Ribosome parts are not shown for clarity. (C), (D) Close-up view of PHZ (C) and KLB (D) interactions with the loops of proteins uL4 and uL22 forming the NPET constriction site. Superposition of the loops of proteins from *E. coli* and *T. thermophilus* is shown. (E) Multiple sequence alignment of the homologous parts of the ribosomal protein uL4 from three bacterial species. Red boxes highlight the amino acid residues of uL4, which either clash with the PHZ (His69Th) or interact with it in our structure (Arg61^{Eco}). (F) Growth of *S. meliloti* Sm1021 and its derivatives expressing plasmid-borne copies of *wt* or mutant *rplD* or *rplV* genes on solid medium without PHZ (left) or with 8 μ M PHZ (right). Only overexpression of *Th*-like version of uL4 protein confers resistance to PHZ in *S. meliloti*.

positions in the presence of PHZ to avoid a possible clash. Taken together, these results allow us to conclude that the amino acid composition of at least the uL4 loop region is a determinant of the PHZ specificity of action. To our knowledge, PHZ is the first example of a species-specific ribosome-targeting inhibitor with antimicrobial activity, whose action depends not only on its ability to get inside the cell but also on the fine structure of its target, the bacterial 70S ribosome.

Chapter summary

PHZ is a linearazole-modified peptide produced by *Rhizobium* sp. Pop5. It is a 27 amino acid long peptide with eightazole cycles installed in the structure. PHZ biosynthesis is guided by the *phzEACBDR* biosynthetic gene cluster and requires the action of a yet unidentified protease for the removal of the leader peptide. PHZ is a narrow-spectrum antibiotic active against the strains of rhizobia closely related to the PHZ-producing strain. PHZ inhibits bacterial translation by the obstruction of the nascent peptide exit tunnel of the large ribosomal subunit. In the ribosome exit tunnel, PHZ adopts a globular fold, which is stabilized by the system of intramolecular interactions. Azole posttranslational modifications play a key role in both the intramolecular interactions and PHZ interactions with the 23S rRNA. The amino acid sequence of the uL4 ribosomal protein determines the species-specificity of PHZ action since certain bulky residues in the loop region of this protein prevent the binding of PHZ to the ribosome.

Chapter 6. Phazolicin import into susceptible cells

6.1 Two transporter systems are inactivated in PHZ resistant *Sinorhizobium meliloti* mutants

The experiments covered in the previous chapter allowed us to identify the bacterial ribosome as the target of PHZ. Antimicrobial peptides acting intracellularly have to pass one and two membranes of the cell envelope (of Gram-positive and Gram-negative bacteria, respectively) to get inside the susceptible cell. Given the relatively large molecular weight and polarity of antimicrobial peptides, which prevents their passage through lipid bilayers, such molecules often rely on the activity of peptide transporters for internalization. With this in mind, we aimed to identify the transmembrane transporters involved in the import of PHZ into the cells of the model strain *S. meliloti* Sm1021, which was shown to be susceptible to PHZ (see section 5.3).

For previously described LAP microcin B17 the spontaneous mutations in the gene of the inner membrane importer were readily identified using the screening on the antibiotic-containing medium [199]. We tried to select PHZ-resistant Sm1021 spontaneous mutants by plating the overnight culture aliquots on solid medium containing 20 μ M PHZ. However, our multiple attempts were unsuccessful; no colonies were recovered. Since a direct approach using *wt* strain did not provide any result, we used a Mariner Himar C9 transposon library constructed in *Sinorhizobium meliloti* Sm1021 in the same screening

setup⁵. This approach allowed us to isolate four resistant colonies growing on the PHZ-containing medium (referred further as mut #1 through mut #4).

Illumina whole-genome sequencing⁶ of the obtained mutants' gDNA followed by the BLAST search using the sequence of the transposon identified the locations of the insertions in their genomes. In three obtained mutants (mut #1 – mut #3) the transposon insertions were mapped in the *yejA* gene, while in mut #4 it was located in the *bacA* gene (**Fig. 6.1.1A**). Chromosomally encoded *yejA* (SMc02829) is a part of the *yejABEF* operon coding for a putative ABC transporter, *yejA* product is predicted to be its periplasmic substrate-binding protein (SBP) [200]. *bacA* (SMb20999) located on the pSymB 1.68 Mbp megaplasmid encodes a promiscuous SbmA-like peptide transporter (SLiPT) of the inner membrane [201]. Interestingly, when we examined the contigs containing the *bacA* gene for mutants #1 - #3, we found that in all three strains there were mutations caused either by insertion or deletion of a single nucleotide, which led to the frameshift in the *bacA* gene rendering its product nonfunctional. In the genome assembly of the mutant #4 the parts of the *yejABEF* operon were found in two different contigs, which contained short sequences of the transposon ends, as if it was inserted in the *yejE* gene (nucleotide position 517). We concluded that in mutant #4 two sequential transposition events could have occurred. In any case, the *yejE* gene ORF was disrupted in this mutant. Thus, all four selected PHZ-

⁵ Transposon library construction was performed by Peter Mergaert and Joy Lachat (Institute for Integrative Biology of the Cell, Gif-sur-Yvette, France).

⁶ NGS library preparation and sequencing were performed at Skoltech Sequencing Core Facilities. Genome assembly and annotation was performed by Dmitry Sutormin (Skoltech).

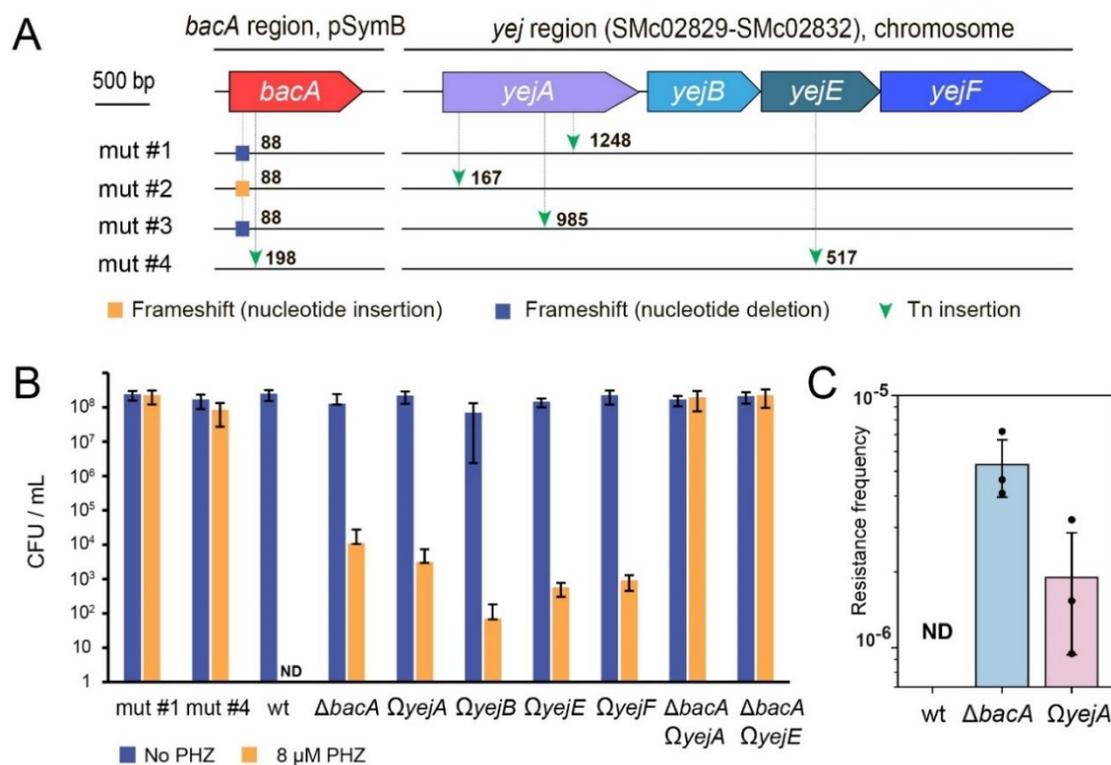


Figure 6.1.1 BacA and YejABEF transporters independently contribute to the PHZ internalization into the cells of *Sinorhizobium meliloti* (A) Mutations identified by whole-genome sequencing of four PHZ-resistant mutants of Sm1021. The numbers indicate the nucleotide position in the gene, where the mutation was located. (B) CFU of Sm1021 growing on the medium with PHZ (8 μ M) and without PHZ. The results shown represent the average of three biological replicates for two obtained PHZ-resistant mutants, as well as genetically constructed single and double mutants in *bacA* and *yejABEF* genes. ND – not detected. (C) Frequency of PHZ resistance acquisition in *wt* Sm1021, *bacA* and *yejA* single gene mutants. The data for three biological replicates are shown. ND – not detected.

resistant lineages turned out to be double mutants with both the BacA and the YejABEF inner membrane uptake systems inactivated. This result explains why we were unable to select the mutants using the *wt* strain in our initial setup since the simultaneous inactivation of the two unrelated transporters is required for the resistance acquisition.

6.2 BacA and YejABEF transporters contribute to the uptake of PHZ

To provide further evidence of the identified transporters' participation in the PHZ uptake we compared the PHZ sensitivity of the previously constructed *S. meliloti* mutants with either BacA ($\Delta bacA$) or YejABEF ($\Omega yejA$, $\Omega yejB$, $\Omega yejE$, and $\Omega yejF$) system inactivated [200], [202] with that of *wt* Sm1021 and PHZ-resistant mutants selected previously (mut #1 and mut #4). We plated serial dilutions of the indicated cultures on the solid medium either containing no antibiotic or supplemented with 8 μ M PHZ and calculated the CFU values (**Fig. 6.1.1B**). As expected, there was no difference in growth on the media with and without PHZ for previously selected double mutants, while no colonies were detected on the antibiotic-containing medium for the *wt* Sm1021. Single mutants were also PHZ-susceptible, however, we observed the growth of several separate colonies for these strains in 0 and 10x dilution spots (**Fig. 6.2.1A**). In a separate experiment, we compared the PHZ MIC values for a representative set of strains, which included *wt* Sm1021 and its derivatives ($\Delta bacA$, $\Omega yejA$, $\Omega yejE$, and mut #1) grown in liquid YEB medium (**Table 6.1**). The results of this experiment were consistent with what we observed on the solid media. While single mutants displayed the same MIC as the *wt* strain (for $\Omega yejA$ and $\Omega yejE$) or two times higher ($\Delta bacA$), the range of the PHZ concentrations tested did not allow us to detect the MIC for mut #1, its value is at least 64 times higher than that for the *wt*. These results show that the inactivation of either import system is not sufficient for the development of the resistance to PHZ in Sm1021.

Additional mutations accumulated in the genes other than *bacA* and *yejABEF* could potentially contribute to the PHZ resistance of the mutants selected by the transposon library screening. To rule out this possibility we constructed double mutants with both *BacA* and *YejABEF* systems inactivated. This was achieved through the generalized transduction with ϕ M12 phage, in which Sm1021 $\Delta bacA$ strain served as a donor, while either Sm1021 $\Omega yejA$ or Sm1021 $\Omega yejE$ were used as acceptor strains. Similar to the Sm1021 mut #1 strain, both obtained double mutants Sm1021 $\Delta bacA \Omega yejA$ and $\Delta bacA \Omega yejE$ grew well on solid PHZ-containing medium (**Fig. 6.1.1B**) and in liquid medium with the highest PHZ concentration tested in the broth microdilution assay (**Table 6.1**). This allows concluding that the inactivation of both *BacA* and *YejABEF* is necessary and sufficient for the acquisition of PHZ resistance in *S. meliloti*.

As noted above, there were several colonies growing in 0 and 10x dilution spots on the PHZ-containing medium for single mutants tested in our CFU assay (**Fig. 6.2.1A**). The growth of these colonies could indicate the selection for the fully resistant double mutants on the single-mutant background. We compared the frequency of spontaneous resistance acquisition in $\Delta bacA$ and $\Omega yejA$ single mutants and the wild-type Sm1021. While no resistant clones were obtained for the *wt* strain as before, the rate of resistance acquisition for single mutants was between 10^{-5} and 10^{-6} (**Fig. 6.1.1C**). To get insights into the genetic basis of resistance acquired on the background of inactivated *yejA*, we isolated six resistant clones of the Sm1021 $\Omega yejA$ strain (two from each independent biological replicate) and amplified their *bacA* genomic regions with specific primers. Subsequent Sanger

sequencing revealed mutations in 5 out of 6 *bacA* amplicons (**Fig. 6.2.1B**). In three cases, single nucleotide insertions or deletions led to premature stop codon formation in the *bacA* ORF. In two other cases, mutations led to single amino acid substitutions (L158R and F162S) in one of the BacA transmembrane α -helices. Since Leu¹⁵⁸ and Phe¹⁶² face the inner part of the membrane (**Fig. 6.2.1C and D**) substituting these residues with charged or polar ones should prevent the proper folding of the transmembrane protein and render it inactive. In the remaining mutant, no mutation in the *bacA* open reading frame was detected, we speculate that a mutation in the promoter or another *bacA* regulatory element could inactivate BacA expression in this clone. Overall, these data confirm that mutations in *bacA* are the primary source of acquiring resistance by the strain with inactivated YejABEF. We assume that the complementary result would be obtained when PHZ resistant mutants were

Table 6.1 Phazolicin and bleomycin MIC values for Sm1021 and its derivatives

Strain	PHZ MIC ($\mu\text{g}\cdot\text{mL}^{-1}$)	BLM MIC ($\mu\text{g}\cdot\text{mL}^{-1}$)
Sm1021	0.738	0.5
Sm1021 $\Delta bacA$	1.47	1
Sm1021 $\Omega yejA$	0.738	1
Sm1021 $\Omega yejE$	0.738	1
Sm1021 mut #1	>47	NA
Sm1021 $\Delta bacA \Omega yejA$	>47	8
Sm1021 $\Delta bacA \Omega yejE$	>47	8
Sm1021 $\Omega tolC$	<0.092	NA

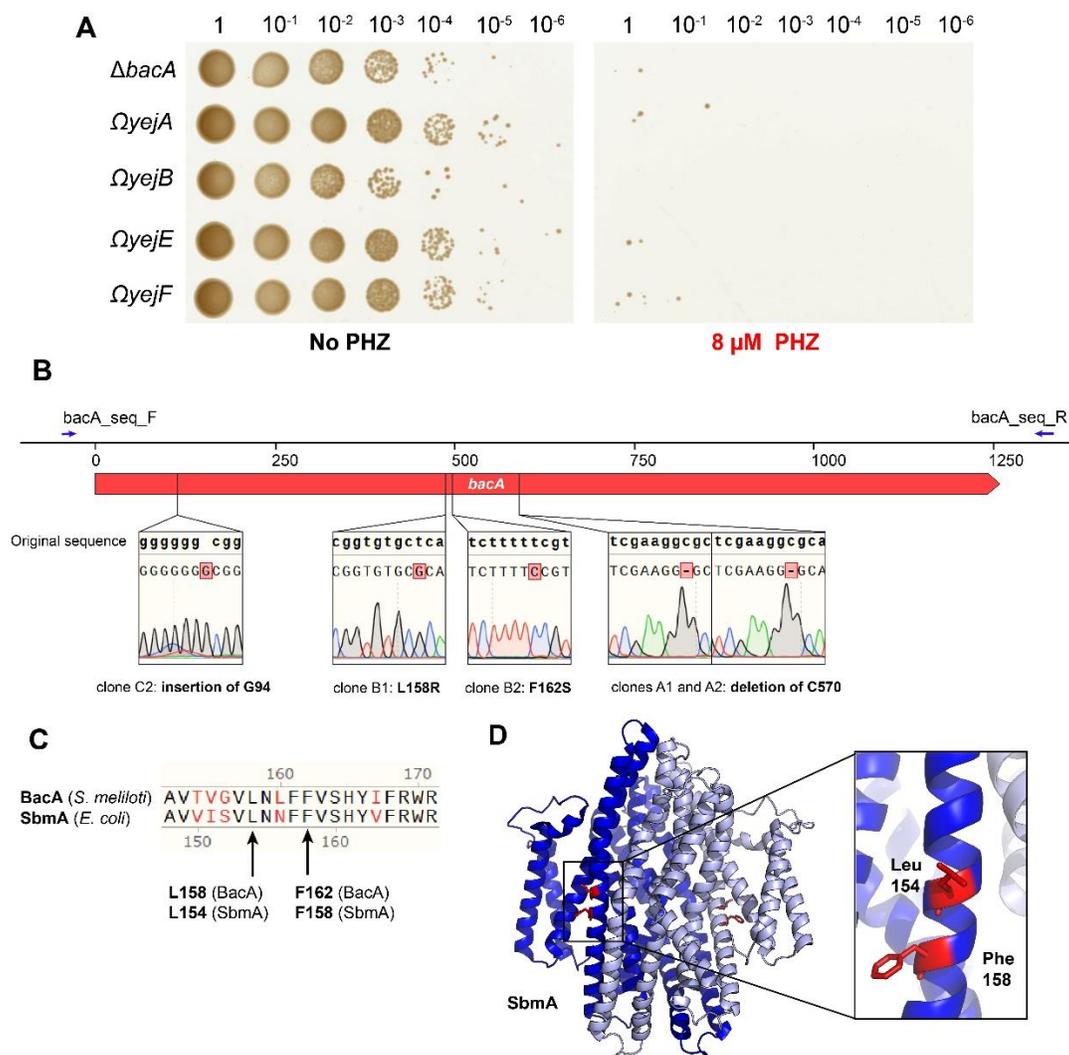


Figure 6.2.1 Genetic basis of the PHZ resistance of the isolates derived from Sm1021 $\Omega yejA$ (A) CFU assay of single Sm 1021 mutants in *yejABEF* and *bacA* genes on the YEB agar plates without PHZ and with 8 μ M PHZ. Note the growth of separate colonies from 0 and 10x dilution spots. (B) The mutations identified by Sanger sequencing in the *bacA* gene of PHZ-resistant mutants selected using Sm 1021 $\Omega yejA$ strain. The numbers indicate the nucleotide position in the gene, the primers used for PCR amplification and sequencing are shown (blue arrows). (C) A fragment of amino acid sequence alignment of BacA (*S. meliloti*) and SbmA (*E. coli*). Note, that both L158 and F162 are conserved between the two studied SLiPTs. (D) The structure of *E. coli* SbmA dimer (subunits are shown in blue and light blue, PDB ID: 7P34 [201]), the amino acids homologous to those undergoing substitutions in PHZ-resistant mutants (clones B1 and B2) are shown in red.

selected from the *bacA* mutant background. However, we did not check this experimentally, since the large size of the *yejABEF* operon makes the identification of second-site mutations using Sanger sequencing alone a more complicated task.

6.3 BacA and YejABEF homologs are capable of PHZ internalization

Genes encoding the transporters homologous to BacA and YejABEF of *S. meliloti* (hereafter, referred to as BacASm and YejABEFSm) are ubiquitous across the genomes of Alpha- and Gammaproteobacteria [203]. To check, if the homologs from other bacteria can internalize PHZ, we constructed pSRK [191] shuttle vector-based plasmids harboring several *bacA/yejABEF* genes from various strains under the control of the inducible *lac* promoter. We selected transporters, which were previously shown to internalize peptide antibiotics (SbmA and YejABEF [203] of *E. coli*, NppA1A2BCD of *Pseudomonas aeruginosa* [204]) or to be required for the establishment of "host-symbiont" (BclA from *Bradyrhizobium* sp. [205]) and "host-pathogen" (BacA^{Ba} from *Brucella abortus* [206]) relationships (**Fig. 6.3.1A**). These constructs along with the empty pSRK vector control were transformed to the PHZ-resistant Sm1021 mut#1 described above. Colony formation on plates with and without PHZ in the presence of the inducer (1 mM IPTG) was monitored. With the exception of *yejABEF*^{Ec}, strains expressing the genes of transporters were susceptible to PHZ, while the strain harboring the empty vector control was resistant, as expected (**Fig. 6.3.1B**). The PHZ-susceptible phenotype reversion observed for all but one strain indicates

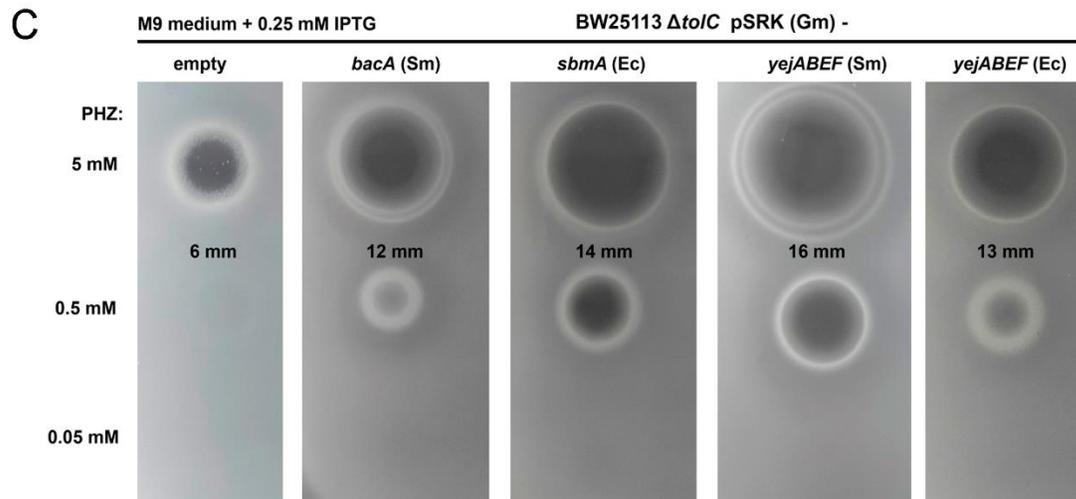
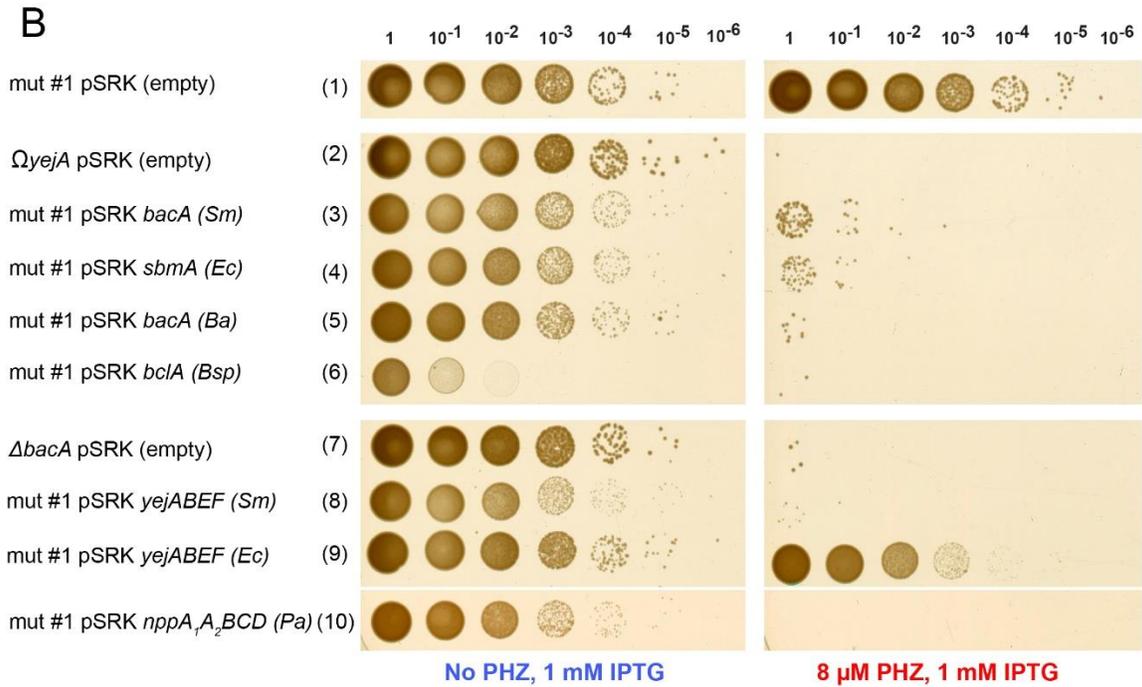
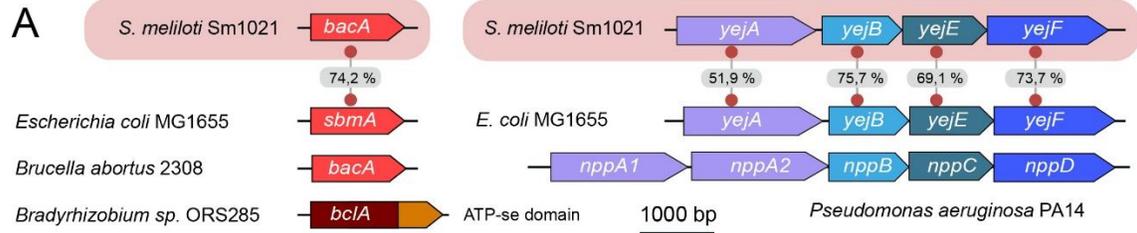


Figure 6.3.1 Resistance to PHZ can be reverted by the episomal expression of BacA, YejABEF or their orthologs from various Alpha- and Gammaproteobacteria. (A) Schematic representations of genes (*bacA*-like) and gene operons (*yejABEF*-like), which were chosen for the expression in PHZ-resistant Sm1021 (mut #1). Numbers on the grey background indicate the amino acid sequence similarity (in%) between the homologous proteins from *S. meliloti* Sm1021 and *E. coli* MG1655. (B) The results of the CFU assay on the Petri dishes with 8 μ M PHZ and without PHZ for the strains of Sm1021 expressing different genes (operons) of transport proteins upon IPTG induction. Note that the CFU count for the *yejABEF*(*Ec*)-expressing strain (lane 9) is approximately 10 times lower in the presence of PHZ compared to the medium without PHZ and this is not observed for the empty plasmid control (lane 1). (C) The inhibition zones from 5 μ L of PHZ spotted in 3 different concentrations (5, 0.5, and 0.05 mM) on top of a lawn of *E. coli* BW25113 Δ *tolC* cells expressing the genes of different transport proteins or containing empty pSRK vector. The sizes of zones for 5 mM PHZ are indicated. Note that the use of high PHZ concentrations was required because of the intrinsic high resistance of *E. coli* to PHZ.

the ability of the tested transporters, except for YejABEF^{Ec}, to internalize PHZ.

The fact that expression of YejABEF^{Ec}, a close homolog of YejABEFSm, led to only a very moderate increase in PHZ-sensitivity (**Fig. 6.3.1B** line (9)), may be due to the lower affinity for PHZ of the periplasmic ABC-transporter subunit YejA^{Ec} or because of poor assembly of the multisubunit transporter in a heterologous Sm1021 host. To distinguish between these possibilities we aimed to test the ability of YejABEF^{Ec} to transport PHZ in its native host. However, *E. coli* is naturally resistant to PHZ in concentrations 100 times higher than those inhibitory for rhizobia. A possible contributing factor to PHZ-resistance of *E. coli* is TolC, which is a major outer membrane multidrug efflux protein that can export various compounds from the cell [193]. Indeed, an *E. coli tolC* deletion mutant has an increased sensitivity to PHZ (see section 5.5). Similarly, the PHZ MIC for a Sm1021 Ω *tolC* mutant is at least eight times lower than for the *wt* (**Table 6.1**). To test the ability of *E. coli* peptide transporters to import PHZ in their native host *in vivo*, we transformed *E. coli*

BW25113 $\Delta tolC$ with pSRK-based *bacA*Sm, *sbmA*^{Ec}, *yejABEF*Sm, or *yejABEF*^{Ec} expression plasmids. Drops of PHZ solution were deposited on lawns of cells harboring pSRK plasmids with indicated genes along with the empty vector control grown on the medium supplemented with IPTG. The appearance of growth inhibition zones was monitored. Interestingly, the expression of any of the four transporters either from *S. meliloti* or *E. coli* led to an increase in the inhibition zone sizes compared to control (**Fig. 6.3.1C**). Thus, the only minor increase in PHZ sensitivity upon expression of *yejABEF*^{Ec} in Sm1021 is more likely caused by protein misfolding or inefficient expression, rather than the inability of *YejA*^{Ec} to bind PHZ.

6.4 BacA and YejABEF are involved in the internalization of the PHZ-unrelated thiazole-containing antibiotic bleomycin

Next, we aimed to determine whether compounds other than PHZ are also internalized via the same pair of *S. meliloti* transporters. Previously, *in vivo* experiments with the *bacA* null mutant demonstrated that *BacA*Sm contributes to the sensitivity of *S. meliloti* to the thiazole-containing peptide-polyketide hybrid antibiotic bleomycin A2 (BLM) [207]. While the direct uptake of BLM by *BacA*Sm has been reported previously [201], the observed partial resistance to BLM of the *bacA* mutant pointed towards the involvement of an additional *BacA*-independent pathway in the BLM internalization [207]. We determined the MICs of BLM against *wt* Sm1021 along with single and double *bacA* and *yejABEF* mutants using a broth microdilution assay. In agreement with published data,

the $\Delta bacA$ Sm1021 mutant was twice less susceptible to BLM compared to the *wt*. Both $\Omega yejA$ and $\Omega yejE$ single mutants had similarly increased resistance to BLM (a MIC of 0.66 μM compared to 0.33 μM for the *wt*, see **Table 6.1**). We, therefore, conclude that the YejABEF provides an alternative pathway for BLM uptake. Remarkably, the double mutants $\Delta bacA \Omega yejA$ and $\Delta bacA \Omega yejE$ were 16 times more resistant to BLM than the *wt* (**Table 6.1**). While these two transporters are largely responsible for the BLM sensitivity of Sm1021, the inhibition of double mutants' growth seen at high concentrations of BLM may be due to the function of yet another low-affinity transport system that remains to be identified or to a low-efficient transporter-independent uptake mechanism.

6.5 Structure of *S. meliloti* YejA

Recently the structure of BacA SLiPT from *S. meliloti* was determined by means of cryo-EM [201], but very little is known about the structure and the mode of ligand binding of YejABEF-like ABC transporters. Therefore, we aimed to determine the structure of the YejABEFSm substrate-binding subunit YejA in a complex with PHZ, since for other previously studied peptide transporters it has been shown that SBPs determine the specificity of transport [208]–[210]. We expressed and purified YejA-CHis6 in *E. coli* and crystallized it in the apo form or with externally added PHZ⁷. In both cases, however,

⁷ Protein expression, purification, and crystallization was performed by me and Dr. Armelle Vigouroux during the long term mobility in the Institute for Integrative Biology of the Cell (I2BC, Université Paris-Saclay, Gif-sur-Yvette, France). Diffraction data were collected on PROXIMA 2 beamline at SOLEIL synchrotron (Saint-Aubin, France) by Dr. Solange Morera and Dr. Armelle Vigouroux. Processing of crystallographic data was performed by Dr. Solange Morera.

in the obtained electron density we observed the association of short peptides with the protein. Our multiple attempts to prepare YejASm sample free from bound peptides included the periplasmic expression of the protein in *E. coli* grown in the peptide-free medium (M9 minimal medium supplemented with casamino acids) and denaturation/renaturation of purified cytoplasmically expressed YejA. Unfortunately, both approaches proved to be unsuccessful.

The structure of YejASm with bound peptides was determined at 1.58 Å resolution (PDB ID: 7Z8E, **Supplementary Table 5**). The numbering used below for the description of residues corresponds to mature YejASm lacking the signal peptide; *i.e.* residue number 1 in YejASm is encoded by codon of Glu31 in *yejASm*. YejASm possesses a typical cluster C fold within the SBP structural classification [211]. The protein consists of two lobes, each formed by a central β -sheet flanked by α -helices (**Fig. 6.5.1A and B**). The bigger lobe consists of residues 1-290 and 558-589 and the smaller one comprises residues 301-552. Two short segments (residues 291-300 and 553-557) serve as hinges connecting the two lobes (**Fig. 6.5.1A and B**, orange). YejASm structure adopts closed conformation due to the binding of random oligopeptides at the interface between the two lobes. According to the results of SSM-EBI search (<http://www.ebi.ac.uk/msd-srv/ssm>) the PDB structures closest to the YejASm are those of the oligopeptide-binding AppA from *Bacillus subtilis* (PDB ID: 1XOC, rmsd 2.27 Å for 448 C α) [212] and the dipeptide-binding DppA from *E. coli* (PDB ID: 1DPP, rmsd 2.22 Å for 437 C α) [203]. Noteworthy, although YejA and these peptide-recognizing SBPs share a similar fold, their oligopeptide-binding sites are distinct.

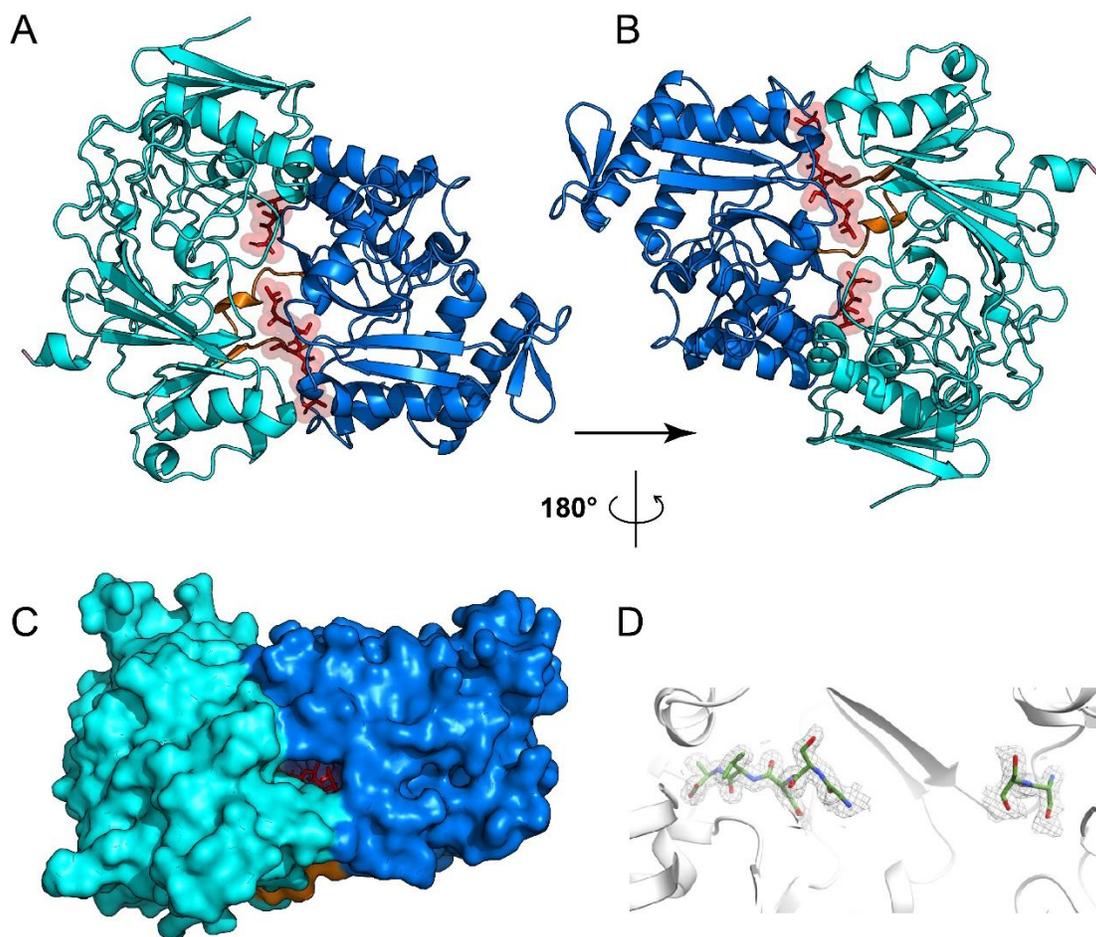


Figure 6.5.1 Crystal structure of YejASm solute-binding protein (A-B) Ribbon diagram showing YejASm in peptide-bound (closed) conformation. The two bound peptides are red (shown in stick), the bigger YejA lobe is light blue, the smaller lobe is blue, hinge regions connecting the two lobes are orange. (C) The surface representation of YejA, coloring as in (A), the terminus of the bound pentapeptide is visible in the groove between the two lobes. (D) Electron density for the bound peptides in the YejASm peptide-binding pocket. Peptides are shown in stick and colored according to the atom type. The 2Fo-Fc electron density map is contoured at 1σ and is shown as grey mesh.

Oligopeptide ligands are bound between the YejASm lobes in our structure (Fig. 6.5.1C). The assignment of the side chains for these peptides was challenging because of the poor quality of the electron density. A dipeptide (SS) and a pentapeptide (GSDVA)

were built, each at different places in the closed interface (**Fig. 6.5.1D**). Electron density linking the two peptides was missing, which may reflect that a population of different peptides is bound between the lobes of SBPs in the crystal. Consequently, the resultant crystal structure is likely an average of YejA molecules bound to different peptides in the crystal lattice. The binding of the pentapeptide to YejASm is almost exclusively mediated by the hydrogen bonds formed between the peptide backbone and the protein (10 interactions out of 12) and that of the dipeptide is only via its backbone, meaning that the binding of the observed peptides is predominantly unspecific.

The multiple peptides bound to YejASm may be derived from the proteins of the heterologous expression host, *E. coli*, or may originate from YejA degradation during overexpression. The recognition of multiple degradation products observed in our YejASm structure has been reported for peptide transporters AppA from *Bacillus subtilis* [212], and OppA from *Lactococcus lactis* [208], though these transporters have not been linked with antimicrobial peptide recognition. Whatever the role of the peptides bound to YejASm, their presence precluded us from solving the structures with our ligand of interest, PHZ. Presumably, the closed conformation of peptide-bound SBPs and the tight binding of the peptides that have been selected from the pool available in the cytoplasm during overexpression did not allow for the exchange of bound peptides with externally added PHZ. However, the nonspecific binding of various peptides by YejA is consistent with its function in the uptake of PHZ, BLM, and other antimicrobial peptides.

Discussion

Here we identified BacA and YejABEF as the two entry points of PHZ into the cells of *S. meliloti*. BacA is a dimer composed of two identical membrane-spanning subunits capable of peptide/H⁺ symport, it adopts the SLiPT fold and is powered by the proton-motive force [201]. YejABEF is a member of the large family of ABC transporters, powered by ATP hydrolysis. While its structure has not been determined yet, based on similarity with other ABC-transporters, it is most likely composed of two transmembrane subunits (YejB and YejE), two nucleotide-binding subunits (YejF₂), and a periplasmic substrate-binding protein (YejA).

Both systems and their close homologs from other microorganisms internalize bioactive compounds. BacA of *S. meliloti* was previously shown to import thiazole-containing antibiotic BLM [207], while SbmA, its ortholog in *E. coli*, is required for the uptake of azole-modified RiPPs microcin B17 [199] and klebsazolicin [76]. PHZ is thus another example of an azole-containing molecule internalized through a SLiPT. The detailed mechanism of peptide recognition by SLiPTs remains unknown. However, available SLiPT structures show the presence of a large cavity where structurally unrelated compounds are thought to bind [201]. The large size of the ligand-binding site contributes to the great promiscuity of SLiPTs, which in addition to azole-containing peptides are known to transport unmodified proline-rich peptides [213], lasso-peptides [189], NCR plant peptides [205], and a number of other natural [214] and artificially designed substrates [215], [216].

NCRs (see section 2.2) comprise one of the most important groups of peptides imported through BacA and YejABEF into the cells of rhizobia. These antimicrobial cationic peptides produced by the plant host target the membrane of bacteria, while their internalization through the unspecific peptide transporters and subsequent degradation by cellular peptidases allows rhizobia to avoid the action of NCRs and develop into functional bacteroids within nodules. Thus, peptide transporters involved in the NCR internalization play a key role in the establishment of the symbiosis between rhizobia and their plant host. Indeed, *S. meliloti* mutant lacking BacA is unable to form functional nodules on the roots of the legume *Medicago truncatula* and dies rapidly in the nodule cells due to the inhibitory action of NCRs [217]. Likewise, we showed that *S. meliloti* strains with the deletions of any one of the *yejABEF* operon genes form highly abnormal hypertrophied cells inside the nodules due to increased susceptibility to NCRs [200]. Similarly, the *yej* genes in pathogenic *Salmonella* and *Brucella* contribute to virulence by protecting bacteria from membrane-targeting antimicrobial peptides produced by the animal host, most likely via their internalization [218], [219].

PHZ is an example of an NCR-unrelated antirhizobial compound internalized via two independent pathways. The dual-entry mode of PHZ by itself dramatically decreases the rate of resistance compared to compounds with a single entry point. Moreover, bacteria that managed to acquire PHZ resistance through mutations in both transporters will be unable to develop functional nodules, for which the uptake of membrane-targeting NCRs is essential. Since passage through symbiosis and massive multiplication inside the legume

nodules, followed by the return of bacteria into the soil at the end of the nodules' lifetime, is a key mechanism of rhizobial spread in the environment, the loss of any one of these transporters will be evolutionary disfavoured.

Chapter summary

Intracellularly acting PHZ is imported into the cells of the PHZ-susceptible bacterium *S. meliloti* Sm1021 via two peptide transporters, BacA (SLiPT) and YejABEF (ABC transporter). This dramatically decreases the rate of PHZ resistance acquisition, since both transporters need to be inactivated to render the bacterium resistant. The mutations in the genes of these transporters should be additionally disfavoured in natural populations since both of them are required for the development of functional symbiosis of *S. meliloti* with legumes, which is the main pathway of multiplication and subsequent spread of these bacteria in soil. BacA and YejABEF homologs from various bacteria can also internalize PHZ. A similar dual mode of uptake was demonstrated for an unrelated thiazole-containing antibiotic bleomycin. The obtained crystal structure of YejASm in complex with peptides bound shows that they are preferentially recognized through the backbone, rather than the side chains, which is consistent with the non-specific mode of peptide uptake by YejABEF, which is hijacked by PHZ and BLM.

Chapter 7. Phazolicin: homologs and the ecological role

7.1 *phz*-like BGC are found across the genomes of proteobacteria

When we performed our initial search for BGCs of putative new LAPs in the publicly available prokaryotic genomes, which resulted in the discovery of the *phz* BGC in the genome of *Rhizobium* sp. Pop5 (see section 5.1), only two of its close homologs were identified in other genomes: one from *Rhizobium* sp. PDO1-076 and the other from *Phyllobacterium myrsinacearum* DSM 5893. Almost three years after this original search we wondered, if the diversity of known *phz*-like BGCs can be expanded based on the newly published genomes. With the information on the PHZ structure, biosynthesis, and the mechanism of action in hand, we performed BLASTP searches using the amino acid sequences of the YcaO domain-containing cyclodehydratase PhzD or phazolicin precursor peptide PhzA as baits. Subsequent manual annotation of the genomic regions surrounding the identified *phzD* and *phzA* homologs allowed us to locate other genes' homologs in each BGC. The search using PhzA as a bait retrieved the clusters located at the contig ends and lacking the full *phzD* gene. The overall composition of BGCs along with the sequence of the precursor were the criteria we used to sort out unrelated clusters.

In total, we found *phz*-type BGCs in the genomes of 14 bacterial strains belonging to four genera (*Rhizobium* (7), *Mesorhizobium* (2), *Pleomorphomonas* (1), and *Phyllobacterium* (3)) of the order Hyphomicrobiales (class Alphaproteobacteria) and one genus (*Kinneretia*) of the order Burkholderiales (class Betaproteobacteria). **Table 7.1**

Table 7.1 Isolation sources for the strains with *phz*-like BGCs in the genome

Strain	Isolation source	BGC on a plasmid?	Accession number, Refs, and comments
<i>Kinneretia</i> sp. XES5	swab of adult <i>Xenopus laevis</i> skin	-	NZ_CP084752, [220]
<i>Mesorhizobium huakuii</i> 583	<i>Oxytropis kamtschatica</i> root nodules	+ (331 kbp)	NZ_CP050298, [221]
<i>Mesorhizobium loti</i> 582	<i>Oxytropis kamtschatica</i> root nodules	+ (452 kbp)	NZ_CP050294, [221]
<i>Phyllobacterium myrsinacearum</i> DSM 5893	NA	NA*	NZ_SGXB01000013
<i>Phyllobacterium calauticae</i> R2-JL	freshwater sediment	NA*	NZ_JAGENB010000002, [222]
<i>Phyllobacterium</i> sp. KW56	nodule (species unknown)	NA*	NZ_JAIQWW010000038
<i>Pleomorphomonas</i> sp. SG524	<i>Sorghum bicolor</i> root samples	NA*	NZ_JAAOYR010000004, [223]
<i>Rhizobium altiplani</i> BR 10423	<i>Mimosa pudica</i> nodules	NA*	NZ_LNCD01000036, [224], lacks <i>phzCBD</i> homologs
<i>Rhizobium herbae</i> HU44	<i>Cajanus cajan</i> root	NA*	NZ_JAEUAO010000003
<i>Rhizobium</i> sp. Pop5	<i>Phaseolus vulgaris</i> root nodule	+ (402 kbp)	NZ_AMCP01000684
<i>Rhizobium</i> sp. RHZ01	soil	NA*	NZ_JACUZZ010000027
<i>Rhizobium</i> sp. RHZ02	soil	NA*	NZ_JACUZX010000024
<i>Rhizobium</i> sp. PDO1-076	<i>Populus deltoides</i> root material	NA*	NZ_AHZC01000156, [225], located at the contig end, lacks part of <i>phzD</i> and <i>phzR</i> homologs
<i>Rhizobium</i> sp. 57MFTsu3.2	NA	NA*	NZ_JDWI01000010

*NA marks the strains, for which the genome assembly is available only in the form of multiple contigs



Figure 7.1.1 Sampling sites of the environmentally isolated strains with *phz*-like biosynthetic gene clusters in the genomes.

summarizes the data on the isolation sources of the strains containing the *phz*-like BGCs in the genome. The geographical locations of their sampling sites, when available, are mapped in **Figure 7.1.1**. As can be seen from the map, the strains harboring *phz*-type BGCs are found worldwide. Most of them were isolated from the root nodules and root material of diverse plants and soil. *Kinneretia* sp. XES5 strain derived from the frog skin is an exception from this trend [220]. However, given that other studied members of this genus are free-living freshwater bacteria [226], it remains unclear, if it is a component of the *Xenopus* skin microbiota, or originates from the freshwater environment. Thus, *phz*-like BGCs are best represented in the genomes of plant symbiotic bacteria capable of nodule formation on the roots of legumes (*Rhizobium*, *Mesorhizobium*) and plant-associated bacteria inhabiting the rhizosphere (*Phyllobacterium*, *Pleomorphomonas*). Interestingly, in

three cases out of four, where the genome assemblies included the complete genomes rather than sets of contigs, *phz*-like BGCs were located on large plasmids (300-450 kbp). It is tempting to speculate, that these BGCs may be the subject of horizontal gene transfer between the bacterial strains.

All identified *phz*-like BGCs share the same gene composition including those coding for the precursor peptide, three proteins required for its post-translational modification, and the ABC-exporter (**Fig. 7.1.2A**). In the BGCs from *Phyllobacterium* genomes, the gene of the MucR family transcriptional regulator found in other clusters is missing, while an additional ORF is found downstream of the *phzE* homolog. We could not propose any function for this putative short protein based on the sequence comparisons with known proteins and structure prediction results. Multiple alignment of the amino acid sequences of the PhzA homologs revealed a significant degree of sequence conservation between the precursors (**Fig. 7.1.2B**). The highest level of conservation was observed in the N-terminal part of the core, which includes the residues undergoing modification and involved in the establishment of the intramolecular π - π stacking system within the PHZ globule and the interactions with the 23S rRNA (see section **5.6**). In four out of eight positions, which undergo cyclization in the course of PHZ maturation we observed the presence of an alternative residue (e.g. Ser instead of Cys, or vice versa), which implies, that the thiazoles and oxazoles are interchangeable for the interactions they make. Other residues, which were shown to be involved in the interactions with the ribosome (Arg5, Arg11, Asp7, Ser8) were also conserved, with only arginines being substituted with lysines

in some peptides. The conservation level was lower for the C-terminal part of the core peptide, which is consistent with our previous observations, that this part of PHZ is occupying the part of the NPET further from the PTC and does not contribute much to the binding of the whole molecule to the ribosome. Taken together, the comparison of the precursors allows us to propose that they undergo modifications similar to those of PHZ and we would expect a similar mode of binding of these mature peptides to the ribosome. The precursors encoded in the *phz*-like BGCs from the genomes of *Phyllobacterium* strains are a notable exception (**Fig. 7.1.2B**). They are on average 15 residues longer and contain an additional region with potentially cyclizable Ser residues located to the N-terminus of the otherwise conserved core part. We predict that the mature products of these clusters contain 3-4 additional azoles and use their N-terminal string of residues to establish additional interactions in the ribosome exit tunnel. However, further research is required to provide support to these predictions.

We constructed a phylogenetic tree using the amino acid sequences of the PhzD homologs from the *phz*-like clusters found (**Fig. 7.1.2B**). Its overall topology shows the divergence of the clusters from the genomes of *Phyllobacterium*, which is supported also by the analysis of the gene composition and the comparison of the precursor peptides (see above). Interestingly, the sequences of both the cyclodehydratase and the precursor peptide from the genome of *Burkholderia*-related *Kinneretia* sp. XES5 were similar to those from the core group of *phz*-like clusters from various rhizobia. This may be another piece of evidence allowing to propose the horizontal gene transfer of *phz*-like BGCs.

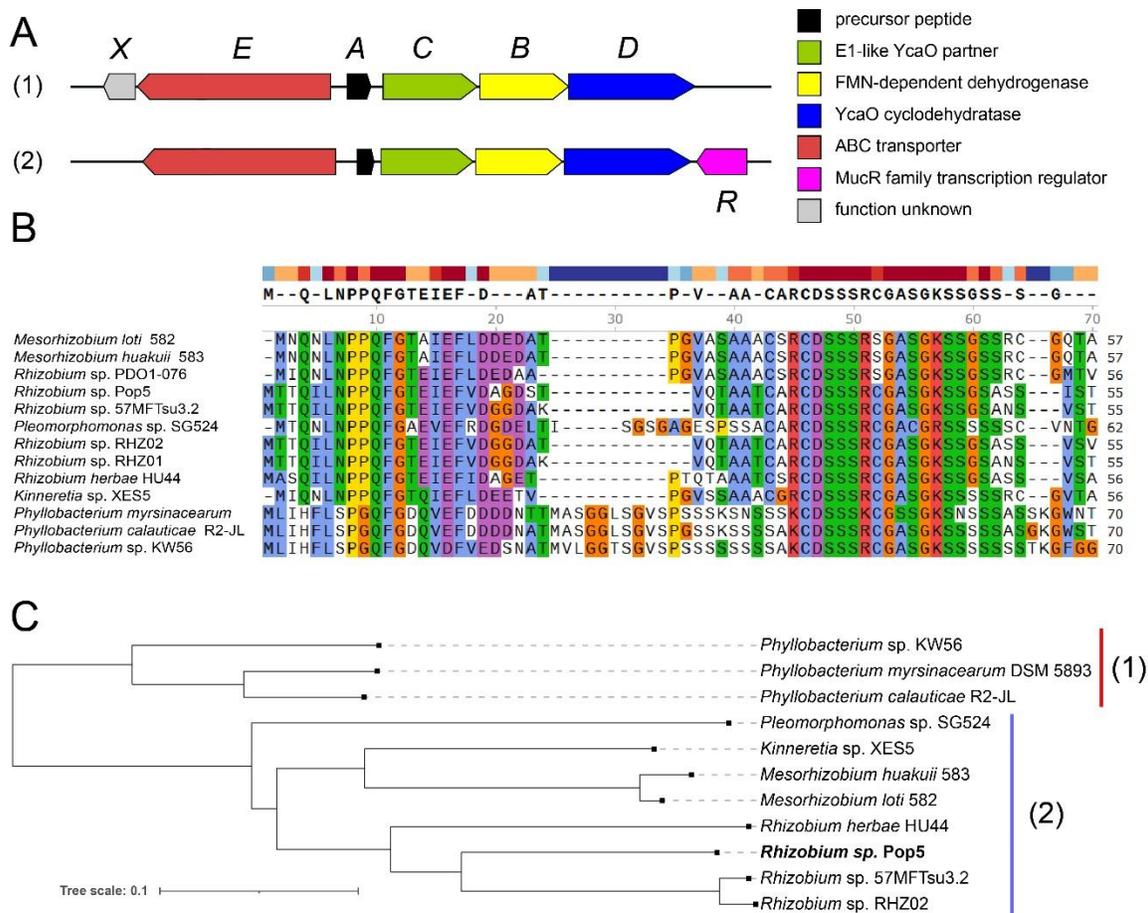


Figure 7.1.2 *phz*-like biosynthetic gene clusters (A) Schematic maps of the *phz*-like BGCs found across the prokaryotic genomes. Variants of gene compositions found in the genus *Phyllobacterium* (1) and other genera (2) are shown. Each arrow indicates one gene, functions of the encoded proteins are listed on the right. (B) Multiple alignment of the amino acid sequences of the precursor peptides encoded in the *phz*-like BGCs. The alignment consensus is shown above, color highlighting is based on the chemical properties of the amino acids and their conservation. (C) Maximum likelihood phylogenetic tree of PhzD homologs, built using PhyML [227]. PhzD sequence from PHZ-producing *Rhizobium* sp. Pop5 is shown in bold. Sequences from the incomplete BGCs of *Rhizobium altiplani* BR 10423 and *Rhizobium* sp. PDO1-076 were not included in the analysis.

7.2 *Mesorhizobium loti* strain 582 produces PHZ-like azole-modified peptides

Our genomic search revealed the presence of *phz*-like BGCs in the genomes of bacteria isolated around the globe. We aimed to check if the expression of any of these clusters results in the production of PHZ-like azole-modified RiPPs. We got access to the two strains of the genus *Mesorhizobium*, which were recently isolated from the nodules of *Oxytropis kamtschatica* – an endemic legume found in the Russian Far East (Kamchatka and Chukotka peninsulas) [221]. The strains *Mesorhizobium loti* 582 and *Mesorhizobium huakuii* 583 were kindly provided by Dr. Vera Safronova (All-Russian Research Institute for Agricultural Microbiology, Saint-Petersburg, Russia). Both strains have *phz*-like clusters located on large plasmids (**Table 7.1**) and encode identical precursors displaying a high degree of sequence similarity to PhzA (**Fig. 7.1.2B** and **7.2.1A**).

We cultivated both strains in liquid RM medium or on Petri plates and performed MALDI-ToF-MS of the whole cells and cultivation medium extracts in search of the PHZ-like metabolites. Only for *M. loti* 582 grown on the solid RM medium, we could detect a mass peak with $m/z = 2562.2$ $[M+H]^+$ (**Fig. 7.2.1B**), which corresponds to the 29 amino acid-long core part of the precursor cut between Ser28 and Ala29 and harboring eight azole cycles (-160 Da). In addition, two smaller peaks with $m/z = 2390.1$ and $m/z = 2491.2$ $[M+H]^+$ could correspond to the shorter forms of the same compound with a dipeptide Thr-Ala removed from the C-terminus, and with one C-terminal or N-terminal Ala residue cleaved off, respectively. Since these compounds were not produced in the liquid medium, their purification for further structural analysis from the slime-producing cells grown on

the solid medium turned out to be challenging. We could not get enough material for the characterization of these molecules even with tandem mass-spectrometry; the optimization of production and purification is required for further work with these putative PHZ homologs. Nevertheless, the match of observed masses with those calculated based on the precursor sequence, along with the presence of additional peaks with an alternative leader cleavage site and/or C-terminal processing allows us to conclude that, under laboratory conditions tested, *Mesorhizobium loti* 582 produces a set of PHZ-likeazole-modified compounds, however, in much smaller quantities compared to *Rhizobium* sp. Pop5.

Since we could not purify the compounds produced by *M. loti* 582 and test them against other bacterial strains as we did previously for PHZ (see section 5.3), we checked the bioactivity of the strain against several rhizobial strains from the genera *Rhizobium*, *Sinorhizobium*, and *Mesorhizobium* in agar overlay assay (Fig. 7.2.1C). *Rhizobium* sp. Pop5 producing PHZ, which is active against *Sinorhizobium* and *Rhizobium* strains, served as a positive control. We observed a clear inhibition zone around the colony of *M. loti* 582 growing on the lawn of *M. ciceri* LMG14989 (Fig. 7.2.1C, red dotted line), while no other tested strain was affected. This assay does not allow us to link the inhibition of the *M. ciceri* LMG14989 growth with the production of the PHZ-like compounds by *M. loti* 582 directly. However, since the antiSMASH search with the whole genome sequence of *M. loti* 582 did not reveal any other BGCs of known antibiotics, it is likely that the growth inhibition of the closely related strain is due to the production of the PHZ-like compounds. To sum up, we here show that *M. loti* 582 isolated in a geographically distant site produces a set of

PHZ-like RiPPs, which may contribute to its antimicrobial activity against the strain from the same genus. Additional research on this strain and the properties of the compounds it produces is required. Of special interest is the basis of the specificity of PHZ and its homologs action on various rhizobial strains.

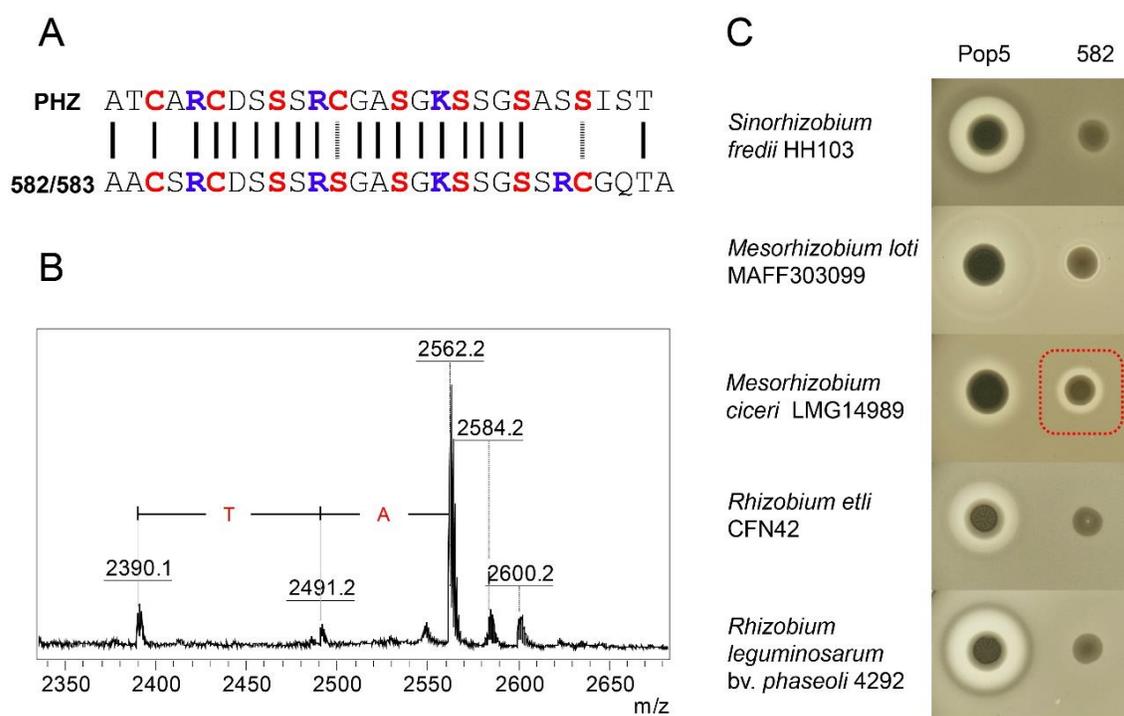


Figure 7.2.1 PHZ-like peptides produced by *Mesorhizobium loti* strain 582 (A) Amino acid sequence alignment of the core parts of PHZ precursor and the precursor encoded in the genomes of the two *Mesorhizobium* strains (582 and 583). Residues converted into azoles in PHZ and those matching them are red, positively charged residues are blue. Black lines show identical positions in the peptides, dashed lines – the positions of synonymous substitutions. (B) MALDI-ToF-MS spectrum of the whole cells of *Mesorhizobium loti* strain 582 featuring the presence of peaks with m/z 2390, 2491 and 2562 [M+H]⁺ corresponding to the modified PHZ-like azole-containing peptides. (C) Growth inhibition (or the lack of thereof) around the drops of *Rhizobium* sp. Pop5 or *Mesorhizobium loti* strain 582 spotted over a lawn of test rhizobial strains. Red dotted line indicates a clear inhibition zone against *M. ciceri* around the drop of *M. loti* 582.

7.3 PHZ-production underlies *Rhizobium* sp. Pop5 competitiveness in culture

The production of antibiotics, including antimicrobial peptides, is considered a powerful weapon allowing the producers to compete with susceptible strains for the ecological niche. The production of another antirhizobial RiPP trifolitoxin (TFX, see section 2.1.6) was previously shown to provide a competitive advantage to the producer in the nodulation of the clover roots and to prevent the nodulation by TFX-sensitive rhizobia [133]. With this in mind, we aimed to check if the ability to synthesize PHZ contributes to the increased competitiveness of *Rhizobium* sp. Pop5. PHZ-nonproducing strain *Rhizobium* sp. Pop5 $\Omega phzD$, whose construction and characterization were described above (see section 5.4), served as a negative control for the experiments described further.

We started with an *in vitro* co-cultivation experiment, in which we prepared 1:1 mixtures of either Pop5 or Pop5 $\Omega phzD$ with either PHZ-sensitive *S. meliloti* Sm1021 or the PHZ-resistant transporter-deficient derivative of the latter strain Sm1021 $\Delta bacA \Omega yejA$ (see section 6.2). Sm1021 and its derivatives are intrinsically resistant to streptomycin, while Pop5 lacks any genomic markers, the use of which would allow us to select these bacteria from the mixture for the analysis. To overcome this we transformed both PHZ-producing and PHZ-nonproducing Pop5 versions with the pIN72 plasmid, which encodes the DsRed fluorescent protein under the control of a constitutive promoter and provides the resulting transformants with resistance to tetracycline. The four obtained mixtures were cultivated for four days in RM medium at 28 °C. Before the start of the experiment and after every 24 hours, CFUs were counted by serial dilutions of the corresponding cultures

on the agar plates with either tetracycline (selective for Pop5 derivatives) or streptomycin (selective for the Sm1021 derivatives) for CFU counting. The experiment was repeated three times using independent starter cultures for the preparation of the initial mixtures.

After 24h of growth, we observed a drop in CFU/mL for the mixed culture of wild-type Sm1021 and wild-type Pop5 on streptomycin-containing medium (selective for Sm1021) for almost 4 orders of magnitude (**Fig. 7.3.1**). No increase in the number of *S. meliloti* CFUs occurred during further cultivation, while the number of *Rhizobium* sp. Pop5 CFUs increased by the end of the experiment compared to the starting time point. In contrast, for the mixtures containing either transporter-deficient Sm1021 or PHZ-nonproducing *Rhizobium* sp. Pop5, the CFU/mL numbers on both selective media increased approximately 10-fold during the first 24 hours of cultivation and remained stable until the end of the experiment. We conclude that in the conditions tested PHZ production alone is sufficient and required for the ability of *Rhizobium* sp. Pop5 to eliminate the sensitive strain from the co-culture, since the mixtures containing PHZ-resistant Sm1021 $\Delta bacA$ $\Omega yejA$ or PHZ-nonproducing derivative of Pop5 do not demonstrate any decrease in the *S. meliloti* CFU/mL upon the course of the co-cultivation.

7.4 The effects of PHZ production and addition are not detected in soil

Although PHZ production was shown to provide a competitive advantage to *Rhizobium* sp. Pop5 grown in rich RM medium, it remained unknown, if the same effect can be observed in conditions that are closer to those found in nature. We aimed to check

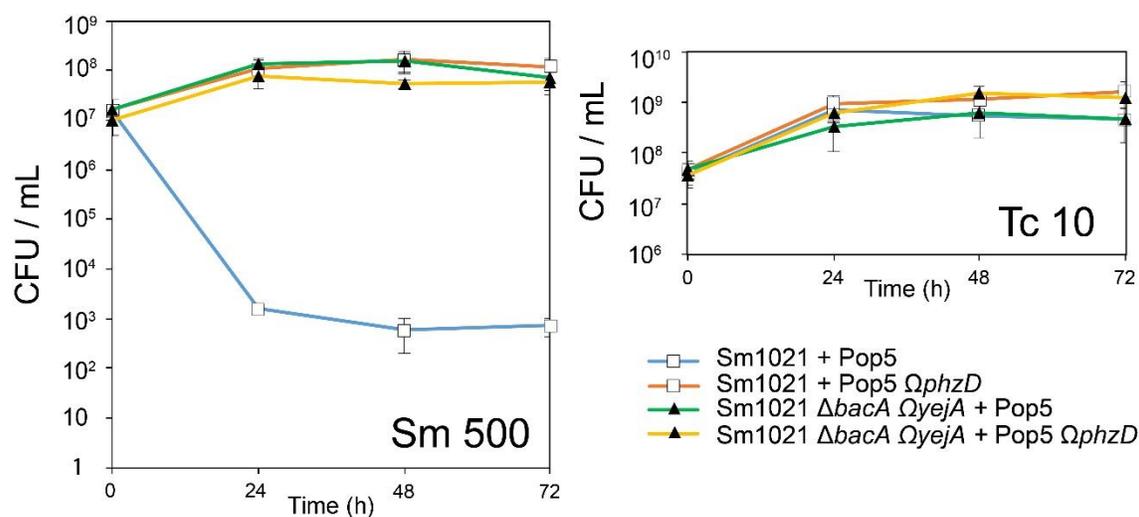


Figure 7.3.1 PHZ production mediates the competitiveness of *Rhizobium* sp. Pop5 in co-culture CFU counts for the aliquots of two-strain mixtures sampled every 24 hours and spotted on plates with selective media (growth on Sm500 reflects the number of *S. meliloti* Sm1021 CFUs, on Tc10, the number of *Rhizobium* sp. Pop5 CFUs).

this by setting up a series of co-cultivation experiments in sterile soil. For these experiments, we opted for *Rhizobium leguminosarum* bv. *phaseoli* 4292 (referred further as *Rlp* 4292) as a model PHZ-sensitive strain. This strain belongs to the same genus as the *Rhizobium* sp. Pop5, and forms the root nodules on the same plant – common beans (*Phaseolus vulgaris*). Moreover, this strain displayed the lowest PHZ MIC among all rhizobial strains tested (see section 5.3). Last, but not least, *Rlp* 4292 carries the genomic resistance to rifampicin, which enabled its isolation from the mixtures containing *Rhizobium* sp. Pop5 or other bacteria sensitive to rifampicin. In the first experiment⁸, we

⁸ Soil preparation for the experiments was performed by Dr. Peter Mergaert. Justine Del Bianco (Master Student at the University Paris-Saclay) was involved in the soil competition experiments with PHZ-producing strains.

inoculated 10 g of farmland soil with $1 \cdot 10^7$ CFU/g of *Rlp* 4292 followed by the addition of *Rhizobium* sp. Pop5 ($1 \cdot 10^7$ or $1 \cdot 10^9$ CFU/g) or its PHZ-nonproducing derivative *Rhizobium* sp. Pop5 $\Omega phzD$ in the same concentrations (see section 4.13.2 for the experimental details). As a negative control, we used soil samples inoculated with *Rlp* 4292, to which mQ was added instead of the *Rhizobium* sp. Pop5 *wt* or $\Omega phzD$. CFU counting on the rifampicin-containing medium selective for *Rlp* 4292 showed that there was no difference in the number of bacteria recovered from the soil samples that were inoculated with PHZ-producing or PHZ-nonproducing strains regardless of the inoculum concentration (**Fig. 7.4.1A**).

Being unable to detect the effect of PHZ on the sensitive bacteria in soil inoculated with water suspensions of the producing strain, we hypothesized that the biosynthesis of PHZ in concentrations required for the competitor elimination may be an energy-consuming process, which requires the presence of easily mobilized nutrients in the environment. In soil, such conditions can occur, for instance, around decomposing organic material. To mimic nutrient-rich conditions we repeated the experiment described above using the high concentration ($1 \cdot 10^9$ CFU/g) of the *Rhizobium* sp. Pop5 and *Rhizobium* sp. Pop5 $\Omega phzD$ inoculums. This time, the resulting soil samples were additionally supplemented either with water or with the rich RM medium, which provided additional sources of nutrients and energy. However, in this setup of the experiment also, the CFU counting did not reveal any considerable drop in the numbers of *Rlp* 4292 recovered from samples inoculated with the PHZ producer compared to the mQ control (**Fig. 7.4.1B**).

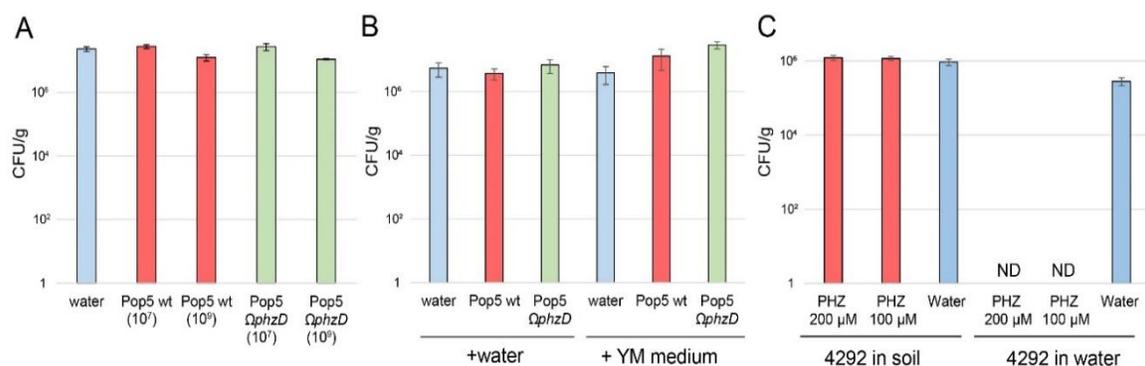


Figure 7.4.1 The effects of PHZ on the growth of the PHZ-susceptible *Rhizobium leguminosarum* bv. *phaseoli* 4292 in soil (A) CFU counts for *Rlp* 4292 grown in soil upon the addition of *Rhizobium* sp. Pop5 or its PHZ-nonproducing derivative in different concentrations. ND – not detected. (B) Same as (A) with the addition of YM rich medium. Pop5 and Pop5 Ω phzD are used in the concentration 1×10^9 CFU/g. (C) The effect of PHZ addition into the soil inoculated with *Rlp* 4292 (left) or the culture of *Rlp* 4292 in water.

This negative result may be explained by either the lack of PHZ production by *Rhizobium* sp. Pop5 in the experimental conditions or by the inability of PHZ produced to inhibit the growth of otherwise sensitive strain in this complex environment. To test the latter conjecture, we set to determine if we were able to observe the activity of PHZ in soil using the purified antibiotic rather than the PHZ-producing strain. We used PHZ in concentrations almost 1000x higher than the MIC for *Rlp* 4292 to treat soil samples inoculated with *Rlp* 4292 and the suspensions of the same bacteria in water. mQ was used instead of PHZ in the same setup as a negative control. While there was no single *Rlp* 4292 colony grown on the selective medium after the PHZ treatment of the water suspension, the numbers of bacteria recovered from PHZ-treated and untreated soil were comparable (Fig. 7.4.1C). This unexpected result may be explained by the relatively large size of the

PHZ molecule and the presence ofazole cycles, which can mediate additional interactions with soil-forming organic polymers and absorption on inorganic clay and silica particles. These complex substrates may bind PHZ, thereby removing the antibiotic from the solution and decreasing its concentration. Thus, we propose, that even if produced by *Rhizobium* sp. Pop5 in soil, PHZ cannot effectively act against sensitive bacteria found nearby, probably due to the physicochemical properties of this modified peptide.

7.5 PHZ production and the competition for root nodulation

In the previous section, we demonstrated that the production of PHZ by the resident population of rhizobia in the soil does not lead to a decrease in the number of susceptible bacteria. Indeed, when inhabiting bulk soil, rhizobia have an oligotrophic lifestyle, possibly not appropriate for energy-consuming PHZ synthesis. In addition, even when the soil environment is richer in nutrients, which can be the case during saprophytic growth of rhizobia, PHZ is inefficient probably because of the absorption on soil material, as discussed above. Moreover, rhizobia constitute a minor fraction of soil microbes (see section 2.2). This implies that it is unlikely that rhizobia are able to successfully compete in this environment via the production of narrow-spectrum antimicrobials like PHZ, since the distances between the cells of the producer and the susceptible strain may require very high production of the antibiotic. This situation changes once rhizobia are in close proximity to plant roots and start to compete for the opportunity to form a nodule. This competition in the rhizosphere occurs in a thin layer surrounding the root, where the

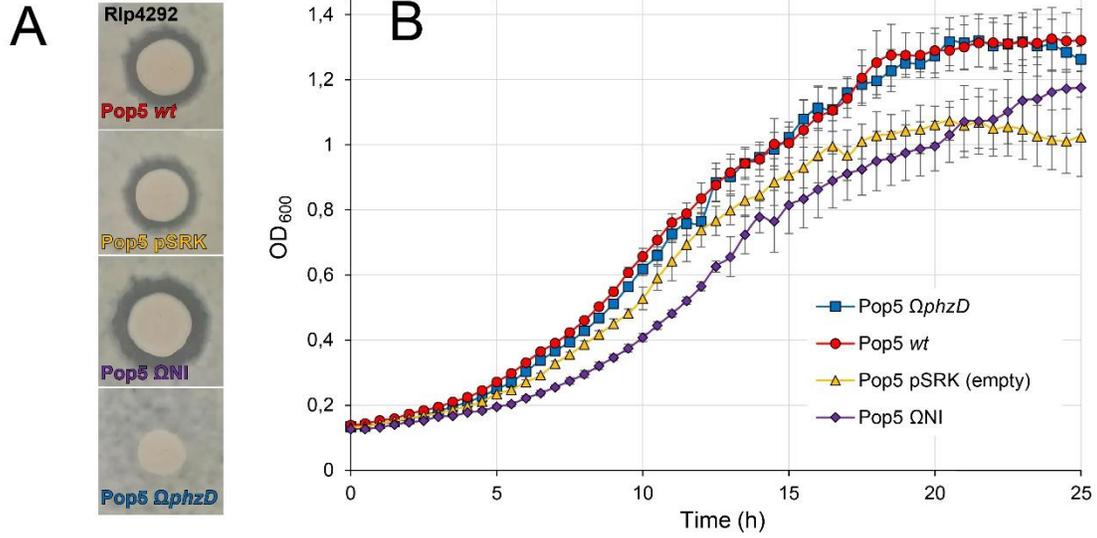
chemical communication between bacteria and plant cells takes place (see section 2.2). We hypothesized, that the ability to produce PHZ may provide a competitive advantage to *Rhizobium* sp. Pop5 during the nodulation of *Phaseolus*, rather than in its free-living stage in soil. To check this experimentally we set to develop a system, which could be used to quantitatively access the nodulation effectiveness of the PHZ-producing and PHZ-nonproducing strains in the presence of the PHZ-sensitive competitor.

Rhizobium sp. Pop5 was isolated from the nodules of wild beans in the south of Mexico. However, it was not previously shown that this strain can effectively nodulate *Phaseolus* and perform nitrogen fixation under laboratory conditions. We inoculated 10-day-old seedlings of *Phaseolus vulgaris* (variety “Magical”) grown in perlite-sand mixture with the suspension of *Rhizobium* sp. Pop5 in water ($OD_{600nm}=0.1$), several plants left noninoculated constituted the negative control group. After 21 days of growth, the plants were checked for the presence of nodules, which were found only on the roots of those inoculated with *Rhizobium* sp. Pop5 (**Fig. 7.5.1A**). The noninoculated plants were small and displayed clear signs of chlorosis, while the plants with root nodules were better developed and were dark green (**Fig. 7.5.1A**), which implies, that the nodules formed by *Rhizobium* sp. Pop5 were functional and provided fixed nitrogen to the plant. The cross-sections of the obtained nodules were stained by a mixture of fluorescent dyes (propidium iodide, SYTO 9, and calcofluor white) and analyzed by microscopy. This revealed the presence of viable bacteria inside the plant cells of the nodules (**Fig. 7.5.1B and C**).



Figure 7.5.1 *Rhizobium* sp. Pop5 forms functional nodules on the roots of *Phaseolus vulgaris*. (A) *Phaseolus vulgaris* “Magical” plants, which were either inoculated with *Rhizobium* sp. Pop5 (right) or left noninoculated (left) after 1 month of growth in the nitrogen-deficient perlite-sand mixture. Note the chlorotic phenotype of the left plant, which is a typical sign of nitrogen starvation. The close-up view of the nodules formed on the roots of the right plant is shown. (B) Root nodule cross-section stained with a mixture of the dyes propidium iodide, SYTO 9, and calcofluor white. Staining with this mixture highlights the nodule bacteria with a green fluorescence signal (SYTO 9) when their membranes are intact and with red fluorescence (propidium iodide) when their membranes are permeable (e.g. dead cells). Plant cell walls are stained blue (calcofluor white), plant cell nuclei are stained red (propidium iodide). (C) A close-up view of a single root nodule cell. Staining as in (B). Note the presence of bacteria inside the plant cell.

Once we demonstrated that *Rhizobium* sp. Pop5 could form functional nitrogen-fixing nodules on the roots of *P. vulgaris* in our laboratory conditions, we aimed to select the combination of test strains for further competition experiments, which would allow us to identify, which strain gave rise to each nodule tested. Indeed, the majority of nodules on legume plants are initiated by a single rhizobium bacterium and are therefore colonized by a clonal population of this strain. Therefore, the use of markers that can distinguish test strains can readily indicate, which strain has induced and colonized nodules. As we expected to test large samples of nodules, we decided to use antibiotic resistance as a marker for the analysis of bacteria recovered from the nodules. As a PHZ-susceptible model strain we selected rifampicin-resistant *Rhizobium leguminosarum* bv. *phaseoli* 4292 (*Rlp* 4292), which nodulates *P. vulgaris* and was successfully inhibited by *Rhizobium* sp. Pop5 in co-cultivation experiments (see section 7.3). PHZ-nonproducing strain Pop5 Ω *phzD* we obtained previously (see section 5.4) is resistant to kanamycin; however, there is no resistance determinant in the *wt* Pop5 strain. We obtained two Kan^R derivatives of this strain, either by making a pVO155*nptIIgfp* insertion between the genes *phzD* and *phzR*, without the disruption of any of these ORFs (referred further as *Rhizobium* sp. Pop5 Ω NI, NI –neutral insert), or by transforming Pop5 with an empty plasmid pSRK (Kan^R). Both obtained strains inhibited the growth of PHZ-susceptible *Rlp* 4292, which reflects their ability to produce PHZ, since no inhibition zone was observed around a spot of Pop5 Ω *phzD* (Fig. 7.5.2A). Next, we examined the growth of the two obtained strains in rich RM medium in comparison to Pop5 *wt* and Pop5 Ω *phzD*. Unexpectedly, the strain



C Noninoculated Rlp4292 Pop5 pSRK Pop5 Δ phzD

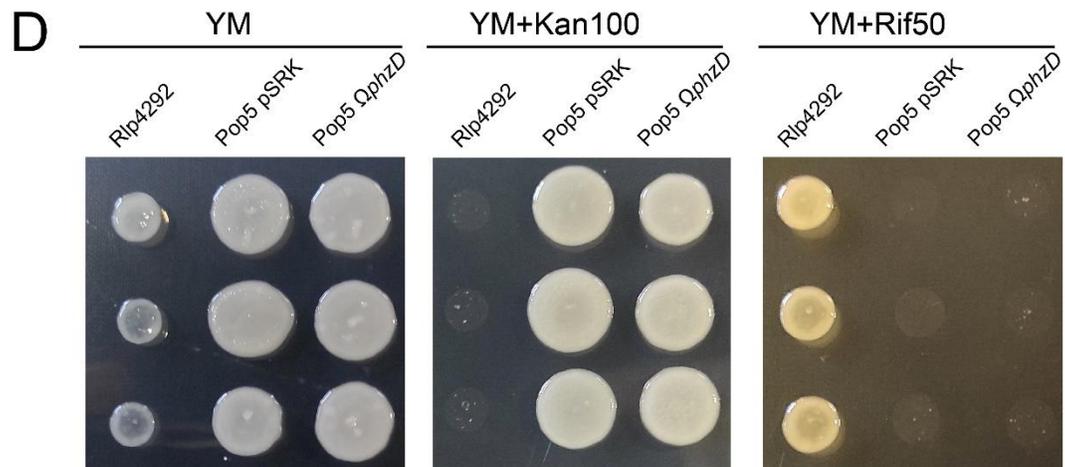
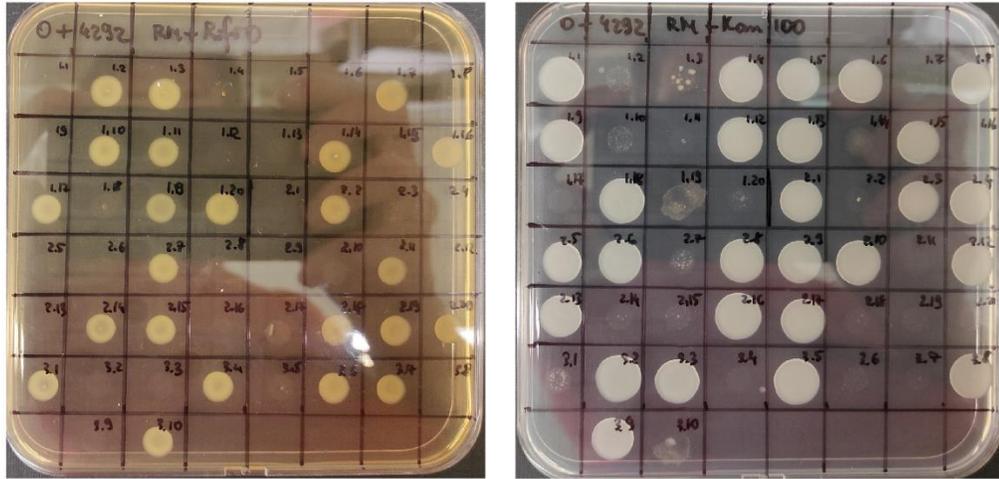


Figure 7.5.2 Selection of strains for nodulation competition experiment. (A) Growth inhibition zones around the spots of *Rhizobium* sp. Pop5 *wt* and three of its derivatives on the lawn of PHZ-susceptible *Rlp* 4292. (B) Growth curves of Pop5 *wt* and its derivatives grown in RM medium in 96-well plates. (C) Roots of noninoculated *P. vulgaris* and three plants inoculated with suspensions of denoted strains. Red arrows point to the nodules formed on the roots of inoculated plants. (D) The growth of bacteria recovered from single nodules on RM medium without antibiotics or containing either Rif (selective for *Rlp* 4292) or Kan (selective for Pop5 derivatives).

harboring the pSRK plasmid was growing better than the one with a genomic insertion, despite the need of the former to replicate plasmids (**Fig. 7.5.2B**). The reason of the growth defect of the latter strain remains unknown but it may indicate that the selected region for pVO155*nptIIgfp* insertion in Pop5 Ω NI is not neutral. Therefore, we decided to proceed to further competition experiments using the pSRK plasmid-bearing strain.

To establish the procedure of bacteria recovery from the nodules we inoculated three groups of *P. vulgaris* seedling with the suspensions of *Rlp* 4292, Pop5 pSRK, and Pop5 Ω *phzD*, while the fourth group remained noninoculated. The examination of the roots of the plants on the 21st day after inoculation revealed the presence of nodules on the roots of the plants from the three inoculated experimental groups, while the noninoculated plants did not carry nodules, as expected (**Fig. 7.5.2C**). The nodules were collected, surface-sterilized by ethanol, and individually crushed in the buffer to preserve the integrity of the nodule bacteria (see section **4.13.5** for the detailed description of this procedure). The obtained suspensions from single nodules were spotted on solid RM medium without antibiotics or supplemented with either Rif or Kan. As can be seen from the **Figure 7.5.2D**, the growth observed on selective media reflected the type of the strain used for plant

A



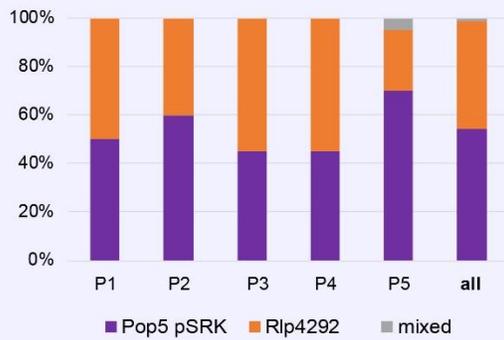
B

Exp 1: 5 plants per group, 20 nodules per plant
 $OD_{600} = 0.1, 50:50$

Pop5 $\Omega phzD$ + Rlp4292



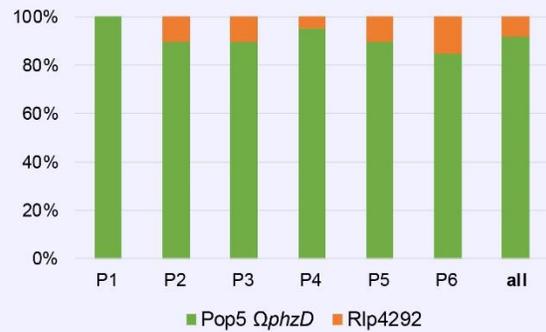
Pop5 pSRK + Rlp4292



C

Exp 2: 6 plants per group, 20 nodules per plant
 $OD_{600} = 0.05, 50:50$

Pop5 $\Omega phzD$ + Rlp4292



Pop5 wt + Rlp4292

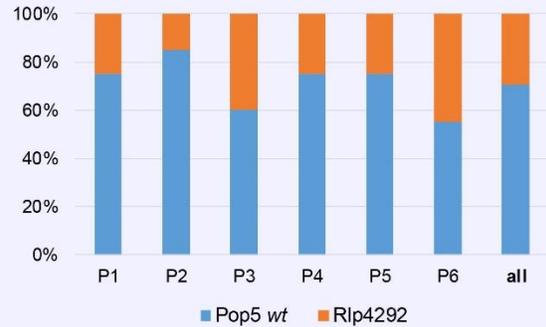


Figure 7.5.3 Plant nodulation competition experiments. (A) Growth of rhizobia recovered from the root nodules on the Petri plates with Rif- (left) and Kan-containing RM medium (right). Representative plates for the plants inoculated with the mixture of *Rlp* 4292 and Pop5 pSRK are shown. Note that clear growth is observed only on one or the other plate, which indicates that the majority of nodules are the result of single-strain infection. (B), (C) Results of the first (B) and second (C) competition experiments. Each bar represents the ratio between the nodules of different origins for a single plant, the last bar shows the average for the sample of plants (n=5 for the first and n=6 for the second experiment). The color code for the strains is universal for panels (B) and (C).

inoculation. Thus, we had a system including three strains (PHZ-susceptible, PHZ-producing, and PHZ-nonproducing), all of which nodulated *P. vulgaris*, and between which we could distinguish according to the growth profile on antibiotic-containing selective media.

With this system in hand, we set up a first nodulation competition experiment, which included two experimental groups (five plants per group, inoculation with 1:1 mixtures of Pop5 Ω *phzD* + *Rlp* 4292 or Pop5 pSRK + *Rlp* 4292, inoculum OD_{600nm}=0.1, 10 mL of inoculum per plant), three control groups with single-strain inoculations (three plants each), and one group of noninoculated plants. Subsequent spotting on selective media of bacterial suspensions obtained from single nodules (we analyzed 20 nodules per plant in each experimental group) demonstrated that the majority of nodules were the result of single-strain infection and we could clearly identify the strain forming each nodule by the growth profile (**Fig. 7.5.3A**). For several nodules, we observed the additional growth of other bacteria on selective media (e.g. **Fig. 7.5.3A** right plate, positions 1.19 and 3.10), however, this could be easily distinguished from that of both rhizobial strains used for inoculation. The results of the comparison between the two experimental groups were

unexpected since in the group inoculated with PHZ-producing strain Pop5 pSRK we observed more nodules formed by the PHZ-sensitive strain *Rlp* 4292 (9 ± 2.55 nodules per plant sample ($n=20$)) compared to the group inoculated with the PHZ-nonproducing derivative Pop5 $\Omega phzD$ (2 ± 1 nodules per plant) (**Fig. 7.5.3B**). We decided to repeat the competition experiment using Pop5 wt strain instead of Pop5 pSRK to eliminate the effect of the cost for plasmid maintenance on the effectiveness of plant nodulation. In this setup, we were unable to identify the mixed nodules containing both strains, as we lacked the resistance determinant for Pop5 wt. However, as we observed in the first experiment, the number of mixed nodules was small, and we further disfavored their formation by decreasing the cell density of the initial inoculum for the second experiment two times ($OD_{600nm}=0.05$). The results of the second experiment were similar to those of the first one. Thus, contrary to our expectations, the number of nodules formed by the PHZ-sensitive strain was higher in the mixture containing the PHZ producer (5.83 ± 2.04 nodules per plant) compared to the mixture with $\Omega phzD$ derivative (1.66 ± 1.03 nodules per plant).

In addition, we performed MALDI-ToF-MS analysis of methanol extracts from nodules formed on the roots of plants from the control groups inoculated by single-strain suspensions of PHZ-producing or PHZ-nonproducing strains. Our analysis did not reveal any mass-peaks that could be attributed to any PHZ-related compounds. This allows us to propose, that although the cell density of bacteria inside the nodules is high, the production of PHZ does not happen in the nodules, or, alternatively, the compound produced is effectively degraded by the action of plant enzymes.

Discussion

Our analysis of publicly available genomes identified *phz*-like BGCs, which likely guide the biosynthesis of PHZ structural and functional homologs, in multiple genomes of plant-associated Alphaproteobacteria sampled around the globe. For one strain of *Mesorhizobium* we were able to identify the production of PHZ-like peptides, which may contribute to the antimicrobial activity of this strain against *Mesorhizobium ciceri*. However, this should be further proved by experiments with purified compounds.

Taking into account the narrow-spectrum activity of PHZ demonstrated earlier, we proposed that PHZ-like compounds may play some role in the interspecies competition of the producers with closely related bacteria at certain stages of the rhizobial lifecycle. To investigate this experimentally, we performed a series of cocultivation experiments with strains grown in rich cultivation media and in soil. Although the PHZ-producing strain effectively decreased the number of susceptible bacteria once grown in the liquid medium, we could not observe the same effect in soil. Moreover, even the addition of high concentrations of purified PHZ to the soil samples did not lead to a decrease in the number of susceptible bacteria recovered. This allows us to propose, that PHZ production in bulk soil is inefficient not only because of the limited supply of energy and nutrients required for PHZ biosynthesis but also because of the chemical nature of PHZ, which likely forms multiple interactions with diverse molecules and particles present in this chemically complex environment.

For previously studied antirrhizobial RiPP trifolitoxin, it was demonstrated that its production allows the producer to outcompete other strains and exclusively nodulate clover in the presence of TFX-sensitive strains (see section **2.1.6**). Interested if the same is true for the PHZ-producing strain Pop5, we developed a system, which allows for high-throughput screening of root nodules formed on *P. vulgaris* by the competing strains. Since we could not distinguish the nodules formed by different strains based on their appearance, as it was made previously for ineffective nodules induced by the TFX producer [128], we used spotting of the nodule homogenates on the media selective for one or another strain.

Remarkably, the results of the competition experiment were opposite of what we expected: the strain lacking the ability to produce PHZ demonstrated higher competitiveness compared to the PHZ-producing one. The same result was obtained in the repeat of the experiment with isogenic strains lacking additional plasmids, whose presence could contribute to slower growth of the strain Pop5 pSRK, which was used as a PHZ-producer in the initial experimental setup. Better performance of the PHZ-nonproducing strain in the competition may be due to the fact, that the nutrients and energy, which are required to produce PHZ, are used in other metabolic pathways and contribute to better growth and more effective nodulation. An increase in fitness for Pop5 $\Omega phzD$, however, was not observed once the strains were grown in rich RM medium, but it does not necessarily mean that *wt* and $\Omega phzD$ strains perform equally well in the rhizosphere, where the available nutrients are limited.

Thus, using the model systems developed we could not observe the positive effects of PHZ production on the competitiveness of Pop5 strain either in soil, or in the rhizosphere in the process of *P. vulgaris* nodulation. It is possible, that the conditions we used do not match those found in the natural habitat of Pop5 strain originating from a tropical forest. Multiple parameters such as the genotype of the host plant, rhizosphere pH, availability of nitrogen, and mineral composition may influence the biosynthesis of PHZ. The production of PHZ was not detected by means of MALDI MS in the extracts from nodules formed by PHZ-producing bacteria as well, suggesting that either PHZ production is suppressed inside the nodules or that the sensitivity of the method was not sufficient for the detection of the amounts of the antibiotic produced.

One other possibility, which we did not test experimentally, is the production of PHZ at the stage of nodule senescence and the dispersion of bacteria in soil, which follows the dying off of the nodules. At this stage, the bacteria are present in high numbers and have the access to the organics released from dead plant tissues. PHZ production may be beneficial at this stage, as producers released from the nodules would be able to successfully spread and compete with sensitive rhizobia found in the proximity of the senescent roots. Further studies are required to identify if the production of PHZ is advantageous for Pop5 at any stage of its lifecycle.

Chapter summary

Bioinformatic analysis reveals *phz*-like BGCs in multiple genomes of plant-associated bacteria isolated from both nodules and root material around the globe. The residues, which are converted into azoles or involved in the interaction with rRNA in PHZ, are conserved in the precursor peptides encoded in these clusters. This allows proposing that the RiPPs produced also target translation by a mechanism similar to that of PHZ. One strain *Mesorhizobium loti* 582 was shown to produce PHZ-like peptides upon the cultivation on the solid medium.

PHZ production determines the ability of Pop5 strain to eliminate a PHZ-sensitive strain from a coculture. However, we were unable to detect any effect of the ability to produce PHZ on the competitiveness of Pop5 in soil or upon the nodulation of cultivated beans (*P. vulgaris*). This may be explained either by the difference in conditions used in the model systems we used and those found in nature or by the fact that PHZ production may provide a competitive advantage at a different stage of the rhizobial lifecycle than the ones we investigated.

Chapter 8. Systematic search for novel LAP BGCs

8.1 Construction of networks of *ycaO*-containing BGCs

Phazolicin BGC was identified in the genome of *Rhizobium* sp. Pop5 as a result of a BLASTP search using the sequence of KlpD YcaO domain-containing cyclodehydratase involved in the biosynthesis of KLB as a bait (see section 8.1). Further, using the amino acid sequences of PhzA precursor peptide and the proteins involved in PHZ biosynthesis, we found and described closely related BGCs encoded in the genomes of proteobacteria (see section 7.1). In an attempt to find novel BGCs⁹ of other previously unknown LAPs we set to perform a systematic search for the clusters guiding the biosynthesis of azol(in)e-containing peptides in publicly available genomic data. Since YcaO domain-containing enzymes play a key role in the installation of azol(in)es into the peptides (see section 2.1.2), we started by retrieving the sequences of YcaO domain-containing enzymes present in genomes from the RefSeq database [181]. Subsequent steps of our pipeline (**Fig. 8.1**) included filtering and clusterization of the obtained sample of YcaO sequences followed by the annotation of genomic regions surrounding the recovered *ycaO* genes (see section

⁹ Search for novel BGCs of azol(in)e-containing RiPPs was performed by me and Dmitry Bikmetov, who prepared custom scripts (see Methods section), the sample of YcaO sequences and calculated the SSN.

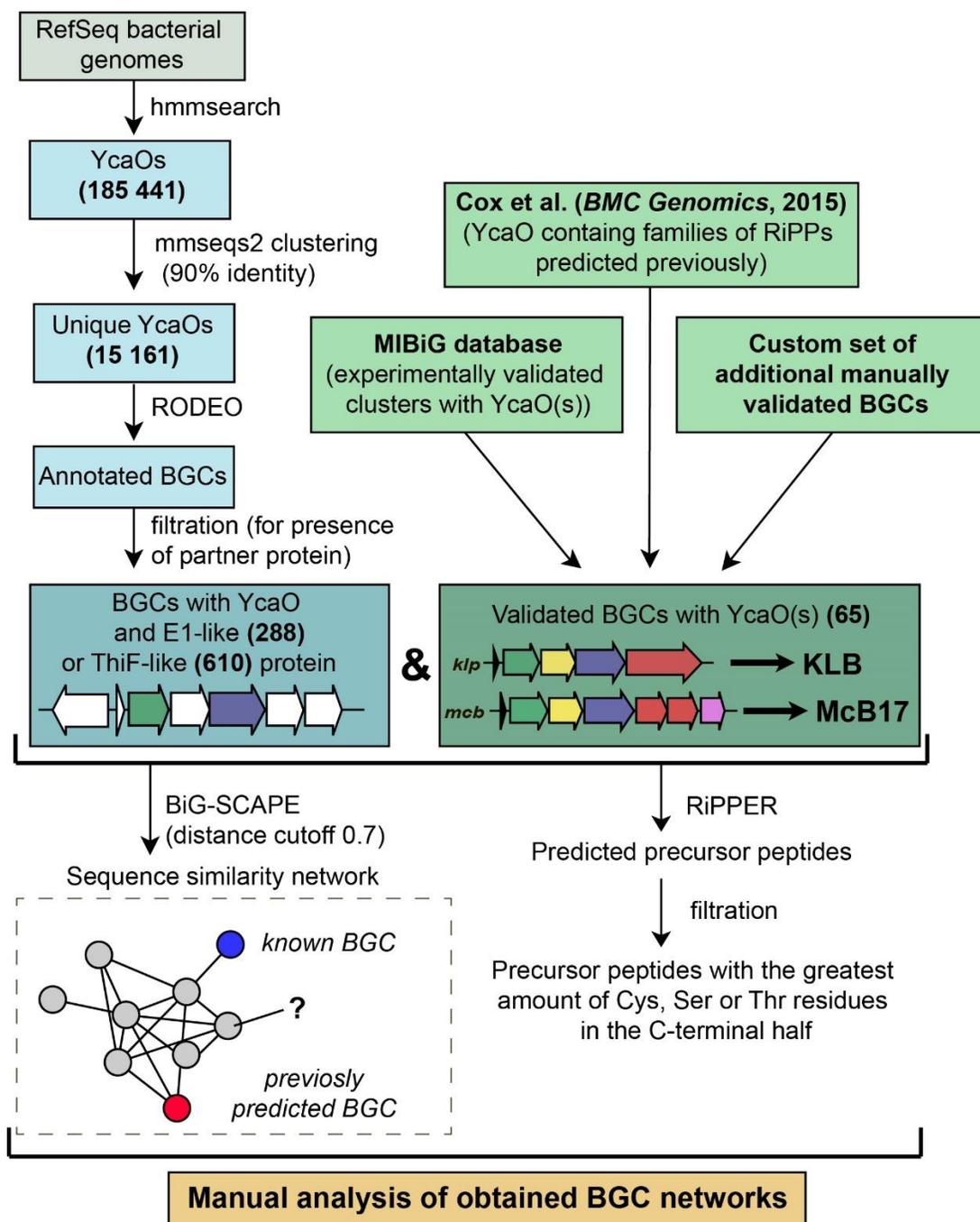


Figure 8.1.1 The pipeline used for the identification of YcaO-containing RiPP BGCs. See section 4.18 for a detailed description of procedures.

4.18 for a detailed description of the procedures). To visualize the diversity and identify families of BGCs we constructed a similarity network of *ycaO*-containing BGCs, which was then analyzed manually. To have a reference for this analysis we composed a curated set of experimentally characterized *ycaO*-containing BGCs (**Supplementary Table 6**), which included those present in MIBiG database [13] or described elsewhere in the literature. This set was also supplemented with the BGCs identified and proposed to guide the biosynthesis of azol(in)e-containing peptides in the previous bioinformatic search (**Supplementary Table 7**, [228]). In our search for novel LAP BGCs, we decided to focus on the clusters containing a gene of E1-like YcaO partner protein, since this is a feature of the majority of experimentally validated LAP clusters. Thus, we did not consider here the BGCs coding for TfuA and ThiF-partners as well as BGCs with standalone YcaOs. The genomic landscape of all azol(in)e-containing peptides was studied by Cox et al. [228], however, since the time of this publication new azoline-containing RiPPs with characterized modes of action (including ribosome-targeting KLB and PHZ) were identified, many more sequenced genomes were deposited in publicly available databases, and improved methods and software became available.

8.2 Analysis of the network of *ycaO*-containing BGCs

Figure 8.2.1 shows the obtained similarity network of *ycaO*-containing BGCs encoding E1-like YcaO partner proteins. BGCs of already characterized compounds from the curated dataset are shown as blue circles. These are the clusters of antibacterial LAPs

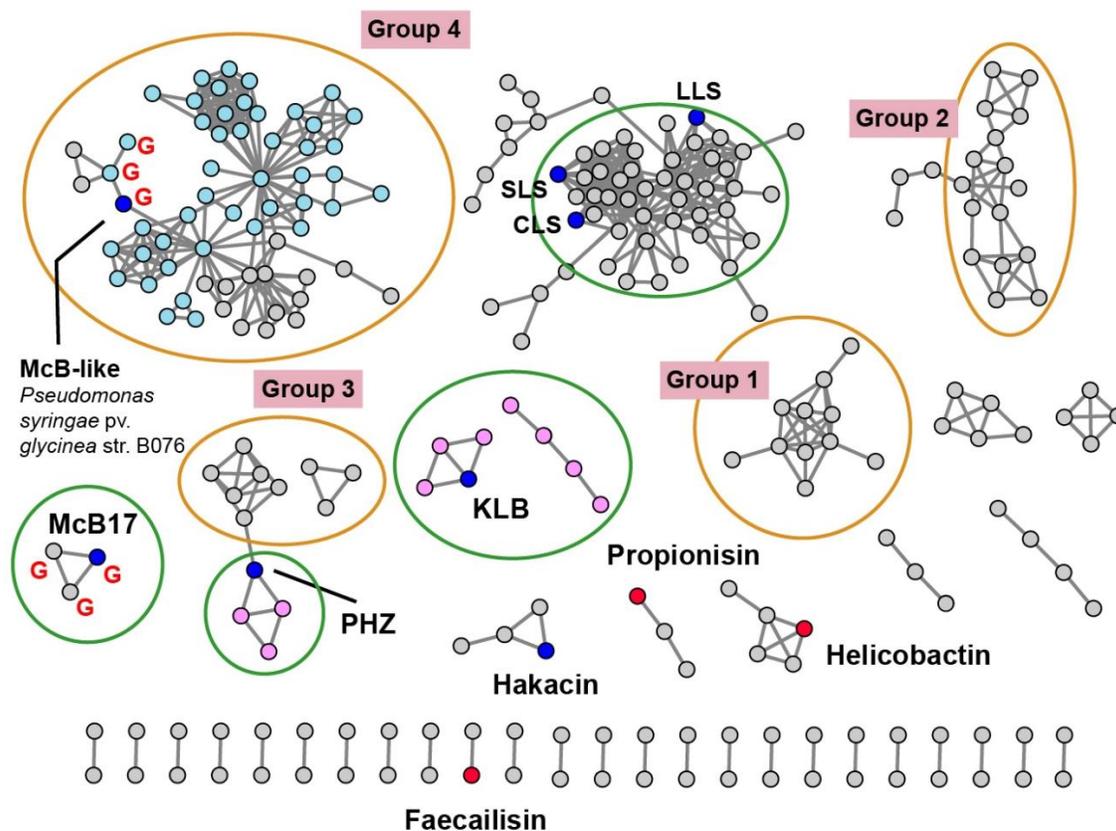


Figure 8.2.1 A similarity network of *ycaO*-containing BGCs containing the genes of E1-like partner proteins. Nodes representing BGCs of experimentally characterized compounds are shown in blue. Nodes representing BGCs of close PHZ and KLB homologs discussed earlier are pink. Nodes representing BGCs analyzed in Cox et al. 2015 are red. BGCs containing *mcbG*-homolog are denoted with red “G”. Groups of clusters discussed further in the text are shown in orange ellipses. Groups of clusters including close homologs of well-characterized compounds are in green ellipses. The light blue color of nodes in the group 4 marks BGCs from genus *Pseudomonas*. Network singletons are not shown. SLS, streptolysin S; LLS, listeriolysin S; CLS, clostridiolysin S; KLB, klebsazolicin; PHZ, phazolicin; McB17, microcin B17.

microcin B17, klebsazolicin and phazolicin, as well as a number of streptolysin S-like RiPPs (clostridilysin S, listeriolysin S) and hakacin, which biosynthesis was studied *in vitro* but the structure of the naturally produced compound remains unknown [229], [230]. The nodes of the network corresponding to the BGCs of close PHZ and KLB homologs, which

were discussed previously (see sections 5.1 and 6.1), are colored pink. The members of several previously predicted families of BGCs were mapped in the obtained network: these include the BGCs of propionisin, helicobactin, and faecailisin [228].

Further, I discuss five groups of BGCs, which encode proteins for the biosynthesis of novel putative LAPs. They attracted our attention during the analysis of the **Figure 8.2.1** network and putative precursor peptides predicted by RiPPER [40].

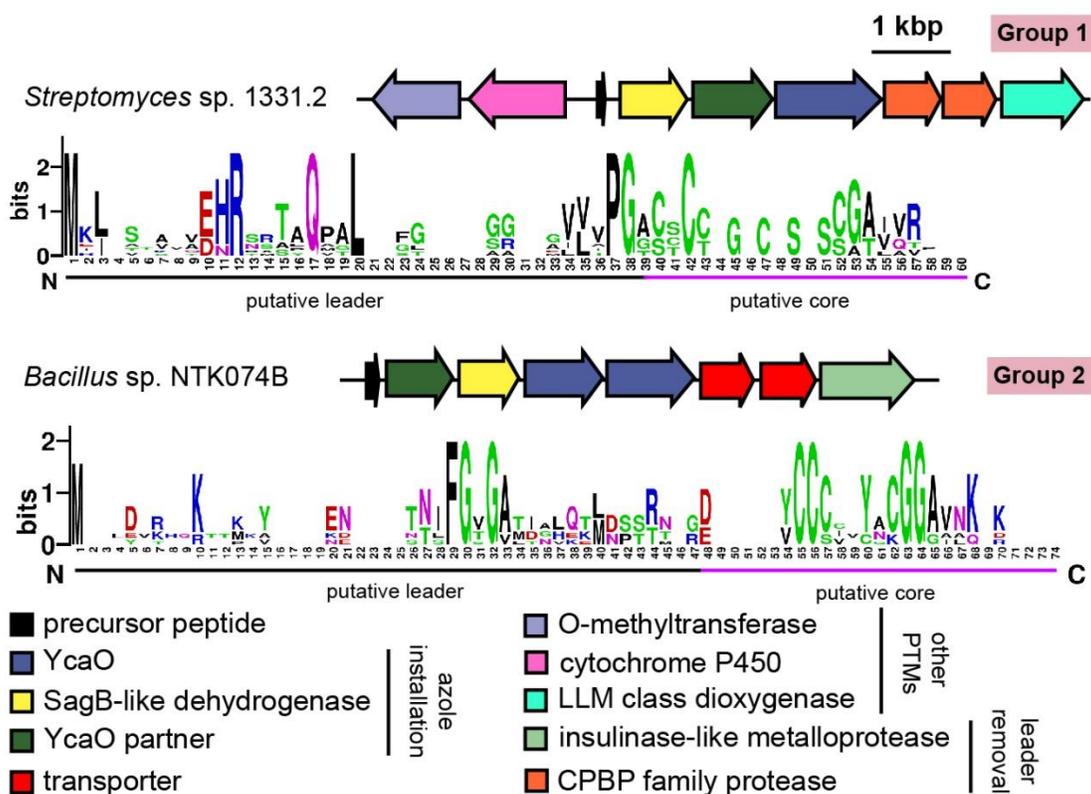


Figure 8.2.2 The two groups of new BGCs of putative LAPs. Schematic maps of BGCs from groups 1 and 2 (see Fig. 8.2.1) are shown. Each arrow indicates one gene, functions of the encoded proteins are listed below. Sequence logos are provided for the alignments of predicted precursor peptides. Putative core and leader parts of the precursors are shown with black and pink lines respectively. LLM, luciferase-like monooxygenase; CPBP, CAAX proteases, and bacteriocin-processing enzymes.

8.2.1 Group 1

First, we aimed to describe the two largest groups of BGCs in our network, which do not contain any experimentally validated or previously predicted clusters (**Fig. 8.2.1**, Groups 1 and 2). Group 1 is represented by BGCs from the genomes of Actinobacteria (genera *Streptomyces*, *Microbispora*, *Microtetraspora*, *Mobiluncus*, etc.). An interesting feature of the BGCs from this group is the presence of the genes coding for the enzymes, which likely catalyze additional tailoring PTMs. These include the genes of putative cytochrome P450 monooxygenase, O-methyltransferase, and LLM class flavin-dependent oxidoreductase (**Fig. 8.2.2**). The PTMs installed by these enzymes are known for RiPPs outside of LAPs. Cytochrome P450 enzymes catalyze a variety of hydroxylations in the biosynthetic pathways of thiopeptides [231], [232] and perform oxidative decarboxylation of the C-terminal azoline of bottromycin [233]. O-methylation of the Asp side chain is also described for bottromycin [234], LLM class flavin-dependent oxidoreductase Cao12 is proposed to introduce D-amino acids into the structure of a recently described lantipeptide-like RiPP cacaoidin [235]. An unusual combination of putative tailoring modifications makes the elucidation of the structure of RiPPs, which biosynthesis is guided by these BGCs an interesting task.

Another unusual feature of the group1 BGCs is the presence of two genes of CPBP (CAAX Proteases and Bacteriocin-Processing enzymes) family proteases (**Fig. 8.2.2**, orange). These are intramembrane metalloproteases, which were shown to be involved in the removal of the leader of SLS-like LAPs [236] and resistance to unmodified antibacterial

peptides produced by *Lactobacillus sakei* and *L. plantarum* [237]. The analysis of precursor peptides of the group 1 BGCs (**Fig. 8.2.2**, sequence logo) revealed the PG(A/G) motif, preceding the C-terminal part rich in Cys and Ser residues (the putative core). The same sequence was shown to be the recognition site of SagE CPBP in SagA – the precursor peptide of streptolysin S [236]. In the case of putative LAPs, encoded by group 1 BGCs one CPBP protease may serve for the removal of the leader, while the other may contribute to the resistance of the producer.

8.2.2 Group 2

The second group of previously undescribed BGCs in our network (**Figure 8.2.1**, Group 2) contains clusters identified in the genomes of Firmicutes (genera *Geobacillus*, *Anoxybacillus*, *Bacillus*, *Paenibacillus*, *Macrococcus*, etc.). These BGCs are distinguished by the presence of two genes coding of YcaO-domain containing enzymes (**Figure 8.2.2**). Among known RiPPs, there are several examples of compounds, whose modification requires the action of two YcaOs installing different types of PTMs. For instance, bottromycin biosynthesis includes the formation of a thiazoline and macrocyclization, catalyzed by two standalone YcaOs [96], [98], while during the modification of thiopeptides GE2270A and sylfomycin two azoline-forming YcaOs modify different positions in the precursor [70], [238]. In the case of group 2 BGCs, there is one E1-like YcaO partner in the cluster, which implies either that the second YcaO acts as a standalone

one and recognizes the precursor in an RRE-independent way, or that both enzymes interact with the same partner.

8.2.3 Group 3 (*Lactazolicins*)

This group of BGCs contains clusters from the representatives of the genus *Lactobacillus*, which form a connected component with PHZ BGC and several BGCs from *Streptococcus* (**Fig. 8.2.1**, Group 3). Analysis of these BGCs and their homologs from the genus *Enterococcus* found in an additional BLAST search showed that they share the same set of genes, which, in addition to modification machinery and export pump homologs of those in PHZ BGC (**Fig. 8.2.3A**, genes *E*, *C*, *B*, and *D*₂), includes three auxiliary genes (**Fig. 8.2.3A**, genes *X*₁, *D*₁, and *X*₂). The product of gene *D*₁ is a second YcaO protein. It is distinct from the product of the *D*₂ gene and lacks C-terminal PxP-motif, found in azoline-forming YcaOs and involved in catalysis [93]. According to the results of HHPred [239] the product of gene *X*₂ is distantly related to ThiF/MccB/PaaA proteins and contains RRE – a domain found in RiPP modification enzymes that bind leader peptides [94]. The presence of the second YcaO and the *X*₂ gene product, which could function either as a partner protein or an independent adenylation enzyme [240], [241], makes additional modifications of the precursor highly probable. We were not able to detect homologs of the *X*₁ gene product among known proteins.

In accordance with the nomenclature recommended for LAPs [23], we named this new group of putative translation inhibitors *lactazolicins*. All lactazolicin clusters encode

83-106 amino acid-long putative precursor peptides with 8-12 repeats of [Cxxx] motif in the N-terminal part of the predicted core segment (**Fig. 8.2.3A**). Heterocycloanthracins (HCAs) represent an already known group of RiPPs, which have a similar pattern of repeated cysteine residues in the core part [39]. However, HCA precursors have [Cxx] motif repeated rather than [Cxxx] and the overall composition of HCA BGCs also differs significantly from that of lactazolicins. Unlike HCAs, where [Cxx]-repeat containing part of the precursor is rich in Gly, the N-terminal [Cxxx] repeat-containing part of lactazolicin core peptides is enriched in positively charged amino acids (Arg, Lys). In the cases of PHZ and proline-rich peptides (which do not belong to RiPPs but also target the ribosome exit tunnel) [242], side chains of positively-charged amino acids take part in the interaction with phosphate groups of rRNA. We hypothesize that lactazolicins also may affect translation.

8.2.4 Group 4 (*mcb*-like BGCs from *pseudomonads*)

Microcin B17 (McB) is a DNA-gyrase-targeting LAP produced by some strains of *E. coli*. The McB BGC contains a set of enzymes similar to those encoded by the KLB and PHZ BGCs and an additional gene *mcbG*, which encodes a pentapeptide repeat protein (PRP) [108], [187]. *McbG* is a DNA mimic that decreases the formation of toxic gyrase-DNA complexes trapped by McB, thus protecting the gyrase in McB-producing cell [243]–[245]. Clusters similar to *mcb* were described in the genomes of several pathogens of *Pseudomonas syringae* and their products also target the gyrase [110].

Figure 8.2.3 BGCs of lactazolicins and *mcb*-like BGCs from pseudomonads (A) Common BGC composition conserved among lactazolicin BGCs, proposed functions of the encoded proteins are listed on the right. RiPP recognition elements (RREs) in C and X₂ genes are depicted. Alignment of precursor peptides of lactazolicins, predicted leader and core parts are shown. Cysteines in the core part are shown on red background, positively charged amino acids of the predicted core are blue, and negatively charged residues are green. **(B)** The composition of *mcb*-like BGC from pseudomonads. *mcbG* homolog is shown in square brackets as it is not present in the majority of clusters from pseudomonads, the genes are colored according to the color scheme in (A). Alignment of precursor peptides predicted with RiPPER, those encoded in PRP gene-containing BGCs are in magenta frame. Potentially cyclized residues are shown with red background, predicted core and leader parts are shown.

A relatively large network of clusters retrieved by our search (**Fig. 8.2.1**, Group 4) contains no characterized representatives except for a *mcb* homolog from *P. syringae* (blue circle). However, only several of these clusters (marked with red letter G) contain a gene coding for a PRP protein. The overall sequence similarity and the distribution of potentially cyclizable residues in precursor peptides from clusters with and without the PRP gene differ (**Fig. 8.2.3B**). Thus, it is highly probable that *mcb*-like clusters lacking the PRP gene encode a RiPP whose target is distinct from DNA gyrase.

8.2.5 Group 5 (*Flavazolicins*)

The last group of putative new unusual LAP BGCs was identified during the analysis of precursor peptides predicted with RiPPER [40]. The precursor peptide identified in the genome of flavobacterium *Algibacter aquaticus* SK-16 (a singlet and therefore not shown at **Fig. 8.2.1**; **Fig. 8.2.4A**) appeared to have resulted from a duplication of a standard leader-core ancestral precursor gene (**Fig. 8.2.4B**). As a result, in a single

ORF there are two putative core sequences rich in Ser and Cys residues separated by an “internal” leader (another leader is N-terminally located) (**Fig. 8.2.4C**). Similar cassette-like arrangement of core peptides have been described for several different groups of

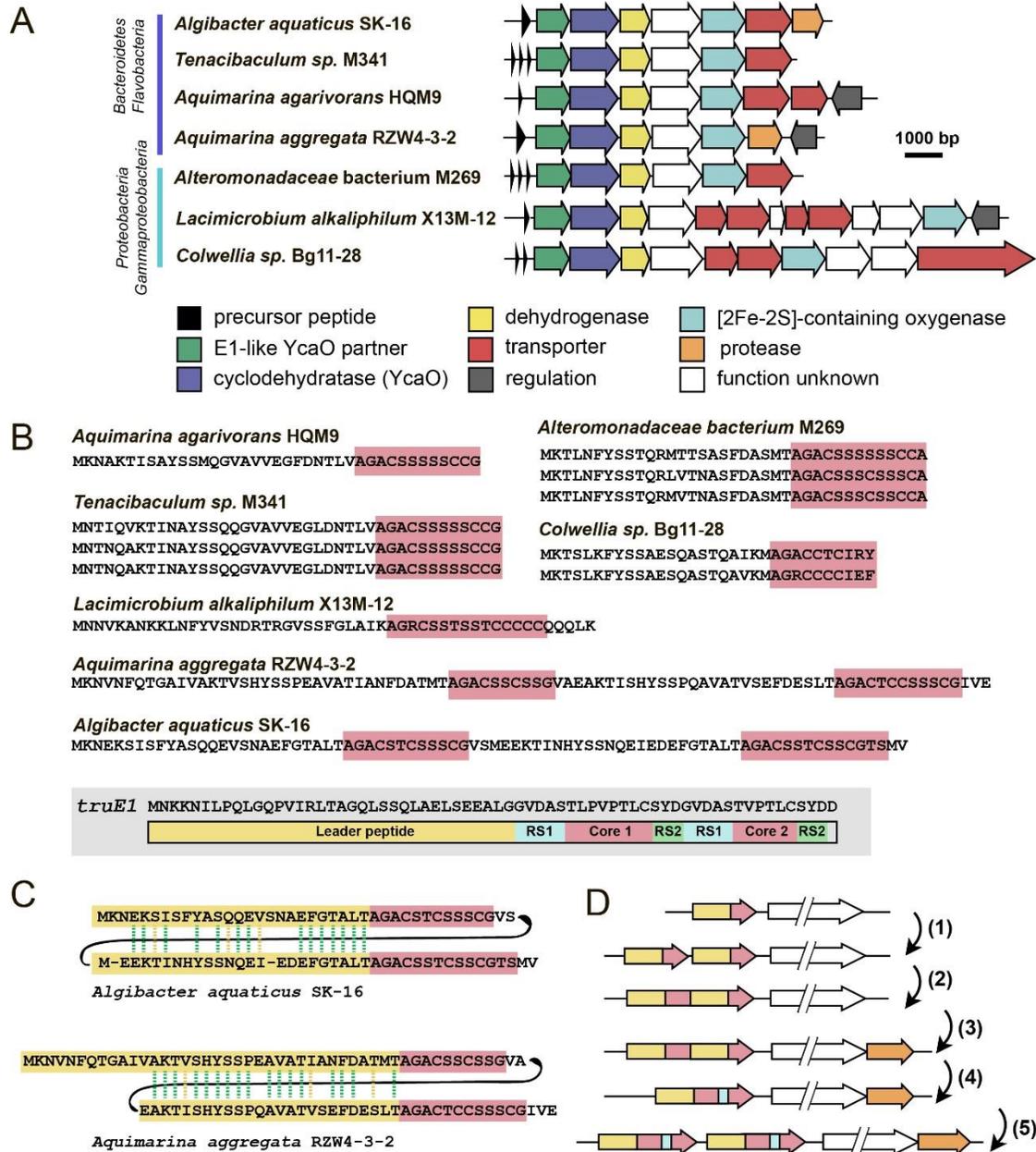


Figure 8.2.4 Flavazolicines. (A) Comparison of biosynthetic gene clusters encoding a putative new group of LAPs found in Flavobacteria and Gammaproteobacteria genomes. Predicted functions of the encoded proteins are listed below. (B) Comparison of precursor peptides of flavazolicines. Conserved core sequences containing cyclizable residues are shown with red background. The precursor peptide sequence of cyanobactins patellins 2 and 3 (*truE1* gene product) is shown for comparison on the gray background. Functional parts of the peptide including leader (yellow), two cores (red), and recognition sequences of peptidases (RS1 and RS2) are shown [45]. (C) Sequences of cassette-containing precursor peptides of flavazolicines showing the conserved positions in two leader sequences. Conserved positions are shown with green dashed lines, synonymous substitutions with yellow dashed lines. (D) A possible scenario in the evolution of cassette-containing peptides. See main text for the explanations.

RiPPs including cyanobactins [45], thiovarsolines [40], orbitides [44], and dikaritins [46] but in all these cases precursors are composed of a single leader, followed by several core peptides, interspersed by signal sequences required for the cleavage of each core at C- and N-termini by dedicated peptidases (**Fig. 8.2.4B** shows, as an example, the sequence of TruE1 – the precursor of patellins 2 and 3, representatives of cyanobactins).

A BLAST search for similar BGCs resulted in the identification of three additional BGCs sharing the same set of modification enzymes in the genomes of *Flavobacteriaceae* closely related to *Algibacter* and three BGCs in the genomes of *Gammaproteobacteria* (**Fig. 8.2.4A**). Interestingly, only two of these clusters contained a fused precursor peptide gene, while others had a set of 1-3 separate ORFs encoding non-fused precursor peptides (**Fig. 8.2.4B**). These different arrangements in the genomes from closely related species provide a glimpse of how the genes of cassette-containing peptides may form out of an independent single short ORF through gene duplication (**Fig. 8.2.4D (1)**), fusion (**Fig. 8.2.4D (2)**), and subsequent reduction of the role of the internal leader to that of a

recognition sequence of proteases (**Fig. 8.2.4D (4)**). Further multiplication of cassette-containing precursor genes may lead to situations found in several cyanobactin clusters (**Fig. 8.2.4D (5)** [45]).

Strikingly, only the BGCs with fused precursors contain an additional gene, which is a predicted protease (**Fig. 8.2.4A**, orange). This enzyme may be involved in the processing required to produce individual modified core parts. The acquisition of an additional protease gene may be the next step after the fusion of two independent ORF in the course of cassette-containing BGC evolution (**Fig. 8.2.4D (3)**). We named the products of this family of BGCs *flavazolicins*. Characterization of the products encoded in these BGCs and establishment of details of their biosynthesis and function appears to be an exciting direction for future work.

Chapter summary

In this chapter we performed a systematic search for novel BGCs of putative LAPs in the publicly available bacterial genomes, focusing on the BGCs including the genes of YcaO domain-containing enzymes and E1-like partner proteins. Our analysis identified the clusters of already known compounds and allowed us to predict several groups of new putative LAPs, for five of which we provide a detailed description of the BGC architecture and encoded precursor peptides. This work expands our understanding of the diversity of LAPs' genomic landscapes outside of well-defined groups and demonstrates how powerful the genome mining approach may be.

Chapter 9. Conclusion

In this work, we discovered and comprehensively characterized a new azole-modified peptide antibiotic phazolicin (PHZ), produced by the plant-symbiotic bacterium *Rhizobium* sp. Pop5. PHZ production is guided by the *phzEACBDR* biosynthetic gene cluster and includes posttranslational installation of eight azole cycles into the PhzA precursor peptide. Mature PHZ is imported into the cells of susceptible bacteria closely related to the producer via BacA and YejABEF transporters of the inner membrane. Such dual mode of internalization dramatically decreases the level of PHZ resistance acquisition. PHZ targets the large ribosomal subunit and inhibits bacterial translation via the obstruction of the ribosome exit tunnel. Azole cycles installed in the PHZ structure mediate its binding to the 23S rRNA. In the ribosome exit tunnel PHZ also interacts with the loop regions of uL4 and uL22 ribosomal proteins; the amino acid sequence of the former determines the species-specificity of the antibiotic action.

phz-like biosynthetic gene clusters are found in multiple Alphaproteobacteria around the globe, which allows proposing the role of the produced compounds in the interspecies competition. Our cocultivation experiments showed that PHZ production is the only mechanism, which allows *Rhizobium* sp. Pop5 to eliminate the susceptible strains from the co-culture. However, in the experimental conditions tested we could not detect a

competitive advantage of PHZ-producing strains in the cocultivation experiments in soil and upon the nodulation of common beans *Phaseolus vulgaris*.

Finally, we performed a systematic search for previously unknown families of gene clusters guiding the production of putative LAPs and described in detail the gene composition and the features of precursor peptides for five of them. These families display features unprecedented across known LAP BGCs such as the presence of multiple genes of YcaO enzymes in one cluster, or the “cassette” organization of the precursor peptides.

Bibliography

- [1] B. Petrovska, "Historical review of medicinal plants' usage," *Pharmacogn. Rev.*, vol. 6, no. 11, p. 1, 2012, doi: 10.4103/0973-7847.95849.
- [2] M. I. Hutchings, A. W. Truman, and B. Wilkinson, "Antibiotics: past, present and future," *Curr. Opin. Microbiol.*, vol. 51, pp. 72–80, Oct. 2019, doi: 10.1016/j.mib.2019.10.008.
- [3] G. D. Wright, "Opportunities for natural products in 21st century antibiotic discovery.," *Nat. Prod. Rep.*, vol. 34, no. 7, pp. 694–701, 2017, doi: 10.1039/c7np00019g.
- [4] R. H. Baltz, "Natural product drug discovery in the genomic era: realities, conjectures, misconceptions, and opportunities," *J. Ind. Microbiol. Biotechnol.*, vol. 46, no. 3–4, pp. 281–299, 2019, doi: 10.1007/s10295-018-2115-4.
- [5] A. R. Carroll, B. R. Copp, R. A. Davis, R. A. Keyzers, and M. R. Prinsep, "Marine natural products.," *Nat. Prod. Rep.*, vol. 38, no. 2, pp. 362–413, 2021, doi: 10.1039/d0np00089b.
- [6] Z. Qin *et al.*, "Formicamycins, antibacterial polyketides produced by *Streptomyces formicae* isolated from African *Tetraoponera* plant-ants.," *Chem. Sci.*, vol. 8, no. 4, pp. 3218–3227, Apr. 2017, doi: 10.1039/c6sc04265a.
- [7] E. Tortorella *et al.*, "Antibiotics from Deep-Sea Microorganisms: Current Discoveries and Perspectives.," *Mar. Drugs*, vol. 16, no. 10, pp. 1–16, Sep. 2018, doi: 10.3390/md16100355.
- [8] L. L. Ling *et al.*, "A new antibiotic kills pathogens without detectable resistance.," *Nature*, vol. 517, no. 7535, pp. 455–9, Jan. 2015, doi: 10.1038/nature14098.
- [9] L. Mahler *et al.*, "Highly parallelized droplet cultivation and prioritization of antibiotic producers from natural microbial communities," *Elife*, vol. 10, pp. 1–23, Mar. 2021, doi: 10.7554/eLife.64774.
- [10] S. S. Terekhov *et al.*, "Ultrahigh-throughput functional profiling of microbiota communities," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 115, no. 38, pp. 9551–9556, 2018, doi: 10.1073/pnas.1811250115.
- [11] N. Ziemert, M. Alanjary, and T. Weber, "The evolution of genome mining in microbes - a review.," *Nat. Prod. Rep.*, vol. 33, no. 8, pp. 988–1005, Aug. 2016, doi: 10.1039/c6np00025h.
- [12] K. Blin *et al.*, "antiSMASH 6.0: improving cluster detection and comparison capabilities," *Nucleic Acids Res.*, vol. 49, no. W1, pp. W29–W35, Jul. 2021, doi: 10.1093/nar/gkab335.
- [13] S. A. Kautsar *et al.*, "MIBiG 2.0: a repository for biosynthetic gene clusters of known function.," *Nucleic Acids Res.*, Oct. 2019, doi: 10.1093/nar/gkz882.
- [14] S. D. Bentley *et al.*, "Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2).," *Nature*, vol. 417, no. 6885, pp. 141–7, May 2002, doi: 10.1038/417141a.
- [15] P. J. Rutledge and G. L. Challis, "Discovery of microbial natural products by activation of silent biosynthetic gene clusters," *Nat. Rev. Microbiol.*, vol. 13, no. 8, pp. 509–523, Aug. 2015, doi: 10.1038/nrmicro3496.
- [16] E. C. O'Neill, M. Schorn, C. B. Larson, and N. Millán-Aguiñaga, "Targeted antibiotic discovery through biosynthesis-associated resistance determinants: target directed genome mining," *Crit. Rev. Microbiol.*, vol. 45, no. 3, pp. 255–277, May 2019, doi: 10.1080/1040841X.2019.1590307.
- [17] A.-C. Van Baelen, P. Robin, P. Kessler, A. Maïga, N. Gilles, and D. Servent, "Structural and Functional Diversity of Animal Toxins Interacting With GPCRs.," *Front. Mol. Biosci.*, vol. 9, p. 811365, 2022, doi: 10.3389/fmolb.2022.811365.
- [18] B. I. Posner and S. A. Laporte, "Cellular signalling: Peptide hormones and growth factors.," *Prog. Brain Res.*, vol. 181, pp. 1–16, 2010, doi: 10.1016/S0079-6123(08)81001-1.
- [19] D. N. McBrayer, C. D. Cameron, and Y. Tal-Gan, "Development and utilization of peptide-based quorum sensing modulators in Gram-positive bacteria.," *Org. Biomol. Chem.*, vol. 18, no. 37, pp. 7273–7290, 2020, doi: 10.1039/d0ob01421d.
- [20] L. C. P. Vilas Boas, M. L. Campos, R. L. A. Berlanda, N. de Carvalho Neves, and O. L. Franco,

- “Antiviral peptides as promising therapeutic drugs.,” *Cell. Mol. Life Sci.*, vol. 76, no. 18, pp. 3525–3542, Sep. 2019, doi: 10.1007/s00018-019-03138-w.
- [21] G. Swayambhu, M. Bruno, A. M. Gulick, and B. A. Pfeifer, “Siderophore natural products as pharmaceutical agents.,” *Curr. Opin. Biotechnol.*, vol. 69, pp. 242–251, 2021, doi: 10.1016/j.copbio.2021.01.021.
- [22] M. Montalbán-López *et al.*, “New developments in RiPP discovery, enzymology and engineering.,” *Nat. Prod. Rep.*, Sep. 2020, doi: 10.1039/d0np00027b.
- [23] P. G. Arnison *et al.*, “Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature.,” *Nat. Prod. Rep.*, vol. 30, no. 1, pp. 108–60, Jan. 2013, doi: 10.1039/c2np20085f.
- [24] S. C. Lu, “Glutathione synthesis.,” *Biochim. Biophys. Acta*, vol. 1830, no. 5, pp. 3143–53, May 2013, doi: 10.1016/j.bbagen.2012.09.008.
- [25] A. Gautam, R. Vyas, and R. Tewari, “Peptidoglycan biosynthesis machinery: a rich source of drug targets.,” *Crit. Rev. Biotechnol.*, vol. 31, no. 4, pp. 295–336, Dec. 2011, doi: 10.3109/07388551.2010.525498.
- [26] R. D. Süßmuth and A. Mainz, “Nonribosomal Peptide Synthesis—Principles and Prospects.,” *Angew. Chemie - Int. Ed.*, vol. 56, no. 14, pp. 3770–3821, 2017, doi: 10.1002/anie.201609079.
- [27] F. Kudo, A. Miyanaga, and T. Eguchi, “Structural basis of the nonribosomal codes for nonproteinogenic amino acid selective adenylation enzymes in the biosynthesis of natural products.,” *J. Ind. Microbiol. Biotechnol.*, vol. 46, no. 3–4, pp. 515–536, Mar. 2019, doi: 10.1007/s10295-018-2084-7.
- [28] H. von Döhren, R. Dieckmann, and M. Pavela-Vrancic, “The nonribosomal code.,” *Chem. Biol.*, vol. 6, no. 10, pp. R273-9, Oct. 1999, doi: 10.1016/s1074-5521(00)80014-9.
- [29] D. F. Ackerley, “Cracking the Nonribosomal Code.,” *Cell Chem. Biol.*, vol. 23, no. 5, pp. 535–537, 2016, doi: 10.1016/j.chembiol.2016.05.001.
- [30] C. T. Walsh, R. V. O’Brien, and C. Khosla, “Nonproteinogenic Amino Acid Building Blocks for Nonribosomal Peptide and Hybrid Polyketide Scaffolds.,” *Angew. Chemie Int. Ed.*, vol. 52, no. 28, pp. 7098–7124, Jul. 2013, doi: 10.1002/anie.201208344.
- [31] L. Xu *et al.*, “Complete genome sequence and comparative genomic analyses of the vancomycin-producing *Amycolatopsis orientalis*.,” *BMC Genomics*, vol. 15, p. 363, May 2014, doi: 10.1186/1471-2164-15-363.
- [32] L. Du, C. Sánchez, M. Chen, D. J. Edwards, and B. Shen, “The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus* ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase.,” *Chem. Biol.*, vol. 7, no. 8, pp. 623–42, Aug. 2000, doi: 10.1016/s1074-5521(00)00011-9.
- [33] X. Yang, P. Feng, Y. Yin, K. Bushley, J. W. Spatafora, and C. Wang, “Cyclosporine Biosynthesis in *Tolypocladium inflatum* Benefits Fungal Adaptation to the Environment.,” *MBio*, vol. 9, no. 5, Nov. 2018, doi: 10.1128/mBio.01211-18.
- [34] M. A. Funk and W. A. van der Donk, “Ribosomal Natural Products, Tailored To Fit.,” *Acc. Chem. Res.*, vol. 50, no. 7, pp. 1577–1586, Jul. 2017, doi: 10.1021/acs.accounts.7b00175.
- [35] D. Tsibulskaya *et al.*, “The Product of *Yersinia pseudotuberculosis* *mcc* Operon Is a Peptide-Cytidine Antibiotic Activated Inside Producing Cells by the TldD/E Protease.,” *J. Am. Chem. Soc.*, vol. 139, no. 45, pp. 16178–16187, 2017, doi: 10.1021/jacs.7b07118.
- [36] D. Ghilarov *et al.*, “The Origins of Specificity in the Microcin-Processing Protease TldD/E.,” *Structure*, vol. 25, no. 10, pp. 1549–1561.e5, Oct. 2017, doi: 10.1016/j.str.2017.08.006.
- [37] D. Xue *et al.*, “Correlational networking guides the discovery of unclustered lanthipeptide protease-encoding genes.,” *Nat. Commun.*, vol. 13, no. 1, p. 1647, Dec. 2022, doi: 10.1038/s41467-022-29325-1.
- [38] J. P. Klinman and F. Bonnot, “Intrigues and Intricacies of the Biosynthetic Pathways for the

- Enzymatic Quinocofactors: PQQ, TTQ, CTQ, TPQ, and LTQ,” *Chem. Rev.*, vol. 114, no. 8, pp. 4343–4365, Apr. 2014, doi: 10.1021/cr400475g.
- [39] D. H. Haft, “A strain-variable bacteriocin in *Bacillus anthracis* and *Bacillus cereus* with repeated Cys-Xaa-Xaa motifs,” *Biol. Direct*, vol. 4, p. 15, Jan. 2009, doi: 10.1186/1745-6150-4-15.
- [40] J. Santos-Aberturas *et al.*, “Uncovering the unexplored diversity of thioamidated ribosomal peptides in Actinobacteria using the RiPPER genome mining tool,” *Nucleic Acids Res.*, vol. 47, no. 9, pp. 4624–4637, May 2019, doi: 10.1093/nar/gkz192.
- [41] I. Zukher *et al.*, “Ribosome-controlled transcription termination is essential for the production of antibiotic microcin C,” *Nucleic Acids Res.*, vol. 42, no. 19, pp. 11891–11902, Oct. 2014, doi: 10.1093/nar/gku880.
- [42] A. Grigoreva *et al.*, “Identification and characterization of andalusicin: N-terminally dimethylated class III lantibiotic from *Bacillus thuringiensis* sv. *andalousiensis*,” *iScience*, vol. 24, no. 5, p. 102480, May 2021, doi: 10.1016/j.isci.2021.102480.
- [43] L. Huo, X. Zhao, J. Z. Acedo, P. Estrada, S. K. Nair, and W. A. Donk, “Characterization of a Dehydratase and Methyltransferase in the Biosynthesis of Ribosomally Synthesized and Post-translationally Modified Peptides in Lachnospiraceae,” *ChemBioChem*, vol. 21, no. 1–2, pp. 190–199, Jan. 2020, doi: 10.1002/cbic.201900483.
- [44] Y. Y. Shim, L. W. Young, P. G. Arnison, E. Gilding, and M. J. T. Reaney, “Proposed systematic nomenclature for orbitides,” *J. Nat. Prod.*, vol. 78, no. 4, pp. 645–52, Apr. 2015, doi: 10.1021/np500802p.
- [45] W. Gu, D. Sardar, E. Pierce, and E. W. Schmidt, “Roads to Rome: Role of Multiple Cassettes in Cyanobactin RiPP Biosynthesis,” *J. Am. Chem. Soc.*, vol. 140, no. 47, pp. 16213–16221, 2018, doi: 10.1021/jacs.8b09328.
- [46] W. Ding, W.-Q. Liu, Y. Jia, Y. Li, W. A. van der Donk, and Q. Zhang, “Biosynthetic investigation of phomopsins reveals a widespread pathway for ribosomal natural products in Ascomycetes,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 13, pp. 3521–6, Mar. 2016, doi: 10.1073/pnas.1522907113.
- [47] G. M. Rubin and Y. Ding, “Recent advances in the biosynthesis of RiPPs from multicore-containing precursor peptides,” *J. Ind. Microbiol. Biotechnol.*, vol. 47, no. 9–10, pp. 659–674, Oct. 2020, doi: 10.1007/s10295-020-02289-1.
- [48] X. Yang *et al.*, “A lanthipeptide library used to identify a protein–protein interaction inhibitor,” *Nat. Chem. Biol.*, vol. 14, no. 4, pp. 375–380, Apr. 2018, doi: 10.1038/s41589-018-0008-5.
- [49] S. Luo and S. H. Dong, “Recent advances in the discovery and biosynthetic study of eukaryotic RiPP natural products,” *Molecules*, vol. 24, no. 8, pp. 1–15, 2019, doi: 10.3390/molecules24081541.
- [50] B. J. Burkhart, C. J. Schwalen, G. Mann, J. H. Naismith, and D. A. Mitchell, “YcaO-Dependent Posttranslational Amide Activation: Biosynthesis, Structure, and Function,” *Chem. Rev.*, vol. 117, no. 8, pp. 5389–5456, Apr. 2017, doi: 10.1021/acs.chemrev.6b00623.
- [51] L. M. Repka, J. R. Chekan, S. K. Nair, and W. A. van der Donk, “Mechanistic Understanding of Lanthipeptide Biosynthetic Enzymes,” *Chem. Rev.*, vol. 117, no. 8, pp. 5457–5520, Apr. 2017, doi: 10.1021/acs.chemrev.6b00591.
- [52] Y. Chen, J. Wang, G. Li, Y. Yang, and W. Ding, “Current Advancements in Sactipeptide Natural Products,” *Front. Chem.*, vol. 9, May 2021, doi: 10.3389/fchem.2021.595991.
- [53] H. Lee, M. Choi, J.-U. Park, H. Roh, and S. Kim, “Genome Mining Reveals High Topological Diversity of ω -Ester-Containing Peptides and Divergent Evolution of ATP-Grasp Macrocyclases,” *J. Am. Chem. Soc.*, vol. 142, no. 6, pp. 3013–3023, Feb. 2020, doi: 10.1021/jacs.9b12076.
- [54] A. R. Weiz *et al.*, “Harnessing the Evolvability of Tricyclic Microviridins To Dissect Protease-Inhibitor Interactions,” *Angew. Chemie Int. Ed.*, vol. 53, no. 14, pp. 3735–3738, Apr. 2014, doi: 10.1002/anie.201309721.
- [55] K. J. Molohon *et al.*, “Structure determination and interception of biosynthetic intermediates for the plantazolicin class of highly discriminating antibiotics,” *ACS Chem. Biol.*, vol. 6, no. 12, pp. 1307–

- 13, Dec. 2011, doi: 10.1021/cb200339d.
- [56] N. Liu *et al.*, “Unique post-translational oxime formation in the biosynthesis of the azolemycin complex of novel ribosomal peptides from *Streptomyces* sp. FXJ1.264.,” *Chem. Sci.*, vol. 7, no. 1, pp. 482–488, Jan. 2016, doi: 10.1039/c5sc03021h.
- [57] T. H. Eyles, N. M. Vior, R. Lacret, and A. W. Truman, “Understanding thioamide biosynthesis using pathway engineering and untargeted metabolomics,” *Chem. Sci.*, vol. 12, no. 20, pp. 7138–7150, 2021, doi: 10.1039/D0SC06835G.
- [58] N. S. van der Velden, N. Kälin, M. J. Helf, J. Piel, M. F. Freeman, and M. Künzler, “Autocatalytic backbone N-methylation in a family of ribosomal peptide natural products,” *Nat. Chem. Biol.*, vol. 13, no. 8, pp. 833–835, Aug. 2017, doi: 10.1038/nchembio.2393.
- [59] A. Ökesli, L. E. Cooper, E. J. Fogle, and W. A. van der Donk, “Nine Post-translational Modifications during the Biosynthesis of Cinnamycin,” *J. Am. Chem. Soc.*, vol. 133, no. 34, pp. 13753–13760, Aug. 2011, doi: 10.1021/ja205783f.
- [60] W. J. K. Crone, N. M. Vior, J. Santos-Aberturas, L. G. Schmitz, F. J. Leeper, and A. W. Truman, “Dissecting Botromycin Biosynthesis Using Comparative Untargeted Metabolomics,” *Angew. Chemie - Int. Ed.*, vol. 55, no. 33, pp. 9639–9643, 2016, doi: 10.1002/anie.201604304.
- [61] M. Purushothaman, S. Sarkar, M. Morita, M. Gugger, E. W. Schmidt, and B. I. Morinaka, “Genome-Mining-Based Discovery of the Cyclic Peptide Tolypamide and TolF, a Ser/Thr Forward O - Prenyltransferase,” *Angew. Chemie Int. Ed.*, vol. 60, no. 15, pp. 8460–8465, Apr. 2021, doi: 10.1002/anie.202015975.
- [62] T. Hamada, S. Matsunaga, G. Yano, and N. Fusetani, “Polytheonamides A and B, highly cytotoxic, linear polypeptides with unprecedented structural features, from the marine sponge, *Theonella swinhoei*,” *J. Am. Chem. Soc.*, vol. 127, no. 1, pp. 110–118, 2005, doi: 10.1021/ja045749e.
- [63] L. C. Foulston and M. J. Bibb, “Microbisporicin gene cluster reveals unusual features of lantibiotic biosynthesis in actinomycetes,” *Proc. Natl. Acad. Sci.*, vol. 107, no. 30, pp. 13461–13466, Jul. 2010, doi: 10.1073/pnas.1008285107.
- [64] J. I. Tietz *et al.*, “A new genome-mining tool redefines the lasso peptide biosynthetic landscape,” *Nat. Chem. Biol.*, vol. 13, no. 5, pp. 470–478, May 2017, doi: 10.1038/nchembio.2319.
- [65] C. Zong, W. L. Cheung-Lee, H. E. Elashal, M. Raj, and A. J. Link, “Albusnodin: an acetylated lasso peptide from *Streptomyces albus*,” *Chem. Commun.*, vol. 54, no. 11, pp. 1339–1342, 2018, doi: 10.1039/C7CC08620B.
- [66] S. Zhu *et al.*, “Insights into the Unique Phosphorylation of the Lasso Peptide Paeninodin.,” *J. Biol. Chem.*, vol. 291, no. 26, pp. 13662–78, Jun. 2016, doi: 10.1074/jbc.M116.722108.
- [67] J. Martins and V. Vasconcelos, “Cyanobactins from cyanobacteria: Current genetic and chemical state of knowledge,” *Mar. Drugs*, vol. 13, no. 11, pp. 6910–6946, 2015, doi: 10.3390/md13116910.
- [68] B. J. Lethbridge *et al.*, “Post translational modifications of Trifolitoxin: a blue fluorescent peptide antibiotic,” *J. Antibiot. (Tokyo)*, vol. 75, no. 3, pp. 125–135, Mar. 2022, doi: 10.1038/s41429-021-00497-0.
- [69] X.-H. Jian *et al.*, “Analysis of YM-216391 Biosynthetic Gene Cluster and Improvement of the Cyclopeptide Production in a Heterologous Host,” *ACS Chem. Biol.*, vol. 7, no. 4, pp. 646–651, Apr. 2012, doi: 10.1021/cb200479f.
- [70] A. Tocchetti *et al.*, “Capturing Linear Intermediates and C-Terminal Variants during Maturation of the Thiopeptide GE2270,” *Chem. Biol.*, vol. 20, no. 8, pp. 1067–1077, Aug. 2013, doi: 10.1016/j.chembiol.2013.07.005.
- [71] J. O. Melby, N. J. Nard, and D. A. Mitchell, “Thiazole/oxazole-modified microcins: complex natural products from ribosomal templates.,” *Curr. Opin. Chem. Biol.*, vol. 15, no. 3, pp. 369–78, Jun. 2011, doi: 10.1016/j.cbpa.2011.02.027.
- [72] T. Prange, A. Ducruix, C. Pascard, and J. Lunel, “Structure of nosipeptide, a polythiazole-containing antibiotic.,” *Nature*, vol. 265, no. 5590, pp. 189–90, Jan. 1977, doi: 10.1038/265189a0.

- [73] L. Franz, U. Kazmaier, A. W. Truman, and J. Koehnke, "Bottromycins - Biosynthesis, synthesis and activity," *Nat. Prod. Rep.*, vol. 38, no. 9, pp. 1659–1683, 2021, doi: 10.1039/d0np00097c.
- [74] A. Carroll *et al.*, "Patellins 1-6 and Trunkamide A: Novel Cyclic Hexa-, Hepta- and Octa-peptides From Colonial Ascidiars, *Lissoclinum* sp.," *Aust. J. Chem.*, vol. 49, no. 6, p. 659, 1996, doi: 10.1071/CH9960659.
- [75] M.-Y. Kim, H. Vankayalapati, K. Shin-Ya, K. Wierzba, and L. H. Hurley, "Telomestatin, a potent telomerase inhibitor that interacts quite specifically with the human telomeric intramolecular g-quadruplex.," *J. Am. Chem. Soc.*, vol. 124, no. 10, pp. 2098–9, Mar. 2002, doi: 10.1021/ja017308q.
- [76] M. Metelev *et al.*, "Klebsazolicin inhibits 70S ribosome by obstructing the peptide exit tunnel.," *Nat. Chem. Biol.*, vol. 13, no. 10, pp. 1129–1136, Oct. 2017, doi: 10.1038/nchembio.2462.
- [77] X. Just-Baringo, F. Albericio, and M. Álvarez, "Thiopeptide antibiotics: retrospective and recent advances.," *Mar. Drugs*, vol. 12, no. 1, pp. 317–51, Jan. 2014, doi: 10.3390/md12010317.
- [78] Q. Zhang and W. Liu, "Biosynthesis of thiopeptide antibiotics and their pathway engineering.," *Nat. Prod. Rep.*, vol. 30, no. 2, pp. 218–26, Feb. 2013, doi: 10.1039/c2np20107k.
- [79] C. J. Schwalen, G. A. Hudson, B. Kille, and D. A. Mitchell, "Bioinformatic Expansion and Discovery of Thiopeptide Antibiotics.," *J. Am. Chem. Soc.*, vol. 140, no. 30, pp. 9494–9501, 2018, doi: 10.1021/jacs.8b03896.
- [80] Q. Zheng, H. Fang, and W. Liu, "Post-translational modifications involved in the biosynthesis of thiopeptide antibiotics.," *Org. Biomol. Chem.*, vol. 15, no. 16, pp. 3376–3390, Apr. 2017, doi: 10.1039/c7ob00466d.
- [81] B. Clough, M. Strath, P. Preiser, P. Denny, and I. R. Wilson, "Thiostrepton binds to malarial plastid rRNA.," *FEBS Lett.*, vol. 406, no. 1–2, pp. 123–5, Apr. 1997, doi: 10.1016/s0014-5793(97)00241-x.
- [82] C. M. Czekster, Y. Ge, and J. H. Naismith, "Mechanisms of cyanobactin biosynthesis.," *Curr. Opin. Chem. Biol.*, vol. 35, pp. 80–88, Dec. 2016, doi: 10.1016/j.cbpa.2016.08.029.
- [83] K. Sivonen, N. Leikoski, D. P. Fewer, and J. Jokela, "Cyanobactins-ribosomal cyclic peptides produced by cyanobacteria," *Appl. Microbiol. Biotechnol.*, vol. 86, no. 5, pp. 1213–1225, 2010, doi: 10.1007/s00253-010-2482-x.
- [84] N. Tanaka, K. Sashikata, H. Yamaguchi, and H. Umezawa, "Inhibition of protein synthesis by bottromycin A2 and its hydrazide.," *J. Biochem.*, vol. 60, no. 4, pp. 405–10, Oct. 1966, doi: 10.1093/oxfordjournals.jbchem.a128451.
- [85] T. Otaka and A. Kaji, "Mode of action of bottromycin A2: effect of bottromycin A2 on polysomes.," *FEBS Lett.*, vol. 153, no. 1, pp. 53–9, Mar. 1983, doi: 10.1016/0014-5793(83)80118-5.
- [86] Z.-F. Pei, M.-J. Yang, K. Zhang, X.-H. Jian, and G.-L. Tang, "Heterologous characterization of mechercharmycin A biosynthesis reveals alternative insights into post-translational modifications for RiPPs.," *Cell Chem. Biol.*, Aug. 2021, doi: 10.1016/j.chembiol.2021.08.005.
- [87] I. Kawewan *et al.*, "Isolation and structure determination of a new cytotoxic peptide, curacozole, from *Streptomyces curacoi* based on genome mining," *J. Antibiot. (Tokyo)*, vol. 72, no. 1, pp. 1–7, Jan. 2019, doi: 10.1038/s41429-018-0105-4.
- [88] K. Amagai *et al.*, "Identification of a gene cluster for telomestatin biosynthesis and heterologous expression using a specific promoter in a clean host," *Sci. Rep.*, vol. 7, no. 1, p. 3382, Dec. 2017, doi: 10.1038/s41598-017-03308-5.
- [89] E. W. Triplett and T. M. Barta, "Trifolitoxin Production and Nodulation Are Necessary for the Expression of Superior Nodulation Competitiveness by *Rhizobium leguminosarum* bv. *trifolii* Strain T24 on Clover.," *Plant Physiol.*, vol. 85, no. 2, pp. 335–42, Oct. 1987, doi: 10.1104/pp.85.2.335.
- [90] B. A. Pfeifer, C. C. C. Wang, C. T. Walsh, and C. Khosla, "Biosynthesis of Yersiniabactin, a Complex Polyketide-Nonribosomal Peptide, Using *Escherichia coli* as a Heterologous Host," *Appl. Environ. Microbiol.*, vol. 69, no. 11, pp. 6698–6702, Nov. 2003, doi: 10.1128/AEM.69.11.6698-6702.2003.
- [91] K. L. Dunbar, J. O. Melby, and D. A. Mitchell, "YcaO domains use ATP to activate amide backbones

- during peptide cyclodehydrations,” *Nat. Chem. Biol.*, vol. 8, no. 6, pp. 569–575, 2012, doi: 10.1038/nchembio.944.
- [92] K. L. Dunbar, J. R. Chekan, C. L. Cox, B. J. Burkhardt, S. K. Nair, and D. A. Mitchell, “Discovery of a new ATP-binding motif involved in peptidic azoline biosynthesis.,” *Nat. Chem. Biol.*, vol. 10, no. 10, pp. 823–9, Oct. 2014, doi: 10.1038/nchembio.1608.
- [93] D. Ghilarov *et al.*, “Architecture of Microcin B17 Synthetase: An Octameric Protein Complex Converting a Ribosomally Synthesized Peptide into a DNA Gyrase Poison.,” *Mol. Cell*, vol. 73, no. 4, pp. 749–762, Feb. 2019, doi: 10.1016/j.molcel.2018.11.032.
- [94] B. J. Burkhardt, G. A. Hudson, K. L. Dunbar, and D. A. Mitchell, “A prevalent peptide-binding domain guides ribosomal natural product biosynthesis.,” *Nat. Chem. Biol.*, vol. 11, no. 8, pp. 564–70, Aug. 2015, doi: 10.1038/nchembio.1856.
- [95] A. Liu *et al.*, “Functional elucidation of TfuA in peptide backbone thioamidation,” *Nat. Chem. Biol.*, vol. 17, no. 5, pp. 585–592, May 2021, doi: 10.1038/s41589-021-00771-0.
- [96] C. J. Schwalen, G. A. Hudson, S. Kosol, N. Mahanta, G. L. Challis, and D. A. Mitchell, “In Vitro Biosynthetic Studies of Bottromycin Expand the Enzymatic Capabilities of the YcaO Superfamily.,” *J. Am. Chem. Soc.*, vol. 139, no. 50, pp. 18154–18157, Dec. 2017, doi: 10.1021/jacs.7b09899.
- [97] S. Gao, Y. Ge, A. F. Bent, U. Schwarz-Linek, and J. H. Naismith, “Oxidation of the Cyanobactin Precursor Peptide Is Independent of the Leader Peptide and Operates in a Defined Order,” *Biochemistry*, vol. 57, no. 41, pp. 5996–6002, Oct. 2018, doi: 10.1021/acs.biochem.8b00835.
- [98] L. Franz, S. Adam, J. Santos-Aberturas, A. W. Truman, and J. Koehnke, “Macroamidine Formation in Bottromycins Is Catalyzed by a Divergent YcaO Enzyme.,” *J. Am. Chem. Soc.*, vol. 139, no. 50, pp. 18158–18161, Dec. 2017, doi: 10.1021/jacs.7b09898.
- [99] D. Y. Travin *et al.*, “Biosynthesis of Translation Inhibitor Klebsazolicin Proceeds through Heterocyclization and N-Terminal Amidine Formation Catalyzed by a Single YcaO Enzyme.,” *J. Am. Chem. Soc.*, vol. 140, no. 16, pp. 5625–5633, Apr. 2018, doi: 10.1021/jacs.8b02277.
- [100] A. H. Russell, N. M. Vior, E. S. Hems, R. Lacret, and A. W. Truman, “Discovery and characterisation of an amidine-containing ribosomally-synthesised peptide that is widely distributed in nature,” *Chem. Sci.*, vol. 12, no. 35, pp. 11769–11778, 2021, doi: 10.1039/D1SC01456K.
- [101] J. Liu, Z. Lin, Y. Li, Q. Zheng, D. Chen, and W. Liu, “Insights into the thioamidation of thiopeptins to enhance the understanding of the biosynthetic logic of thioamide-containing thiopeptides,” *Org. Biomol. Chem.*, vol. 17, no. 15, pp. 3727–3731, 2019, doi: 10.1039/C9OB00402E.
- [102] D. D. Nayak, N. Mahanta, D. A. Mitchell, and W. W. Metcalf, “Post-translational thioamidation of methyl-coenzyme M reductase, a key enzyme in methanogenic and methanotrophic Archaea.,” *Elife*, vol. 6, no. 1, pp. 1–18, Sep. 2017, doi: 10.7554/eLife.29218.
- [103] Z. L. Watson *et al.*, “Structure of the bacterial ribosome at 2 Å resolution,” *Elife*, vol. 9, Sep. 2020, doi: 10.7554/eLife.60482.
- [104] E. M. Molloy *et al.*, “Identification of the minimal cytolytic unit for streptolysin S and an expansion of the toxin family,” *BMC Microbiol.*, vol. 15, p. 141, 2015, doi: 10.1186/s12866-015-0464-y.
- [105] D. H. Haft, M. K. Basu, and D. A. Mitchell, “Expansion of ribosomally produced natural products: a nitrile hydratase- and Nif11-related precursor family.,” *BMC Biol.*, vol. 8, p. 70, May 2010, doi: 10.1186/1741-7007-8-70.
- [106] W. M. Parks, A. R. Bottrill, O. A. Pierrat, M. C. Durrant, and A. Maxwell, “The action of the bacterial toxin, microcin B17, on DNA gyrase.,” *Biochimie*, vol. 89, no. 4, pp. 500–7, Apr. 2007, doi: 10.1016/j.biochi.2006.12.005.
- [107] J. L. Vizán, C. Hernández-Chico, I. del Castillo, and F. Moreno, “The peptide antibiotic microcin B17 induces double-strand cleavage of DNA mediated by E. coli DNA gyrase.,” *EMBO J.*, vol. 10, no. 2, pp. 467–76, Feb. 1991, doi: 10.1002/j.1460-2075.1991.tb07969.x.
- [108] J. G. Heddle *et al.*, “The antibiotic microcin B17 is a DNA gyrase poison: characterisation of the mode of inhibition.,” *J. Mol. Biol.*, vol. 307, no. 5, pp. 1223–34, Apr. 2001, doi:

- 10.1006/jmbi.2001.4562.
- [109] F. Collin and A. Maxwell, “The Microbial Toxin Microcin B17: Prospects for the Development of New Antibacterial Agents,” *J. Mol. Biol.*, vol. 431, no. 18, pp. 3400–3426, Aug. 2019, doi: 10.1016/j.jmb.2019.05.050.
- [110] M. Metelev, M. Serebryakova, D. Ghilarov, Y. Zhao, and K. Severinov, “Structure of microcin B-like compounds produced by *Pseudomonas syringae* and species specificity of their antibacterial action,” *J. Bacteriol.*, vol. 195, no. 18, pp. 4129–37, Sep. 2013, doi: 10.1128/JB.00665-13.
- [111] J. Lee, Y.-J. J. Cho, J. Y. Yang, Y.-J. J. Jung, S. G. Hong, and O.-S. S. Kim, “Complete genome sequence of *Pseudomonas antarctica* PAMC 27494, a bacteriocin-producing psychrophile isolated from Antarctica,” *J. Biotechnol.*, vol. 259, no. January, pp. 15–18, Oct. 2017, doi: 10.1016/j.jbiotec.2017.08.013.
- [112] R. Scholz *et al.*, “Plantazolicin, a novel microcin B17/streptolysin S-like natural product from *Bacillus amyloliquefaciens* FZB42,” *J. Bacteriol.*, vol. 193, no. 1, pp. 215–24, Jan. 2011, doi: 10.1128/JB.00784-10.
- [113] K. J. Molohon *et al.*, “Plantazolicin is an ultra-narrow spectrum antibiotic that targets the *Bacillus anthracis* membrane,” *ACS Infect. Dis.*, vol. 2, no. 3, pp. 207–220, Mar. 2016, doi: 10.1021/acsinfecdis.5b00115.
- [114] M. Takuma *et al.*, “Heterologous production of coryneazolicin in *Escherichia coli*,” *J. Antibiot. (Tokyo)*, vol. 72, no. 11, pp. 800–806, 2019, doi: 10.1038/s41429-019-0212-x.
- [115] L. Chopra, G. Singh, K. Kumar Jena, and D. K. Sahoo, “Sonorensin: A new bacteriocin with potential of an anti-biofilm agent and a food biopreservative,” *Sci. Rep.*, vol. 5, pp. 1–13, 2015, doi: 10.1038/srep13412.
- [116] L. Chopra, G. Singh, V. Choudhary, and D. K. Sahoo, “Sonorensin: an antimicrobial peptide, belonging to the heterocycloanthracin subfamily of bacteriocins, from a new marine isolate, *Bacillus sonorensis* MT93,” *Appl. Environ. Microbiol.*, vol. 80, no. 10, pp. 2981–90, May 2014, doi: 10.1128/AEM.04259-13.
- [117] H. Onaka, H. Tabata, Y. Igarashi, Y. Sato, and T. Furumai, “Goadsporin, a chemical substance which promotes secondary metabolism and morphogenesis in streptomycetes. I. Purification and characterization,” *J. Antibiot. (Tokyo)*, vol. 54, no. 12, pp. 1036–44, Dec. 2001, doi: 10.7164/antibiotics.54.1036.
- [118] Y. Igarashi *et al.*, “Goadsporin, a chemical substance which promotes secondary metabolism and Morphogenesis in streptomycetes. II. Structure determination,” *J. Antibiot. (Tokyo)*, vol. 54, no. 12, pp. 1045–53, Dec. 2001, doi: 10.7164/antibiotics.54.1045.
- [119] H. Onaka, M. Nakaho, K. Hayashi, Y. Igarashi, and T. Furumai, “Cloning and characterization of the goadsporin biosynthetic gene cluster from *Streptomyces* sp. TP-A0584,” *Microbiology*, vol. 151, no. 12, pp. 3923–3933, 2005.
- [120] M. Suzuki *et al.*, “Isolation and structure determination of new linear azole-containing peptides spongiicolazolicins A and B from *Streptomyces* sp. CWH03,” *Appl. Microbiol. Biotechnol.*, vol. 105, no. 1, pp. 93–104, Jan. 2021, doi: 10.1007/s00253-020-11016-w.
- [121] D. L. Higashi *et al.*, “Activation of band 3 mediates group A *Streptococcus* streptolysin S-based beta-haemolysis,” *Nat. Microbiol.*, vol. 1, no. 2, p. 15004, Feb. 2016, doi: 10.1038/nmicrobiol.2015.4.
- [122] V. Nizet *et al.*, “Genetic locus for streptolysin S production by group A streptococcus,” *Infect. Immun.*, vol. 68, no. 7, pp. 4245–54, Jul. 2000, doi: 10.1128/IAI.68.7.4245-4254.2000.Updated.
- [123] E. M. Molloy, P. D. Cotter, C. Hill, D. A. Mitchell, and R. P. Ross, “Streptolysin S-like virulence factors: the continuing saga,” *Nat. Rev. Microbiol.*, vol. 9, no. 9, pp. 670–81, Sep. 2011, doi: 10.1038/nrmicro2624.
- [124] J. J. Quereda *et al.*, “Listeriolysin S Is a Streptolysin S-Like Virulence Factor That Targets Exclusively Prokaryotic Cells In Vivo,” *MBio*, vol. 8, no. 2, pp. 1–15, Apr. 2017, doi: 10.1128/mBio.00259-17.

- [125] J. Meza-Torres *et al.*, “Listeriolysin S: A bacteriocin from *Listeria monocytogenes* that induces membrane permeabilization in a contact-dependent manner,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 118, no. 40, pp. 1–11, 2021, doi: 10.1073/pnas.2108155118.
- [126] P. D. Cotter *et al.*, “Listeriolysin S, a novel peptide haemolysin associated with a subset of lineage I *Listeria monocytogenes*,” *PLoS Pathog.*, vol. 4, no. 9, p. e1000144, Jan. 2008, doi: 10.1371/journal.ppat.1000144.
- [127] D. J. Gonzalez *et al.*, “Clostridiolysin S, a post-translationally modified biotoxin from *Clostridium botulinum*,” *J. Biol. Chem.*, vol. 285, no. 36, pp. 28220–8, Sep. 2010, doi: 10.1074/jbc.M110.118554.
- [128] E. A. Schwinghamer and R. P. Belkengren, “Inhibition of rhizobia by a strain of *Rhizobium trifolii*: Some properties of the antibiotic and of the strain,” *Arch. Mikrobiol.*, vol. 64, no. 2, pp. 130–145, 1968, doi: 10.1007/BF00406972.
- [129] E. W. Triplett, “Isolation of genes involved in nodulation competitiveness from *Rhizobium leguminosarum* bv. *trifolii* T24,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 85, no. 11, pp. 3810–4, Jun. 1988, doi: 10.1073/pnas.85.11.3810.
- [130] B. Breil, J. Borneman, and E. W. Triplett, “A newly discovered gene, *tfuA*, involved in the production of the ribosomally synthesized peptide antibiotic trifolitolxin,” *J. Bacteriol.*, vol. 178, no. 14, pp. 4150–6, Jul. 1996, doi: 10.1128/jb.178.14.4150-4156.1996.
- [131] A. J. Scupham and E. W. Triplett, “Determination of the amino acid residues required for the activity of the anti-rhizobial peptide antibiotic trifolitolxin,” *J. Appl. Microbiol.*, vol. 100, no. 3, pp. 500–507, Mar. 2006, doi: 10.1111/j.1365-2672.2005.02803.x.
- [132] Y. Li and S. Rebuffat, “The manifold roles of microbial ribosomal peptide-based natural products in physiology and ecology,” *J. Biol. Chem.*, vol. 295, no. 1, pp. 34–54, 2020, doi: 10.1074/jbc.REV119.006545.
- [133] E. W. Triplett, “Construction of a Symbiotically Effective Strain of *Rhizobium leguminosarum* bv. *trifolii* with Increased Nodulation Competitiveness,” *Appl. Environ. Microbiol.*, vol. 56, no. 1, pp. 98–103, Jan. 1990, doi: 10.1128/aem.56.1.98-103.1990.
- [134] A. H. Bosworth, B. T. Breil, and E. W. Triplett, “Production of the anti-rhizobial peptide, trifolitolxin, in sterile soils by *Rhizobium leguminosarum* bv. *trifolii* T24,” *Soil Biol. Biochem.*, vol. 25, no. 7, pp. 829–832, Jul. 1993, doi: 10.1016/0038-0717(93)90083-N.
- [135] P. Poole, V. Ramachandran, and J. Terpolilli, “Rhizobia: from saprophytes to endosymbionts,” *Nat. Rev. Microbiol.*, vol. 16, no. 5, pp. 291–303, May 2018, doi: 10.1038/nrmicro.2017.171.
- [136] B. M. Hoffman, D. Lukoyanov, Z.-Y. Yang, D. R. Dean, and L. C. Seefeldt, “Mechanism of Nitrogen Fixation by Nitrogenase: The Next Stage,” *Chem. Rev.*, vol. 114, no. 8, pp. 4041–4062, Apr. 2014, doi: 10.1021/cr400641x.
- [137] B. A. Geddes and I. J. Oresnik, “Physiology, genetics, and biochemistry of carbon metabolism in the alphaproteobacterium *Sinorhizobium meliloti*,” *Can. J. Microbiol.*, vol. 60, no. 8, pp. 491–507, Aug. 2014, doi: 10.1139/cjm-2014-0306.
- [138] H. M. Fischer, “Genetic regulation of nitrogen fixation in rhizobia,” *Microbiol. Rev.*, vol. 58, no. 3, pp. 352–386, Sep. 1994, doi: 10.1128/mr.58.3.352-386.1994.
- [139] C. Elmerich, B. L. Dreyfus, G. Reyssset, and J. P. Aubert, “Genetic analysis of nitrogen fixation in a tropical fast-growing *Rhizobium*,” *EMBO J.*, vol. 1, no. 4, pp. 499–503, 1982, doi: 10.1002/j.1460-2075.1982.tb01197.x.
- [140] B. Horvath, C. W. Bachem, J. Schell, and A. Kondorosi, “Host-specific regulation of nodulation genes in *Rhizobium* is mediated by a plant-signal, interacting with the *nodD* gene product,” *EMBO J.*, vol. 6, no. 4, pp. 841–8, Apr. 1987, doi: 10.1002/j.1460-2075.1987.tb04829.x.
- [141] M. C. Peck, R. F. Fisher, and S. R. Long, “Diverse Flavonoids Stimulate NodD1 Binding to *nod* Gene Promoters in *Sinorhizobium meliloti*,” *J. Bacteriol.*, vol. 188, no. 15, pp. 5417–5427, Aug. 2006, doi: 10.1128/JB.00376-06.

- [142] J. Dénarié, F. Debellé, and J. C. Promé, “Rhizobium lipo-chitooligosaccharide nodulation factors: signaling molecules mediating recognition and morphogenesis.,” *Annu. Rev. Biochem.*, vol. 65, no. 1, pp. 503–35, Jun. 1996, doi: 10.1146/annurev.bi.65.070196.002443.
- [143] Q. Wang, J. Liu, and H. Zhu, “Genetic and Molecular Mechanisms Underlying Symbiotic Specificity in Legume-Rhizobium Interactions,” *Front. Plant Sci.*, vol. 9, Mar. 2018, doi: 10.3389/fpls.2018.00313.
- [144] B. Relić, “Biological Activity of Rhizobium sp. NGR234 Nod-factors on *Macroptilium atropurpureum*,” *Mol. Plant-Microbe Interact.*, vol. 6, no. 6, p. 764, 1993, doi: 10.1094/MPMI-6-764.
- [145] L. Walker, B. Lagunas, and M. L. Gifford, “Determinants of Host Range Specificity in Legume-Rhizobia Symbiosis,” *Front. Microbiol.*, vol. 11, Nov. 2020, doi: 10.3389/fmicb.2020.585749.
- [146] R. M. Lima, S. Kylarová, P. Mergaert, and É. Kondorosi, “Unexplored Arsenals of Legume Peptides With Potential for Their Applications in Medicine and Agriculture,” *Front. Microbiol.*, vol. 11, no. June, pp. 1–8, Jun. 2020, doi: 10.3389/fmicb.2020.01307.
- [147] H. Tiricz *et al.*, “Antimicrobial nodule-specific cysteine-rich peptides induce membrane depolarization-associated changes in the transcriptome of *Sinorhizobium meliloti*,” *Appl. Environ. Microbiol.*, vol. 79, no. 21, pp. 6737–46, Nov. 2013, doi: 10.1128/AEM.01791-13.
- [148] G. Maróti, J. A. Downie, and É. Kondorosi, “Plant cysteine-rich peptides that inhibit pathogen growth and control rhizobial differentiation in legume nodules,” *Curr. Opin. Plant Biol.*, vol. 26, pp. 57–63, Aug. 2015, doi: 10.1016/j.pbi.2015.05.031.
- [149] J. S. Reader *et al.*, “Major biocontrol of plant tumors targets tRNA synthetase.,” *Science*, vol. 309, no. 5740, p. 1533, Sep. 2005, doi: 10.1126/science.1116841.
- [150] J.-G. Kim *et al.*, “Bases of biocontrol: sequence predicts synthesis and mode of action of agrocin 84, the Trojan horse antibiotic that controls crown gall.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 23, pp. 8846–51, Jun. 2006, doi: 10.1073/pnas.0602965103.
- [151] W. Wright, J. Little, F. Liu, and R. Chakraborty, “Isolation and structural identification of the trihydroxamate siderophore vicibactin and its degradative products from *Rhizobium leguminosarum* ATCC 14479 bv. trifolii.,” *Biometals*, vol. 26, no. 2, pp. 271–83, Apr. 2013, doi: 10.1007/s10534-013-9609-3.
- [152] A. Dudnik, L. Bigler, and R. Dudler, “Production of proteasome inhibitor syringolin A by the endophyte *Rhizobium* sp. strain AP16.,” *Appl. Environ. Microbiol.*, vol. 80, no. 12, pp. 3741–8, Jun. 2014, doi: 10.1128/AEM.00395-14.
- [153] I. A. Osterman *et al.*, “Sorting Out Antibiotics’ Mechanisms of Action: a Double Fluorescent Protein Reporter for High-Throughput Screening of Ribosome and DNA Biosynthesis Inhibitors.,” *Antimicrob. Agents Chemother.*, vol. 60, no. 12, pp. 7481–7489, Dec. 2016, doi: 10.1128/AAC.02117-16.
- [154] Y. S. Polikanov *et al.*, “Amicoumacin A Inhibits Translation by Stabilizing mRNA Interaction with the Ribosome,” *Mol. Cell*, vol. 56, no. 4, pp. 531–540, 2014, doi: 10.1016/j.molcel.2014.09.020.
- [155] L. Ferrières *et al.*, “Silent mischief: bacteriophage Mu insertions contaminate products of *Escherichia coli* random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery.,” *J. Bacteriol.*, vol. 192, no. 24, pp. 6418–27, Dec. 2010, doi: 10.1128/JB.00621-10.
- [156] A. L. Goodman *et al.*, “Identifying genetic determinants needed to establish a human gut symbiont in its habitat.,” *Cell Host Microbe*, vol. 6, no. 3, pp. 279–89, Sep. 2009, doi: 10.1016/j.chom.2009.08.003.
- [157] A. M. Bolger, M. Lohse, and B. Usadel, “Trimmomatic: a flexible trimmer for Illumina sequence data,” *Bioinformatics*, vol. 30, no. 15, pp. 2114–2120, Aug. 2014, doi: 10.1093/bioinformatics/btu170.
- [158] A. Bankevich *et al.*, “SPAdes: A new genome assembly algorithm and its applications to single-cell

- sequencing,” *J. Comput. Biol.*, vol. 19, no. 5, pp. 455–477, 2012, doi: 10.1089/cmb.2012.0021.
- [159] C. Camacho *et al.*, “BLAST+: Architecture and applications,” *BMC Bioinformatics*, vol. 10, pp. 1–9, 2009, doi: 10.1186/1471-2105-10-421.
- [160] V. Oke and S. R. Long, “Bacterial genes induced within the nodule during the Rhizobium-legume symbiosis,” *Mol. Microbiol.*, vol. 32, no. 4, pp. 837–849, 1999, doi: 10.1046/j.1365-2958.1999.01402.x.
- [161] T. M. Finan, B. Kunkel, G. F. De Vos, and E. R. Signer, “Second symbiotic megaplasmid in *Rhizobium meliloti* carrying exopolysaccharide and thiamine synthesis genes,” *J. Bacteriol.*, vol. 167, no. 1, pp. 66–72, Jul. 1986, doi: 10.1128/jb.167.1.66-72.1986.
- [162] T. M. Finan, E. Hartweg, K. LeMieux, K. Bergman, G. C. Walker, and E. R. Signer, “General transduction in *Rhizobium meliloti*,” *J. Bacteriol.*, vol. 159, no. 1, pp. 120–124, Jul. 1984, doi: 10.1128/jb.159.1.120-124.1984.
- [163] T. Kazmierczak *et al.*, “Specific Host-Responsive Associations Between *Medicago truncatula* Accessions and *Sinorhizobium* Strains,” *Mol. Plant-Microbe Interact.*, vol. 30, no. 5, pp. 399–409, May 2017, doi: 10.1094/MPMI-01-17-0009-R.
- [164] D. N. Mastronarde, “Automated electron microscope tomography using robust prediction of specimen movements,” *J. Struct. Biol.*, vol. 152, no. 1, pp. 36–51, Oct. 2005, doi: 10.1016/j.jsb.2005.07.007.
- [165] N. Biyani *et al.*, “Focus: The interface between data collection and data processing in cryo-EM,” *J. Struct. Biol.*, vol. 198, no. 2, pp. 124–133, 2017, doi: 10.1016/j.jsb.2017.03.007.
- [166] J. Zivanov *et al.*, “New tools for automated high-resolution cryo-EM structure determination in RELION-3,” *Elife*, vol. 7, 2018, doi: 10.7554/eLife.42166.
- [167] A. Rohou and N. Grigorieff, “CTFFIND4: Fast and accurate defocus estimation from electron micrographs,” *J. Struct. Biol.*, vol. 192, no. 2, pp. 216–21, Nov. 2015, doi: 10.1016/j.jsb.2015.08.008.
- [168] E. F. Pettersen *et al.*, “UCSF Chimera - a visualization system for exploratory research and analysis,” *J. Comput. Chem.*, vol. 25, no. 13, pp. 1605–12, Oct. 2004, doi: 10.1002/jcc.20084.
- [169] J. Wang, “Experimental charge density from electron microscopic maps,” *Protein Sci.*, vol. 26, no. 8, pp. 1619–1626, Aug. 2017, doi: 10.1002/pro.3198.
- [170] J. Noeske, M. R. Wasserman, D. S. Terry, R. B. Altman, S. C. Blanchard, and J. H. D. Cate, “High-resolution structure of the *Escherichia coli* ribosome,” *Nat. Struct. Mol. Biol.*, vol. 22, no. 4, pp. 336–41, Apr. 2015, doi: 10.1038/nsmb.2994.
- [171] P. D. Adams *et al.*, “PHENIX: a comprehensive Python-based system for macromolecular structure solution,” *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 66, no. Pt 2, pp. 213–21, Feb. 2010, doi: 10.1107/S0907444909052925.
- [172] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan, “Features and development of Coot,” *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 66, no. Pt 4, pp. 486–501, Apr. 2010, doi: 10.1107/S0907444910007493.
- [173] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, and G. R. Hutchison, “Avogadro: an advanced semantic chemical editor, visualization, and analysis platform,” *J. Cheminform.*, vol. 4, no. 1, p. 17, Dec. 2012, doi: 10.1186/1758-2946-4-17.
- [174] S.-L. You and J. W. Kelly, “The total synthesis of bistratamides F–I,” *Tetrahedron*, vol. 61, no. 1, pp. 241–249, Jan. 2005, doi: 10.1016/j.tet.2004.10.024.
- [175] J. J. Almagro Armenteros *et al.*, “SignalP 5.0 improves signal peptide predictions using deep neural networks,” *Nat. Biotechnol.*, vol. 37, no. 4, pp. 420–423, Apr. 2019, doi: 10.1038/s41587-019-0036-z.
- [176] W. Kabsch, “XDS,” *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 66, no. Pt 2, pp. 125–32, Feb. 2010, doi: 10.1107/S0907444909047337.
- [177] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read,

- “Phaser crystallographic software,” *J. Appl. Crystallogr.*, vol. 40, no. 4, pp. 658–674, Aug. 2007, doi: 10.1107/S0021889807021206.
- [178] L. M. K. Dassama, G. E. Kenney, S. Y. Ro, E. L. Zielazinski, and A. C. Rosenzweig, “Methanobactin transport machinery,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 46, pp. 13027–13032, 2016, doi: 10.1073/pnas.1603578113.
- [179] P. Emsley and K. Cowtan, “Coot: model-building tools for molecular graphics,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 60, no. 12, pp. 2126–2132, Dec. 2004, doi: 10.1107/S0907444904019158.
- [180] G. Bricogne *et al.*, “BUSTER version 2.10.0,” *Glob. Phasing Ltd, Cambridge, UK*, 2011.
- [181] N. A. O’Leary *et al.*, “Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation,” *Nucleic Acids Res.*, vol. 44, no. D1, pp. D733–45, Jan. 2016, doi: 10.1093/nar/gkv1189.
- [182] M. Mirdita, M. Steinegger, and J. Söding, “MMseqs2 desktop and local web server app for fast, interactive sequence searches,” *Bioinformatics*, vol. 35, no. 16, pp. 2856–2858, Aug. 2019, doi: 10.1093/bioinformatics/bty1057.
- [183] J. I. T. Tietz *et al.*, “A new genome mining tool redefines the lasso peptide biosynthetic landscape,” *Nat. Chem. Biol.*, vol. In press., no. February, 2017, doi: 10.1038/nchembio.2319.
- [184] K. Blin *et al.*, “antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline,” *Nucleic Acids Res.*, vol. 47, no. W1, pp. W81–W87, Jul. 2019, doi: 10.1093/nar/gkz310.
- [185] J. C. Navarro-Muñoz *et al.*, “A computational framework for systematic exploration of biosynthetic diversity from large-scale genomic data,” *bioRxiv*, p. 445270, Jan. 2018, doi: 10.1101/445270.
- [186] P. Shannon *et al.*, “Cytoscape: a software environment for integrated models of biomolecular interaction networks,” *Genome Res.*, vol. 13, no. 11, pp. 2498–504, Nov. 2003, doi: 10.1101/gr.1239303.
- [187] Y. M. Li, J. C. Milne, L. L. Madison, R. Kolter, and C. T. Walsh, “From peptide precursors to oxazole and thiazole-containing peptide antibiotics: microcin B17 synthase,” *Science*, vol. 274, no. 5290, pp. 1188–93, Nov. 1996, doi: 10.1126/science.274.5290.1188.
- [188] K. J. Molohon *et al.*, “Plantazolicin Is an Ultranarrow-Spectrum Antibiotic That Targets the Bacillus anthracis Membrane,” *ACS Infect. Dis.*, vol. 2, no. 3, pp. 207–220, 2016, doi: 10.1021/acsinfecdis.5b00115.
- [189] R. A. Salomón and R. N. Farías, “Microcin 25, a novel antimicrobial peptide produced by Escherichia coli,” *J. Bacteriol.*, vol. 174, no. 22, pp. 7428–35, Nov. 1992, doi: 10.1128/jb.174.22.7428-7435.1992.
- [190] A. Kerr and K. Htay, “Biological control of crown gall through bacteriocin production,” *Physiol. Plant Pathol.*, vol. 4, no. 1, pp. 37–44, Jan. 1974, doi: 10.1016/0048-4059(74)90042-3.
- [191] S. R. Khan, J. Gaines, R. M. Roop, and S. K. Farrand, “Broad-Host-Range Expression Vectors with Tightly Regulated Promoters and Their Use To Examine the Influence of TraR and TraM Expression on Ti Plasmid Quorum Sensing,” *Appl. Environ. Microbiol.*, vol. 74, no. 16, pp. 5053–5062, Aug. 2008, doi: 10.1128/AEM.01098-08.
- [192] M. C. Garrido, M. Herrero, R. Kolter, and F. Moreno, “The export of the DNA replication inhibitor Microcin B17 provides immunity for the host cell,” *EMBO J.*, vol. 7, no. 6, pp. 1853–62, Jun. 1988, doi: 10.1002/j.1460-2075.1988.tb03018.x.
- [193] H. I. Zgurskaya, G. Krishnamoorthy, A. Ntrel, and S. Lu, “Mechanism and Function of the Outer Membrane Channel TolC in Multidrug Resistance and Physiology of Enterobacteria,” *Front. Microbiol.*, vol. 2, no. Sep, p. 189, 2011, doi: 10.3389/fmicb.2011.00189.
- [194] A. Metlitskaya *et al.*, “Aspartyl-tRNA synthetase is the target of peptide nucleotide antibiotic Microcin C,” *J. Biol. Chem.*, vol. 281, no. 26, pp. 18033–42, Jun. 2006, doi: 10.1074/jbc.M513174200.
- [195] S. E. Heffron and F. Journak, “Structure of an EF-Tu complex with a thiazolyl peptide antibiotic

- determined at 2.35 Å resolution: atomic basis for GE2270A inhibition of EF-Tu.,” *Biochemistry*, vol. 39, no. 1, pp. 37–45, Jan. 2000, doi: 10.1021/bi9913597.
- [196] S. Arenz *et al.*, “A combined cryo-EM and molecular dynamics approach reveals the mechanism of ErmBL-mediated translation arrest,” *Nat. Commun.*, vol. 7, no. May, p. 12026, 2016, doi: 10.1038/ncomms12026.
- [197] D. Bulkley, C. A. Innis, G. Blaha, and T. A. Steitz, “Revisiting the structures of several antibiotics bound to the bacterial ribosome,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 40, pp. 17158–63, Oct. 2010, doi: 10.1073/pnas.1008685107.
- [198] J. Noeske, J. Huang, N. B. Olivier, R. A. Giacobbe, M. Zambrowski, and J. H. D. Cate, “Synergy of streptogramin antibiotics occurs independently of their effects on translation,” *Antimicrob. Agents Chemother.*, vol. 58, no. 9, pp. 5269–5279, 2014, doi: 10.1128/AAC.03389-14.
- [199] M. Laviña, A. P. Pugsley, and F. Moreno, “Identification, mapping, cloning and characterization of a gene (*sbmA*) required for microcin B17 action on *Escherichia coli* K12,” *J. Gen. Microbiol.*, vol. 132, no. 6, pp. 1685–93, Jun. 1986, doi: 10.1099/00221287-132-6-1685.
- [200] Q. Nicoud *et al.*, “Sinorhizobium meliloti Functions Required for Resistance to Antimicrobial NCR Peptides and Bacteroid Differentiation,” *MBio*, vol. 12, no. 4, p. e0089521, Jul. 2021, doi: 10.1128/mBio.00895-21.
- [201] D. Ghilarov *et al.*, “Molecular mechanism of *SbmA*, a promiscuous transporter exploited by antimicrobial peptides,” *Sci. Adv.*, vol. 7, no. 37, p. eabj5363, Sep. 2021, doi: 10.1126/sciadv.abj5363.
- [202] G. P. Ferguson, R. M. Roop, and G. C. Walker, “Deficiency of a *Sinorhizobium meliloti* *BacA* mutant in alfalfa symbiosis correlates with alteration of the cell envelope,” *J. Bacteriol.*, vol. 184, no. 20, pp. 5625–32, Oct. 2002, doi: 10.1128/JB.184.20.5625-5632.2002.
- [203] M. Novikova *et al.*, “The *Escherichia coli* *Yej* transporter is required for the uptake of translation inhibitor microcin C,” *J. Bacteriol.*, vol. 189, no. 22, pp. 8361–5, Nov. 2007, doi: 10.1128/JB.01028-07.
- [204] D. Pletzer *et al.*, “The *Pseudomonas aeruginosa* PA14 ABC Transporter *NppA1A2BCD* Is Required for Uptake of Peptidyl Nucleoside Antibiotics,” *J. Bacteriol.*, vol. 197, no. 13, pp. 2217–2228, Jul. 2015, doi: 10.1128/JB.00234-15.
- [205] I. Guefrachi *et al.*, “*Bradyrhizobium* *BclA* Is a Peptide Transporter Required for Bacterial Differentiation in Symbiosis with *Aeschynomene* Legumes,” *Mol. Plant. Microbe. Interact.*, vol. 28, no. 11, pp. 1155–66, Nov. 2015, doi: 10.1094/MPMI-04-15-0094-R.
- [206] K. LeVier, R. W. Phillips, V. K. Grippe, R. M. Roop, and G. C. Walker, “Similar requirements of a plant symbiont and a mammalian pathogen for prolonged intracellular survival,” *Science*, vol. 287, no. 5462, pp. 2492–3, Mar. 2000, doi: 10.1126/science.287.5462.2492.
- [207] S. Wehmeier *et al.*, “Internalization of a thiazole-modified peptide in *Sinorhizobium meliloti* occurs by *BacA*-dependent and -independent mechanisms,” *Microbiology*, vol. 156, no. 9, pp. 2702–2713, 2010, doi: 10.1099/mic.0.039909-0.
- [208] R. P.-A. Berntsson *et al.*, “The structural basis for peptide selection by the transport receptor *OppA*,” *EMBO J.*, vol. 28, no. 9, pp. 1332–40, May 2009, doi: 10.1038/emboj.2009.65.
- [209] M. M. Rahman, M. A. Machuca, M. F. Khan, C. K. Barlow, R. B. Schittenhelm, and A. Roujeinikova, “Molecular Basis of Unexpected Specificity of ABC Transporter-Associated Substrate-Binding Protein *DppA* from *Helicobacter pylori*,” *J. Bacteriol.*, vol. 201, no. 20, 2019, doi: 10.1128/JB.00400-19.
- [210] M. K. Doeven, R. Abele, R. Tampé, and B. Poolman, “The binding specificity of *OppA* determines the selectivity of the oligopeptide ATP-binding cassette transporter,” *J. Biol. Chem.*, vol. 279, no. 31, pp. 32301–7, Jul. 2004, doi: 10.1074/jbc.M404343200.
- [211] G. H. Scheepers, J. A. Lycklama a Nijeholt, and B. Poolman, “An updated structural classification of substrate-binding proteins,” *FEBS Lett.*, vol. 590, no. 23, pp. 4393–4401, Dec. 2016, doi:

- 10.1002/1873-3468.12445.
- [212] V. M. Levdikov, E. V. Blagova, J. A. Brannigan, L. Wright, A. A. Vagin, and A. J. Wilkinson, “The structure of the oligopeptide-binding protein, AppA, from *Bacillus subtilis* in complex with a nonapeptide,” *J. Mol. Biol.*, vol. 345, no. 4, pp. 879–92, Jan. 2005, doi: 10.1016/j.jmb.2004.10.089.
- [213] M. Mattiuzzo *et al.*, “Role of the *Escherichia coli* SbmA in the antimicrobial activity of proline-rich peptides,” *Mol. Microbiol.*, vol. 66, no. 1, pp. 151–163, Oct. 2007, doi: 10.1111/j.1365-2958.2007.05903.x.
- [214] A. Ichige and G. C. Walker, “Genetic analysis of the *Rhizobium meliloti* bacA gene: Functional interchangeability with the *Escherichia coli* sbmA gene and phenotypes of mutants,” *J. Bacteriol.*, vol. 179, no. 1, pp. 209–216, 1997, doi: 10.1128/jb.179.1.209-216.1997.
- [215] A. Ghosal, A. Vitali, J. E. M. Stach, and P. E. Nielsen, “Role of SbmA in the Uptake of Peptide Nucleic Acid (PNA)-Peptide Conjugates in *E. coli*,” *ACS Chem. Biol.*, vol. 8, no. 2, pp. 360–367, Feb. 2013, doi: 10.1021/cb300434e.
- [216] S. E. Puckett *et al.*, “Bacterial Resistance to Antisense Peptide Phosphorodiamidate Morpholino Oligomers,” *Antimicrob. Agents Chemother.*, vol. 56, no. 12, pp. 6147–6153, Dec. 2012, doi: 10.1128/AAC.00850-12.
- [217] A. F. Haag *et al.*, “Protection of *Sinorhizobium* against host cysteine-rich antimicrobial peptides is critical for symbiosis,” *PLoS Biol.*, vol. 9, no. 10, p. e1001169, Oct. 2011, doi: 10.1371/journal.pbio.1001169.
- [218] S. M. Eswarappa, K. K. Panguluri, M. Hensel, and D. Chakravorty, “The yejABEF operon of *Salmonella* confers resistance to antimicrobial peptides and contributes to its virulence,” *Microbiology*, vol. 154, no. Pt 2, pp. 666–678, Feb. 2008, doi: 10.1099/mic.0.2007/011114-0.
- [219] Z. Wang, P. Bie, J. Cheng, L. Lu, B. Cui, and Q. Wu, “The ABC transporter YejABEF is required for resistance to antimicrobial peptides and the virulence of *Brucella melitensis*,” *Sci. Rep.*, vol. 6, p. 31876, 2016, doi: 10.1038/srep31876.
- [220] D. T. Hudson, P. A. Chapman, R. C. Day, X. C. Morgan, and C. W. Beck, “Complete Genome Sequences of *Kinneretia* sp. Strain XES5, *Shinella* sp. Strain XGS7, and *Vogesella* sp. Strain XCS3, Isolated from *Xenopus laevis* Skin,” *Microbiol. Resour. Announc.*, vol. 10, no. 50, p. e0105021, Dec. 2021, doi: 10.1128/MRA.01050-21.
- [221] V. I. Safronova *et al.*, “Rhizobial microsymbionts of Kamchatka oxytropis species possess genes of the Type III and VI secretion systems, which can affect the development of symbiosis,” *Mol. Plant-Microbe Interact.*, vol. 33, no. 10, pp. 1232–1241, 2020, doi: 10.1094/MPMI-05-20-0114-R.
- [222] J. J. M. Lustermans, J. J. Bjerg, A. Schramm, and I. P. G. Marshall, “*Phyllobacterium calauticae* sp. nov. isolated from a microaerophilic veil transversed by cable bacteria in freshwater sediment,” *Antonie Van Leeuwenhoek*, vol. 114, no. 11, pp. 1877–1887, Nov. 2021, doi: 10.1007/s10482-021-01647-y.
- [223] D. A. Pelletier *et al.*, “Genome Sequences of 42 Bacteria Isolated from *Sorghum bicolor* Roots,” *Microbiol. Resour. Announc.*, vol. 9, no. 37, Sep. 2020, doi: 10.1128/MRA.00736-20.
- [224] A. C. Baraúna *et al.*, “*Rhizobium altiplani* sp. nov., isolated from effective nodules on *Mimosa pudica* growing in untypically alkaline soil in central Brazil,” *Int. J. Syst. Evol. Microbiol.*, vol. 66, no. 10, pp. 4118–4124, Oct. 2016, doi: 10.1099/ijsem.0.001322.
- [225] S. D. Brown *et al.*, “Draft genome sequence of *Rhizobium* sp. strain PDO1-076, a bacterium isolated from *Populus deltoides*,” *J. Bacteriol.*, vol. 194, no. 9, pp. 2383–4, May 2012, doi: 10.1128/JB.00198-12.
- [226] M. Gomila, J. Pinhassi, E. Falsen, E. R. B. Moore, and J. Lalucat, “*Kinneretia asaccharophila* gen. nov., sp. nov., isolated from a freshwater lake, a member of the *Rubrivivax* branch of the family *Comamonadaceae*,” *Int. J. Syst. Evol. Microbiol.*, vol. 60, no. 4, pp. 809–814, Apr. 2010, doi: 10.1099/ijss.0.011478-0.
- [227] S. Guindon, J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel, “New Algorithms

- and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0,” *Syst. Biol.*, vol. 59, no. 3, pp. 307–321, Mar. 2010, doi: 10.1093/sysbio/syq010.
- [228] C. L. Cox, J. R. Doroghazi, and D. A. Mitchell, “The genomic landscape of ribosomal peptides containing thiazole and oxazole heterocycles,” *BMC Genomics*, vol. 16, p. 778, Jan. 2015, doi: 10.1186/s12864-015-2008-0.
- [229] J. O. Melby, K. L. Dunbar, N. Q. Trinh, and D. A. Mitchell, “Selectivity, directionality, and promiscuity in peptide processing from a bacillus sp. Al Hakam cyclodehydratase,” *J. Am. Chem. Soc.*, vol. 134, no. 11, pp. 5309–5316, 2012, doi: 10.1021/ja211675n.
- [230] K. L. Dunbar and D. A. Mitchell, “Insights into the mechanism of peptide cyclodehydrations achieved through the chemoenzymatic generation of amide derivatives,” *J. Am. Chem. Soc.*, vol. 135, no. 23, pp. 8692–8701, 2013, doi: 10.1021/ja4029507.
- [231] X. Bai, H. Guo, D. Chen, Q. Yang, J. Tao, and W. Liu, “Isolation and structure determination of two new nosiheptide-type compounds provide insights into the function of the cytochrome P450 oxygenase NocV in nocathiacin biosynthesis,” *Org. Chem. Front.*, vol. 7, no. 3, pp. 584–589, 2020, doi: 10.1039/C9QO01328H.
- [232] Q. Zheng, S. Wang, R. Liao, and W. Liu, “Precursor-Directed Mutational Biosynthesis Facilitates the Functional Assignment of Two Cytochromes P450 in Thioestrepton Biosynthesis,” *ACS Chem. Biol.*, vol. 11, no. 10, pp. 2673–2678, Oct. 2016, doi: 10.1021/acscchembio.6b00419.
- [233] S. Adam, L. Franz, M. Milhim, R. Bernhardt, O. V. Kalinina, and J. Koehnke, “Characterization of the Stereoselective P450 Enzyme BotCYP Enables the In Vitro Biosynthesis of the Bottromycin Core Scaffold,” *J. Am. Chem. Soc.*, vol. 142, no. 49, pp. 20560–20565, Dec. 2020, doi: 10.1021/jacs.0c10361.
- [234] L. Huo, S. Rachid, M. Stadler, S. C. Wenzel, and R. Müller, “Synthetic biotechnology to study and engineer ribosomal bottromycin biosynthesis,” *Chem. Biol.*, vol. 19, no. 10, pp. 1278–87, Oct. 2012, doi: 10.1016/j.chembiol.2012.08.013.
- [235] F. Román-Hurtado, M. Sánchez-Hidalgo, J. Martín, F. Ortiz-López, and O. Genilloud, “Biosynthesis and Heterologous Expression of Cacaoidin, the First Member of the Lanthidin Family of RiPPs,” *Antibiotics*, vol. 10, no. 4, p. 403, Apr. 2021, doi: 10.3390/antibiotics10040403.
- [236] T. Maxson *et al.*, “HIV Protease Inhibitors Block Streptolysin S Production,” *ACS Chem. Biol.*, vol. 10, no. 5, pp. 1217–1226, May 2015, doi: 10.1021/cb500843r.
- [237] M. Kjos, L. Snipen, Z. Salehian, I. F. Nes, and D. B. Diep, “The abi proteins and their involvement in bacteriocin self-immunity,” *J. Bacteriol.*, vol. 192, no. 8, pp. 2068–76, Apr. 2010, doi: 10.1128/JB.01553-09.
- [238] Y. Du, Y. Qiu, X. Meng, J. Feng, J. Tao, and W. Liu, “A Heterotrimeric Dehydrogenase Complex Functions with 2 Distinct YcaO Proteins to Install 5 Azole Heterocycles into 35-Membered Sulfomycin Thiopeptides,” *J. Am. Chem. Soc.*, vol. 142, no. 18, pp. 8454–8463, May 2020, doi: 10.1021/jacs.0c02329.
- [239] J. Söding, A. Biegert, and A. N. Lupas, “The HHpred interactive server for protein homology detection and structure prediction,” *Nucleic Acids Res.*, vol. 33, no. Web Server issue, pp. W244–8, Jul. 2005, doi: 10.1093/nar/gki408.
- [240] S. V Ghodge, K. A. Biernat, S. J. Bassett, M. R. Redinbo, and A. A. Bowers, “Post-translational Claisen Condensation and Decarboxylation en Route to the Bicyclic Core of Pantocin A,” *J. Am. Chem. Soc.*, vol. 138, no. 17, pp. 5487–90, 2016, doi: 10.1021/jacs.5b13529.
- [241] S.-H. Dong *et al.*, “Biosynthesis of the RiPP trojan horse nucleotide antibiotic microcin C is directed by the N-formyl of the peptide precursor,” *Chem. Sci.*, vol. 10, no. 8, pp. 2391–2395, Feb. 2019, doi: 10.1039/c8sc03173h.
- [242] M. G. Gagnon, R. N. Roy, I. B. Lomakin, T. Florin, A. S. Mankin, and T. A. Steitz, “Structures of proline-rich peptides bound to the ribosome reveal a common mechanism of protein synthesis inhibition,” *Nucleic Acids Res.*, vol. 44, no. 5, pp. 2439–2450, 2016, doi: 10.1093/nar/gkw018.

- [243] S. S. Hegde *et al.*, “A fluoroquinolone resistance protein from *Mycobacterium tuberculosis* that mimics DNA.,” *Science*, vol. 308, no. 5727, pp. 1480–3, Jun. 2005, doi: 10.1126/science.1110699.
- [244] M. W. Vetting, S. S. Hegde, M. Wang, G. a Jacoby, D. C. Hooper, and J. S. Blanchard, “Structure of QnrB1, a plasmid-mediated fluoroquinolone resistance factor.,” *J. Biol. Chem.*, vol. 286, no. 28, pp. 25265–73, Jul. 2011, doi: 10.1074/jbc.M111.226936.
- [245] Ł. Mazurek *et al.*, “Pentapeptide repeat protein QnrB1 requires ATP hydrolysis to rejuvenate poisoned gyrase complexes,” *Nucleic Acids Res.*, vol. 49, no. 3, pp. 1581–1596, Feb. 2021, doi: 10.1093/nar/gkaa1266.
- [246] A. M. Cosme, A. Becker, M. R. Santos, L. A. Sharypova, P. M. Santos, and L. M. Moreira, “The outer membrane protein TolC from *Sinorhizobium meliloti* affects protein secretion, polysaccharide biosynthesis, antimicrobial resistance, and symbiosis.,” *Mol. Plant. Microbe. Interact.*, vol. 21, no. 7, pp. 947–57, Jul. 2008, doi: 10.1094/MPMI-21-7-0947.
- [247] C. M. Kurylo *et al.*, “Genome Sequence and Analysis of *Escherichia coli* MRE600, a Colicinogenic, Nonmotile Strain that Lacks RNase I and the Type I Methyltransferase, EcoKI,” *Genome Biol. Evol.*, vol. 8, no. 3, pp. 742–752, Mar. 2016, doi: 10.1093/gbe/evw008.
- [248] T. Baba *et al.*, “Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection.,” *Mol. Syst. Biol.*, vol. 2, p. 2006.0008, Jan. 2006, doi: 10.1038/msb4100050.
- [249] A. C. Vergunst, A. H. Meijer, S. A. Renshaw, and D. O’Callaghan, “*Burkholderia cenocepacia* creates an intramacrophage replication niche in zebrafish embryos, followed by bacterial dissemination and establishment of systemic infection,” *Infect. Immun.*, vol. 78, no. 4, pp. 1495–1508, 2010, doi: 10.1128/IAI.00743-09.
- [250] V. B. Chen *et al.*, “MolProbity: all-atom structure validation for macromolecular crystallography,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 66, no. 1, pp. 12–21, Jan. 2010, doi: 10.1107/S0907444909042073.

Appendix

Supplementary Tables

Supplementary Table 1. Bacterial strains used in the study

Strain	Resistance	Description	Reference
<i>S. meliloti</i> Sm1021	Sm ^R	<i>Sinorhizobium meliloti</i> wt	Common laboratory strain
<i>S. meliloti</i> Sm1021 Δ <i>bacA</i>	Sm ^R , Sp ^R	Sm1021 <i>bacA654::Spc</i> (Δ <i>bacA</i> null mutant)	[202]
<i>S. meliloti</i> Sm1021 Ω <i>yejA</i>	Sm ^R , Km ^R	<i>yejA</i> plasmid insertion mutant	[200]
<i>S. meliloti</i> Sm1021 Ω <i>yejB</i>	Sm ^R , Km ^R	<i>yejB</i> plasmid insertion mutant	This study
<i>S. meliloti</i> Sm1021 Ω <i>yejE</i>	Sm ^R , Km ^R	<i>yejE</i> plasmid insertion mutant	[200]
<i>S. meliloti</i> Sm1021 Ω <i>yejF</i>	Sm ^R , Km ^R	<i>yejF</i> plasmid insertion mutant	[200]
<i>S. meliloti</i> Sm1021 Ω <i>tolC</i>	Sm ^R , Km ^R	<i>tolC</i> plasmid insertion mutant	[246]
<i>S. meliloti</i> Sm1021 Δ <i>bacA</i> Ω <i>yejA</i>	Sm ^R , Km ^R , Sp ^R	Double mutant in <i>yejA</i> and <i>bacA</i> obtained through ϕ M12 phage transduction	This study
<i>S. meliloti</i> Sm1021 Δ <i>bacA</i> Ω <i>yejE</i>	Sm ^R , Km ^R , Sp ^R	Double mutant in <i>yejE</i> and <i>bacA</i> obtained through ϕ M12 phage transduction	This study
<i>S. meliloti</i> Sm1021 mut #1 – #4	Sm ^R , Km ^R	Phazolicin-resistant mutants selected using transposon library screening	This study
<i>Rhizobium</i> sp. Pop5	-	Phazolicin-producing strain (natural isolate)	This study
<i>Rhizobium</i> sp. Pop5 Ω <i>phzD</i>	Km ^R	Mutant with pVO plasmid insertion in <i>phzD</i> gene, unable to produce mature phazolicin	This study
<i>Rhizobium</i> sp. Pop5 Ω NI	Km ^R	Mutant with pVO plasmid insertion between the genes <i>phzD</i> and <i>phzR</i> , produces PHZ, has growth	This study
<i>Mesorhizobium loti</i> strain 582	-	Encodes <i>phz</i> -like BGC	[221]
<i>Mesorhizobium huakuii</i> strain 583	-	Encodes <i>phz</i> -like BGC	[221]
<i>E. coli</i> Rosetta 2 (DE3) pLysS	Cm ^R	The strain used for heterologous protein expression	Novagen
<i>E. coli</i> BL21 (DE3)	-	The strain used for heterologous protein expression	Novagen
<i>E. coli</i> BW25113	-	The strain used for in vivo reporter system in <i>E. coli</i>	Common laboratory strain
<i>E. coli</i> DH5 α	-	The strain used for molecular cloning of all constructs	Common laboratory strain

<i>E. coli</i> MFDpir Δ dapA	-	Donor strain used for transposon library creation, auxotroph for diaminopimelic acid synthesis	[155]
<i>E. coli</i> MRE600	-	Purification of ribosomes for cryo-EM	[247]
<i>E. coli</i> BW25113 Δ tolC	Km ^R	Keio collection strain with the deletion of <i>tolC</i> gene encoding the major outer membrane efflux pump	[248]
<i>Rhizobium leguminosarum</i> bv. <i>viciae</i> 3841	Sm ^R	Proteobacteria: Alphaproteobacteria	Strains from the laboratory strain collection
<i>Rhizobium leguminosarum</i> bv. <i>phaseoli</i> 4292	Rf ^R	Proteobacteria: Alphaproteobacteria	
<i>Rhizobium leguminosarum</i> bv. <i>phaseoli</i> RCR 3622	Sm ^R	Proteobacteria: Alphaproteobacteria	
<i>Rhizobium etli</i> DSM 11541	-	Proteobacteria: Alphaproteobacteria	
<i>Rhizobium tibeticum</i> DSM 21102	-	Proteobacteria: Alphaproteobacteria	
<i>Agrobacterium tumefaciens</i> C58C1	Rf ^R	Proteobacteria: Alphaproteobacteria	
<i>Agrobacterium rhizogenes</i> ARQUA1	Sm ^R	Proteobacteria: Alphaproteobacteria	
<i>Mesorhizobium loti</i> MAFF303099	-	Proteobacteria: Alphaproteobacteria	
<i>Mesorhizobium thianshanense</i> HAMB1 3372	-	Proteobacteria: Alphaproteobacteria	
<i>Azorhizobium caulinodans</i> ORS571	Cb ^R	Proteobacteria: Alphaproteobacteria	
<i>Sinorhizobium fredii</i> HH103	-	Proteobacteria: Alphaproteobacteria	
<i>Sinorhizobium medicae</i> WSM419	Cm ^R	Proteobacteria: Alphaproteobacteria	
<i>Pantoea ananatis</i> PA4	-	Proteobacteria: Gammaproteobacteria	
<i>Pseudomonas putida</i> KT2440	-	Proteobacteria: Gammaproteobacteria	
<i>Bacillus subtilis</i> 168	-	Firmicutes: Bacilli	
<i>Bacillus cereus</i> ATCC 4342	-	Firmicutes: Bacilli	
<i>Arthrobacter</i> sp. ATCC21022	-	Actinobacteria: Actinomycetia	
<i>Microtetraspora glauca</i> NRRL B-3735	-	Actinobacteria: Actinomycetia	

Ap – ampicillin, Cb – carbenicillin, Cm – chloramphenicol, Km – kanamycin, Rf – rifampicin, Sm – streptomycin, Sp – spectinomycin

Supplementary Table 2. Vectors used in the study

Vector	Resistance	Description	Reference
pIN72	Tc ^R	Constitutive production of the DsRed fluorescent protein	[249]
pET Duet-1	Ap ^R	Vector for the co-expression of two genes in <i>E. coli</i> (T7- <i>lac</i> promoters)	Novagen
pET Duet <i>phzA phzCBD</i>	Ap ^R	Expression of the PHZ precursor peptide and azole installation complex	This study
pET29b(+)	Km ^R	Expression vector for C-terminally 6His-tagged proteins in <i>E. coli</i>	Novagen
pET29 <i>yejA</i> (Sm) CHis6	Km ^R	Expression of C-terminally 6His-tagged YejA Sm in <i>E. coli</i>	This study
pSAM_Ec	Ap ^R	Plasmid for <i>Mariner</i> transposon library creation	[156]
pRK600	Cm ^R	Helper plasmid for conjugal DNA transfer	[161]
pVO155 <i>nptIIgfp</i>	Ap ^R	Insertional gene inactivation in rhizobia	[160]
pVO155 <i>nptIIgfp phzD</i>	Ap ^R	Insertional inactivation of <i>phzD</i> in <i>Rhizobium</i> sp. Pop5	This study
pVO155 <i>nptIIgfp</i> NI	Ap ^R	Insertional of Km ^R cassette between the genes <i>phzD</i> and <i>phzR</i> in <i>Rhizobium</i> sp. Pop5 (construction of the “neutral insertion” control strain for competition experiments)	This study
pDualrep2	Ap ^R	Reporter plasmid with two fluorescent proteins, which biosynthesis is activated upon ribosome stalling and SOS-response in the cell	[153]
pSRK	Gm ^R or Km ^R	Broad-host range vector for inducible (<i>lac</i> promoter) protein expression, oriV (pBBR5 derivative)	[191]
pSRK <i>phzD</i>	Gm ^R	<i>phzD</i> from <i>Rhizobium</i> sp. Pop5	This study
pSRK <i>phzE</i>	Km ^R	<i>phzE</i> from <i>Rhizobium</i> sp. Pop5	This study
pSRK <i>rplD</i>	Km ^R	wt <i>rplD</i> from Sm1021 (codes for uL4 ribosomal protein)	This study
pSRK <i>rplD</i> (K65A)	Km ^R	<i>rplD</i> , point mutation K65A	This study
pSRK <i>rplD</i> (G68H)	Km ^R	<i>rplD</i> , point mutation G68H	This study
pSRK <i>rplV</i>	Km ^R	wt <i>rplV</i> from Sm1021 (codes for uL22 ribosomal protein)	This study
pSRK <i>rplV</i> (K90R)	Km ^R	<i>rplV</i> , point mutation K90R	This study
pSRK <i>bacA</i> (Sm)	Gm ^R	<i>bacA</i> from <i>Sinorhizobium meliloti</i> Sm1021	This study
pSRK <i>bclA</i> (Bsp)	Gm ^R	<i>bclA</i> from <i>Bradyrhizobium</i> sp. ORS 285	This study
pSRK <i>sbmA</i> (Ec)	Gm ^R	<i>sbmA</i> from <i>Escherichia coli</i> MG1655	This study
pSRK <i>bacA</i> (Ba)	Gm ^R	<i>bacA</i> from <i>Brucella abortus</i> 2308	This study
pSRK <i>yejABEF</i> (Sm)	Gm ^R	<i>yejABEF</i> from <i>Sinorhizobium meliloti</i> Sm1021	This study
pSRK <i>yejABEF</i> (Ec)	Gm ^R	<i>yejABEF</i> from <i>Escherichia coli</i> MG1655	This study
pSRK <i>nppA₁A₂BCD</i> (Pa)	Gm ^R	<i>nppA₁A₂BCD</i> from <i>Pseudomonas aeruginosa</i> PA14	This study

Ap – ampicillin, Cm – chloramphenicol, Gm – gentamycin, Km – kanamycin, Tc – tetracycline

Supplementary Table 3. Oligonucleotides used in the study

Primer name	Primer sequence (5'-3')	Purpose
bacA_Sm_pSRK_GA_F	ataacaatttcacacaggaacagcatatgt tccaatccttcttcccc	Molecular cloning of <i>bacA</i> (<i>Sinorhizobium meliloti</i> Sm1021), <i>bacA</i> (<i>Brucella abortus</i>), <i>bclA</i> (<i>Bradyrhizobium</i> sp. ORS285) and <i>sbmA</i> (<i>Escherichia coli</i> MG1655) genes into the pSRK plasmid by Gibson Assembly (<i>bacA</i> Sm) or restriction ligation protocol (others).
bacA_Sm_pSRK_GA_R	cgaggtcgacggatcgatattacagagcca gctcttcc	
bacA_Ba_NdeI_F	attataCATATGtttgcgctcatttttcccc g	
bacA_Ba_XbaI_R	atattaTCTAGAtcagctcgccccctggttc	
bclA_Bsp_NdeI_F	attattaCATATGaacaatttgcgctcgacc c	
bclA_Bsp_XbaI_R	attataaTCTAGActactcggcgccaccgc c	
sbmA_Ec_NdeI_F	attattaCATATGtttaagtcttttttccca aagc	
sbmA_Ec_XbaI_R	atattaTCTAGAtttagctcaaggatgggtt acttc	
pSRK_GA_F	atgctgtttcctgtgtgaaattg	
pSRK_GA_R	tatcgataccgtcgacctcg	
yejA_Sm_pSRK_GA_F	ataacaatttcacacaggaacagcataatg ccaaacttctgcaggaccg	Molecular cloning of <i>yejA</i> (<i>Sinorhizobium meliloti</i> Sm1021) gene into the pSRK plasmid by Gibson Assembly.
yejA_Sm_pSRK_GA_R	cgaggtcgacggatcgatattcattttgcag ccgtgtttttcg	
yejA_Ec_NdeI_F	taatattaaCATATGattgtgcgcatactgc	Molecular cloning of <i>yejABEF</i> genes (<i>E. coli</i> MG1655) into the pSRK plasmid.
yejF_Ec_XbaI_R	ataattaTCTAGAtcagctcaacgccagtag ctg	
yejE_Ec_mut_F	tcatcctgcgtcacatgttgccctaatgccat g	Internal NdeI site elimination from the <i>yejE</i> ^{Ec} gene.
yejE_Ec_mut_R	tggcattaggcaacatgtgacgcaggatgat act	
nppA1_Pa_NdeI_F	taatattaaCATATGcgtcgcctctccttc	Molecular cloning of <i>nppA1A2BCD</i> genes (<i>P. aeruginosa</i> PA14) into the pSRK plasmid.
nppD_Pa_SacI_R	attataaGAGCTCtcagttttccgcgcttgc c	
yejA_Sm_NoSP_NdeI_F	atattattaCATATGgaggaacaaccgctctg gcacc	Molecular cloning of <i>yejA</i> gene (<i>S. meliloti</i> Sm1021) into the pET19b plasmid.
yejA_Sm_XhoI_R	attaattCTCGAGttttgcagccgtgtttt cgacc	
yejA_Ec_NoSP_NcoI_F	catgCCATGGctatcaaggaaagctatgcc	Molecular cloning of <i>yejA</i> gene (<i>E. coli</i> MG1655) into the pEHsTEV plasmid.
yejA_Ec_XhoI_R	ccgCTCGAGctactctccctgtttgctgg	
bacA_seq_F	gcatcaggaggcaagtccttg	Amplification of Sm1021 <i>bacA</i> gene region for subsequent amplicon Sanger sequencing
bacA_seq_R	gaggcgttgccgattatcgag	
phzC_1F	atgttttcggtttccccggttcgtac	

phzB_1R	gcattaattctcctccggataggcaaagg	Verification of the pVO155 plasmid insertion into the <i>phzD</i> gene
phzB_2F	accatcgagtttcccgatg	
phzD_2R	gctctcaagctaaagcaaaataaggc	
phzD_pVO_SalI_F	attatatGTCGACcgagatcgtgtgacc c	Cloning of the <i>phzD</i> gene internal 541 bp-long fragment into pVO155
phzD_pVO_XbaI_R	attattaTCTAGAcgccaagaccttcgatag c	
NI_pVO_SalI_F	attatatGTCGACgtttcgagttgggtgtcg gcg	Cloning of the 581 bp-long fragment including the parts of <i>phzD</i> and <i>phzR</i> genes into pVO155
NI_pVO_XbaI_R	attattaTCTAGAccgttcctatcgatcggt cg	
phzD_NdeI_F	atattatCATATGcaacggtcatatcgc	Cloning of <i>phzD</i> gene into pSRK plasmid
phzD_HindIII_R	gctgttAAGCTTttatgaaaatggcatgggc	
rplD_pSRK_GA_F	ataacaatttcacacaggaaacagcatatgg atctcaccgtcaaaacc	
rplD_pSRK_GA_R	cgaggtcgacggatcgcgatcatttgaacc gctcctccagag	Molecular cloning of the <i>rplD</i> gene into the pSRK plasmid (Gibson Assembly)
rplD_G68H_F	tacaagcagaagcatacgggccg	Site mutagenesis in the <i>rplD</i> gene (Gly49His)
rplD_G68H_R	cgcgcccgatgcttctgcttgatc	
rplD_K65A_F	gcgccaagatgtacgcgagaagggt	Site mutagenesis in the <i>rplD</i> gene (Lys65Ala)
rplD_K65A_R	cgtacccttctgcgcgtacatcttgg	
rplV_pSRK_GA_F	ataacaatttcacacaggaaacagcatatgg gcaaggcaaaagcc	Molecular cloning of the <i>rplV</i> gene into the pSRK plasmid (Gibson Assembly)
rplV_pSRK_GA_R	cgaggtcgacggatcgcgatattatgcggcct ccccttgg	
rplD_K90R_F	gcttttggtggcaggtcgcgatcgtg	Site mutagenesis in the <i>rplV</i> gene (Lys90Ala)
rplD_K90R_R	cacgatcgacctgccaacaaaagc	
phzE_pSRK_GA_F	ataacaatttcacacaggaaacagcatatgg gtaaatctgaaagcg	Molecular cloning of the <i>phzE</i> gene into the pSRK plasmid (Gibson Assembly)
phzE_pSRK_GA_R	cgaggtcgacggatcgcgatcatttggag caaacg	
phzA_8xHis_NcoI_F	atataataaCCATGGctcatcaccatcat caccatcaccatacgcagattctgaatc cg	Molecular cloning of the <i>phzA</i> gene into the pET Duet plasmid (Forward primer includes a sequence coding for 8xHis-tag)
phzA_HindIII_R	tatatatatataAAGCTTcaggtcgaaatcg agctggcc	
phzC_BglII_F	ttatattaatAGATCTattttcggtttcccc g	Molecular cloning of the <i>phzCBD</i> genes into the pET Duet plasmid
phzD_KpnI_R	ttaatataaGGTACcttatgaaaatggc atgggctcc	

Supplementary Table 4. Cryo-EM data collection, refinement and validation statistics

<i>Eco50S</i> and <i>PHZ</i> (EMDB-20638) (PDB 6U48)	
Data collection and processing	
Magnification	215,000x
Voltage (kV)	300
Electron exposure (e ⁻ /Å ²)	30
Defocus range (μm)	-0.6 to -2.0
Pixel size (Å)	0.5568 (processed at 1.1136)
Symmetry imposed	none
Initial particle images (no.)	112,130
Final particle images (no.)	65,393
Map resolution (Å)	2.87
FSC threshold	0.143
Map resolution range (Å)	2.5-7
Refinement	
Initial model used (PDB code)	4YBB
Model resolution (Å)	2.7
FSC threshold	0.143
Map sharpening <i>B</i> factor (Å ²)	n/a*
Model composition	
Non-hydrogen atoms	90950
Protein residues	3356
Ligands	1
<i>B</i> factors (Å ²)	
Protein	33.20
Ligand	14.14
Nucleic Acid	52.39
R.m.s. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	0.946
Validation	
MolProbity score	1.58
Clashscore	3.46
Poor rotamers (%)	0.77
Ramachandran plot	
Favored (%)	93.26
Allowed (%)	6.32
Disallowed (%)	0.43

*Input charge density map for refinement was prepared as described in [169].

Supplementary Table 5. Crystallographic data and refinement parameters

	YejA^{Sm#}
PDB code	7Z8E
Crystallization conditions	14% PEG 8K, 0.1 M Tris-HCl pH 8.5, 0.2 M MgCl ₂
Beamline	SOLEIL-PX2
Wavelength (Å)	0.9793
Z _a	1
	<i>P2₁2₁2₁</i>
Space group	<i>a</i> = 59.9
Cell parameters (Å, °)	<i>b</i> = 73.7
	<i>c</i> = 140.8
Resolution (Å)	65.34-1.58 (1.67-1.58)
No. of observed reflections	765515 (40211)
No. of unique reflections	86259 (4780)
Completeness spherical (%)	100 (100)
Mean I/σ(I)	11.3 (0.6)
Completeness spherical Staraniso (%)	85.7 (22.6)
Completeness ellipsoidal Staraniso (%)	95.5 (59.7)
R _{merge} (%)	10.2 (164)
R _{pim} (%)	3.6 (53.8)
Mean I/σ(I) after Staraniso	13.1 (1.4)
CC _{1/2}	0.99 (0.54)
R _{cryst} (%)	16.9
R _{free} (%)	20.2
rms bond deviation (Å)	0.01
rms angle deviation (°)	0.93
Average B (Å ²)	
Protein	25
Peptide 1/2	27/29
Solvent	35
^a Clashscore	0.83
MolProbity score	0.86
^a Ramachandran plot (%)	
Favoured	98.48
Outliers	0

Values for the highest resolution shell are in parentheses

CC_{1/2} = percentage of correlation between intensities from random half-dataset

^aCalculated with MolProbity [250]

Numbers in italic account for statistical values after ellipsoidal mask application by Staraniso.

A dataset collected from a crystal, which diffracted anisotropically to 1.702 Å along *a**, 1.645 Å along *b** and 1.577 Å along *c**

Supplementary Table 6. Selected characteristics of experimentally validated BGCs of azol(in)e-modified RiPPs

Name	Subclass	Organism	MIBiG id	YcaO accession(s)	Reference (PubMed ID or DOI)
Bottromycin A2	Bottromycin	<i>Streptomyces bottropensis</i>	BGC0000468.1	WP_005486705.1, WP_020115555.1	19115340, 6337880, 7014241, 10.1039/C2SC21183A
Bottromycin A2	Bottromycin	<i>Streptomyces scabiei</i> 87.22	BGC0001157.1	WP_013003263.1, WP_037729686.1	19115340, 770464, 10.1039/C2SC21190D
Bottromycin A2	Bottromycin	<i>Streptomyces</i> sp. BC16019	BGC0000469.1	AFV25483.1, AFV25482.1	23021914
Bottromycin D	Bottromycin	<i>Streptomyces</i> sp. WMMB 272	BGC0000470.1	AFU90403.1, AFU90402.1	22984777
Aeruginosamide B	Cyanobactin	<i>Microcystis aeruginosa</i> PCC 9432	BGC0000483.1	WP_002754530.1	23911585
Microcyclamide	Cyanobactin	<i>Microcystis aeruginosa</i> NIES-298	BGC0000473.1	WP_103113147.1	18245249
Microcyclamide	Cyanobactin	<i>Microcystis aeruginosa</i> PCC 7806	BGC0000474.1	WP_002749237.1	18245249
Patellamide A	Cyanobactin	<i>Prochloron didemni</i>	BGC0000475.1	AAY21153.1	15883371
Patellin 2	Cyanobactin	uncultured <i>Prochloron</i> sp. 06037A	BGC0000477.1	ACA04490.1	18425112
Tenuencyclamide A	Cyanobactin	<i>Nostoc spongiaeforme</i> var. tenue str. Carmeli	BGC0000480.1	ACA04483.1	18425112
Trichamide	Cyanobactin	<i>Trichodesmium erythraeum</i> IMS101	BGC0000481.1	WP_011611942.1	16751554
Viridisamide A	Cyanobactin	<i>Oscillatoria nigro-viridis</i> PCC 7112	BGC0000471.1	WP_015177263.1	23911585
Azolemycin	LAP	<i>Streptomyces</i> sp. FXJ1.264		AMQ23506.1	28791101
Clostridiolysin S	LAP	<i>Clostridium botulinum</i> A str. ATCC 3502	BGC0000564.1	WP_011948242.1	20581111, 21315972
Clostridiolysin S	LAP	<i>Clostridium sporogenes</i> ATCC 15579	BGC0001170.1	WP_003488393.1	18375757, 19286651, 20581111, 21822292

Coryneazolicin	LAP	<i>Corynebacterium urealyticum</i> DSM 7111	BGC0001174.1	WP_012360880.1	21950656
Goadsporin	LAP	<i>Streptomyces</i> sp. TP-A0584	BGC0000565.1	BAE46919.2	16339937
Hakacin	LAP	<i>Bacillus cereus</i> VD214		WP_000512733.1	26462797
Heterocycloanthracin	LAP	<i>Bacillus thuringiensis</i> Al Hakam		ABK84467.1	26024319
Klebsazolicin	LAP	<i>Klebsiella pneumoniae</i> subsp. <i>ozaenae</i>	BGC0001638.1	WP_077257196.1	28846667
Listeriolysin S	LAP	<i>Listeria monocytogenes</i> serotype 4b F2365	BGC0001171.1	WP_003740559.1	18375757, 18787690, 19286651, 21075895, 21822292, 24606727
Microcin B17	LAP	<i>Escherichia coli</i>	BGC0000568.1	WP_001528602.1	19413755
Microcin B17-like	LAP	<i>Pseudomonas syringae</i> pv. <i>glycinea</i> B076		WP_004661048.1	26462797
Phazolicin	LAP	<i>Rhizobium</i> sp. Pop5		WP_008531785.1	31594941
Plantazolicin	LAP	<i>Bacillus amyloliquefaciens</i> subsp. <i>plantarum</i> FZB42	BGC0000569.1	WP_015239353.1	20971906, 21950656, 21568297, 23823732, 23878226, 24085393
Plantazolicin	LAP	<i>Bacillus pumilus</i> ATCC 7061	BGC0001173.1	WP_003211025.1	21950656, 23878226, 24085393
Sonorensin	LAP	<i>Bacillus sonorensis</i> SRCM101395		WP_006636650.1	24610839
Streptolysin S	LAP	<i>Streptococcus pyogenes</i> M1 GAS	BGC0000566.1	WP_002990434.1	10858242, 15819624, 18375757, 19286651, 21822292, 22001374
Trifolitoxin	LAP	<i>Rhizobium leguminosarum</i> bv. <i>trifolii</i> T24	BGC0000567.1	P42727.1	8509324
Neothioviridamide	Thioamitide	<i>Streptomyces</i> sp.	BGC0001696.1	BBC15204.1	29381067
Thioholgamide A	Thioamitide	<i>Streptomyces malaysiense</i>	BGC0001802.1	WP_046417233.1	28981254
Thiovarsolin	Thioamitide	<i>Streptomyces cinnamoneus</i> ATCC 12686		WP_071967420.1	30916321
Thioviridamide	Thioamitide	<i>Streptomyces olivoviridis</i>	BGC0000625.1	BAN83923.1	23995943
Berninamycin	Thiopeptide	<i>Streptomyces bernensis</i>	BGC0001472.1	WP_051709136.1	23650400

Cyclothiazomycin A	Thiopeptide	<i>Streptomyces hygroscopicus</i>	BGC0000603.1	ACS50130.1	24937678, 24769844, 20154110, 2071486
Cyclothiazomycin B	Thiopeptide	<i>Streptomyces mobaraensis</i> NBRC 13819	BGC0001145.1	WP_004943462.1, WP_004943471.1	24937678, 21885289, 17010619
Cyclothiazomycin C	Thiopeptide	<i>Streptomyces</i> sp. NRRL WC-3908	BGC0001146.1	AID54694.1	24937678
GE2270	Thiopeptide	<i>Planobispora rosea</i>	BGC0001155.1	WP_141703552.1, AGY49581.1	23932526, 24598591, 19338336
GE2270A	Thiopeptide	<i>Nonomurea</i> sp. WU8817	BGC0000604.1	ACS83760.1, ACS83772.1	19338336
GE37468	Thiopeptide	<i>Streptomyces</i> sp. ATCC 55365	BGC0000605.1	AEM00626.1	21788474
Kocurin	Thiopeptide	<i>Kocuria rosea</i>	BGC0001640.1	WP_109243203.1	28942758
Lactazole	Thiopeptide	<i>Streptomyces lactacystinaeus</i>	BGC0000606.1	BAO57438.1	24768308
Lactocillin	Thiopeptide	<i>Lactobacillus gasseri</i> JV-V03	BGC0000628.1	WP_003649848.1	25215495
Micrococcin P1	Thiopeptide	<i>Micrococcus caseolyticus</i>	BGC0000607.1	AIU53945.1	25313391
Nocathiacin	Thiopeptide	<i>Nocardia</i> sp. ATCC 202099	BGC0000609.1	ADR01083.1	20473441
Nosiheptide	Thiopeptide	<i>Streptomyces actuosus</i>	BGC0000610.1	WP_110629801.1	19678698
Nosiheptide	Thiopeptide	<i>Streptomyces</i> sp.	BGC0001707.1	AQM75233.1	
Radamycin	Thiopeptide	<i>Streptomyces globisporus</i> subsp. <i>globisporus</i>	BGC0001753.1	WP_030578081.1, WP_030578068.1	10.1016/j.tetlet.2017.12.056
Saalfelduracin	Thiopeptide	<i>Amycolatopsis saalfeldensis</i>	BGC0001471.1	WP_091614694.1, WP_091614389.1	29983054, 29507203,
Sch 18640	Thiopeptide	<i>Micromonospora arborensis</i>	BGC0001473.1	WP_110565442.1, WP_110565440.1	29983054, 10.1021/ja00407a047
Siomycin	Thiopeptide	<i>Streptomyces sioyaensis</i>	BGC0000611.1	WP_129248666.1	19246004
Thiocillin I	Thiopeptide	<i>Bacillus cereus</i> ATCC 14579	BGC0000612.1	WP_011110463.1	19196969, 19246004
Thiomuracin	Thiopeptide	<i>Nonomurea</i> sp. Bp3714-39	BGC0000613.1	ACS83787.1, ACS83786.1	19338336

Thiopeptin	Thiopeptide	Streptomyces tateyamensis	BGC0001474.1	WP_110669060.1, WP_146259077.1	29983054, 5045467
Thiostrepton	Thiopeptide	Streptomyces laurentii	BGC0000614.1	ACN52298.1	19265401
TP-1161	Thiopeptide	Nocardioopsis sp. TFS65-07	BGC0000615.1	WP_017535900.1, WP_017535895.1	20562278
Curacozole	Polyazole cyclopeptide	Streptomyces curacoi DSM 40107		WP_079051299.1	30310179
Telomestatin	Polyazole cyclopeptides	Streptomyces anulatus	BGC0001797.1	BBA31819.1	28611443

Supplementary Table 7. Selected characteristics of BGCs of azol(in)e-modified RiPPs predicted in [228]

Name	Organism	YcaO accession
NHLP-Burk	Burkholderia pseudomallei 354a	EIF80112.1
Coryneazolysin	Corynebacterium diphtheriae DSM 43988	ERA50922.1
Faecailisin	Enterococcus faecalis NJ44	KAJ85909.1
Haloazolisin	Haloferax denitrificans ATCC 35960	WP_004044676.1
Helicobactin	Helicobacter pylori GAM201Ai	WP_001934599.1
Propionisin	Propionibacterium acnes SK137	ADD99586.1
Thermoacidophisin	Sulfolobus acidocaldarius N8	WP_011277428.1