

**Making and Breaking Antibiotics –  
Structural Studies of Bottromycin Biosynthesis  
and  
a Mechanism of Albicidin Resistance**

Dissertation  
zur Erlangung des Grades  
des Doktors der Naturwissenschaften  
der Naturwissenschaftlich-Technischen Fakultät der Universität des Saarlandes

von

**Asfandyar Sikandar**

Saarbrücken

2020

Diese Arbeit entstand unter der Anleitung von Jun.-Prof. Jesko Köhnke in der Fachrichtung 8.2, Pharmazeutische Biotechnologie der Naturwissenschaftlich-Technischen Fakultät III der Universität des Saarlandes von September 2015 bis Februar 2020.

**Tag des Kolloquiums:** 13. Oktober 2020

**Dekan:** Prof. Dr. Jörn Eric Walter

**Berichterstatter:** Juniorprof. Dr. J. Köhnke  
Prof. Dr. Rolf Müller  
Prof. Dr. James H. Naismith

**Akad. Mitglied:** Dr. Maksym Myronovskyi

**Vorsitz:** Prof. Dr. Uli Kazmaier

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

## Acknowledgement

I would like to begin by thanking my supervisor, Jun.-Prof. Jesko Köhnke, for the opportunity to work in his group. You gave me the freedom to pursue as many side projects as I wanted, something which may not have been possible in other labs - in hindsight I could have managed with considerably less crystallization plates. The last four years in your lab have been the most exciting and successful period of my research career so far, and for that I am truly grateful to you. I am also grateful for my excellent collaborators and their teams: Prof. Rolf Müller, Jun.-Prof. Alexander Titz, Prof. Andriy Luzhetskyy and Prof. Olga Kalinina. I would also like to thank the members of my thesis committee: Prof. Rolf Müller (co-supervisor), Dr. Antoine Abou Fayad, Dr. Emilia Oueis and Dr. Walter Hofer, for their guidance and critique of my work.

During my time at HIPS I have been fortunate to work with a number of fantastic colleagues. I would like to thank every one of them for their help during the course of my PhD, especially: Dr. Chengzhang Fu, Dr. Bastian Schnell, Maja Remskar, Dr. Francisco Marcos-Torres, Alexander Popoff, Katarina Cirnski, Joachim Hug, Dr. Domen Pogorevc and Dr. Nestor Zaburannyi. Working at HIPS would not have been the same without all of you. I am also grateful to Patrick Haack for his help with Mass-Spec.

Finally, I would like to thank my family - Baba, Ammi, Shmyalla and Zarrar - and Maria. Decision to leave home to pursue education outside Pakistan has not been an easy one. I cannot express in words how much the love and support from each of you has meant to me over all these years. It goes without saying that none of this would have been possible without all of you.

## **Abstract**

The emergence of antibacterial resistance is a major public health threat that warrants urgent attention. The discovery of new antibacterial natural products and unravelling of their biosynthesis, coupled with the study of resistance mechanisms employed by bacteria to evade antibiotics will be vital in ensuring that we maintain a viable antibiotic development pipeline. This thesis covers two aspects of antibiotic research: deciphering bottromycin biosynthesis and AlbA mediated antibiotic resistance.

Bottromycins are potent antibacterial peptides that bind a novel bacterial target: The A-site of the ribosome. Bottromycin biosynthesis is still not fully understood, especially the role of two hydrolases, BotAH and BotH. The in-depth functional and structural characterization studies presented here demonstrate that BotAH is responsible for a key proteolytic step, whereas BotH is an atypical hydrolase-like enzyme responsible for the post-translational epimerization of L-Asp to D-Asp in bottromycin biosynthesis.

AlbA is a resistance protein that had previously been reported to neutralize a potent antibacterial compound, albicidin. The structural and functional studies of AlbA presented here not only reveal the mode of albicidin binding, but also the underlying mechanism of its modification by AlbA, resulting in loss of potency. The data also demonstrate that AlbA constitutes an autoregulated antibiotic resistance system, present in a wide variety of pathogenic bacteria.

## **Zusammenfassung**

Die Entstehung von Antibiotikaresistenzen ist eine ernste Gefahr für die Gesundheit der Weltbevölkerung und bedarf dringender Aufmerksamkeit. Die Entdeckung neuer Naturstoffe mit antibiotischer Wirkung und ein Entschlüsseln ihrer Biosynthese zusammen mit einem besseren Verständnis der Resistenzmechanismen, die Bakterien entwickeln um sich zur Wehr zu setzen, ist essentiell um sicherzustellen, dass wir eine tragfähige Strategie zur Entwicklung neuer Antibiotika haben. Diese Doktorarbeit beschäftigt sich mit zwei Aspekten der Antibiotikaforschung: Einem besseren Verständnis der Botromycin Biosynthese und dem Resistenzprotein AlbA.

Botromycine sind potente antibakterielle Peptide und ihre Biosynthese ist noch nicht vollständig erforscht, besonders die Rolle der beiden Hydrolasen BotAH und BotH. Die detaillierte funktionelle und strukturelle Charakterisierung der beiden Proteine ist Teil dieser Arbeit. BotAH ist für einen wichtigen proteolytischen Schritt in der Biosynthese verantwortlich, wohingegen BotH eine sehr atypische Hydrolase ist und die post-translationale Umwandlung von l-Asp in d-Asp katalysiert.

AlbA neutralisiert den potenten antibiotischen Naturstoff Albicidin. Die strukturelle und funktionelle Untersuchung von AlbA zeigte nicht nur auf wie AlbA Albicidin bindet und es dabei modifiziert, sondern auch das AlbA ein autoinduzierbares antibiotisches Resistenzprotein ist, das in vielen pathogenen Bakterien zu finden ist.

## Publications

The dissertation is composed of the following published or submitted manuscripts:

1. **Sikandar, A.\***, Franz, L.\*, Melse, O.V., Antes, I., and Köhnke, J., *J. Am. Chem. Soc.* **2019**, 141, 25, 9748-9752.
2. **Sikandar, A.\***, Franz, L.\*, Sebastian, A., Aberturas, J.S., Horbal, L., Luzhetskyy, A., Truman, A., Kalinina, O.V., and Köhnke, J., The post-translational amino acid epimerase BotH defines a new group of atypical hydrolase-like enzymes (**In revision**).
3. **Sikandar, A.**, Cirnski, K., Testolin, G., Volz, C., Brönstrup, M., Kalinina, O.V., Müller, R., and Köhnke, J., *J. Am. Chem. Soc.* **2018**, 140, 48, 16641-16649.

The co-authored publications which are not part of the dissertation are:

1. Beshr, G., **Sikandar, A.**, Jemiller, E.M., Klymiuk, N., Hauck, D., Wagner, S., Wolf, E., Köhnke, J. and Titz, A., *J. Biol. Chem.* **2017**, 292, 48, 19935-19951.
2. Kjaerulff, L., **Sikandar, A.**, Zaburannyi, N., Adam, S., Hermann, J., Köhnke, J. and Müller, R., *ACS Chem. Biol.* **2017**, 12, 11, 2837-2841.
3. Fu, C.\*, **Sikandar, A.\***, Donner, J.\*, Zaburannyi, N., Hermann, J., Reck, M., Wagner-Döbler, I., Köhnke, J. and Müller, R., *Nat. Commun.* **2017**, 4, 6, 2449-2455.
4. Kany, A.M.\*, **Sikandar, A.\***, Yahiaoui, S., Haupenthal, J., Walter, I., Empting, M., Köhnke, J. and Hartman, R.W., *ACS Infect. Dis.* 2018, 4, 6, 988-987.
5. Kany, A.M., **Sikandar, A.**, Yahiaoui, S., Haupenthal, J., Walter, I., Empting, M., Köhnke, J. and Hartman, R.W., *ACS Chem. Biol.* **2018**, 13, 9, 2449-2455.
6. **Sikandar, A.** and Köhnke, J., *Nat. Prod. Rep.* **2019**, 141, 25, 9748-9752.
7. Keller, L., Canuto, K.M., Liu, C., Suzuki, B.M., Almatili, J., **Sikandar, A.**, Naman, C.B., Glukhov, E., Luo, D., Duggan, B.M., Luesch, H., Köhnke, J., O'Donoghue, A.J., and Gerwick, W.H., *ACS Chem. Biol.*, **2020**, XXXX, XXX, XXXXXX.

(\*: Authors contributed equally to the work)

## Patents

1. WO2019122400. Inventors: Müller, R., Lukezic, T., Remskar, M., Zaburannyi, N., Bader, C., **Sikandar, A.**, and Köhnke, J..

# Table of Contents

<b>Chapter 1: Introduction.....</b>	<b>1</b>
<b>1 Drug Discovery and Natural Products.....</b>	<b>1</b>
<b>1.1 Introduction to Natural Products.....</b>	<b>1</b>
<b>1.2 Microbial Natural Product and Antibacterial Development .....</b>	<b>1</b>
<b>1.3 Biosynthesis of Microbial Natural Products.....</b>	<b>4</b>
1.3.1 Polyketide Synthase .....	5
1.3.2 Nonribosomal Peptide Synthetase.....	6
1.3.3 Ribosomally Synthesized and Post-Translationally Modified Peptide.....	9
1.3.3.1 Bottromycins .....	11
<b>2 Mechanisms of Antibiotic Resistance .....</b>	<b>17</b>
<b>2.1 Introduction.....</b>	<b>17</b>
<b>2.2 Overview of Antibiotic Resistance Mechanisms.....</b>	<b>17</b>
2.2.1 Thiostrepton-induced protein A .....	18
<b>3 Outline of the thesis.....</b>	<b>21</b>
<b>4 Reference.....</b>	<b>23</b>
<b>Chapter 2: The Thiazole-Specific Amydohydrolase PurAH is the Gatekeeper of Bottromycin Biosynthesis .....</b>	<b>30</b>
<b>1 Introduction .....</b>	<b>31</b>
<b>2 Results and Discussion .....</b>	<b>32</b>
<b>3 Supplementary Information.....</b>	<b>39</b>
<b>3.1 Cloning, protein expression and purification .....</b>	<b>39</b>
3.1.1 Cloning, Expression and Purification of PurCD and IpoC.....	39
3.1.2 Cloning of PurAH wt and mutants .....	39
3.1.3 Cloning, Expression and Purification BotA and BotA point mutants .....	39
3.1.4 Protein expression and purification of PurAH wt and mutants .....	39
<b>3.2 Biochemistry .....</b>	<b>40</b>
3.2.1 Heterocyclization reaction by IpoC.....	40
3.2.2 Macrocyclization reaction by PurCD .....	40
3.2.3 Follower cleavage by PurAH .....	41
<b>3.3 Mass spectrometry .....</b>	<b>42</b>
3.3.1 LC-ESI-MS analysis .....	42
3.3.2 MS <sup>2</sup> Fragmentation .....	44
<b>3.4 Crystallization and structure determination .....</b>	<b>44</b>
<b>4 References .....</b>	<b>45</b>
<b>Chapter 3: The Post-translational Amino Acid Epimerase BotH Defines a New Group of Atypical alpha/beta-hydrolase-fold Enzymes .....</b>	<b>67</b>
<b>1 Introduction .....</b>	<b>68</b>
<b>2 Results and Discussion .....</b>	<b>70</b>
<b>2.1 BotH is an unusual member of the ABH superfamily .....</b>	<b>70</b>
<b>2.2 BotH catalyzes the epimerization of 3b and 3a .....</b>	<b>70</b>
<b>2.3 Structure of the BotH-3a complex.....</b>	<b>72</b>



2.4	BothH has relaxed substrate specificity .....	72
2.5	BothH binds bottromycins, which act as inhibitors of epimerization .....	74
2.6	Evolution of BothH and its distribution amongst bacterial BGCs .....	75
3	Conclusion.....	75
4	Supplementary Information.....	78
4.1	Methods.....	78
4.1.1	Protein expression and purification .....	78
4.1.2	Production of seleno-methionine (SeMet) BothH.....	78
4.1.3	Crystallization and structure determination.....	79
4.1.4	Microscale Thermophoresis (MST) .....	79
4.1.5	Marfey derivatization .....	80
4.2	Enzymatic reaction of BothH with 3 .....	80
4.3	Bottromycin extraction and purification .....	81
4.4	<i>o</i> production and purification of 3.....	81
4.5	Spontaneous epimerization of 3.....	82
4.6	Pull-down of bottromycin A2 and 3a/b using BothH.....	82
4.7	Incorporation of a deuterium at Asp C $\alpha$ of 3.....	82
4.8	Reaction of BothH with derivatives of 3.....	83
4.9	Orthosteric inhibition of BothH epimerization by bottromycin A2 .....	83
4.10	LC-MS and MS <sup>2</sup> analysis.....	84
4.11	Heterologous expression of btmH and btmA in <i>Streptomyces coelicolor</i> M145 .....	85
4.12	Bioassays .....	85
4.13	Bioinformatics analysis.....	86
5	References .....	87
<b>Chapter 4: Adaptation of a Bacterial Multidrug Resistance System Revealed by the Structure and Function of AlbA .....</b>		
<b>116</b>		
1	Introduction .....	117
2	Results .....	119
2.1	AlbA is a structural homolog of TipA-S .....	119
2.2	AlbAwt promotes the cyclization of albicidin.....	122
2.3	Rationalizing AlbA activity .....	125
2.4	Cyclization decreases albicidin activity, but AlbA is insufficient to protect cells from albicidin.....	126
2.5	Evolution of AlbA and evidence for a more efficient homolog in <i>K. pneumoniae</i> .....	128
2.6	AlbA defines a new TipA-like family .....	129
3	Discussion .....	131
4	Supplementary Information.....	134
4.1	Methods.....	134
4.1.1	Cloning and mutagenesis .....	134
4.1.2	Protein expression and purification .....	134
4.1.3	Crystallization and structure determination.....	135

4.1.4	Affinity measurements .....	135
4.1.5	MS and MS <sup>2</sup> analyses.....	136
<b>4.2</b>	<b>Biochemical assays .....</b>	<b>137</b>
<b>4.3</b>	<b>Testing of biological activity.....</b>	<b>137</b>
<b>4.4</b>	<b>Determining MIC values for AlbA-overexpressing E.coli.....</b>	<b>137</b>
<b>4.5</b>	<b>Growth curves .....</b>	<b>137</b>
<b>4.6</b>	<b>RT-qPCR .....</b>	<b>137</b>
4.6.1	RNA extraction .....	137
4.6.2	cDNA synthesis.....	138
4.6.3	Quantitative real-time PCR .....	138
<b>4.7</b>	<b>Phylogenetic analysis .....</b>	<b>139</b>
<b>4.8</b>	<b>Synthesis of albicidin .....</b>	<b>139</b>
<b>4.9</b>	<b>NMR spectra.....</b>	<b>142</b>
4.9.1	Allyl protected albicidin.....	142
4.9.2	Albicidin.....	143
4.9.3	1, 1-adequate .....	145
<b>5</b>	<b>References .....</b>	<b>146</b>
<b>Chapter 5: Discussion .....</b>		<b>170</b>
<b>1</b>	<b>Functional and Structural Characterization of BotAH and BotH .....</b>	<b>170</b>
<b>2</b>	<b>Functional and Structural Characterization of AlbA.....</b>	<b>175</b>
<b>3</b>	<b>Concluding Remarks.....</b>	<b>179</b>
<b>4</b>	<b>References .....</b>	<b>180</b>
<b>Additional Publications.....</b>		<b>184</b>
Photorhabdus luminescens lectin A (PIIA) .....		184
Thioholgamides .....		201
Carolacton inhibits folate-dependent C1 metabolism .....		206
Thiol-based LasB inhibitors .....		215
Hydroxamic acid-based LasB inhibitors .....		225
The role of protein-protein interactions in RiPPs.....		232
Tutuilamides A-C.....		245



# Chapter 1: Introduction

## 1 Drug Discovery and Natural Products

### 1.1 Introduction to Natural Products

Natural products (NPs) are secondary metabolites produced by organisms that are not directly involved in the growth and development of the producing organism<sup>1</sup>. These secondary metabolites, however, are often produced to gain evolutionary advantages e.g. by acting as chemical weapons against other organisms in the environment<sup>2,3</sup>. They are by design bioactive, thus they have been an outstanding source of therapeutics in the mostly non-natural target organism homo sapiens. NPs have diverse biological activities, and apart from antibacterial, antifungal and antiviral activities, these compounds are also being used as immunosuppressants, anticancer, cholesterol-lowering-and antiparasitic agents - just to mention a few<sup>4-6</sup>. Moreover, such compounds are also used in the agriculture sector as herbicides, insecticides and fungicides<sup>7</sup>. Suffice to say, NPs have played a dominant role in human progress.

### 1.2 Microbial Natural Product and Antibacterial Development

Historically, plant-derived NPs were a major source of medicinal preparations; by the early 1900's 80 % of all the drugs in use were obtained from plant sources<sup>8-10</sup>. However, the discovery of penicillin from *Penicillium notatum* by Alexander Fleming in 1928 resulted in a paradigm shift from plants to microorganism as a source of NPs, especially antibiotics<sup>11</sup>. For the next 20 years, termed 'golden age' of antibiotic discovery, extensive mining of microbes as sources of antibiotics resulted in the identification, followed by development for clinical use, of a number of novel antibiotics (Figure 1)<sup>12-14</sup>. As the identification of new antibiotics slowed down, this simple and cost-effective screening platform appeared to have exhausted its sources by late 1960s<sup>13</sup>. The repeated isolation of the same NPs from different microorganisms, coupled with the rise in antibacterial resistance to NPs, resulted in the divestment in NP research, and a general shift in focus towards numerous high-throughput screening (HTS) programmes aimed to discover new synthetic antibiotics<sup>15</sup>. It was generally assumed that only a few antibacterial secondary metabolites were left to be identified, and this would require massive screening programmes<sup>16-18</sup>. Therefore, the HTS programmes were considered a better alternative to the

mining of microbes for new antibacterial agents. This strategy proved largely unsuccessful as only two new synthetic antibiotics entered the clinic in the last 40 years (Figure 1)<sup>14</sup>.

By the late 1990s it became clear that drug discovery platforms were unable to keep up with antibiotic resistance. The pharmaceutical industry responded by developing new high-tech drug discovery platforms combining genomics, combinatorial chemistry, HTS and rational drug design<sup>15,19</sup>. One such platform relied on identification of conserved essential proteins in bacteria through genomics that would then serve as targets of HTS and rational drug design. This strategy was successful in the identification of inhibitors of targets, but these compounds failed to make it beyond early-stage leads. The inhibitors suffered from two major drawbacks: they lacked reasonable spectra of activity against clinically important pathogens, and the ability to sufficiently penetrate the bacterial cell wall, especially in Gram-negative bacteria, to reach their targets. Between 1995-2001 GSK alone ran 70 HTS campaigns, costing approx. US\$1 million per campaign, which resulted in the identification of only five leads - none of which progressed to clinical trials<sup>19</sup>. Lack of success in finding new antibacterial compounds and the low returns on investment resulted in a decline in research and development of antibiotics, with many large pharmaceutical companies leaving the field altogether.

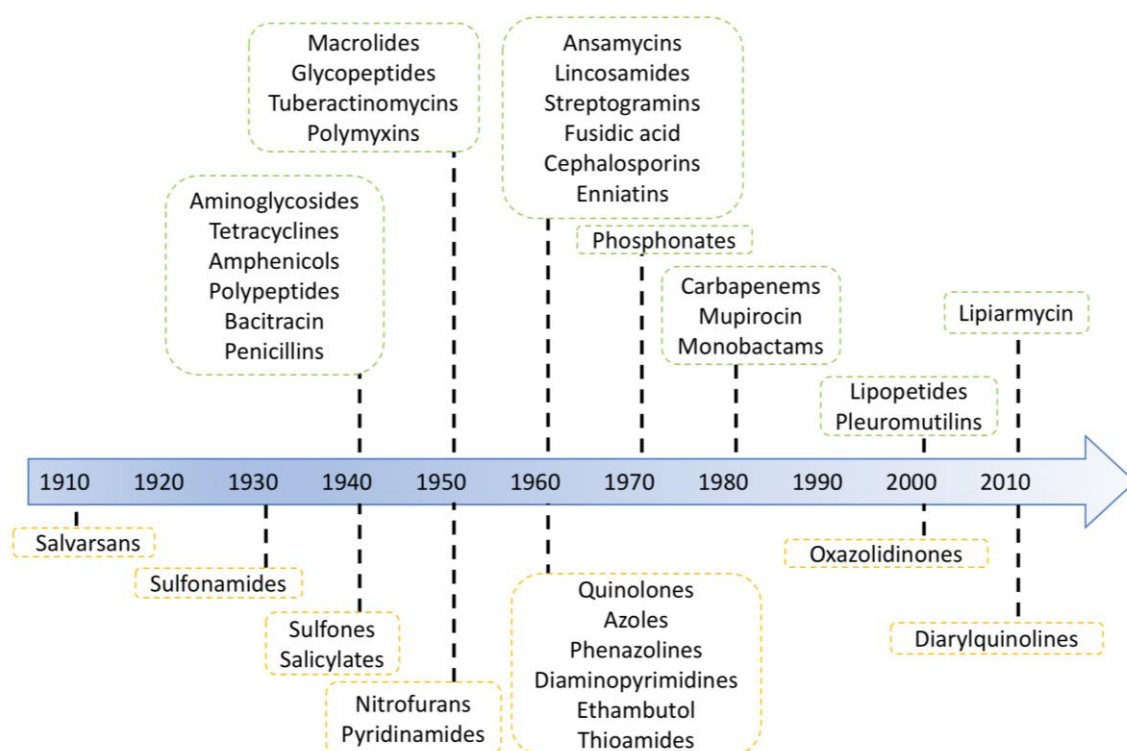


Figure 1. Timeline showing the decade a new class of antibiotic was introduced in the clinic. The source of antibiotic is depicted: green dotted boxes represent antibiotics from microbes, whereas the orange dotted boxes represent synthetic antibiotics. Azoles and Phenazines synthesis was inspired by a NP<sup>13</sup>.

In retrospect, one of the major reasons why non-NP drug discovery platforms have had limited success is due to a lack of structural diversity and complexity covered by synthetic screens<sup>15,19</sup>. It comes as no surprise that this is one area where nature excels at<sup>20-24</sup>. NPs produced by microorganisms also represent pre-validated set of compounds that have been optimized by nature as a result of billions of years of natural selection for optimal interaction with a wide variety of biological macromolecules - so-called privileged structures<sup>24,25</sup>. Therefore, NPs possess physicochemical properties that are crucial for biological activity e.g. NPs are more likely to act as substrates for one or more of many transport systems presents in bacteria to reach the intracellular targets - a quality synthetic compounds usually lack. It is due to these physicochemical properties that NP-based libraries, in comparison to synthetic libraries, have been shown to yield higher-quality leads with a significantly higher hit rate<sup>5,20</sup>. Above all, and in contrast to earlier assumptions, microorganisms as a reservoir of NPs are far from exhausted<sup>22,26,27</sup>. During the ‘golden era’ screening was rather limited in terms of biological diversity as it mostly focused on soil-dwelling bacteria, primarily bacteria of the order Actinomycetales<sup>13,27,28</sup>. Marine ecosystems represent a vast, untapped resource of novel NPs; a recent foray into marine environments has led to the discovery of a plethora of NPs from sponges and corals<sup>29</sup>. Interestingly, it appears that symbiotic microbes associated with marine invertebrates are responsible for the production of these compounds<sup>29</sup>. Other underexploited resources include proteobacteria - such as myxobacteria, pseudomonads, burkholderias - firmicutes and fungi etc. In addition, genome sequencing has revealed that the capacity of organisms for NP biosynthesis has been largely underestimated – on average each strain can produce 20 - 40 secondary metabolites, many of which have not yet been characterized<sup>27</sup>. The biosynthetic machinery (or biosynthetic gene cluster (BCG); explained later) responsible for synthesis of these metabolites is poorly expressed under standard laboratory growth conditions. The importance of these so-called ‘silent or cryptic clusters’ has been recognized by the research community and a number of methods have been develop to activate them, resulting in the identification of new NPs with potent antibacterial properties<sup>30,31</sup>. Given the advantages described above, along with the realization that only a small fraction of the microbial world has been explored, and the recent advances in genome miming and activation of silent clusters, NPs still represent one of the most important sources of antibacterial compounds.

To summarize, NPs are an integral part of antibacterial research and development, and the value of these compounds can be gauged from the fact that currently the vast majority of antibacterial drugs in use today are either NPs or their derivatives - of the nine representative classes of

antibiotics of the modern era shown in Figure 2, six belong to NPs, with only three a result of synthetic chemistry<sup>32</sup>.

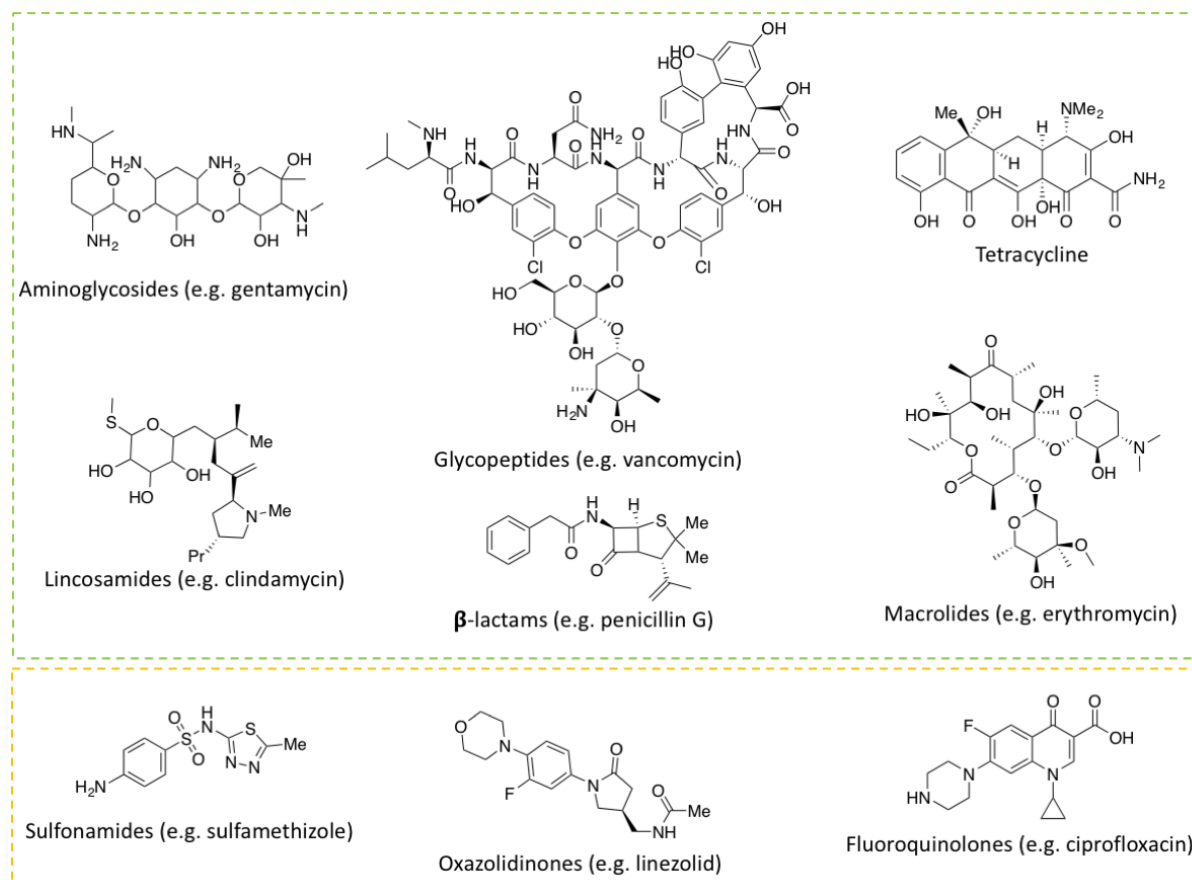


Figure 2. Representative classes of antibiotics of the modern era. The dotted boxes are colored per antibiotic class: green = NP and orange = synthetic.

### 1.3 Biosynthesis of Microbial Natural Products

NPs are produced by a series of proteins encoded by genes, usually clustered in the microbial genome, forming a BGC<sup>33</sup>. This paradigm facilitates their identification and characterization, which has been further streamlined by the advances in next-generation DNA sequencing and bioinformatics tools<sup>34-37</sup>. BGCs are commonly classified based on their product(s), which despite their enormous structural diversity is in principal based on a variation of a few predominant pathways, namely the non-ribosomal peptide synthetase (NRPS), the polyketide synthase (PKS), and the ribosomally synthesized and post-translationally modified peptide (RiPP) pathways.

A detailed discussion of microbial NP biosynthetic pathways is outside the scope of this thesis, and therefore, the readers are directed to a number of reviews on this topic<sup>38-47</sup>.

### 1.3.1 Polyketide Synthase

Polyketide NPs are produced by enzymes called polyketide synthases (PKSs), which are structurally and functionally similar to the fatty acid synthase (FAS) machinery<sup>48,49</sup>. PKS use some of the simplest biosynthetic building blocks: coenzyme-A (CoA)-activated starter units (e.g. acetyl-coenzyme A) and extender unit (e.g. malonyl-coenzyme A), and based on their architecture can be divided into three types (I-III): Type I PKSs are organized into modules, composed of multifunctional enzymes acting non-iteratively; type II PKSs consists of multifunctional enzymes that act iteratively, and type III, the simplest of the PKSs, are formed by iteratively acting homodimeric enzymes<sup>41,50</sup>. A few notable PKS derived NPs are shown in Figure 3, depicting the structural diversity exhibited by this class of NP<sup>51-54</sup>.

In bacteria, type I polyketide NPs are the most common. As mentioned earlier they are organized into modules, wherein each module is responsible for the addition, and thereby extension of polyketide chain, by an extender unit<sup>48,49</sup>. Generally speaking each elongation step requires the action of minimally three core domains: keto synthase (KS), acyltransferase (AT), and acyl carrier protein (ACP) (Figure 3). The AT recruits the extender unit, which is transferred to the 4-phosphopantetheine (Ppant) attached to the ACP domain via phosphopantetheinyl transferases (PPTases). The ACP then delivers the building block to KS domain that catalyzes the decarboxylative Claisen-like condensation between the extender unit and the growing polyketide chain. The ACP domain can then either transfer the processed intermediate to the downstream KS of the downstream module for further elongation or to a chain-releasing thioesterase (TE) domain, usually present at the end of the assembly line, for release of the product. The Ppant of ACP functions as a flexible arm, able to channel the attached nascent polyketide to additional processing domains that may be present between the AT and ACP domains to generate vastly diverse and complex polyketides e.g. ketoreductase (KR), dehydratase (DH), and enolreductase (ER) domains, which are responsible for keto group modification. In addition, various starter and extender units can be utilized based on the AT domain substrate specificity<sup>55,56</sup>. Further post-PKS modifications such as oxidation, glycosylation, hydroxylation, amination etc. contribute to the impressive structural diversity observed for PKS derived NPs<sup>57,58</sup>.



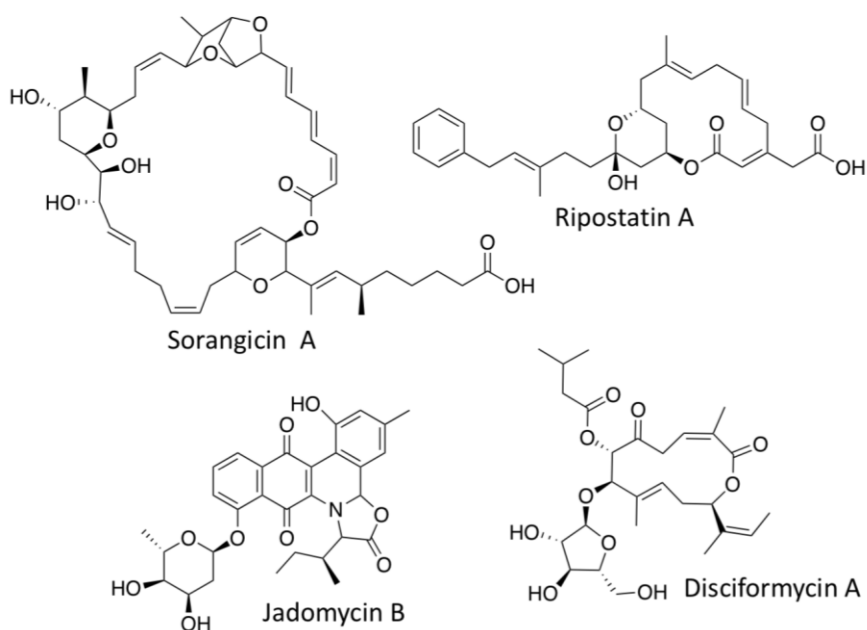
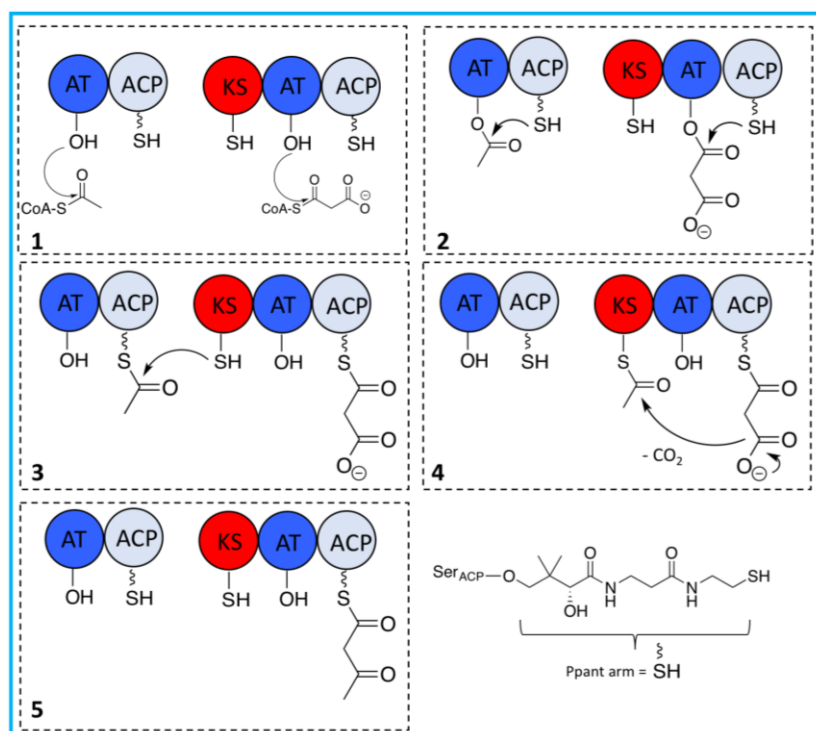


Figure 3. PKS NP biosynthesis and structural diversity. (Top) The schematic reaction steps for one elongation step in PKS pathway is shown in blue box. (1) Loading of an activated precursor (starter or extender) units onto acyltransferase (AT) domain; (2) Transfer of precursor unit to phosphopantetheine arm (Ppant), depicted as a wavy line, attached to a serine of the acyl carrier protein (ACP); (3) Transfer of starter unit to ketosynthase (KS) domain; (4) Condensation of starter unit with extender unit catalysed by KS domain; (5) Extension of chain by one starter unit. (Bottom) A few NPs synthesized by PKS pathways are displayed.

### 1.3.2 Nonribosomal Peptide Synthetase

Nonribosomal peptides are synthesized on large multidomain enzyme complexes, nonribosomal peptide synthetases (NRPSs), that are usually composed of modules working in an assembly line synthetic logic; wherein each module within an NRPS is responsible for the

incorporation of a single amino acid into the growing chain<sup>42,43,59</sup>. These modules are themselves comprised of domains that catalyse the individual steps of peptide synthesis, which proceeds in a N- to C-terminal direction. Dissociated NRPSs consisting of stand-alone modules and domains have also been reported, which are frequently involved in the generation and delivery of building blocks to the modular NRPS assembly line. Molecules made by NRPS can be linear, cyclic or branched-cyclic in nature, usually ranging from 3 - 15 amino acids in length<sup>60</sup>. Structures of a few antibacterial NPs synthesized by NRPSs, including the last resort antibiotic vancomycin, are shown in Figure 4<sup>61-65</sup>.

Despite the structural complexity observed in nonribosomal peptides, NRPS modules are formed by three core domains: adenylation (A), condensation (C) and peptidyl carrier protein (PCP) (Figure 4)<sup>42,44</sup>. The A domain acts a gatekeeper by selecting and activating the amino acid to be loaded, as the aminoacyl-AMP, and transfers it to the paired PCP domain by forming a thioester linkage with the terminal thiol of a phosphopantetheine (Ppant; shown as a wavy bond attached to SH in Figure 4). The Ppant acts as a flexible arm carrying the bound amino acid towards the condensation domain (C). At the same time another PCP-bound amino acid, carried by a downstream module, is directed towards the C domain, which can now carry out the peptide bond formation between the PCP-bound amino acids. During this process the amino acid is also transferred from one module to another in the assembly line (Figure 4). Just like in PKSs, a terminal thioesterase (TE) domain is usually responsible for the release of the peptide either by hydrolysis or internal cyclization in bacteria. By decoupling peptide synthesis from the ribosome, NRPSs - more specifically the A domains- are able to access proteinogenic and non-proteinogenic amino acids; this coupled with the coordinated action of a set of dedicated additional domains and tailoring enzymes catalysing e.g. epimerization (E), oxidation (Ox), methylation (MT), halogenation (Hal) and cyclization (Cy) etc., is the source of structural diversity observed in NRPS derived NPs<sup>42,44</sup>.

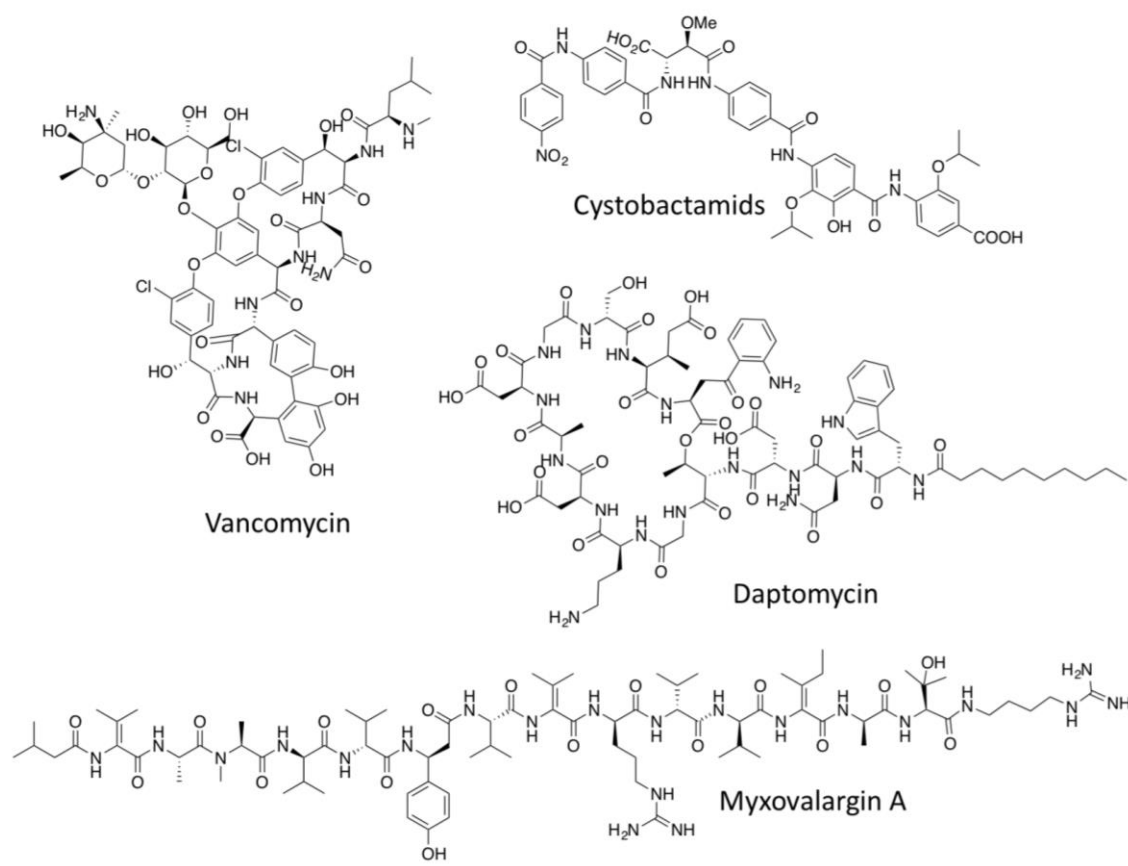
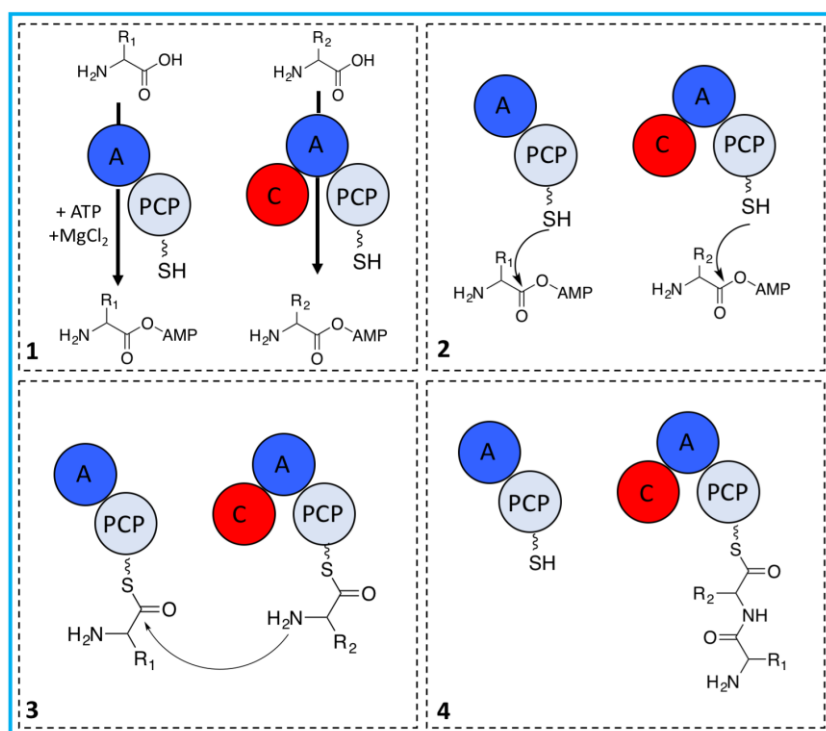


Figure 4. NRPS NP biosynthesis and structural diversity. (Top) The schematic representation of the NRPS elongation cycle. (1) Activation of amino acid by adenylation (A) domain; (2) Transfer of the activated amino acid onto the peptidyl carrier protein (PCP) linked P<sub>ant</sub> arm (shown as wavy bond attached to SH); (3) Peptide bond formation between PCP-linked amino acids is carried out by the condensation (C) domain; (4) The PCP-linked elongated peptide is then ready to be delivered to the downstream C domain for next round of elongation cycle. The scheme features a loading unit (A-PCP) which usually provides the initial amino acid in NRPS pathway. (Bottom) A few NPs produced by NRPS pathways are displayed.

### 1.3.3 Ribosomally Synthesized and Post-Translationally Modified Peptide

Ribosomally synthesized and post-translationally modified peptides (RiPPs) are synthesized as precursor peptides that undergo post-translational modification(s) (PTM) by one or more enzymes to give rise to a mature peptide with structural elements crucial for bioactivity, increased metabolic stability and improved cellular uptake etc<sup>46,66</sup>. Due to their very simple, yet elegant biosynthesis logic (explained below), the investigation of RiPPs has overturned the long-standing view in NP research of the requirement of large genomes to produce architecturally complex natural products<sup>67</sup>. Based on their biosynthetic machinery and structural features, RiPPs are grouped into different families (> 20 RiPP families); a few of these peptides are mentioned in Figure 5<sup>68-72</sup>.

Irrespective of the family, they all share several common biosynthetic features<sup>46,68,73</sup>. As mentioned earlier, all RiPPs start as precursor peptides synthesized on a ribosome. These precursor peptides, ranging from 20 - 110 amino acids in length, consist of two obligatory parts: leader and the core peptide. The leader peptide facilitates recognition by the tailoring enzyme(s) installing a particular modification in the core peptide. In some cases, e.g. bottromycins (discussed later), a follower peptide instead of a leader peptide is present; the follower peptide serves the same role as the leader peptide in terms of recognition. In most cases, modification(s) (known as early-stage or primary modification) is followed by the proteolytic removal of the leader or the follower peptide, and the mature NP is exported. However, tailoring enzymes do exist in some RiPP pathways that further modify the core peptide after the removal of the leader/follower – referred to as ‘late-stage or secondary’ modification. In leader-peptide dependent tailoring enzymes the leader peptide binding region and the active site is present on different parts of the enzyme, allowing RiPP enzymes to be specific for a substrate yet promiscuous in terms of the core peptide. As a result of this many RiPP pathways have highly variable core peptides, but retain nearly identical leader peptides. This makes RiPP a very attractive family of NPs for bioengineering to generate large chemical libraries for bioactivity screening e.g. by simple mutagenesis of the core peptide residues<sup>74,75</sup>.

The remainder of this section of the thesis will be dedicated to one particular family of RiPP, bottromycins. For additional information regarding other RiPP families and the recent advances in the field of RiPP biosynthesis, I direct the reader towards recent reviews<sup>45-46, 68, 73</sup>.

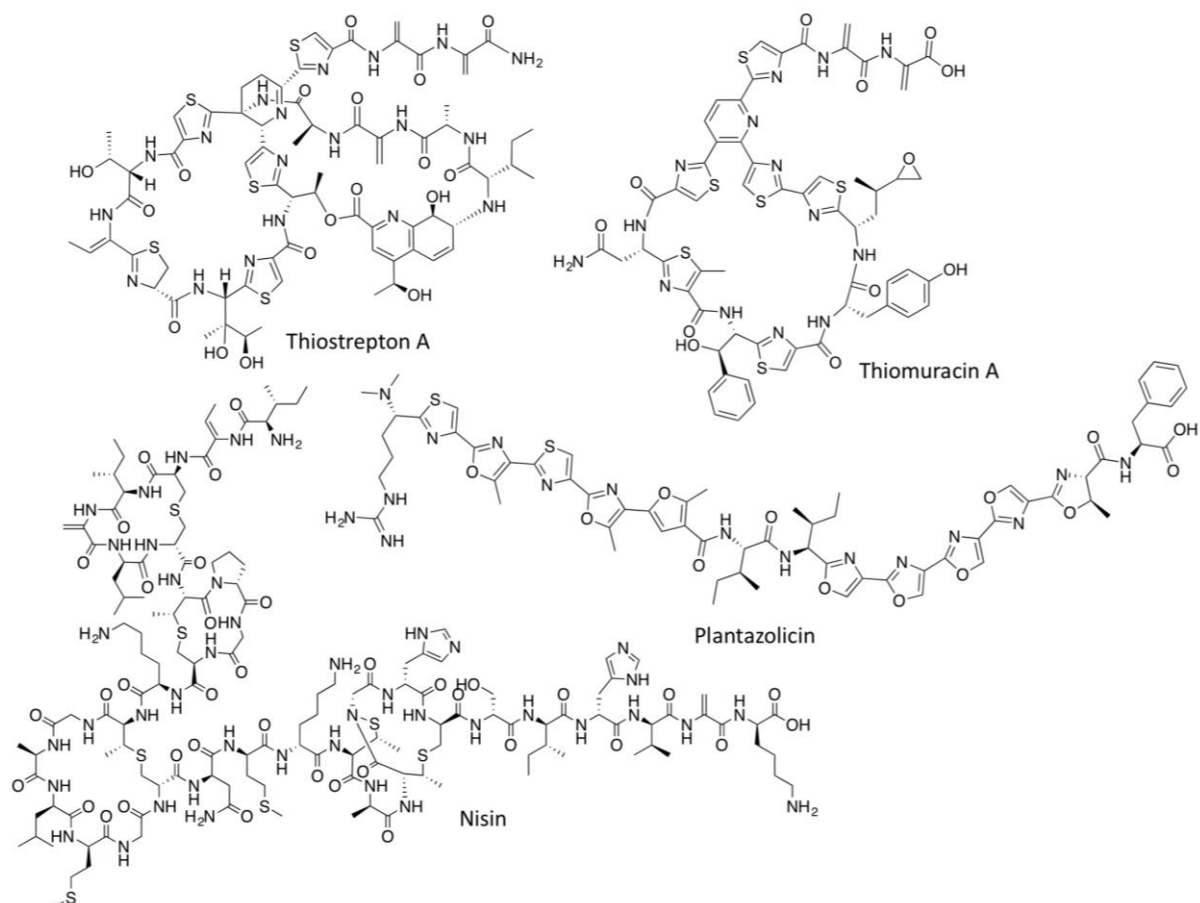
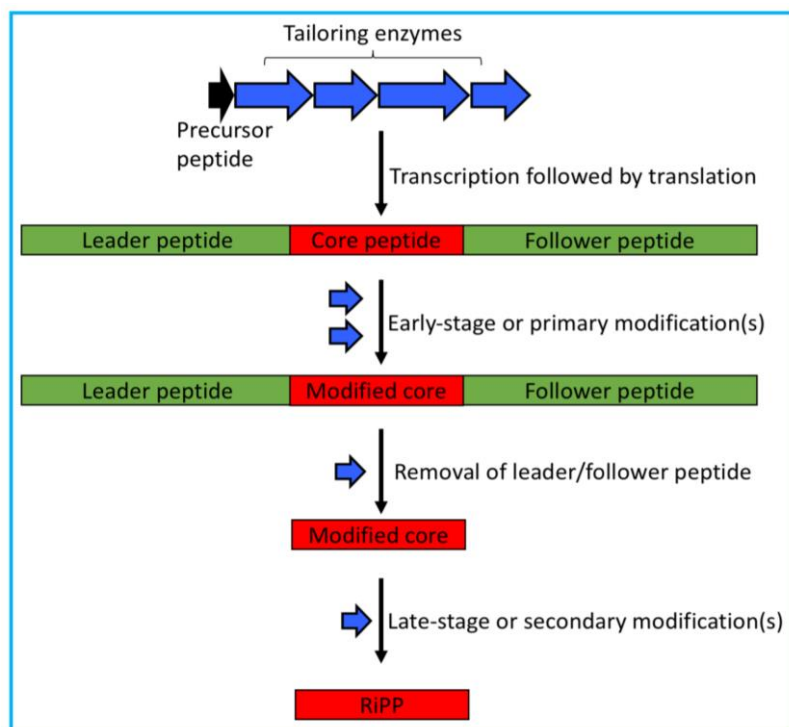


Figure 5. RiPP NP biosynthesis and structural diversity. (Top) The schematic representation of the general strategy for RiPP biosynthesis. The precursor peptide is composed of a leader or follower peptide and a core peptide. Early stage or primary modifications are installed by tailoring enzymes that typically bind precursor peptide through recognition sequences present in the leader or follower peptide, which is followed by the modification of the core peptide. Late-stage or secondary modifications are installed by tailoring enzymes that act on the core peptide independently of the leader or follower peptide. (Bottom) A few NPs produced by RiPP pathways are displayed.

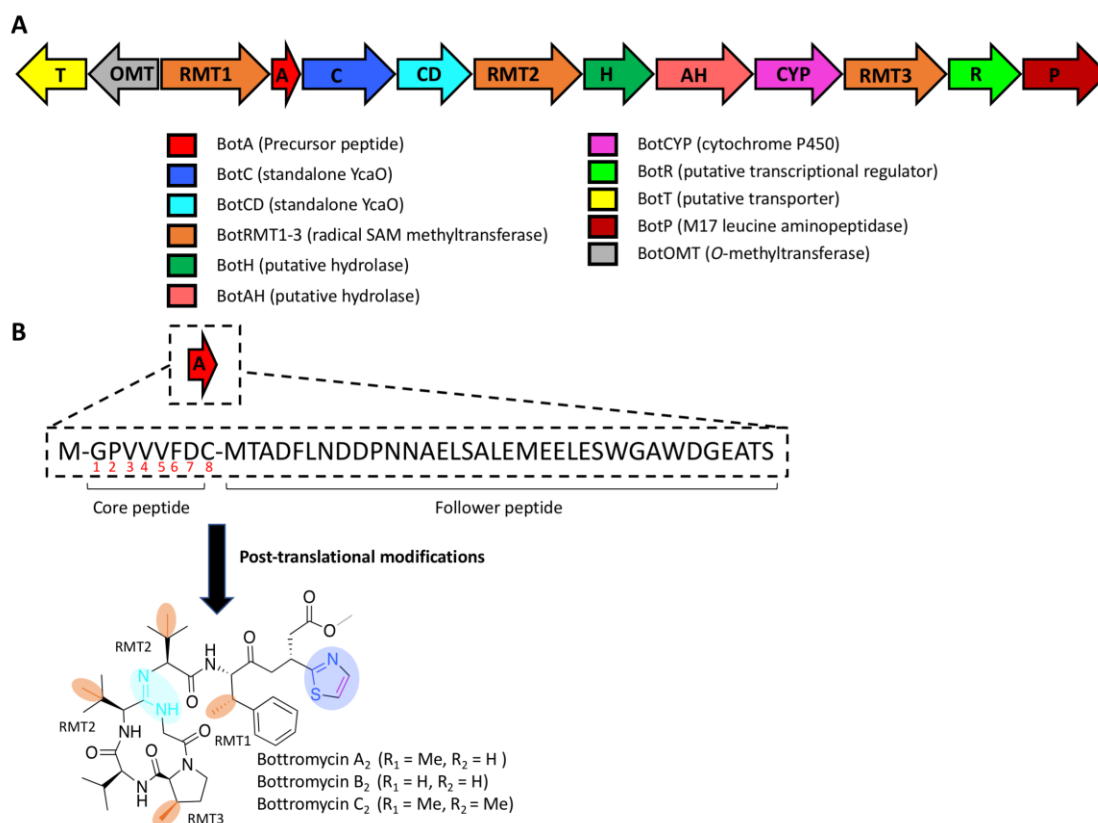
### 1.3.3.1 Bottromycins

Bottromycin was first isolated from fermentation broth of *Streptomyces bottropensis* by Waisvisz et al. in 1957, and subsequently found to be an antibacterial peptide with activity against Gram-positive bacteria and mycoplasma<sup>76-79</sup>. This was followed by the isolation of bottromycin A<sub>2</sub>, B<sub>2</sub>, and C<sub>2</sub>, several years later by Umezawa's group from *Streptomyces* No. 3668-L2, with A<sub>2</sub> being the main and the most active component<sup>80</sup>. Mode of action studies revealed these NPs inhibit bacterial protein synthesis by binding a novel target: amino acyl-tRNA binding site (A site) on the 50S ribosome<sup>81-85</sup>. As this site is not addressed by any antibiotic in use, unsurprisingly, no cross resistance was observed<sup>86</sup>. Later on, it was also shown that their antibacterial activity extends to methicillin resistance *Staphylococcus aureus* (MRSA) and vancomycin-resistant *Enterococci* (VRE), with minimum inhibitors concentrations (MIC) less than 2 µg ml<sup>-1</sup>, making them a very promising lead compound with a novel target for the development of antibiotics<sup>86</sup>. However, the progress was hampered due to the problems associated with structure elucidation, chemical synthesis and derivatization of bottromycins. The structure elucidation process involved numerous revisions, and ultimately led to the assignment shown in Figure 6<sup>80,86-92</sup>. The absolute configuration was confirmed by total synthesis in 2009, followed by the determination of the three-dimensional solution structure of bottromycin A<sub>2</sub> in 2011<sup>86,93</sup>. The structure of bottromycin consist of eight amino acids that are extensively modified, giving rise to structural features including a macrocyclic amidine, which is unique and so far, has only been reported for bottromycins; rare β- methylated amino acids residues, and a terminal thiazole (Figure 6).

Despite the established total synthesis, the structure activity relationship (SAR) studies and chemical synthesis of potent derivatives of bottromycin have proven to be an uphill task due to difficulties associated with chemical synthesis and derivatization of naturally occurring bottromycins<sup>86,94-96</sup>. Synthetic approaches can be complemented with rational biosynthetic engineering for the generation of novel derivatives; however, one of the bottlenecks for the implementation of this combinatorial approach is the lack of understanding of their biosynthesis<sup>97</sup>. The first major breakthrough was the identification of the BGCs responsible for the synthesis of bottromycins, A<sub>2</sub>-C<sub>2</sub>, and a new analog of A<sub>2</sub>, bottromycin D, in 2012<sup>98-101</sup>. The BGC's identification, coupled with gene mutation/deletion experiments and heterologous expression, demonstrated that bottromycins are synthesized as a RiPP, with the sequence of both, precursor peptide and tailoring enzymes, being conserved in several *Streptomyces* species.

Bottromycins are synthesized from the same core peptide and differ only in the methylation pattern, except for Bottromycin D, which harbours a substitution at position 3 (valine to alanine) in the core peptide.

The gene cluster from *Streptomyces* sp. BC16019 is shown in Figure 4a<sup>98</sup>. It contains 13 open reading frames (ORF); the currently proposed function of each gene and the order of biosynthesis based on the untargeted metabolomics approach using mass spectral networking, gene inactivation and *in vitro* functional characterization is shown in Figure 7<sup>98,101,102</sup>. The first step involves the removal of N-terminal methionine by a peptidase, BotP, which shares sequence homology with a number of M17 leucine aminopeptidase (LAP). As expected, the crystal structure of BotP revealed it to be highly similar to other M17 LAPs, and the subsequent *in vitro* characterization using recombinantly produced precursor peptide, BotA, confirmed its role in bottromycin biosynthesis i.e. the removal of the N-terminal methionine residue<sup>103</sup>. Gene inactivation studies have shown that the three radical-S-adenosylmethionine (SAM), BotRMT1-3, are responsible for the methylation on  $\beta$ -carbons of proline, valines and phenylalanine. These deletion experiments resulted in significantly reduced bottromycin production which can be explained by the inability of later stage tailoring enzymes to process demethylated substrates<sup>98,101,102</sup>. This implicates RMT1-3 as early-stage modification enzymes, which is in an agreement with metabolomic studies. Moreover, differing quantities of demethylated bottromycins produced also indicated that the order of  $\beta$ -methylation is likely to be: RMT1 (phenylalanine), RMT2 (valines) and RMT3 (proline)<sup>101</sup>. However, in metabolomic studies the observation of methylated valines when analysing RMT1 knockout strain, indicates that the order based on production yields may not be correct<sup>102</sup>. Besides this the observation of shunt products in BotC knockout corresponding to peptides with methylated valines also imply that heterocyclization (explained below) precedes methylation of proline and phenylalanine residues in bottromycin biosynthesis.



**Figure 6. Bottromycin biosynthesis and structure.** (A) Bottromycin BGC from *Streptomyces bottropensis* (genes not drawn to scale). (B) The sequence of the precursor peptide, BotA, is shown. The core peptide is identical in all known producers of bottromycins, except for bottromycin D (not shown) which contains a point mutation at position 3 (valine to alanine). The follower peptide sequence is also almost identical in all known producers (not shown). The core peptide undergoes extensive post-translational modifications installed by enzymes present in the BGC to yield mature bottromycins. Key structural features are color coded: macrocyclic amidine (cyan),  $\beta$ -methylation (orange), and a terminal thiazole (blue).

The bottromycin BGC encodes two YcaO proteins, BotC and BotCD, hypothesized to be involved in macrocyclization and heterocyclization<sup>98-101</sup>. This was based on the functional characterization of YcaO proteins present in other RiPP families; these proteins have been shown to catalyse the formation of oxazolines or thiazolines via the activation of backbone amide carbonyl by pyrophosphorylation or phosphorylation. All previously characterized YcaOs were shown to require the presence of an additional protein, belonging to E1 ubiquitin activating like superfamily, either as an N-terminal fusion to the YcaO or as a discrete protein<sup>104</sup>. This additional protein aids catalysis, through the presence of a motif on it, known as the RiPP precursor peptide recognition element (RRE), that supports the binding of the precursor peptide to the YcaO active site. Pairwise sequence alignments of YcaO proteins revealed the presence of key ATP/Mg<sup>+2</sup> binding residues in BotC and CD. However, RRE elements were not found in BotC and CD, making them unusual standalone YcaOs. Subsequent *in vitro* work confirmed that BotP processed precursor peptide is preferred substrate for BotC, which catalyses the formation of the heterocycle (thiazoline), followed by macrocyclization by BotCD<sup>105,106</sup>. Upon extended incubation BotCD was also shown to be able to reopen the



macrocycle<sup>106</sup>. Both enzymatic transformations require the presence of the follower peptide, but unlike other YcaOs, BotC and CD do not require an additional enzyme for activity- this is in contrast to metabolomic studies, which hinted towards the concerted role of BotCD and a putative amidohydrolase, BotAH, in macrocyclization<sup>102</sup>. Based on these findings, it is likely that one of the two remaining putative hydrolases present in the BGC, namely BotAH or BotH, is responsible for the removal of the follower peptide after cyclization. This leaves the other hydrolase as the mystery enzyme of the BGC.

Generally, thiazole-containing RiPPs encode a flavin-dependent dehydrogenase that is responsible for the oxidative decarboxylation of thiazolines to thiazoles<sup>107</sup>. However, such an enzyme is not present in the bottromycin BGC, and instead a cytochrome P450 enzyme, BotCYP, was assumed to perform this role. Although unusual, cytochrome P450 enzymes have been reported to catalyse thiazole formation in other non-RiPP NP biosynthesis e.g. the plant alkaloid camalexin<sup>108,109</sup>. Subsequently, metabolic studies confirmed that BotCYP is indeed responsible for the conversion of thiazoline to thiazole in bottromycin biosynthesis, and this conversion follows the removal of the follower peptide<sup>102</sup>. This functional assignment is based on the observation of carboxylated *O*-desmethyl bottromycin A<sub>2</sub> and B<sub>2</sub> as two major compounds in BotCYP knockout strain. Interestingly, two peaks were observed for each *O*-desmethyl-compound by liquid chromatography- mass spectrometry (LC-MS) with identical fragmentation pattern on tandem MS. As bottromycins contain a non-proteogenic D-Asp at position 7, these peaks were assumed to reflect the epimers at the aspartate residue. Deuterium labelling experiment was carried out to assess whether Asp  $\alpha$ -proton is exchangeable or not<sup>102</sup>. The exchangeable protons on decarboxylated *O*-desmethyl bottromycins were replaced with deuterium, and the thiazoline was hydrolyzed back to a Cys residue using dilute aq. DCl. The reopening of the heterocycle was postulated to trap a deuterium in the Asp  $\alpha$ -position, as the exchange of proton at this position is not possible in the absence of thiazoline. Subsequent exchange of protons with H<sub>2</sub>O followed by LC-MS analysis revealed specific incorporation of a deuterium at Asp7, confirming that Asp  $\alpha$ -proton is exchangeable. The relative ratio of the two epimers was also observed to change over time. Taken together these findings led the authors to propose that epimerization in bottromycin pathways is non-enzymatic and spontaneous<sup>102</sup>. Since imine-enamine tautomerization that is responsible for the exchange of Asp  $\alpha$ -proton is disfavoured once the thiazole has been formed, the non-enzymatic conversion of L- to D-Asp probably takes place before the formation of the thiazole<sup>102,110</sup>. In addition, the presence of carboxylated *O*-desmethyl bottromycins also revealed that *O*-methylation follows

BotCYP mediated thiazole formation. The bottromycin BGC encodes an *O*-methyltransferase, which has been shown *in vitro* to carry out aspartate *O*-methylation to yield bottromycin A<sub>2</sub> in the presence of S-adenosylmethionine (SAM)<sup>102</sup>.

The remaining two genes of the BGC encode a putative transcriptional regulator (BotR) and a putative multidrug transporter (BotT) that are not directly involved in the biosynthesis of bottromycins. Both of these proteins await functional characterization; however, BotT is likely to be involved in export of, and self-resistance to, bottromycins as its overexpression in a heterologous host resulted in 20-fold increase in production of bottromycin<sup>98</sup>. BotR shares weak sequence similarity to the ArsR/SmtB family of transcriptional repressors, and unexpectedly, the deletion of BotR did not have any effect on the yield of bottromycins<sup>101</sup>. Therefore, further work is required to determine the precise role of BotT and BotR in bottromycin biosynthesis.

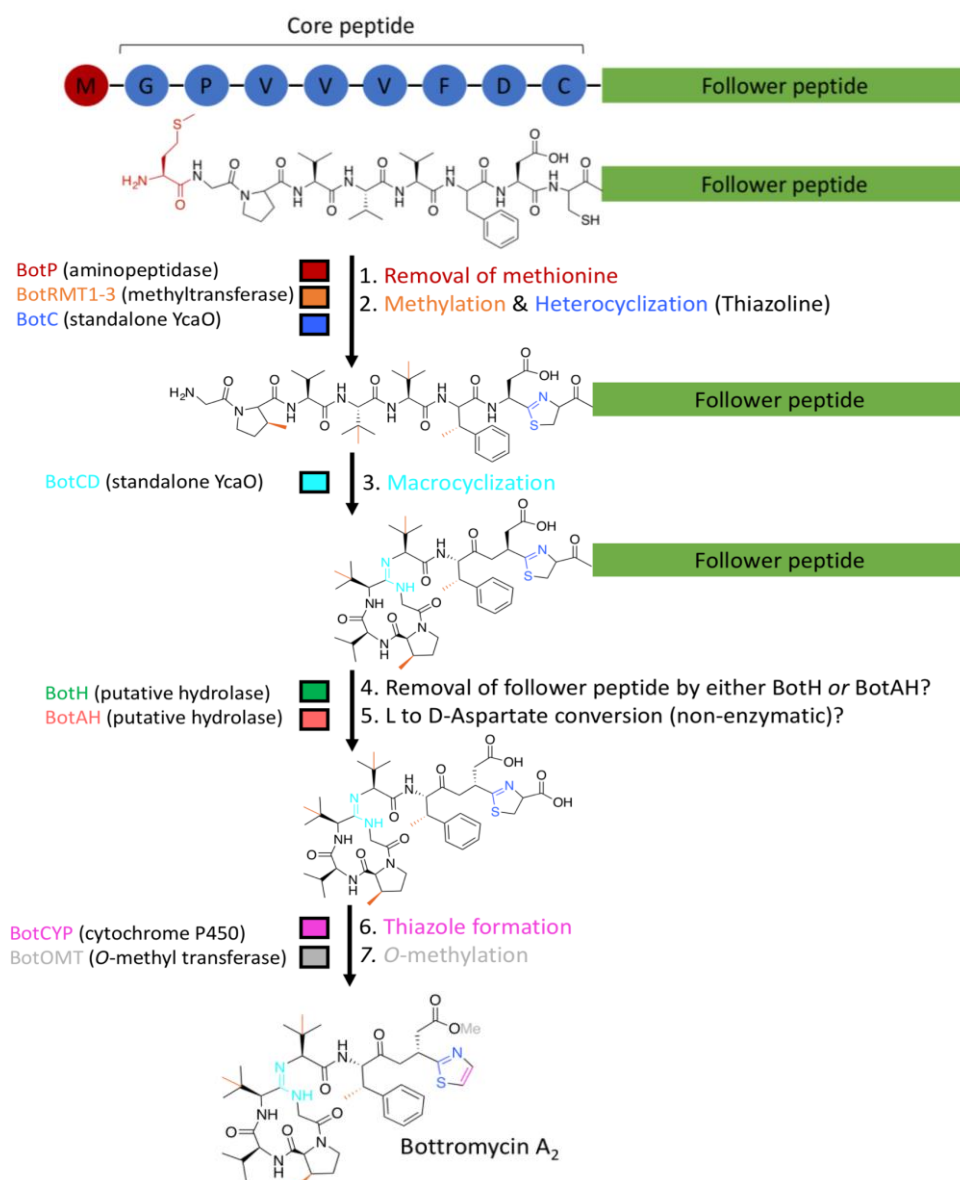


Figure 7. Bottromycin biosynthesis as proposed by *in vivo* and *in vitro* studies.

To sum up, the identification of bottromycins BGCs laid down the foundation for understanding the biosynthesis of bottromycin. Subsequent *in vivo* and *in vitro* studies have helped to elucidate or shed light on the function of most of the proteins involved in biosynthesis. However, further work is needed, especially related to the *in vitro* functional characterization of putative hydrolases present in the BGC. RiPP biosynthetic logic dictates that one of the hydrolases is responsible for removal of the follower peptide - a role previously assigned to BotH as BotAH was implicated in macrocyclization<sup>102</sup>. However, *in vitro* studies revealed that heterocyclization and macrocyclization are carried out by BotC and BotCD independently<sup>105,106</sup>. This predicament of having one additional hydrolase than to what is seemingly required for biosynthesis needs to be resolved to fully understand bottromycins biosynthesis. It is unlikely that the bottromycin BGCs contain highly conserved genes, *botH* and *botAH*, without any role in the pathway; although it could be argued that one of these hydrolases is not essential for biosynthesis of mature bottromycins, but plays a yet unknown post-maturation role. An additional area which needs more attention is the structural characterization of biosynthetic enzymes in the presence and absence of substrate or product to help shed light on the substrate specificity of each enzyme. So far only one enzyme of bottromycin BGC, namely BotP, has been structurally characterized<sup>103</sup>. The hallmark of RiPP tailoring enzymes is the spatial separation of substrate recognition and catalysis that allows the enzymes to process a wide variety of substrates. The structural data can help to better understand the substrate malleability of each enzyme in bottromycin biosynthesis, which will undoubtedly help the rational derivatization of bottromycins with enhanced antibacterial activity.

## **2 Mechanisms of Antibiotic Resistance**

### **2.1 Introduction**

The emergence of antibacterial resistance is a major public health issue of the 21<sup>st</sup> century with some infections now effectively untreatable<sup>111</sup>. According to conservative estimates by the Centers for Disease Control and Prevention infections caused by drug resistant bacteria are responsible for more than 35,000 deaths per year in the United States alone, with some reports estimating it to reach 10 million deaths per year globally by 2050<sup>112,113</sup>. The situation is worsened by a paucity of robust antibiotic pipeline, with the withdrawal of the large pharmaceutical companies from this field being one of the major reasons. The dangers of a post-antibiotic era have been acknowledged by policymakers, and a number of new initiatives have been launched e.g. Global Antibiotic Research and Development Partnership (GARDP) and Combating Antibiotic-Resistant Bacteria Biopharmaceutical Accelerator (CARB-X), to turn the tide against antibiotic resistance; however, it is too early to comment on the success of these initiatives to counter the faltering antibiotic discovery pipeline<sup>114,115</sup>. Given that bacteria are rapidly developing resistance to antibiotics, an in-depth study of resistance mechanism(s) employed by bacteria can help to extend the lifetime of an antibiotics by guiding the development of modifications that can help evade resistance mechanism. This strategy has been successful in the development of ‘generations’ of antibiotics over the past decades<sup>116</sup>. The resurgence of interest in antibiotic discovery and developments, coupled with a better understanding of bacterial resistance mechanisms, will be vital in ensuring we stay ahead of drug resistance bacteria.

### **2.2 Overview of Antibiotic Resistance Mechanisms**

The concept of antibacterial resistance is not new, and given the adaptability of bacteria it is an expected result of the interaction of organisms with their environment<sup>116</sup>. As most of the antibiotics are NPs, and thus, bacteria sharing the same ecological niche with other organisms producing antimicrobials have developed mechanism(s) to counter these molecules; this is referred to as natural resistance. Of greater concern are cases of acquired resistance via horizontal gene transfer or mutation of pre-existing genes that confers resistance to bacterial population that was initially susceptible to the antibiotic<sup>117,118</sup>. Irrespective of the nature of resistance, antimicrobial resistance mechanisms in bacteria can be broadly divided into four main categories: (1) limiting entry of the antibiotic; (2) mutation of the target protein; (3) inactivation of the antibiotic, and (4) expulsion of the antibiotic<sup>117-120</sup>. Presence of one or more

of these mechanisms in bacteria can lead to multidrug resistance (MDR)<sup>120</sup>. MDR systems respond to and inactivate antibacterial compounds with diverse structures and targets; however, a minimal structural motif is usually required for recognition by such systems<sup>121</sup>. Generally, antibiotic resistance affords a gain of function, which may impart a fitness cost<sup>122,123</sup>. Moreover, considering that the antibiotic resistance is most often only transiently advantageous, bacteria tend to modulate resistance mechanism at the transcriptional or translational level, thereby allowing it to adapt to its external environment<sup>124</sup>. One elegant and efficient system that incorporates autoregulation and MDR is thiostrepton-induced protein A (TipA).

### 2.2.1 Thiostrepton-induced protein A

Thiostrepton induced protein A (TipA), first identified in *Streptomyces lividans*, belongs to the family of mercuric ion resistance (MerR)- like transcriptional regulators which include MDR transcriptional regulators B1tR, BmrR and Mta<sup>125,126</sup>. The characteristic feature of this family is the similar N-terminal helix-turn-helix (HTH) motif responsible for DNA binding, which is connected by a long coiled-coil linker to a diverse C-terminal effector binding domain that recognizes a wide variety of ligands ranging from metal ions to large antibiotics<sup>126</sup>. However, unlike other MerR regulators, the *tipA* gene, which is expressed in response to thiostrepton-like antibiotics, encodes two alternate in-frame translation products: the long minor form TipA<sub>L</sub> and the predominant short form TipA<sub>S</sub>, which constitutes the C-terminal part TipA<sub>L</sub> (Figure 8)<sup>126,127</sup>. While TipA<sub>L</sub> follows the basic architecture of MerR-like regulators, TipA<sub>S</sub> on the other hand consist only of the effector binding domain (Figure 8). The effector domain present in both TipA<sub>L</sub> and TipA<sub>S</sub> interacts with a variety of a thiopeptide-like antibiotics such as thiostrepton, nosiheptide and promothiocin A<sup>128,129</sup>. Recently, TipA<sub>S</sub> apo and complex solution structures and dynamic studies have allowed the proposal of a mechanism for the TipA MDR system (Figure 8)<sup>130,131</sup>. TipA<sub>S</sub> consists of an unfolded N-terminal region and a C-terminal, globin-like  $\alpha$ -helical structure with a deep hydrophobic antibiotic binding cleft. The ligand-binding residue Cys214, which forms a covalent bond with the dehydroalanine residue present in the tail region of thiopeptides, is located at the bottom of this cleft (Figure 8a)<sup>130</sup>. In the case of thiostrepton which contains two dehydroalanine residues, Dha15 and Dha16, the data suggests that Dha15 is most likely involved in covalent bond formation with Cys214. However, the presence of Dha is not a necessary condition for the recognition by TipA, and conserved structural features of thiostrepton-like antibiotics have been implicated in recognition<sup>128,131</sup>. The first step of antibiotic interaction involves the recognition of these features by hydrophobic residues present at the entrance of antibiotic binding cleft and at the N-terminus of TipA<sub>S</sub>. The interaction of the

antibiotic with the N-terminus induces significant conformational change resulting in the transformation of the N-terminus from partially flexible to all-helical structure (Figure 8a), that now encloses the bound antibiotic, thereby forming a very stable protein-ligand complex (Figure 8a)<sup>131,132</sup>. In addition, antibiotic binding also stabilizes a helix connected to the N-terminus; extending it by two turns in the direction of the N-terminus (Figure 8a slate colored helix). These interactions also position the Dha residue present in the tail of thiostrepton-like antibiotics such that it can form a covalent bonding with Cys214. As TipA<sub>S</sub> is connected to the DNA binding domain in the transcriptional regulator, TipA<sub>L</sub>, it is likely that the transformation of TipA<sub>S</sub>'s N-terminus upon antibiotic binding constitutes a mechanical signal that leads to the observed higher affinity to the *tipA* promoter, which then binds to it as a dimer and causes its untwisting or unwinding<sup>131</sup>. This binding event has also been reported to increase the affinity of RNA polymerase (RNAP) for the promoter by more than 10-fold, possibly an outcome of more stable TipA<sub>L</sub> - antibiotic - RNAP complex, initiating transcription of *tipA*<sup>130,132</sup>. However, further studies are needed to fully understand the underlying mechanism of *tipA* activation by TipA<sub>L</sub>. In short, because of its two distinct function: (1) sequestration and thereby inactivation of thiostrepton-like antibiotics by TipA<sub>S</sub>, and (2) autoregulation of *tipA* gene by TipA<sub>L</sub> in the presence thiostrepton-like antibiotics, TipA can be considered as an autoregulated MDR system. TipA-like proteins are also reported to be present in the genomes of many bacteria, including pathogens like *Bacillus anthracis*, *Streptococcus mitis*, *Enterobacter faecalis* and *Listeria monocytogenes*<sup>130</sup>. At present none of these proteins have been functionally or structurally characterized, and it would be interesting to see what assortment of metabolites are bound and inactivated by them.

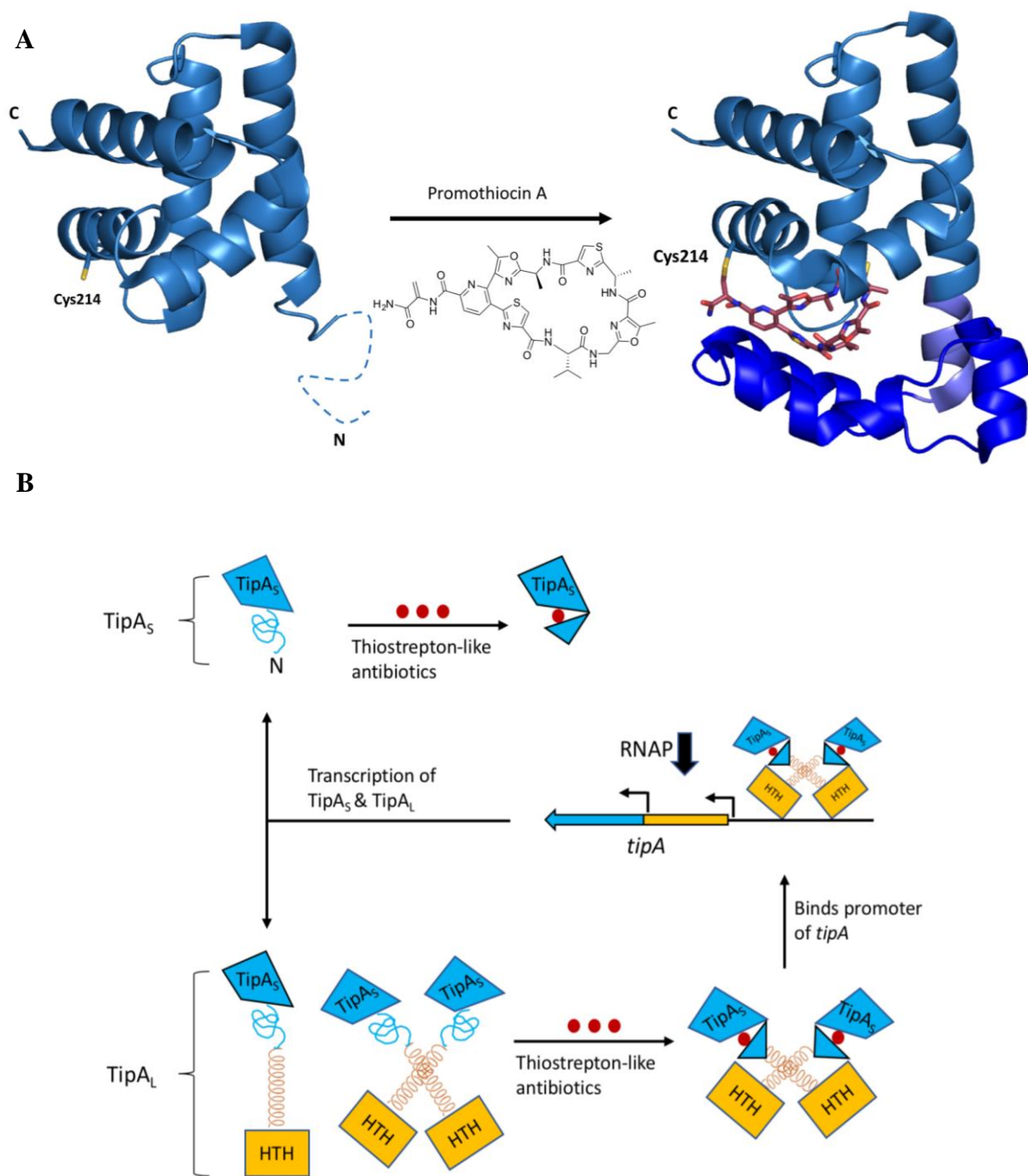


Figure 8. Basic logic of TipA multidrug resistance system. (A) Structure of TipA<sub>S</sub> and TipA<sub>S</sub> antibiotic complex structure in ribbon representation. Thiostrepton-like antibiotic, promothiocin A, and antibiotic binding residue Cys214 are shown as sticks with carbon atoms maroon, oxygen atoms red, nitrogen atoms blue, and sulphur as yellow. The flexible N-terminal residues that undergo significant conformational change upon ligand binding are shown as a dashed line. The ligand-induced conformational change leads to the formation of additional helices which are color coded: extension of an existing helix (slate) and new helices (dark blue) (B) The presence of thiostrepton-like antibiotics induce the expression of tipA gene which encodes two alternate in-frame products: TipA<sub>S</sub> and TipA<sub>L</sub>. The covalent binding of the antibiotic by TipA<sub>S</sub> leads to transformation of the flexible N-terminal to a helical structure enclosing the antibiotic. At the same time the antibiotic induced conformational change promotes TipA<sub>L</sub> dimerization and induction of tipA gene.

### 3 Outline of the thesis

The main focus on this work was to structurally and functionally characterize three proteins: BotH and BotAH, putative hydrolases present in bottromycin BCG; and AlbA, an albicidin antibiotic resistance protein found in *Klebsiella oxytoca*. Chapter 2 and 3 are dedicated to understanding the role BotAH and BotH in bottromycin biosynthesis; whereas, chapter 4 deals with AlbA.

Bottromycins are a RiPP whose biosynthesis is not yet fully understood, especially the role of BotAH and BotH. Recently, work done by Franz, L. et al. and Schwalen, C.J. et al. on the *in vitro* reconstitution of macrocyclization in bottromycin biosynthesis revealed that BotCD is able to install this modification without the need of an additional enzyme, and that this modification is reversible. This was in contrast to *in vivo* data which hinted towards a concerted role of BotAH and BotCD in macrocyclization. Therefore, in order to resolve the disparity between the *in vivo* and *in vitro* data we set out to understand the role of BotAH and BotH. The data presented confirms the role of BotAH in bottromycin biosynthesis i.e. the removal of the follower peptide after macrocyclization. Moreover, once the follower peptide has been removed by BotAH, macrocycle cannot be reopened by BotCD. This observation helps to remove the disparity between the *in vitro* and *in vivo* data and places BotAH as the gatekeep between the early stage and late stage modifications installed on the precursor peptide. The crystal structure of BotAH is also reported, which helped shed light on its substrate specificity and mechanism of action.

In chapter 3 the role of BotH in bottromycin biosynthesis is discussed. The data presented demonstrates that this enzyme is responsible for the conversion of L-aspartate to D-aspartate with relaxed substrate specificity. The high-resolution crystal structures of BotH and BotH in complex with its substrate sheds light on the mechanism of epimerization. In addition, BotH can also bind the natural products bottromycins with high affinity, and the mode of binding is explained by the complex structures of three bottromycins bound to BotH. Subsequent, *in vivo* and *in vitro* data hints towards an additional role of BotH in bottromycin biosynthesis: negative-feedback regulation of bottromycin biosynthesis. An in-silico analysis also reveals the presence of BotH-like proteins in putative BGCs encoding diverse NPs including RiPP, NRPS and PKS, which may also be able to epimerize non-Asp stereocenters via the unprecedented mechanism we proposed for BotH.



Albicidin is a NP produced by phytopathogen *Xanthomonas albilineans*, which was reported as potent antibiotic against Gram-positive and in particular Gram-negative bacteria. Owing to its unique structural properties and potent antibacterial activity, it is considered a lead structure. Therefore, it is important to understand the resistance mechanisms that already exist in nature against this interesting antibiotic. Previously three resistance proteins have been identified against albicidin: AlbA, AlbB and AlbD. Chapter 4 is dedicated to understanding the mechanism(s) employed by AlbA to neutralize albicidin. We report high-resolution crystal structures of AlbA and the AlbA-albicidin complex, which explain the mode of binding of albicidin. Structural analysis also revealed AlbA be a homolog of TipAs; it appears that an internal gene duplication gave rise to two pseudo-symmetric TipAs dimers which constitute AlbA. Since the TipA system consists of two protein isoforms: short version (TipAs) and long version (TipAL), we had a closer look at the AlbA sequence. The original annotation of AlbA in *Klebsiella oxytoca* was incomplete, and an alternate inframe start codon located upstream was found to give rise to two in-frame translation products: short version (AlbAs) and the long version (AlbAL). Unexpectedly, unlike TipAs, AlbAs was found to modify the antibiotic. This AlbAs-mediated modification of albicidin leads to a loss of biological activity, extending our understanding of AlbAs, which was previously thought to neutralize albicidin by sequestering the compound by forming a very stable complex. Subsequent, extensive mutational studies and crystal structures allowed us to propose a mechanism for this chemical modification. Finally, we also tested if the AlbA system, akin to TipA system, can neutralize compounds structurally similar to albicidin. To this end we tested cystobactimid - a potent antibiotic produced by myxobacteria. Interestingly, cystobactimid, despite being able to interreact with AlbA did not upregulate the transcription of the *alba* gene. This hints towards a different mode of binding for cystobactamid, compared to albicidin, which is currently being investigated.

## 4 Reference

- 1 Williams, D. H., Stone, M. J., Hauck, P. R. & Rahman, S. K. Why are secondary metabolites (natural products) biosynthesized? *J Nat Prod* **52**, 1189-1208, doi:10.1021/np50066a001 (1989).
- 2 Maplestone, R. A., Stone, M. J. & Williams, D. H. The evolutionary role of secondary metabolites--a review. *Gene* **115**, 151-157, doi:10.1016/0378-1119(92)90553-2 (1992).
- 3 Firn, R. D. & Jones, C. G. The evolution of secondary metabolism - a unifying model. *Mol Microbiol* **37**, 989-994, doi:10.1046/j.1365-2958.2000.02098.x (2000).
- 4 Dias, D. A., Urban, S. & Roessner, U. A historical overview of natural products in drug discovery. *Metabolites* **2**, 303-336, doi:10.3390/metabo2020303 (2012).
- 5 Koehn, F. E. & Carter, G. T. The evolving role of natural products in drug discovery. *Nat Rev Drug Discov* **4**, 206-220, doi:10.1038/nrd1657 (2005).
- 6 Garg, N. *et al.* Natural products as mediators of disease. *Nat Prod Rep* **34**, 194-219, doi:10.1039/c6np00063k (2017).
- 7 Demain, A. L. & Sanchez, S. Microbial drug discovery: 80 years of progress. *J Antibiot (Tokyo)* **62**, 5-16, doi:10.1038/ja.2008.16 (2009).
- 8 Sneader, W. *Drug Prototypes and Their Exploitation*. (John Wiley, 1996).
- 9 Siddiqui, A. a., Iram, F., Siddiqui, S., Sahu, K. Role of natural products in drug discovery process. *Int. J. Drug Dev. Res* **6**, 172-204 (2014).
- 10 Pham, J. V. *et al.* A Review of the Microbial Production of Bioactive Natural Products and Biologics. *Front Microbiol* **10**, 1404, doi:10.3389/fmicb.2019.01404 (2019).
- 11 Fleming, A. The discover of penicillin. *Br. Med. Bull.* **2**, 4-5 (1944).
- 12 Ribeiro da Cunha, B., Fonseca, L. P. & Calado, C. R. C. Antibiotic Discovery: Where Have We Come from, Where Do We Go? *Antibiotics (Basel)* **8**, doi:10.3390/antibiotics8020045 (2019).
- 13 Hutchings, M. I., Truman, A. W. & Wilkinson, B. Antibiotics: past, present and future. *Curr Opin Microbiol* **51**, 72-80, doi:10.1016/j.mib.2019.10.008 (2019).
- 14 Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. *J Nat Prod* **79**, 629-661, doi:10.1021/acs.jnatprod.5b01055 (2016).
- 15 Lewis, K. Platforms for antibiotic discovery. *Nat Rev Drug Discov* **12**, 371-387, doi:10.1038/nrd3975 (2013).
- 16 Zengler, K. *et al.* Cultivating the uncultured. *Proc Natl Acad Sci U S A* **99**, 15681-15686, doi:10.1073/pnas.252630999 (2002).
- 17 Monciardini, P., Iorio, M., Maffioli, S., Sosio, M. & Donadio, S. Discovering new bioactive molecules from microbial sources. *Microb Biotechnol* **7**, 209-220, doi:10.1111/1751-7915.12123 (2014).
- 18 Wohlleben, W., Mast, Y., Stegmann, E. & Ziemert, N. Antibiotic drug discovery. *Microb Biotechnol* **9**, 541-548, doi:10.1111/1751-7915.12388 (2016).
- 19 Payne, D. J., Gwynn, M. N., Holmes, D. J. & Pompliano, D. L. Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat Rev Drug Discov* **6**, 29-40, doi:10.1038/nrd2201 (2007).
- 20 Li, J. W. & Vederas, J. C. Drug discovery and natural products: end of an era or an endless frontier? *Science* **325**, 161-165, doi:10.1126/science.1168243 (2009).
- 21 Weissman, K. J. & Leadlay, P. F. Combinatorial biosynthesis of reduced polyketides. *Nat Rev Microbiol* **3**, 925-936, doi:10.1038/nrmicro1287 (2005).
- 22 Herrmann, J. *et al.* Strategies for the Discovery and Development of New Antibiotics from Natural Products: Three Case Studies. *Curr Top Microbiol Immunol* **398**, 339-363, doi:10.1007/82\_2016\_498 (2016).

- 23 Chen, Y., Garcia de Lomana, M., Friedrich, N. O. & Kirchmair, J. Characterization of  
the Chemical Space of Known and Readily Obtainable Natural Products. *J Chem Inf  
Model* **58**, 1518-1532, doi:10.1021/acs.jcim.8b00302 (2018).
- 24 Rodrigues, T., Reker, D., Schneider, P. & Schneider, G. Counting on natural products  
for drug design. *Nat Chem* **8**, 531-541, doi:10.1038/nchem.2479 (2016).
- 25 Evans, B. E. *et al.* Methods for drug discovery: development of potent, selective, orally  
effective cholecystokinin antagonists. *J Med Chem* **31**, 2235-2246,  
doi:10.1021/jm00120a002 (1988).
- 26 Wenzel, S. C. & Muller, R. The impact of genomics on the exploitation of the  
myxobacterial secondary metabolome. *Nat Prod Rep* **26**, 1385-1407,  
doi:10.1039/b817073h (2009).
- 27 Wright, G. D. Opportunities for natural products in 21(st) century antibiotic discovery.  
*Nat Prod Rep* **34**, 694-701, doi:10.1039/c7np00019g (2017).
- 28 Lewis, K. Antibiotics: Recover the lost art of drug discovery. *Nature* **485**, 439-440,  
doi:10.1038/485439a (2012).
- 29 Banakar, S. P., Karthik, L. & Li, Z. in *Symbiotic Microbiomes of Coral Reefs Sponges  
and Corals* (ed Zhiyong Li) 505-526 (Springer Netherlands, 2019).
- 30 Rutledge, P. J. & Challis, G. L. Discovery of microbial natural products by activation  
of silent biosynthetic gene clusters. *Nat Rev Microbiol* **13**, 509-523,  
doi:10.1038/nrmicro3496 (2015).
- 31 Mao, D., Okada, B. K., Wu, Y., Xu, F. & Seyedsayamdost, M. R. Recent advances in  
activating silent biosynthetic gene clusters in bacteria. *Curr Opin Microbiol* **45**, 156-  
163, doi:10.1016/j.mib.2018.05.001 (2018).
- 32 Rossiter, S. E., Fletcher, M. H. & Wuest, W. M. Natural Products as Platforms To  
Overcome Antibiotic Resistance. *Chem Rev* **117**, 12415-12474,  
doi:10.1021/acs.chemrev.7b00283 (2017).
- 33 Medema, M. H. *et al.* Minimum Information about a Biosynthetic Gene cluster. *Nat  
Chem Biol* **11**, 625-631, doi:10.1038/nchembio.1890 (2015).
- 34 Skinnider, M. A., Merwin, N. J., Johnston, C. W. & Magarvey, N. A. PRISM 3:  
expanded prediction of natural product chemical structures from microbial genomes.  
*Nucleic Acids Res* **45**, W49-W54, doi:10.1093/nar/gkx320 (2017).
- 35 Blin, K. *et al.* antiSMASH 5.0: updates to the secondary metabolite genome mining  
pipeline. *Nucleic Acids Res* **47**, W81-W87, doi:10.1093/nar/gkz310 (2019).
- 36 Giani, A. M., Gallo, G. R., Gianfranceschi, L. & Formenti, G. Long walk to genomics:  
History and current approaches to genome sequencing and assembly. *Comput Struct  
Biotechnol J* **18**, 9-19, doi:10.1016/j.csbj.2019.11.002 (2020).
- 37 de Los Santos, E. L. C. NeuRiPP: Neural network identification of RiPP precursor  
peptides. *Sci Rep* **9**, 13406, doi:10.1038/s41598-019-49764-z (2019).
- 38 Scott, T. A., Piel, J. The hidden enzymology of bacterial natural product biosynthesis.  
*Nat Rev Chem* **3**, 404-425, doi:10.1038/s41570-019-0107-1 (2019).
- 39 Walsh, C. T. & Fischbach, M. A. Natural products version 2.0: connecting genes to  
molecules. *J Am Chem Soc* **132**, 2469-2493, doi:10.1021/ja909118a (2010).
- 40 Yuzawa, S., Backman, T. W. H., Keasling, J. D. & Katz, L. Synthetic biology of  
polyketide synthases. *J Ind Microbiol Biotechnol* **45**, 621-633, doi:10.1007/s10295-  
018-2021-9 (2018).
- 41 Shen, B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase  
paradigms. *Curr Opin Chem Biol* **7**, 285-295, doi:10.1016/s1367-5931(03)00020-6  
(2003).
- 42 Sussmuth, R. D. & Mainz, A. Nonribosomal Peptide Synthesis-Principles and  
Prospects. *Angew Chem Int Ed Engl* **56**, 3770-3821, doi:10.1002/anie.201609079  
(2017).

- 43 Reimer, J. M., Haque, A. S., Tarry, M. J. & Schmeing, T. M. Piecing together nonribosomal peptide synthesis. *Curr Opin Struct Biol* **49**, 104-113, doi:10.1016/j.sbi.2018.01.011 (2018).
- 44 Bloudoff, K. & Schmeing, T. M. Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. *Biochim Biophys Acta Proteins Proteom* **1865**, 1587-1604, doi:10.1016/j.bbapap.2017.05.010 (2017).
- 45 Luo, S. & Dong, S. H. Recent Advances in the Discovery and Biosynthetic Study of Eukaryotic RiPP Natural Products. *Molecules* **24**, doi:10.3390/molecules24081541 (2019).
- 46 Arnison, P. G. *et al.* Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* **30**, 108-160, doi:10.1039/c2np20085f (2013).
- 47 Payne, J. A., Schoppet, M., Hansen, M. H. & Cryle, M. J. Diversity of nature's assembly lines - recent discoveries in non-ribosomal peptide synthesis. *Mol Biosyst* **13**, 9-22, doi:10.1039/c6mb00675b (2016).
- 48 Herbst, D. A., Townsend, C. A. & Maier, T. The architectures of iterative type I PKS and FAS. *Nat Prod Rep* **35**, 1046-1069, doi:10.1039/c8np00039e (2018).
- 49 Chen, H. & Du, L. Iterative polyketide biosynthesis by modular polyketide synthases in bacteria. *Appl Microbiol Biotechnol* **100**, 541-557, doi:10.1007/s00253-015-7093-0 (2016).
- 50 Morita, H., Wong, C. P. & Abe, I. How structural subtleties lead to molecular diversity for the type III polyketide synthases. *J Biol Chem* **294**, 15121-15136, doi:10.1074/jbc.REV119.006129 (2019).
- 51 Irschik, H. *et al.* Analysis of the sorangicin gene cluster reinforces the utility of a combined phylogenetic/retrobiosynthetic analysis for deciphering natural product assembly by trans-AT PKS. *ChemBiochem* **11**, 1840-1849, doi:10.1002/cbic.201000313 (2010).
- 52 Fu, C. *et al.* Solving the Puzzle of One-Carbon Loss in Ripostatin Biosynthesis. *Angew Chem Int Ed Engl* **56**, 2192-2197, doi:10.1002/anie.201609950 (2017).
- 53 Sharif, E. U. & O'Doherty, G. A. Biosynthesis and Total Synthesis Studies on The Jadomycin Family of Natural Products. *European J Org Chem* **2012**, doi:10.1002/ejoc.201101609 (2012).
- 54 Surup, F. *et al.* Disciformycins A and B: 12-membered macrolide glycoside antibiotics from the myxobacterium *Pyxidicoccus fallax* active against multiresistant staphylococci. *Angew Chem Int Ed Engl* **53**, 13588-13591, doi:10.1002/anie.201406973 (2014).
- 55 Moore, B. S. & Hertweck, C. Biosynthesis and attachment of novel bacterial polyketide synthase starter units. *Nat Prod Rep* **19**, 70-99, doi:10.1039/b003939j (2002).
- 56 Chan, Y. A., Podevels, A. M., Kevany, B. M. & Thomas, M. G. Biosynthesis of polyketide synthase extender units. *Nat Prod Rep* **26**, 90-114, doi:10.1039/b801658p (2009).
- 57 Olano, C., Mendez, C. & Salas, J. A. Post-PKS tailoring steps in natural product-producing actinomycetes from the perspective of combinatorial biosynthesis. *Nat Prod Rep* **27**, 571-616, doi:10.1039/b911956f (2010).
- 58 Rix, U., Fischer, C., Remsing, L. L. & Rohr, J. Modification of post-PKS tailoring steps through combinatorial biosynthesis. *Nat Prod Rep* **19**, 542-580, doi:10.1039/b103920m (2002).
- 59 Strieker, M., Tanovic, A. & Marahiel, M. A. Nonribosomal peptide synthetases: structures and dynamics. *Curr Opin Struct Biol* **20**, 234-240, doi:10.1016/j.sbi.2010.01.009 (2010).

- 60 Martínez-Núñez, M. A., López, V.E.L.y. Nonribosomal peptides synthetases and their applications in industry. *Sustain Chem Process* **4**, doi:0.1186/s40508-016-0057-6 (2016).
- 61 Hubbard, B. K. & Walsh, C. T. Vancomycin assembly: nature's way. *Angew Chem Int Ed Engl* **42**, 730-765, doi:10.1002/anie.200390202 (2003).
- 62 Baumann, S. *et al.* Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity. *Angew Chem Int Ed Engl* **53**, 14605-14609, doi:10.1002/anie.201409964 (2014).
- 63 Miao, V. *et al.* Daptomycin biosynthesis in *Streptomyces roseosporus*: cloning and analysis of the gene cluster and revision of peptide stereochemistry. *Microbiology* **151**, 1507-1523, doi:10.1099/mic.0.27757-0 (2005).
- 64 Irschik, H., Gerth, K., Kemmer, T., Steinmetz, H. & Reichenbach, H. The myxovalargins, new peptide antibiotics from *Myxococcus fulvus* (Myxobacterales). I. Cultivation, isolation, and some chemical and biological properties. *J Antibiot (Tokyo)* **36**, 6-12, doi:10.7164/antibiotics.36.6 (1983).
- 65 Schaberle, T. F., Lohr, F., Schmitz, A. & König, G. M. Antibiotics from myxobacteria. *Nat Prod Rep* **31**, 953-972, doi:10.1039/c4np00011k (2014).
- 66 Velasquez, J. E. & van der Donk, W. A. Genome mining for ribosomally synthesized natural products. *Curr Opin Chem Biol* **15**, 11-21, doi:10.1016/j.cbpa.2010.10.027 (2011).
- 67 Dunbar, K. L. & Mitchell, D. A. Revealing nature's synthetic potential through the study of ribosomal natural product biosynthesis. *ACS Chem Biol* **8**, 473-487, doi:10.1021/cb3005325 (2013).
- 68 Hetrick, K. J. & van der Donk, W. A. Ribosomally synthesized and post-translationally modified peptide natural product discovery in the genomic era. *Curr Opin Chem Biol* **38**, 36-44, doi:10.1016/j.cbpa.2017.02.005 (2017).
- 69 Kelly, W. L., Pan, L. & Li, C. Thiostrepton biosynthesis: prototype for a new family of bacteriocins. *J Am Chem Soc* **131**, 4327-4334, doi:10.1021/ja807890a (2009).
- 70 Morris, R. P. *et al.* Ribosomally synthesized thiopeptide antibiotics targeting elongation factor Tu. *J Am Chem Soc* **131**, 5946-5955, doi:10.1021/ja900488a (2009).
- 71 Scholz, R. *et al.* Plantazolicin, a novel microcin B17/streptolysin S-like natural product from *Bacillus amyloliquefaciens* FZB42. *J Bacteriol* **193**, 215-224, doi:10.1128/JB.00784-10 (2011).
- 72 Kuipers, O. P., Beerthuyzen, M. M., Siezen, R. J. & De Vos, W. M. Characterization of the nisin gene cluster nisABTCIPR of *Lactococcus lactis*. Requirement of expression of the nisA and nisI genes for development of immunity. *Eur J Biochem* **216**, 281-291, doi:10.1111/j.1432-1033.1993.tb18143.x (1993).
- 73 Ortega, M. A. & van der Donk, W. A. New Insights into the Biosynthetic Logic of Ribosomally Synthesized and Post-translationally Modified Peptide Natural Products. *Cell Chem Biol* **23**, 31-44, doi:10.1016/j.chembiol.2015.11.012 (2016).
- 74 Gu, W. & Schmidt, E. W. Three Principles of Diversity-Generating Biosynthesis. *Acc Chem Res* **50**, 2569-2576, doi:10.1021/acs.accounts.7b00330 (2017).
- 75 Sardar, D. & Schmidt, E. W. Combinatorial biosynthesis of RiPPs: docking with marine life. *Curr Opin Chem Biol* **31**, 15-21, doi:10.1016/j.cbpa.2015.11.016 (2016).
- 76 J.M. Waisvisz, M. G. v. d. H., B. te Nijnhuis. The structure of the sulfur-containing moiety of bottromycin. *J. Am. Chem. Soc.* **80**, 383-385 (1957).
- 77 Nakamura, S., Omura, S., Nishimura, T., Tanaka, N. & Umezawa, H. Derivatives of bottromycin A2 and their biological activity. *J Antibiot (Tokyo)* **20**, 162-166 (1967).
- 78 Tanaka, N., Nishimura, T., Nakamura, S. & Umezawa, H. Activity of bottromycin against *Mycoplasma gallisepticum*. *J Antibiot (Tokyo)* **21**, 75-76, doi:10.7164/antibiotics.21.75 (1968).

- 79 Miller, B. M., Stapley, E. O. & Woodruff, H. B. Antimycoplasmal activity of the bottromycin complex and its production by *Streptomyces canadensis*. *Antimicrob Agents Chemother (Bethesda)* **7**, 407-414 (1967).
- 80 Nakamura, S., Yajima, T., Lin, Y. & Umezawa, H. Isolation and characterization of bottromycins A2, B2, C2. *J Antibiot (Tokyo)* **20**, 1-5 (1967).
- 81 Otaka, T. & Kaji, A. Mode of action of bottromycin A2: effect of bottromycin A2 on polysomes. *FEBS Lett* **153**, 53-59, doi:10.1016/0014-5793(83)80118-5 (1983).
- 82 Otaka, T. & Kaji, A. Mode of action of bottromycin A2: effect on peptide bond formation. *FEBS Lett* **123**, 173-176, doi:10.1016/0014-5793(81)80280-3 (1981).
- 83 Otaka, T. & Kaji, A. Mode of action of bottromycin A2. Release of aminoacyl- or peptidyl-tRNA from ribosomes. *J Biol Chem* **251**, 2299-2306 (1976).
- 84 Kinoshita, T. & Tanaka, N. On the site of action of bottromycin A2. *J Antibiot (Tokyo)* **23**, 311-312, doi:10.7164/antibiotics.23.311 (1970).
- 85 Lin, Y. C. & Tanaka, N. Mechanism of action of bottromycin in polypeptide biosynthesis. *J Biochem* **63**, 1-7, doi:10.1093/oxfordjournals.jbchem.a128735 (1968).
- 86 Shimamura, H. *et al.* Structure determination and total synthesis of bottromycin A2: a potent antibiotic against MRSA and VRE. *Angew Chem Int Ed Engl* **48**, 914-917, doi:10.1002/anie.200804138 (2009).
- 87 Nakamura, S., Chikaike, T., Yonehara, H. & Umezawa, H. Structures of Bottromycins a and B. *J Antibiot (Tokyo)* **18**, 60-61 (1965).
- 88 Nakamura, S., Tanaka, N. & Umezawa, H. Bottromycin A1, A2 and their structures. *J Antibiot (Tokyo)* **19**, 10-12 (1966).
- 89 Takahashi, Y., Naganawa, H., Takita, T., Umezawa, H. & Nakamura, S. The revised structure of bottromycin A2. *J Antibiot (Tokyo)* **29**, 1120-1123, doi:10.7164/antibiotics.29.1120 (1976).
- 90 Schipper, D. The revised structure of bottromycin A2. *J Antibiot (Tokyo)* **36**, 1076-1077, doi:10.7164/antibiotics.36.1076 (1983).
- 91 Kaneda, M. Studies on bottromycins. I. <sup>1</sup>H and <sup>13</sup>C NMR assignments of bottromycin A2, the main component of the complex. *J Antibiot (Tokyo)* **45**, 792-796, doi:10.7164/antibiotics.45.792 (1992).
- 92 Kaneda, M. Studies on bottromycins. II. Structure elucidation of bottromycins B2 and C2. *J Antibiot (Tokyo)* **55**, 924-928, doi:10.7164/antibiotics.55.924 (2002).
- 93 Gouda, H. *et al.* Three-dimensional solution structure of bottromycin A2: a potent antibiotic active against methicillin-resistant *Staphylococcus aureus* and vancomycin-resistant Enterococci. *Chem Pharm Bull (Tokyo)* **60**, 169-171, doi:10.1248/cpb.60.169 (2012).
- 94 Ackermann, S., Lerchen, H. G., Habich, D., Ullrich, A. & Kazmaier, U. Synthetic studies towards bottromycin. *Beilstein J Org Chem* **8**, 1652-1656, doi:10.3762/bjoc.8.189 (2012).
- 95 Kobayashi, Y. *et al.* Bottromycin derivatives: efficient chemical modifications of the ester moiety and evaluation of anti-MRSA and anti-VRE activities. *Bioorg Med Chem Lett* **20**, 6116-6120, doi:10.1016/j.bmcl.2010.08.037 (2010).
- 96 Yamada, T. *et al.* Synthesis and Evaluation of Antibacterial Activity of Bottromycins. *J Org Chem* **83**, 7135-7149, doi:10.1021/acs.joc.8b00045 (2018).
- 97 Fischbach, M. & Voigt, C. A. Prokaryotic gene clusters: a rich toolbox for synthetic biology. *Biotechnol J* **5**, 1277-1296, doi:10.1002/biot.201000181 (2010).
- 98 Huo, L., Rachid, S., Stadler, M., Wenzel, S. C. & Muller, R. Synthetic biotechnology to study and engineer ribosomal bottromycin biosynthesis. *Chem Biol* **19**, 1278-1287, doi:10.1016/j.chembiol.2012.08.013 (2012).
- 99 Hou, Y. *et al.* Structure and biosynthesis of the antibiotic bottromycin D. *Org Lett* **14**, 5050-5053, doi:10.1021/ol3022758 (2012).

- 100 Gomez-Escribano, J. P., Song, L., Bibb, M. J. & Challis, G. L. Posttranslational  $\beta$ -methylation and macrolactamidation in the biosynthesis of the bottromycin complex of ribosomal peptide antibiotics. *Chemical Science* **3**, 3522-3525, doi:10.1039/C2SC21183A (2012).
- 101 Crone, W. J. K., Leeper, F. J. & Truman, A. W. Identification and characterisation of the gene cluster for the anti-MRSA antibiotic bottromycin: expanding the biosynthetic diversity of ribosomal peptides. *Chemical Science* **3**, 3516-3521, doi:10.1039/C2SC21190D (2012).
- 102 Crone, W. J. *et al.* Dissecting Bottromycin Biosynthesis Using Comparative Untargeted Metabolomics. *Angew Chem Int Ed Engl* **55**, 9639-9643, doi:10.1002/anie.201604304 (2016).
- 103 Mann, G. *et al.* Structure and Substrate Recognition of the Bottromycin Maturation Enzyme BotP. *Chembiochem* **17**, 2286-2292, doi:10.1002/cbic.201600406 (2016).
- 104 Burkhart, B. J., Schwalen, C. J., Mann, G., Naismith, J. H. & Mitchell, D. A. YcaO-Dependent Posttranslational Amide Activation: Biosynthesis, Structure, and Function. *Chem Rev* **117**, 5389-5456, doi:10.1021/acs.chemrev.6b00623 (2017).
- 105 Schwalen, C. J. *et al.* In Vitro Biosynthetic Studies of Bottromycin Expand the Enzymatic Capabilities of the YcaO Superfamily. *J Am Chem Soc* **139**, 18154-18157, doi:10.1021/jacs.7b09899 (2017).
- 106 Franz, L., Adam, S., Santos-Aberturas, J., Truman, A. W. & Koehnke, J. Macroamidine Formation in Bottromycins Is Catalyzed by a Divergent YcaO Enzyme. *J Am Chem Soc* **139**, 18158-18161, doi:10.1021/jacs.7b09898 (2017).
- 107 Melby, J. O., Nard, N. J. & Mitchell, D. A. Thiazole/oxazole-modified microcins: complex natural products from ribosomal templates. *Curr Opin Chem Biol* **15**, 369-378, doi:10.1016/j.cbpa.2011.02.027 (2011).
- 108 Böttcher, C. *et al.* The Multifunctional Enzyme CYP71B15 (PHYTOALEXIN DEFICIENT3) Converts Cysteine-Indole-3-Acetonitrile to Camalexin in the Indole-3-Acetonitrile Metabolic Network of *Arabidopsis thaliana*. *The Plant Cell* **21**, 1830-1845, doi:10.1105/tpc.109.066670 (2009).
- 109 Klein, A. P., Anarat-Cappillino, G. & Sattely, E. S. Minimum set of cytochromes P450 for reconstituting the biosynthesis of camalexin, a major *Arabidopsis* antibiotic. *Angew Chem Int Ed Engl* **52**, 13625-13628, doi:10.1002/anie.201307454 (2013).
- 110 Hao Liu, E. J. T. Synthesis of the (E)-dehydrobutyrine-thiazoline-proline-leucine fragment of vioprolides B and D. *Tetrahedron Letters* **54**, 3150-3153 (2013).
- 111 WHO. Antimicrobial resistance: global report on surveillance. (World Health Organization, 2014).
- 112 CDC. Antibiotic Resistance Threats in the United States. (Centers for Disease Control and Prevention, Atlanta, GA, 2013).
- 113 O'Neill, J. Review on Antimicrobial Resistance Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations. (London, 2014).
- 114 Piddock, L. J. & Gardp. The Global Antibiotic Research and Development Partnership (GARDP): a not-for-profit antibiotic development organisation. *Lancet Infect Dis* **18**, 1304-1305, doi:10.1016/S1473-3099(18)30661-3 (2018).
- 115 Outtersson, K. *et al.* Accelerating global innovation to address antibacterial resistance: introducing CARB-X. *Nat Rev Drug Discov* **15**, 589-590, doi:10.1038/nrd.2016.155 (2016).
- 116 Wright, G. D. & Poinar, H. Antibiotic resistance is ancient: implications for drug discovery. *Trends Microbiol* **20**, 157-159, doi:10.1016/j.tim.2012.01.002 (2012).
- 117 Reygaert, W. C. An overview of the antimicrobial resistance mechanisms of bacteria. *AIMS Microbiol* **4**, 482-501, doi:10.3934/microbiol.2018.3.482 (2018).

- 118 Munita, J. M. & Arias, C. A. Mechanisms of Antibiotic Resistance. *Microbiol Spectr* **4**, doi:10.1128/microbiolspec.VMBF-0016-2015 (2016).
- 119 Dever, L. A. & Dermody, T. S. Mechanisms of bacterial resistance to antibiotics. *Arch Intern Med* **151**, 886-895 (1991).
- 120 Nikaido, H. Multidrug resistance in bacteria. *Annu Rev Biochem* **78**, 119-146, doi:10.1146/annurev.biochem.78.082907.145923 (2009).
- 121 Wade, H. MD recognition by MDR gene regulators. *Curr Opin Struct Biol* **20**, 489-496, doi:10.1016/j.sbi.2010.06.003 (2010).
- 122 Melnyk, A. H., Wong, A. & Kassen, R. The fitness costs of antibiotic resistance mutations. *Evol Appl* **8**, 273-283, doi:10.1111/eva.12196 (2015).
- 123 Bjorkman, J. & Andersson, D. I. The cost of antibiotic resistance from a bacterial perspective. *Drug Resist Updat* **3**, 237-245, doi:10.1054/drup.2000.0147 (2000).
- 124 Depardieu, F., Podglajen, I., Leclercq, R., Collatz, E. & Courvalin, P. Modes and modulations of antibiotic resistance gene expression. *Clin Microbiol Rev* **20**, 79-114, doi:10.1128/CMR.00015-06 (2007).
- 125 Murakami, T., Holt, T. G. & Thompson, C. J. Thiostrepton-induced gene expression in *Streptomyces lividans*. *J Bacteriol* **171**, 1459-1466, doi:10.1128/jb.171.3.1459-1466.1989 (1989).
- 126 Brown, N. L., Stoyanov, J. V., Kidd, S. P. & Hobman, J. L. The MerR family of transcriptional regulators. *FEMS Microbiol Rev* **27**, 145-163, doi:10.1016/S0168-6445(03)00051-2 (2003).
- 127 Holmes, D. J., Caso, J. L. & Thompson, C. J. Autogenous transcriptional activation of a thiostrepton-induced gene in *Streptomyces lividans*. *EMBO J* **12**, 3183-3191 (1993).
- 128 Chiu, M. L. *et al.* Broad spectrum thiopeptide recognition specificity of the *Streptomyces lividans* TipAL protein and its role in regulating gene expression. *J Biol Chem* **274**, 20578-20586, doi:10.1074/jbc.274.29.20578 (1999).
- 129 Chiu, M. L. *et al.* Characterization of the covalent binding of thiostrepton to a thiostrepton-induced protein from *Streptomyces lividans*. *Biochemistry* **35**, 2332-2341, doi:10.1021/bi952073e (1996).
- 130 Kahmann, J. D. *et al.* Structural basis for antibiotic recognition by the TipA class of multidrug-resistance transcriptional regulators. *EMBO J* **22**, 1824-1834, doi:10.1093/emboj/cdg181 (2003).
- 131 Habazettl, J. *et al.* Structural basis and dynamics of multidrug recognition in a minimal bacterial multidrug resistance system. *Proc Natl Acad Sci U S A* **111**, E5498-5507, doi:10.1073/pnas.1412070111 (2014).
- 132 Chiu, M. L., Viollier, P. H., Katoh, T., Ramsden, J. J. & Thompson, C. J. Ligand-induced changes in the *Streptomyces lividans* TipAL protein imply an alternative mechanism of transcriptional activation for MerR-like proteins. *Biochemistry* **40**, 12950-12958, doi:10.1021/bi010328k (2001).



## **Chapter 2: The Thiazole-Specific Amydohydrolase PurAH is the Gatekeeper of Bottromycin Biosynthesis**

*Asfandyar Sikandar, Laura Franz, Okke Melse, Iris Antes, and Jesko Koehnke*

Reproduced with permission from *J. Am. Chem. Soc.* 2019, 141, 25, 9748-9752

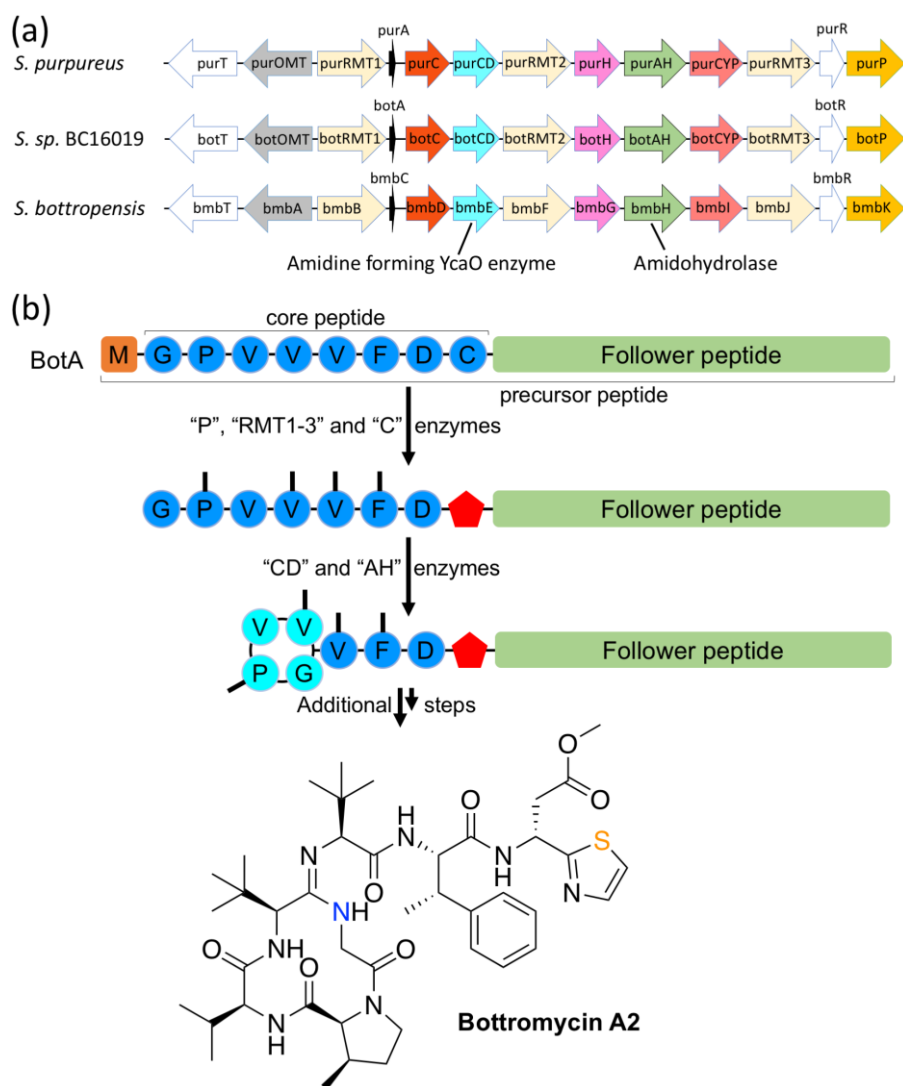
Publication Date: June 7, 2019

<https://doi.org/10.1021/jacs.8b12231>

**Copyright © 2019 American Chemical Society**

## 1 Introduction

Bottromycins<sup>1,2</sup> are natural product antibiotics with activity against problematic human pathogens such as Methicillin-resistant *Staphylococcus aureus* (MRSA)<sup>3,4</sup>. They bind to the A-site of the prokaryotic 50S ribosome, which is a novel antibiotic target<sup>5-7</sup>. Bottromycins belong to the growing family of ribosomally synthesized and post-translationally modified peptides (RiPPs), and their biosynthesis and total synthesis have received increasing attention<sup>8-12</sup>. They are derived from the precursor peptide BotA, which undergoes a series of post-translational enzymatic tailoring steps, the order of which has been proposed based on an untargeted metabolomics approach using mass spectral networking (Scheme 1, we will use *Streptomyces* sp. BC16019 nomenclature)<sup>13</sup>: First, the N-terminal methionine is removed by a leucyl-amino peptidase (BotP), followed by heterocyclization of the BotA cysteine residue to thiazoline by the YcaO enzyme BotC and C $\beta$ -methylation by radical methyl transferases 1-3. Next, a second YcaO enzyme, BotCD, was reported to act together with the metallo-dependent amidohydrolase BotAH in macroamidine formation<sup>13</sup>. Removal of the follower peptide by the  $\alpha/\beta$ -hydrolase BotH, successive oxidative decarboxylation of the thiazoline to a thiazole (BotCYP) and O-methylation of an aspartate (BotOMT) complete bottromycin biosynthesis.



Scheme 1. (a) A gene cluster highly homologous in sequence and organization to those of confirmed bottromycin producers *Streptomyces* sp. BC16019 and *S. bottropensis* was found in *S. purpureus* (top). Intergenic regions not drawn to scale. (b) Bottromycin biosynthesis as proposed by untargeted metabolomics<sup>13</sup>. Methylation by radical methyl transferases (RMT1-3) is shown as black lines, heterocyclized cysteine as a red pentagon and the macroamide in cyan.

## 2 Results and Discussion

*In vitro* work has confirmed the assigned functions for BotP<sup>13,14</sup>, BotC<sup>8,9</sup> and BotCD<sup>8,9</sup>. It was demonstrated that the BotCD homologs BmbE and PurCD alone are sufficient for macroamide formation *in vitro*<sup>8,9</sup>. For PurCD, macroamide formation was reversible: It catalyzes both, macroamide formation and its reopening<sup>8</sup>.

To probe the role of the amidohydrolase in bottromycin biosynthesis *in vitro*, we attempted to express BotAH (insoluble). The close homolog PurAH (72 % sequence identity to BotAH) from *Streptomyces purpureus* (Scheme 1) could be expressed and purified. We first tested possible substrates and incubated PurAH with precursor peptide BotA, BotA<sup>P</sup> (N-terminal methionine

removed), BotA<sup>PC</sup> (BotA<sup>P</sup> with heterocyclized cysteine) and BotA<sup>PCCD</sup> (macrocyclized BotA<sup>PC</sup>) (Figure 1a).

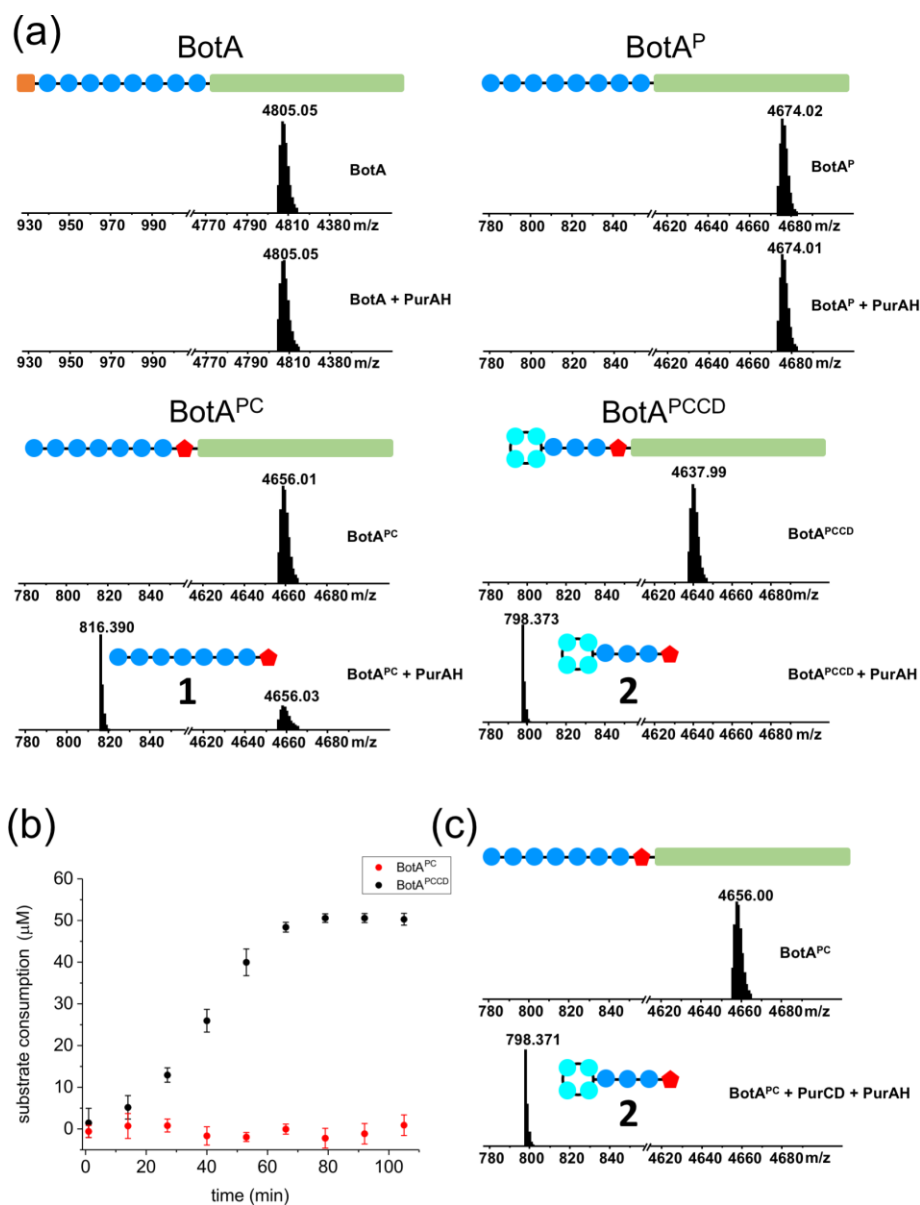


Figure 1. LC-ESI-MS characterization of PurAH reactions. (a) Incubation of PurAH with BotA, BotA<sup>P</sup>, BotA<sup>PC</sup>, and BotA<sup>PCCD</sup> (5 μM enzyme, 50 μM substrate, 37 °C, 16 h). Only BotA<sup>PC</sup> and BotA<sup>PCCD</sup> were substrates of PurAH. Colors correspond to scheme 1. (b) Time-course of BotA<sup>PC</sup> and BotA<sup>PCCD</sup> (50 μM) cleavage by PurAH (0.2 μM) in the presence of Co<sup>2+</sup> (100 μM). (c) When PurAH is added to a macrocyclization reaction, the reaction goes to completion and 2 is formed.

All samples were analyzed by liquid chromatography-electrospray ionization mass spectrometry (LC-ESI-MS). We observed no activity on BotA or BotA<sup>P</sup>. For the BotA<sup>PC</sup> reaction, we observed two new masses, which corresponded to the heterocyclized BotA core peptide (1, Figure S1A) (m/z calc<sub>mono.</sub>: 816.3840 Da, observed: 816.3814 Da, error -3.2 ppm) and the follower peptide (m/z calc<sub>mono.</sub>: 3857.62 Da, observed: 3857.63 Da, error 2.6 ppm), but the reaction was incomplete. For the BotA<sup>PCCD</sup> reaction, the substrate was consumed completely and two new masses appeared: One corresponded to the heterocyclized, macroamidine-

containing BotA core peptide (**2**, Figure S1b), while the other corresponded to the follower peptide ( $m/z$  calc<sub>mono.</sub>: 798.3734 Da, observed: 798.3706 Da, error -3.5 ppm and  $m/z$  calc<sub>mono.</sub>: 3857.62 Da, observed: 3858.63 Da, error 2.6 ppm), respectively. The identity of **1** and **2** was confirmed by MS<sup>2</sup> (Figure S2, Tables S1 and S2). These data demonstrate that PurAH is responsible for removing the follower peptide during bottromycin biosynthesis. Since the protein is annotated as a metal-dependent amidohydrolase, we dialyzed purified PurAH extensively with EDTA to remove endogenous metal ions and set up reactions of PurAH with BotA<sup>PCCD</sup> using different divalent metal ions (Figure S3). No activity was observed for CdCl<sub>2</sub> and NiCl<sub>2</sub>, while addition of FeCl<sub>2</sub> or MgCl<sub>2</sub> accelerated the background rate slightly. MnCl<sub>2</sub> << ZnCl<sub>2</sub> < CoCl<sub>2</sub> each led to appreciable turnover, with Co<sup>2+</sup> giving best activity. A time course experiment with 0.2 μM PurAH and 50 μM BotA<sup>PC</sup> or BotA<sup>PCCD</sup> revealed that the reaction using BotA<sup>PCCD</sup> was complete after 80 min, while no product formation was detectable in the same time frame using BotA<sup>PC</sup> (Figure 1b). PurAH is thus highly selective for the heterocyclized, macrocyclized intermediate BotA<sup>PCCD</sup> and removes the follower peptide.

With the function of the amidohydrolase established we investigated its involvement in macroamidine formation. The macroamidine forming YcaO enzyme PurCD converts BotA<sup>PC</sup> into BotA<sup>PCCD</sup> in an ATP/Mg<sup>2+</sup>-dependent reaction for which the BotA follower peptide is essential<sup>8,9</sup>. We have shown that this reaction is reversible – purified BotA<sup>PCCD</sup> incubated with PurCD and ATP/MgCl<sub>2</sub> will be converted back to BotA<sup>PC</sup>, now lacking the macrocycle (Figure S4)<sup>8</sup>. Given the very strong preference of PurAH for BotA<sup>PCCD</sup> and the dependence of PurCD activity on the presence of the follower peptide that is removed by PurAH, we hypothesized that PurAH may prevent re-opening of the macrocycle. First, we purified **2** and incubated it with PurCD and ATP/MgCl<sub>2</sub>. Even after extensive incubation times at 37 °C, we did not observe reopening of the macroamidine to yield **1** (Figure S4). BotA<sup>PCCD</sup> on the other hand converted readily back to BotA<sup>PC</sup> (Figure S4). When BotA<sup>PC</sup> is incubated with PurCD under optimized conditions, the turnover to BotA<sup>PCCD</sup> does not exceed 70 %.<sup>8</sup> However, when PurAH is added to the reaction, we observe complete conversion of BotA<sup>PC</sup> to **2**, indicating complete conversion of BotA<sup>PC</sup> (macroamidine formation) and subsequent removal of the follower peptide (Figure 1c). These data rationalize why the amidohydrolase was reported as essential for macroamidine formation *in vivo*<sup>13</sup> – knocking out PurAH likely leads to reopening of the macroamidine and prevented downstream processing. This places PurAH as the gatekeeper of bottromycin biosynthesis, that removes the follower peptide once all enzymes requiring the follower have acted on the precursor peptide.

To understand how PurAH selects for the macrocycle in BotA<sup>PCCD</sup>, we determined the PurAH crystal structure to 1.73 Å resolution. All data collection and refinement statistics can be found in Table S3. The refined model contained one protomer in the asymmetric unit and includes residues 9 – 74 and 78 – 460 (Figure S5). Two Zn<sup>2+</sup> ions are coordinated at the active site: Zn1 by His210 and His229, Zn2 by His94, His96 and Asp348. Lys183 has been carboxylated as commonly observed in amidohydrolases<sup>15</sup>, with each oxygen coordinating one Zn<sup>2+</sup> (Figure S5). The distorted trigonal bipyramidal coordination of the two Zn<sup>2+</sup> is completed by ordered water molecules, one of which bridges the two Zn<sup>2+</sup> ions. This bridging water may be an activated hydroxyl known to facilitate catalysis in amidohydrolases<sup>16-19</sup>. We generated a sequence similarity network (SSN) for PurAH (Figure S6a) and found it to be part of a small node that exclusively contained PurAH homologs from all known bottromycin biosynthetic gene clusters. Mapping the sequences of this node onto the PurAH structure using ConSurf<sup>20-23</sup>, we found residues surrounding the active site to be highly conserved (Figure 2a). In contrast, the closest structural homologs identified by a DALI<sup>24</sup> search show virtually no sequence conservation at the active site (Figure 2b). PurAH contains an extended, wide binding site that may have evolved specifically to accept branched cyclic peptide substrates (Figure S6b).

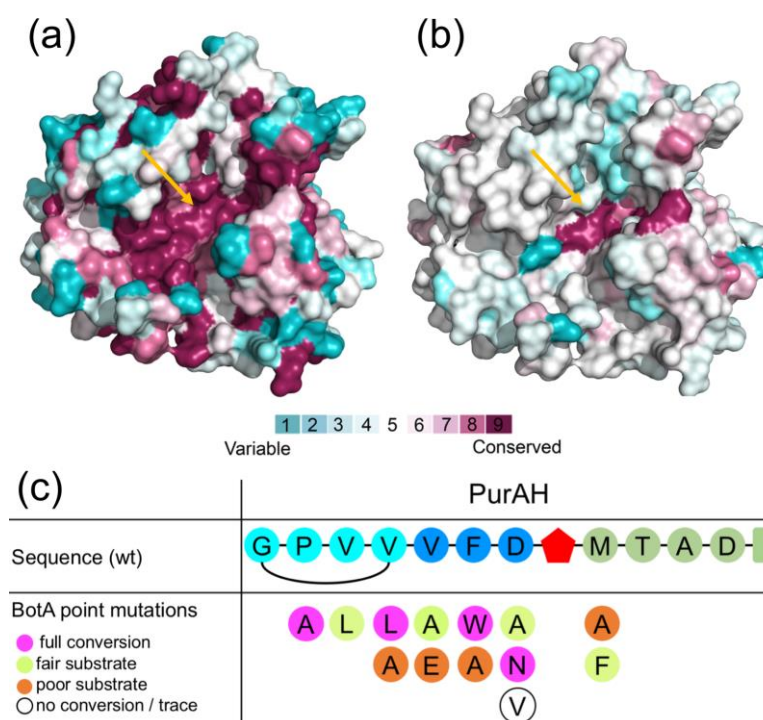
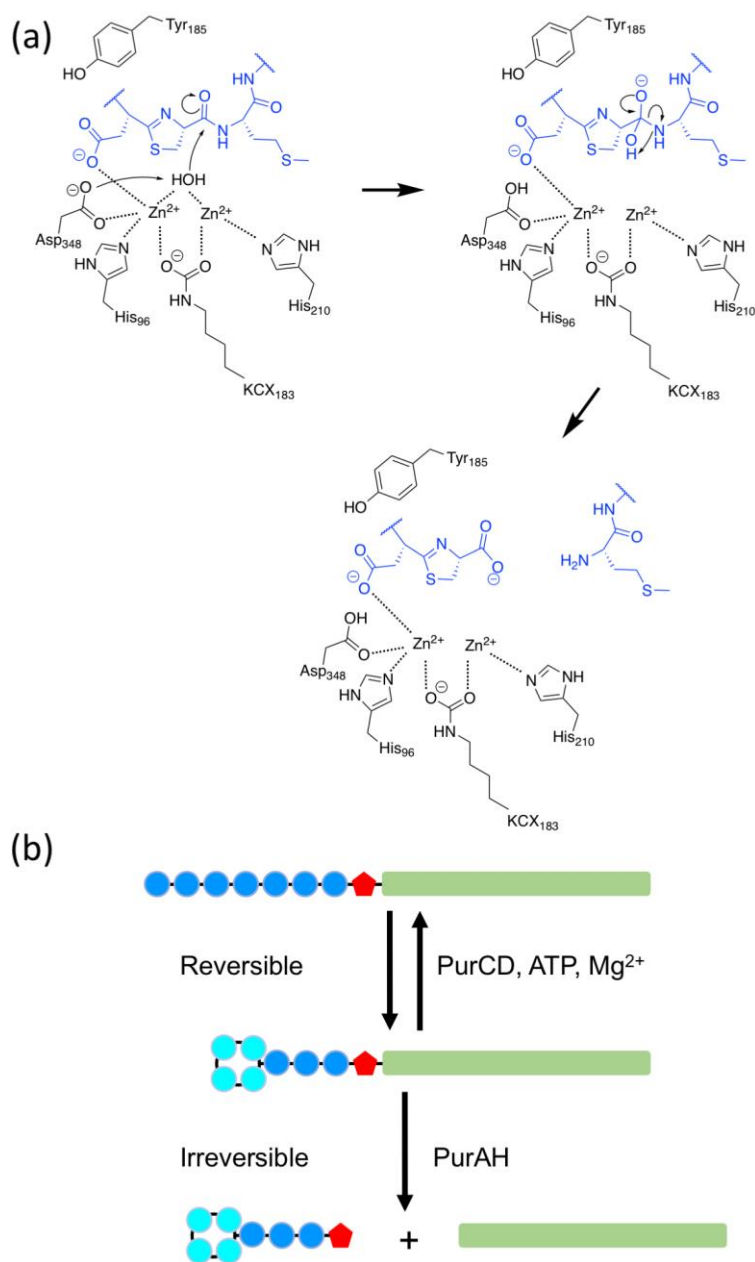


Figure 2. (a) and (b) Consurf maps showing the conservation of residues around the PurAH active site (arrow) for (a) PurAH homologs from other bottromycin biosynthetic gene clusters (PurAH containing node, Figure S6a) and (b) the closest structural homologs identified by a DALI search (Figure S6b). (c) Schematic view of the conversion of 13 BotA mutants designed to test PurAH promiscuity.

To rationalize PurAH's selectivity, we sought to determine the complex crystal structure of PurAH with its substrate, but extensive screening failed in delivering suitable conditions. In

most RiPP systems the core peptide sequence is at least to some degree variable, but in all bottromycin biosynthetic gene clusters identified to date the core peptide sequence is fully conserved<sup>25-27</sup> (with the exception of a single-amino acid change found in one cluster<sup>28</sup>). This is at odds with the substrate promiscuity observed for bottromycin biosynthetic enzymes in vitro thus far<sup>8,9,14</sup>. To probe the promiscuity of PurAH and the fidelity for macrocyclized intermediate, we generated a series of mutant precursor peptides (Figure 2c). The degree of conversion was estimated by monitoring the consumption of substrate (see SOI for details). Data for the mutations are provided in Figures S7 – S20. BotA mutants Pro2Ala, Val4Leu, Phe6Trp, and Asp7Asn were found to be good substrates of PurAH (Figure 2c). The mutants Val3Leu, Val5Ala, and Asp7Ala were fair and Val4Ala, Val5Glu, and Phe6Ala were poor substrates of PurAH. Asp7Val could not be processed, which may implicate Asp7 in metal coordination. While presence of a heterocycle within the S1 site (Using standard protease nomenclature) of PurAH was essential for activity, the S1' site is more flexible, as demonstrated by partial conversion of BotA mutants Met9Ala and Met9Phe. In all cases where turnover was observed, only the mass for heterocyclized, macrocyclized product could be detected. These data demonstrate that PurAH is promiscuous in the context of single amino-acid changes, while retaining a high fidelity for macrocyclized substrate.



*Scheme 2. (a) Proposed mechanism of follower peptide removal by PurAH. It is unclear which residue stabilizes the oxyanion.  $Zn^{2+}$ -coordinating residues His94 and His229 have been omitted for clarity (b) Proposed role of PurAH in bottromycin biosynthesis. Removal of the follower peptide (green) by PurAH prevents reopening of the macroamide (cyan) and thus drives biosynthesis.*

It has been demonstrated for amidohydrolases distantly related to PurAH that the aspartate residue involved in metal coordination at the active site (Asp348 in PurAH) and a tyrosine residue in proximity of the active site were involved in catalysis (e.g.<sup>29</sup>). In PurAH, the mutation Asp348Asn resulted in a loss of catalytic activity (Figure S21). These data may implicate Asp348 as the base in catalysis, but we cannot exclude disruption of metal binding. The only



tyrosine residue in the immediate vicinity of the active-site Zn<sup>2+</sup>-ions was Tyr185, and in PurAH the mutant Tyr185Phe showed impaired substrate processing (Figure S21).

We have demonstrated that removal of the follower peptide in bottromycin biosynthesis is catalyzed by PurAH. Since the preceding macroamidine formation is reversible, this enzyme has a profound effect on bottromycin biosynthesis by cooperating with the YcaO enzyme PurCD to ensure efficient macrocyclization. This resolves the apparent disparity between *in vitro*<sup>8,9</sup> and *in vivo*<sup>13</sup> data and places PurAH as the gatekeeper between primary and secondary modification steps (Scheme 2). It also adds to our understanding of the YcaO superfamily: Thioamide-forming YcaO enzymes appear to require a TfuA protein for activity<sup>30</sup>, while YcaO enzymes installing heterocycles in linear azolic peptides require an E1-like protein<sup>31,32</sup> and their activity appears in some cases coupled to an FMN-dependent oxidase<sup>33,34</sup>. Despite the cooperation between PurCD and PurAH, we were unable to detect complex formation *in vitro* using various methods (data not shown). This raises questions regarding the potential colocalization of RiPP enzymes within the producing organism, which have only been addressed in a very limited number of RiPP systems and require intensive further study. The selectivity of PurAH for macrocyclized precursor peptides coupled with its substrate promiscuity will be important in future efforts to derivatize bottromycins.

## **3 Supplementary Information**

### **3.1 Cloning, protein expression and purification**

#### **3.1.1 Cloning, Expression and Purification of PurCD and IpoC**

PurCD and IpoC were purified as described previously.<sup>8</sup>

#### **3.1.2 Cloning of PurAH wt and mutants**

Full-length, codon-optimized *purAH* (Accession No. WP\_019887083.1, Eurofins Genomics) was cloned into pET-28b (Novagen). Presence of the insert was confirmed by restriction digestion and DNA sequencing before being transformed into Rosetta™ (DE3). Mutations that lead to the exchange of one amino acid in the PurAH sequence were introduced with the site-directed mutagenesis (SDM) primers shown in Table S5 using the reagents and protocol of the Quick Change II kit (Agilent Technologies). The introduction of the mutations was verified by DNA sequencing.

#### **3.1.3 Cloning, Expression and Purification BotA and BotA point mutants**

BotA, BotA<sup>P</sup> and BotA<sup>P</sup> variants were cloned, expressed and purified as described previously.<sup>8</sup> BotA<sup>P</sup> variants with one altered amino acid were generated by using mutation carrying primers. Peptide concentrations were determined at a Spectrophotometer (NanoDrop2000, Thermo Fisher Scientific) using the extinction coefficient calculated from the amino acid sequence by the ExPASy ProtParam Server<sup>35</sup> (<http://web.expasy.org/protparam/>).

#### **3.1.4 Protein expression and purification of PurAH wt and mutants**

A single colony was picked into LB liquid medium containing the appropriate antibiotics (50 µg / mL Kanamycin and / or 34 µg / mL Chloramphenicol) to make an overnight culture. The overnight culture was inoculated 1 to 100 into fresh LB medium supplemented with antibiotics and was grown at 37 °C until the optical density (OD<sub>600</sub>) reached 0.6; protein expression was then induced by the addition of 1 mM IPTG and the cells grown at 16 °C overnight. Cells were harvested by centrifuging the culture at 6,000 x g at 4 °C for 15 min and the cell pellets frozen at -80 °C.

The cell pellets were resuspended in lysis buffer (20 mM Tris pH 8.0, 200 mM NaCl, 20 mM imidazole, 10 % glycerol (w/v) and 3 mM DTT) supplemented with 0.4 mg DNase per gram of wet cell pellet and cOmplete EDTA-free protease-inhibitor tablets (Roche). The cells were lysed by sonication and cell debris removed by centrifugation (40,000 x g, 4 °C, 20 min). The supernatant was loaded onto a pre-equilibrated (lysis buffer) 5 mL Histrap HP column (GE Healthcare) and was washed with 20 column volumes of lysis buffer. The protein was eluted from the column with 250 mM imidazole and loaded onto a gel filtration column (HiLoad 16/600 Superdex 200 pg, GE Healthcare) pre-equilibrated in 50 mM Tris-Cl, pH 8.5, 200 mM NaCl, 10 % (w/v) glycerol and 1mM TCEP. The fractions of the highest purity were pooled together and concentrated to 5-8 mg mL<sup>-1</sup>.

## **3.2 Biochemistry**

### **3.2.1 Heterocyclization reaction by IpoC**

For heterocyclisation reactions by IpoC, 50 µM substrate (BotA<sup>P</sup> or BotA<sup>P</sup> variants) was incubated with 5 µM IpoC and 5 mM ATP/MgCl<sub>2</sub> in reaction buffer (50 mM Tris pH 8.5, 200 mM NaCl, 10 % glycerol) for 12 h at 37 °C.

For further use, heterocyclized BotA<sup>P</sup> (BotA<sup>PC</sup>) was immediately purified by size-exclusion chromatography on a pre-equilibrated Superdex S30 column (GE Healthcare) using reaction buffer and concentrated using 5 kDa cutoff filters (Merk Millipore).

Reactions of BotA<sup>P</sup> variants with IpoC were set up and analyzed in triplicates. Turnover was analyzed by LC-MS. For LC-MS analyses, reactions were stopped and enzymes were precipitated by the addition of ACN.

### **3.2.2 Macrocyclization reaction by PurCD**

For macrocyclisation reactions by PurCD, 50 µM substrate (BotA<sup>PC</sup>) was incubated with 5 µM PurCD and 10 mM ATP/MgCl<sub>2</sub> in reaction buffer (50 mM Tris pH 8.5, 200 mM NaCl, 10 % glycerol) for 2 h at room temperature. Macroamidine formation was determined by LC-MS analysis.

Product (BotA<sup>PCCD</sup>) was immediately purified for further experiments by size-exclusion chromatography on a pre-equilibrated Superdex S30 column (GE Healthcare) using reaction buffer. BotA<sup>PCCD</sup> was concentrated using 5 kDa cutoff filters (Merck Millipore). As we observe macroamidine re-opening by PurCD, the purification product is a mixture of BotA<sup>PC</sup> and

BotA<sup>PCCD</sup>. Thus, this mixture was further purified by HPLC to obtain pure BotA<sup>PCCD</sup>. Purification was performed on a Dionex Ultimate 3000 RSLC system using a Xbridge C18, 4.6 x 100 mm, 5 µm dp column equipped with a C18 precolumn (Waters). Aliquots of 15 µL were separated by a gradient from (A) 10 mM NH<sub>4</sub>HCO<sub>3</sub> in H<sub>2</sub>O to (B) ACN at a flow rate of 1 mL/min and 45 °C. The gradient was initiated by a 1 min isocratic step at 5 % B, followed by an increase to 19.7 % B in 2 min, to 21.0 % B in 24 min and to 95 % B in 0.5 min. After a 2 min step at 95 % B the system was re-equilibrated to the initial conditions (5 % B). BotA<sup>PCCD</sup> containing fractions (confirmed by LC-MS) were pooled and ACN evaporated on a rotary evaporator. Finally, the residue was injected on a Superdex S30 column (GE Healthcare) using reaction buffer.

The ability of PurCD to use different IpoC-cyclised BotA<sup>P</sup> variants as substrate was tested. 50 µM BotA<sup>P</sup> variant or wt (as control) was first incubated with 5 µM IpoC and 10 mM ATP/MgCl<sub>2</sub> in reaction buffer for 12 h at 37 °C. 5 µM PurCD was added and samples were incubated for further 2 h at 37 °C. The addition of ACN was used to stop the reaction and precipitate the enzymes. The conversion was analyzed by LC-MS. Reactions were set up and analyzed in triplicates.

To test if macroamidine re-opening by PurCD occurs after the follower is cleaved, 50 µM purified BotA<sup>PCCD</sup> or BotA<sup>PCCDAH</sup> was incubated with 10 µM PurCD, 10 mM ATP/MgCl<sub>2</sub> in macrocyclisation buffer (50 mM Tris pH 8.5, 200 mM NaCl, 10 % glycerol) at room temperature for 16 h. Controls without the addition of enzymes were also set up. Reactions were stopped by the addition of ACN and analyzed by LC-MS.

### 3.2.3 Follower cleavage by PurAH

To test possible substrates for PurAH, 50 µM BotA peptide (BotA, BotA<sup>P</sup>, BotA<sup>PC</sup> or BotA<sup>PCCD</sup>) was incubated with 5 µM PurAH and 100 µM CoCl<sub>2</sub> in reaction buffer at 37 °C for 16 h. The reactions were set up in triplicates. After the 16 h incubation time, reactions were stopped and enzymes were precipitated by the addition of ACN and analyzed by LC-MS.

The activity of EDTA-treated PurAH was tested in the presence of no metal ions and using different metal ions. 50 µM BotA<sup>PCCD</sup> was incubated with 0.2 µM PurAH and 100 µM ZnCl<sub>2</sub>, CoCl<sub>2</sub>, CdCl<sub>2</sub>, FeCl<sub>2</sub>, MgCl<sub>2</sub>, MnCl<sub>2</sub>, NiCl<sub>2</sub> or no metal ions, respectively, at 37 °C. Reactions were set up in triplicates and stopped after an incubation time of 45 min by the addition of ACN. Turnover was analyzed by LC-MS.

The ability of PurAH to cleave off the follower of BotA<sup>P</sup> variants was tested. 50 µM BotA<sup>P</sup> variant or wt (as control) were first incubated with 5 µM IpoC and 10 mM ATP/MgCl<sub>2</sub> in

reaction buffer for 12 h at 37 °C. 5  $\mu\text{M}$  PurCD and 2.5  $\mu\text{M}$  PurAH were added and samples were incubated for further 2 h at 37 °C. The reactions were stopped and enzymes were precipitated by the addition of ACN and the conversion was analyzed by LC-MS. Reactions were set up and analyzed in triplicates.

To determine the initial conversion rate of BotA<sup>PC</sup> and BotA<sup>PCCD</sup> by PurCD, 50  $\mu\text{M}$  substrate was incubated with 0.2  $\mu\text{M}$  PurAH and 100  $\mu\text{M}$  CoCl<sub>2</sub>. Samples were prepared without addition of enzyme and pre-incubated at 37 °C. After addition of enzyme, samples were incubated in the LC sample holder at 37 °C and 1  $\mu\text{l}$  aliquots were automatically taken and analyzed every 13 min by the LC-MS system in the time range from 1 to 105 min. Reactions were set up and analyzed in triplicates.

The turnover of PurAH wt, PurAH Y185F and PurAH D348N was compared. 50  $\mu\text{M}$  BotA<sup>P</sup> wt was first incubated with 5  $\mu\text{M}$  IpoC and 10 mM ATP/MgCl<sub>2</sub> in reaction buffer for 12 h at 37 °C. 5  $\mu\text{M}$  PurCD was added and samples were incubated for further 1.5 h at 37 °C. The reaction solution was split up and 2.5  $\mu\text{M}$  PurAH (wt or mutant) was added. After 0, 30, 60, 120, 240 and 480 min, the reactions were stopped and enzymes were precipitated by the addition of two volumes of ACN.

To produce and purify BotA<sup>PCCDAH</sup>, 50  $\mu\text{M}$  BotA<sup>PC</sup> was incubated with 5  $\mu\text{M}$  PurCD, 0.2  $\mu\text{M}$  PurAH, 10 mM ATP/MgCl<sub>2</sub>, 100  $\mu\text{M}$  CoCl<sub>2</sub> in reaction buffer at 37 °C overnight. The reaction solution was applied on a Chromabond C18 Hydra column (6 mL, 2000 mg, Macherey-Nagel), pre-equilibrated with dest. H<sub>2</sub>O. After washing with 2 CV dest. H<sub>2</sub>O, bound peptides (**2** and Follower peptide) were eluted separately by increasing the ACN concentration in 5 % steps. Fractions containing pure BotA<sup>PCCDAH</sup> (analyzed by LC-MS) were combined and all solvent evaporated on a rotary evaporator. Finally, **2** was dissolved in reaction buffer.

### **3.3 Mass spectrometry**

#### **3.3.1 LC-ESI-MS analysis**

All measurements to analyze the mass of processed or unprocessed BotA peptides were performed on a Dionex Ultimate 3000 RSLC system using a Xbridge C18, 100 x 2.1 mm, 3.5  $\mu\text{m}$  dp column equipped with a C18 precolumn (Waters). Samples of 1  $\mu\text{L}$  were separated by a gradient from (A) 10 mM NH<sub>4</sub>HCO<sub>3</sub> to (B) ACN at a flow rate of 600  $\mu\text{L}/\text{min}$  and 45 °C. The gradient was initiated by a 0.5 min isocratic step at 5 % B, followed by an increase to 15 % B in 1 min, to 27 % B in 8.5 min and to 95 % B in 0.5 min. After a 2 min step at 95 % B the

system was re-equilibrated to the initial conditions (5 % B). UV spectra were recorded by a DAD in the range from 200 to 600 nm.

For MS measurements on maXis-2 UHR-TOF mass spectrometer (Bruker Daltonics), the LC flow was split 1:8 before entering the mass spectrometer using the Apollo ESI source. In the source region, the temperature was set to 250 °C, the capillary voltage was 4000 V, the dry-gas flow was 10.0 L/min and the nebulizer was set to 30 psi. After the generated ions passed the quadrupole with a low cut-off at 150 m/z they were trapped in the collision cell for 100  $\mu$ s and then transferred within 10  $\mu$ s through the hexapole into the ICR cell. Data were recorded in the mass range from 250 to 2500 m/z. Peaks in the MS-spectra are labelled with the observed monoisotopic masses.

For quantification of the time course experiments, the consumption of the substrates (BotA<sup>PC</sup> and BotA<sup>PCCD</sup>, respectively) was quantified. The peak areas of the [M+H]<sup>4+</sup>, [M+H]<sup>5+</sup> and [M+H]<sup>6+</sup> ions (which were the predominant ions) from BotA<sup>PC</sup> or BotA<sup>PCCD</sup> (see Table S4, EIC width  $\pm$  5 ppm) were integrated and concentrations were calculated using a calibration curve obtained from serial dilutions of the respective purified BotA (BotA<sup>PC</sup> or BotA<sup>PCCD</sup>) using the TASQ 1.1 software (Bruker Daltonics). The relationship between the peak areas and the peptide concentrations are mostly linear in the tested concentration range (2-75  $\mu$ M) and no saturation of the detector was not observed (as an example the calibration curve for BotA<sup>PCCD</sup> is shown in Figure S26).

For the comparison of the activity of EDTA-treated PurAH in presence of different metal ions, the area of the reaction product **2** ([M+H]<sup>+</sup>=799.3807  $\pm$  0.05 Da) was integrated and compared. For **2**, a linear relationship of peak areas and concentrations were tested (concentration range 2.6-100  $\mu$ M) and verified (Figure S27), which allows a comparison of the formation of **2** by comparison of the peak areas.

To compare the turnover of the BotA<sup>PCCD</sup> wt by PurAH wt and mutants, the peak areas of the reaction product **2** ([M+H]<sup>+</sup>=799.3807  $\pm$  0.05 Da) were integrated and compared.

For estimation (relative differences) of the turnover of the BotA substrate mutants by PurAH wt (categorized as “good” (turnover comparable to wild-type (wt) substrate), “fair” (> 50 % turnover compared to wt substrate), (> 50% turnover compared to wt substrate) “poor” , and “no turnover/trace”), the sum of substrate peak areas of the [M+H]<sup>4+</sup>, [M+H]<sup>5+</sup> and [M+H]<sup>6+</sup> ions from the respective BotA<sup>PC</sup> and BotA<sup>PCCD</sup> mutants (see Table S4) before and after addition of PurAH wt were compared. A linear relationship between the peak areas and the peptide substrate concentrations (in the assay concentrations range) was confirmed: Dilutions of the enzyme assay solutions (wt and mutant peptides) and dilutions of purified

BotA<sup>PC</sup> wt and BotA<sup>PCCD</sup> wt show a linear relationship between the dilution factor and the peak area. We thus conclude that the inference of relative differences is feasible.

### 3.3.2 MS<sup>2</sup> Fragmentation

MS<sup>2</sup> fragmentation experiments were executed by direct infusion MS using a solarix XR (7T) FT-ICR mass spectrometer (Bruker Daltonics) equipped with an Apollo ESI source. The source region temperature was set to 220 °C, the capillary voltage was 4500 V, the dry gas flow was 4.0 L min<sup>-1</sup> and the nebulizer was set to 1.0 bar. The collision energy was ramped from 10 to 22 V until sufficient dissociation of the target molecule was observed. MS<sup>2</sup> spectra were obtained by collision-induced dissociation (CID) in the collision cell.

## 3.4 Crystallization and structure determination

Crystals of PurAH were obtained at 18 °C in 3.0 - 4.2 M sodium formate. The crystals were cryoprotected in mother liquor supplemented with 35 % glycerol and flash cooled in liquid nitrogen. Data was collected at ESRF (Beamlines: ID23-1 and ID30A-3). To solve the PurAH structure, a single-wavelength anomalous dispersion (SAD) data set and a high-resolution dataset were collected. Data were processed using XDS<sup>36</sup> and POINTLESS<sup>37</sup>, AIMLESS<sup>38</sup> and truncate<sup>39</sup> implemented in ccp4<sup>40</sup>. The structure was solved using the SAD dataset with the Auto-Rickshaw<sup>41</sup> pipeline's MR-SAD phasing method<sup>42-49</sup>. The model was manually rebuilt in COOT<sup>50</sup> and refined using PHENIX<sup>42</sup>. The structure was validated using *MolProbity*<sup>51</sup>, and all images presented were created using PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.). Interaction diagrams were created using PoseView<sup>52,53</sup> and plotting was performed with Matplotlib v2.1.0.<sup>54</sup>

## 4 References

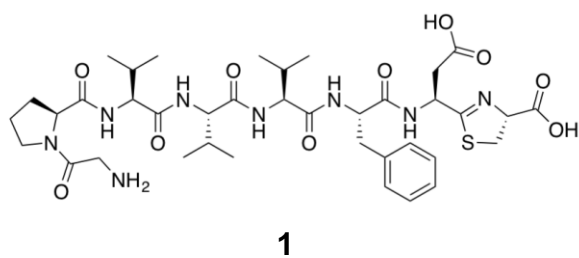
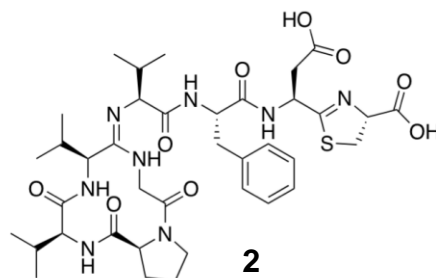
- 1 Nakamura, S. Isolation and characterization of bottromycins A and B. *J. Antibiotics, Ser. A* **18**, 47-52 (1965).
- 2 Waisvisz, J. M. Bottromycin. I. A new sulfurcontaining antibiotic. *J. Am. Chem. Soc.* **79**, 4520-4521 (1957).
- 3 Sowa, S. *et al.* Susceptibility of methicillin-resistant *Staphylococcus aureus* clinical isolates to various antimicrobial agents. *Hiroshima J Med Sci* **40**, 137-144 (1991).
- 4 Shimamura, H. *et al.* Structure determination and total synthesis of bottromycin A2: a potent antibiotic against MRSA and VRE. *Angew Chem Int Ed Engl* **48**, 914-917, doi:10.1002/anie.200804138 (2009).
- 5 Otaka, T. & Kaji, A. Mode of action of bottromycin A2. Release of aminoacyl- or peptidyl-tRNA from ribosomes. *J Biol Chem* **251**, 2299-2306 (1976).
- 6 Otaka, T. & Kaji, A. Mode of action of bottromycin A2: effect on peptide bond formation. *FEBS Lett* **123**, 173-176 (1981).
- 7 Otaka, T. & Kaji, A. Mode of action of bottromycin A2: effect of bottromycin A2 on polysomes. *FEBS Lett* **153**, 53-59 (1983).
- 8 Franz, L., Adam, S., Santos-Aberturas, J., Truman, A. W. & Koehnke, J. Macroamidine Formation in Bottromycins Is Catalyzed by a Divergent YcaO Enzyme. *J Am Chem Soc* **139**, 18158-18161, doi:10.1021/jacs.7b09898 (2017).
- 9 Schwalen, C. J. *et al.* In Vitro Biosynthetic Studies of Bottromycin Expand the Enzymatic Capabilities of the YcaO Superfamily. *J Am Chem Soc* **139**, 18154-18157, doi:10.1021/jacs.7b09899 (2017).
- 10 Horbal, L., Marques, F., Nadmid, S., Mendes, M. V. & Luzhetskyy, A. Secondary metabolites overproduction through transcriptional gene cluster refactoring. *Metab Eng* **49**, 299-315, doi:10.1016/j.ymben.2018.09.010 (2018).
- 11 Eyles, T. H., Vior, N. M. & Truman, A. W. Rapid and Robust Yeast-Mediated Pathway Refactoring Generates Multiple New Bottromycin-Related Metabolites. *ACS Synth Biol* **7**, 1211-1218, doi:10.1021/acssynbio.8b00038 (2018).
- 12 Yamada, T. *et al.* Synthesis and Evaluation of Antibacterial Activity of Bottromycins. *J Org Chem* **83**, 7135-7149, doi:10.1021/acs.joc.8b00045 (2018).
- 13 Crone, W. J. *et al.* Dissecting Bottromycin Biosynthesis Using Comparative Untargeted Metabolomics. *Angew Chem Int Ed Engl* **55**, 9639-9643, doi:10.1002/anie.201604304 (2016).
- 14 Mann, G. *et al.* Structure and Substrate Recognition of the Bottromycin Maturation Enzyme BotP. *Chembiochem* **17**, 2286-2292, doi:10.1002/cbic.201600406 (2016).
- 15 Jimenez-Morales, D., Adamian, L., Shi, D. & Liang, J. Lysine carboxylation: unveiling a spontaneous post-translational modification. *Acta Crystallogr D Biol Crystallogr* **70**, 48-57, doi:10.1107/S139900471302364X (2014).
- 16 Concha, N. O., Rasmussen, B. A., Bush, K. & Herzberg, O. Crystal structure of the wide-spectrum binuclear zinc beta-lactamase from *Bacteroides fragilis*. *Structure* **4**, 823-836 (1996).
- 17 Concha, N. O. *et al.* Crystal structure of the IMP-1 metallo beta-lactamase from *Pseudomonas aeruginosa* and its complex with a mercaptocarboxylate inhibitor: binding determinants of a potent, broad-spectrum inhibitor. *Biochemistry* **39**, 4288-4298 (2000).
- 18 Ullah, J. H. *et al.* The crystal structure of the L1 metallo-beta-lactamase from *Stenotrophomonas maltophilia* at 1.7 Å resolution. *J Mol Biol* **284**, 125-136, doi:10.1006/jmbi.1998.2148 (1998).



- 19 Bounaga, S., Laws, A. P., Galleni, M. & Page, M. I. The mechanism of catalysis and the inhibition of the *Bacillus cereus* zinc-dependent beta-lactamase. *Biochem J* **331** (Pt 3), 703-711 (1998).
- 20 Ashkenazy, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* **44**, W344-350, doi:10.1093/nar/gkw408 (2016).
- 21 Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* **38**, W529-533, doi:10.1093/nar/gkq399 (2010).
- 22 Landau, M. *et al.* ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res* **33**, W299-302, doi:10.1093/nar/gki370 (2005).
- 23 Glaser, F. *et al.* ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* **19**, 163-164 (2003).
- 24 Holm, L. & Laakso, L. M. Dali server update. *Nucleic Acids Res* **44**, W351-355, doi:10.1093/nar/gkw357 (2016).
- 25 Crone, W. J. K., Leeper, F. J. & Truman, A. W. Identification and characterisation of the gene cluster for the anti-MRSA antibiotic bottromycin: expanding the biosynthetic diversity of ribosomal peptides. *Chemical Science* **3**, 3516-3521, doi:10.1039/C2SC21190D (2012).
- 26 Gomez-Escribano, J. P., Song, L., Bibb, M. J. & Challis, G. L. Posttranslational  $\beta$ -methylation and macrolactamidation in the biosynthesis of the bottromycin complex of ribosomal peptide antibiotics. *Chemical Science* **3**, 3522-3525, doi:10.1039/C2SC21183A (2012).
- 27 Huo, L., Rachid, S., Stadler, M., Wenzel, S. C. & Muller, R. Synthetic biotechnology to study and engineer ribosomal bottromycin biosynthesis. *Chem Biol* **19**, 1278-1287, doi:10.1016/j.chembiol.2012.08.013 (2012).
- 28 Hou, Y. *et al.* Structure and biosynthesis of the antibiotic bottromycin D. *Org Lett* **14**, 5050-5053, doi:10.1021/ol3022758 (2012).
- 29 Fast, W. & Tipton, P. A. The enzymes of bacterial census and censorship. *Trends Biochem Sci* **37**, 7-14, doi:10.1016/j.tibs.2011.10.001 (2012).
- 30 Mahanta, N., Liu, A., Dong, S., Nair, S. K. & Mitchell, D. A. Enzymatic reconstitution of ribosomal peptide backbone thioamidation. *Proc Natl Acad Sci U S A* **115**, 3030-3035, doi:10.1073/pnas.1722324115 (2018).
- 31 Dunbar, K. L., Melby, J. O. & Mitchell, D. A. YcaO domains use ATP to activate amide backbones during peptide cyclodehydrations. *Nat Chem Biol* **8**, 569-575, doi:10.1038/nchembio.944 (2012).
- 32 Dunbar, K. L. *et al.* Discovery of a new ATP-binding motif involved in peptidic azoline biosynthesis. *Nat Chem Biol* **10**, 823-829, doi:10.1038/nchembio.1608 (2014).
- 33 Melby, J. O., Li, X. & Mitchell, D. A. Orchestration of enzymatic processing by thiazole/oxazole-modified microcin dehydrogenases. *Biochemistry* **53**, 413-422, doi:10.1021/bi401529y (2014).
- 34 Li, Y. M., Milne, J. C., Madison, L. L., Kolter, R. & Walsh, C. T. From peptide precursors to oxazole and thiazole-containing peptide antibiotics: microcin B17 synthase. *Science* **274**, 1188-1193 (1996).
- 35 Wilkins, M. R. *et al.* Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* **112**, 531-552 (1999).
- 36 Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132, doi:10.1107/S09074444909047337 (2010).

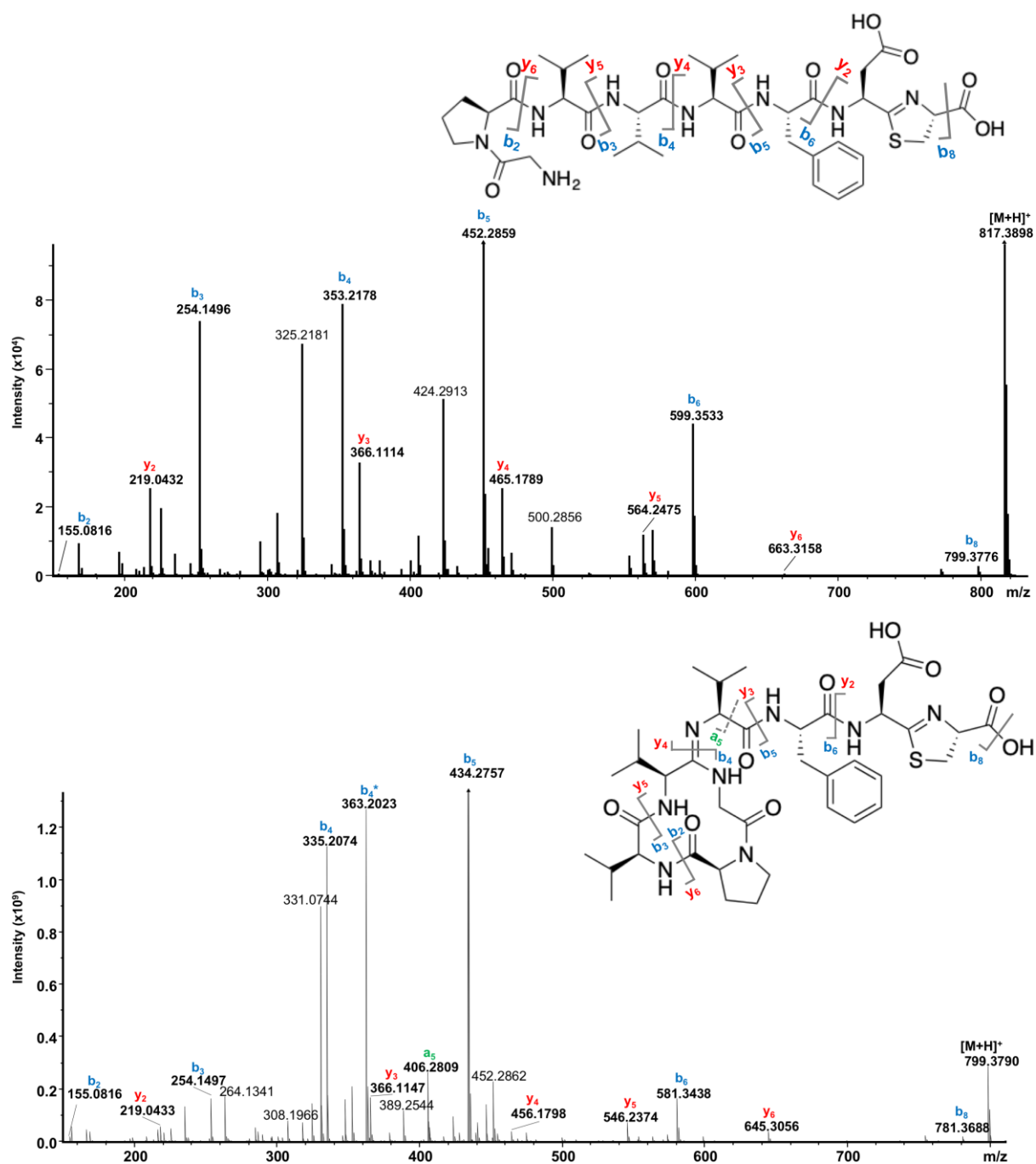
- 37 Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr D Biol Crystallogr* **67**, 282-292, doi:10.1107/S090744491003982X (2011).
- 38 Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204-1214, doi:10.1107/S0907444913000061 (2013).
- 39 French, S. & Wilson, K. On the treatment of negative intensity observations. *Acta Crystallographica Section A* **34**, 517-525, doi:doi:10.1107/S0567739478001114 (1978).
- 40 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242, doi:10.1107/S0907444910045749 (2011).
- 41 Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. On the combination of molecular replacement and single-wavelength anomalous diffraction phasing for automated structure determination. *Acta Crystallogr D Biol Crystallogr* **65**, 1089-1097, doi:10.1107/S0907444909029643 (2009).
- 42 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).
- 43 Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallographica Section D* **59**, 2023-2030, doi:doi:10.1107/S0907444903017694 (2003).
- 44 Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**, 905-921 (1998).
- 45 Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D* **62**, 1002-1011, doi:doi:10.1107/S0907444906022116 (2006).
- 46 McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674, doi:10.1107/S0021889807021206 (2007).
- 47 Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355-367, doi:10.1107/S0907444911001314 (2011).
- 48 Otwinowski, Z. Maximum likelihood refinement of heavy atom parameters. *Isomorphons Replacement and Anomalous Scattering*, 80-86 (1991).
- 49 Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* **66**, 479-485, doi:10.1107/S0907444909038360 (2010).
- 50 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).
- 51 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21, doi:10.1107/S0907444909042073 (2010).
- 52 Fricker, P. C., Gastreich, M. & Rarey, M. Automated Drawing of Structural Molecular Formulas under Constraints. *Journal of Chemical Information and Computer Sciences* **44**, 1065-1078, doi:10.1021/ci049958u (2004).
- 53 Stierand, K., Rarey, M. & Maaß, P. C. Molecular complexes at a glance: automated generation of two-dimensional complex diagrams. *Bioinformatics* **22**, 1710-1716, doi:10.1093/bioinformatics/btl150 (2006).

- 54 Hunter, J. D. Matplotlib: A 2D graphics environment. *Computing in science & engineering* **9**, 90-95 (2007).
- 55 Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402 (1997).
- 56 Gerlt, J. A. *et al.* Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim Biophys Acta* **1854**, 1019-1037, doi:10.1016/j.bbapap.2015.04.015 (2015).
- 57 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).
- 58 Durrant, J. D., Votapka, L., Sorensen, J. & Amaro, R. E. POVME 2.0: An Enhanced Tool for Determining Pocket Shape and Volume Characteristics. *J Chem Theory Comput* **10**, 5047-5056, doi:10.1021/ct500381c (2014).

**A****B****C**

MTADFLNDDPNNAELSALEMEELESWGAWDGEATS

*Figure S1: A Chemical structure of 1 (MW: 816.38 Da)- the modified BotA core peptide after methionine removal (BotP), heterocyclization (IpoC) and follower removal (PurAH). B Chemical structure of 2 (MW: 798.37 Da)- the modified BotA core peptide after methionine removal (BotP), heterocyclization (IpoC), macroamide formation (PurCD) and follower removal (PurAH). C Single-letter amino acid sequence of the BotA follower peptide (MW: 3857.62 Da).*



**Figure S2:** MS<sup>2</sup> fragmentation spectra for **1** (top) and **2** (bottom). A list of the associated fragments can be found in tables S1 and S2, respectively.

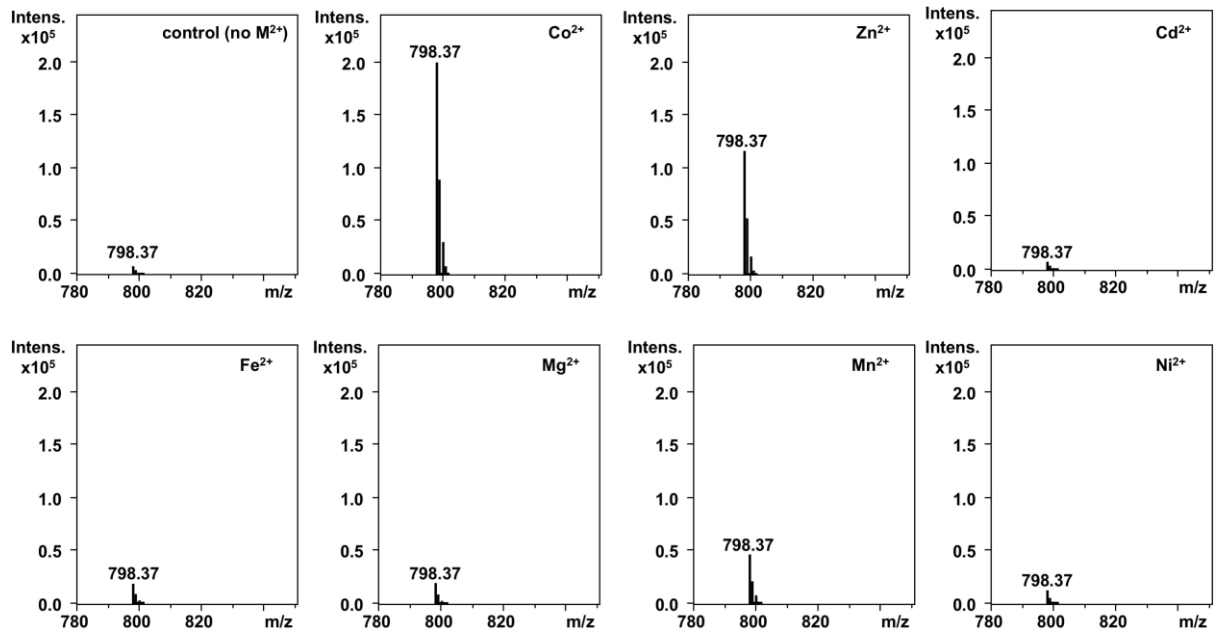


Figure S3: Metal-dependency of PurAH-catalyzed follower removal analysed by LC-MS. As can be seen, addition of  $\text{Co}^{2+}$  or  $\text{Zn}^{2+}$  gives greatest activity.

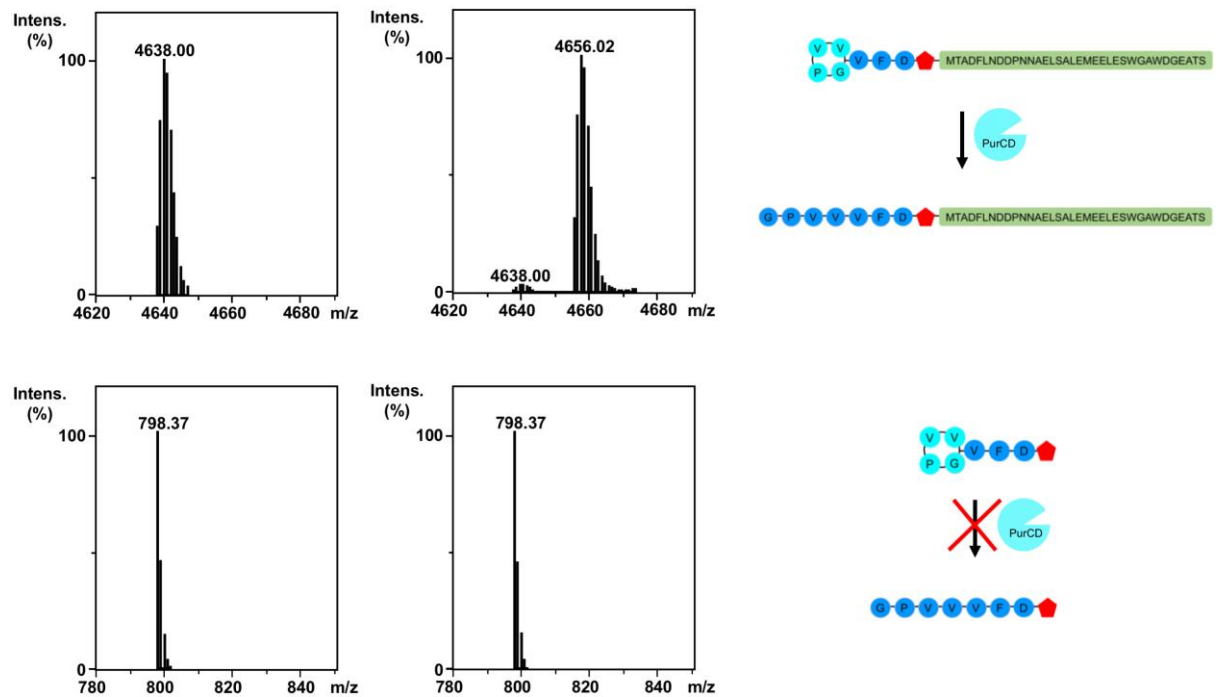


Figure S4: The PurCD,  $\text{Mg}^{2+}$  and ATP-dependent reopening of the macroamidine in BotA is follower-dependent. Analysis by LC-MS.

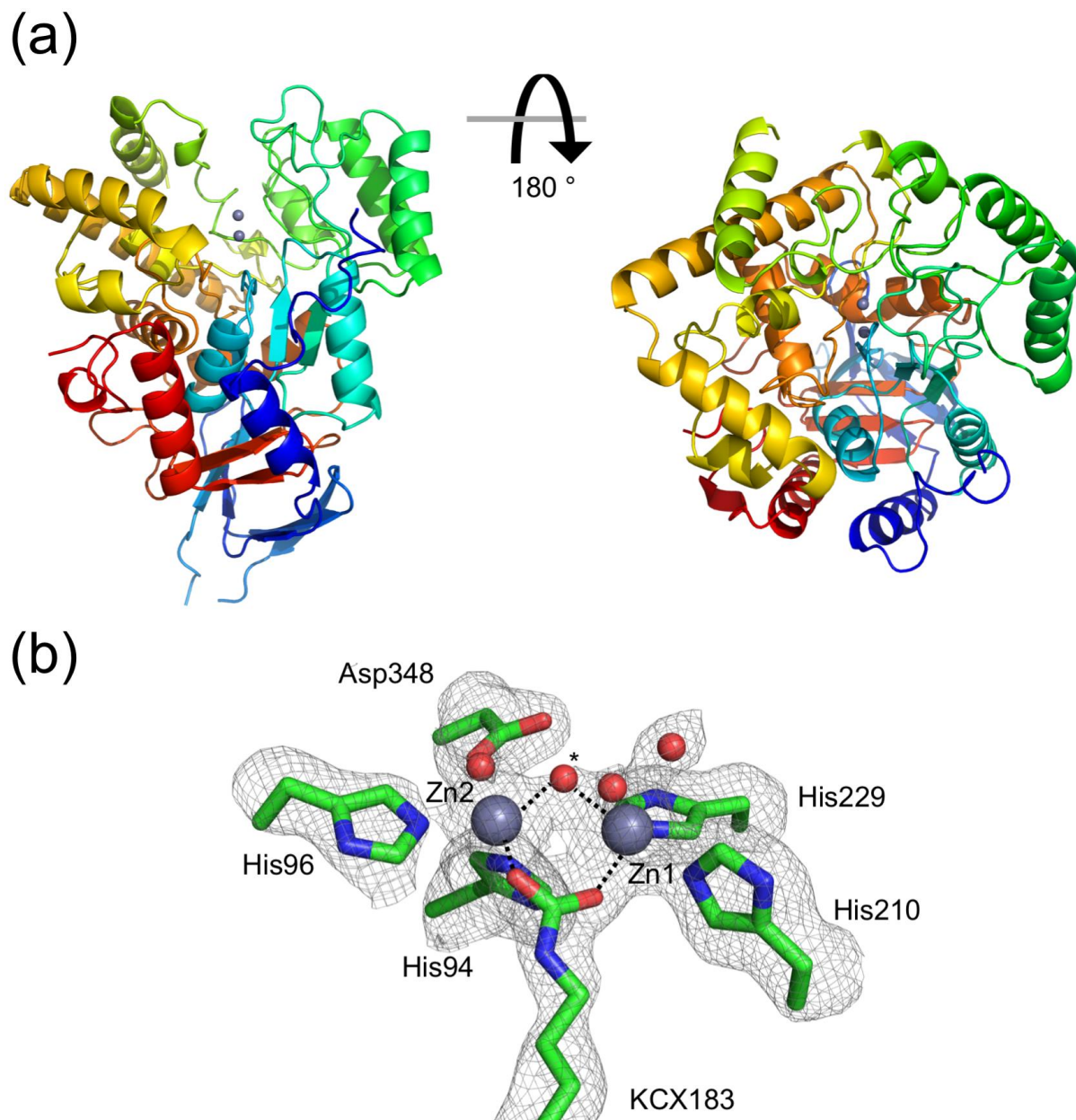


Figure S5: (a) Overall structure of PurAH. Protein is shown as a cartoon representation using the standard rainbow color scheme,  $Zn^{2+}$  ions are shown as grey spheres. (b) Coordination of  $Zn^{2+}$  ions at the active site. Water molecules are shown as red spheres, selected hydrogen bonds are shown as dashed lines. The bridging water molecule between the two  $Zn^{2+}$  ions is marked by an asterisk. The difference electron density ( $F_o - F_c$  contoured at  $3\sigma$  with phases calculated from a model which was refined after deletion of the shown residues, ions and waters) is shown as a grey isomesh.

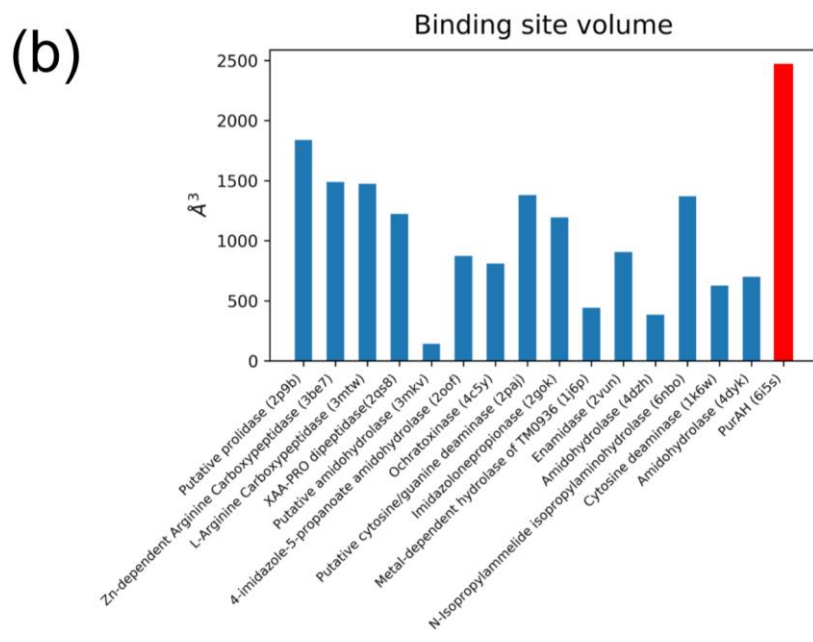
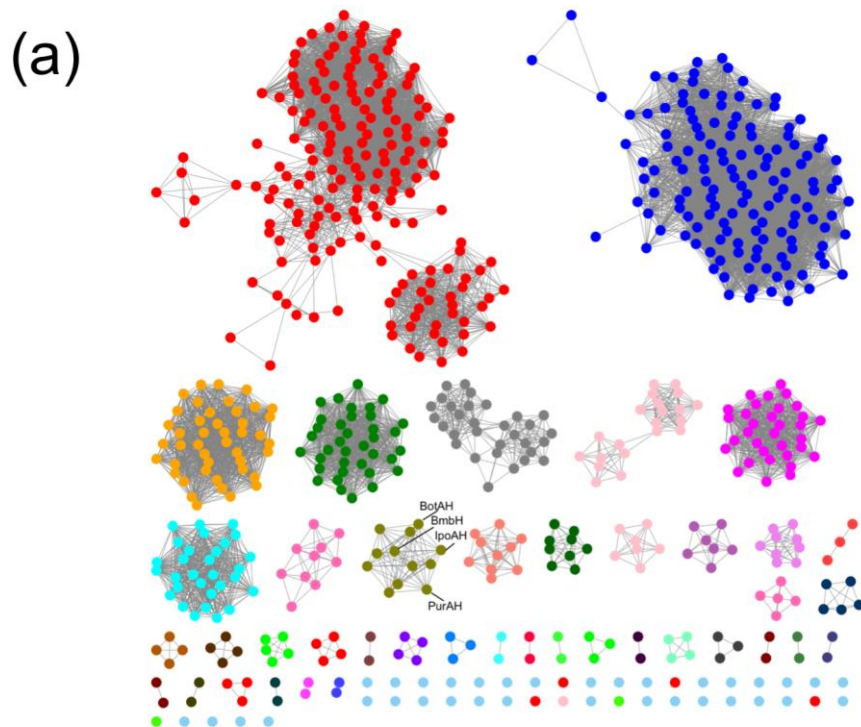


Figure S6: (a) Sequence similarity network (SSN) generated using the PurAH amino acids sequence. After one iteration of PSI-BLAST<sup>55</sup> the closest 1000 sequences were downloaded (e-value cut-off of  $2e-10$ ). The SSN was generated using these sequences on the server of the enzyme function initiative using the enzyme similarity tool<sup>56</sup>. The cut-offs used were sequence maximum/minimum lengths of 650/250 and an alignment score of 80. The SSN was visualized using Cytoscape<sup>57</sup>. PurAH and other sequences from bottromycin gene clusters form a node. No experimentally validated function has been assigned to any of the proteins that are members of other nodes (or singlets). (b) Binding pocket volumes determined by POVME 2.0<sup>58</sup> for 15 proteins identified as closest structural homologs of PurAH from a DALI<sup>24</sup> search. PurAH is colored red. The pocket volumes were determined using a gridspacing of  $1\text{\AA}$  and applying the 'ConvexHullExclusion' function to remove points outside the binding pocket.



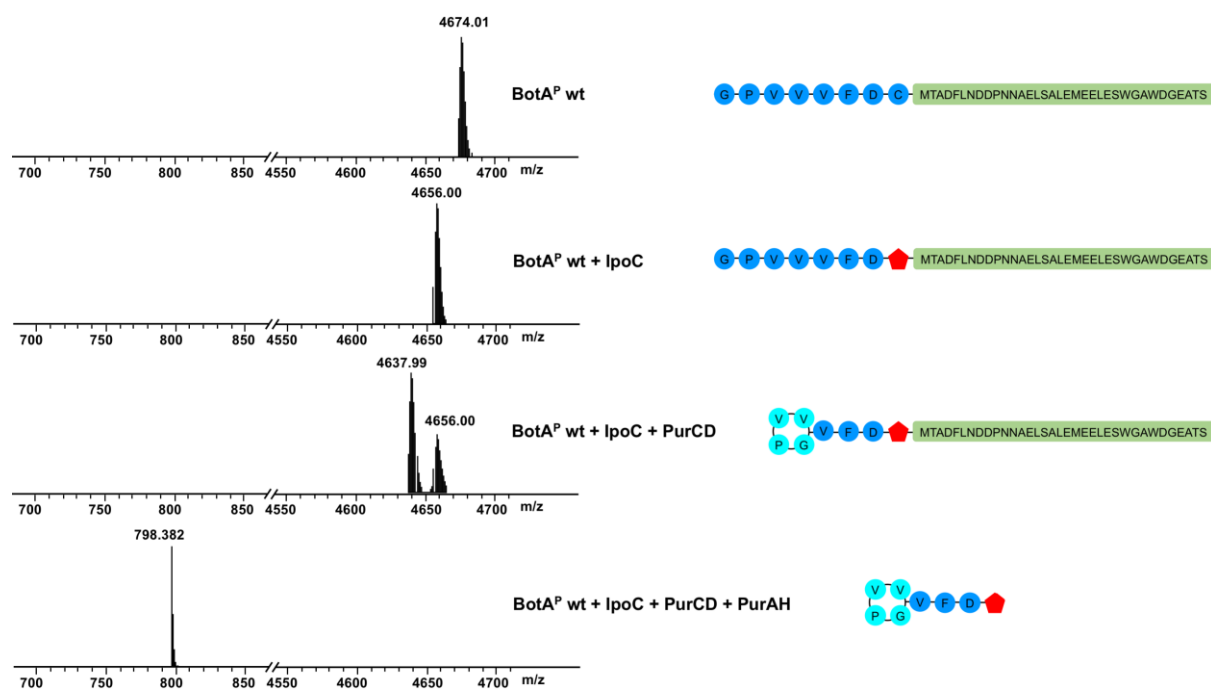


Figure S7: Processing of BotA<sup>P</sup> by IpoC, PurCD and PurAH. To heterocyclize the bottromycin core peptide cysteine to thiazoline we previously reported using the close BotC homolog IpoC from *S. ipomoeae*. The PurCD reaction is not unidirectional but an equilibrium, in which PurCD catalyzes both, macroamide formation and its reopening. Addition of PurAH cleaves the follower and drives macroamide formation to completion. Analysis by LC-MS.

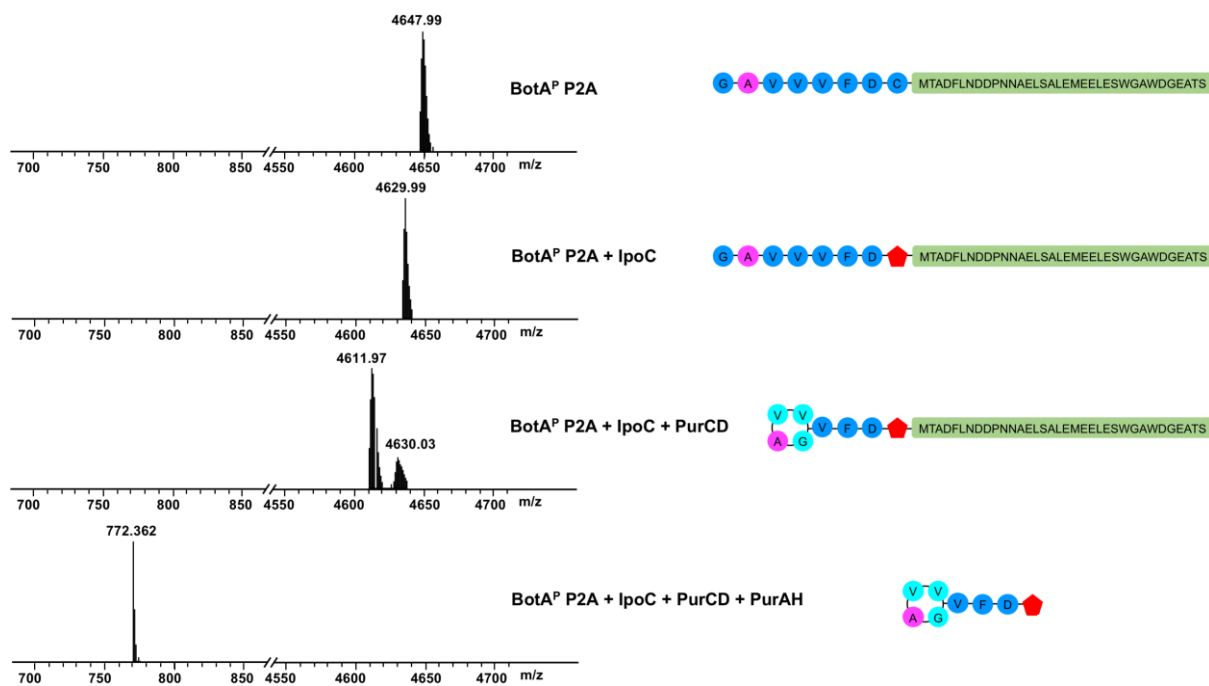


Figure S8: Processing of  $BotA^{P,P2A}$  by  $IpoC$ ,  $PurCD$  and  $PurAH$ . Turnover goes complete when  $PurAH$  is added (as seen for the wild-type  $BotA^P$ ). Analysis by LC-MS.

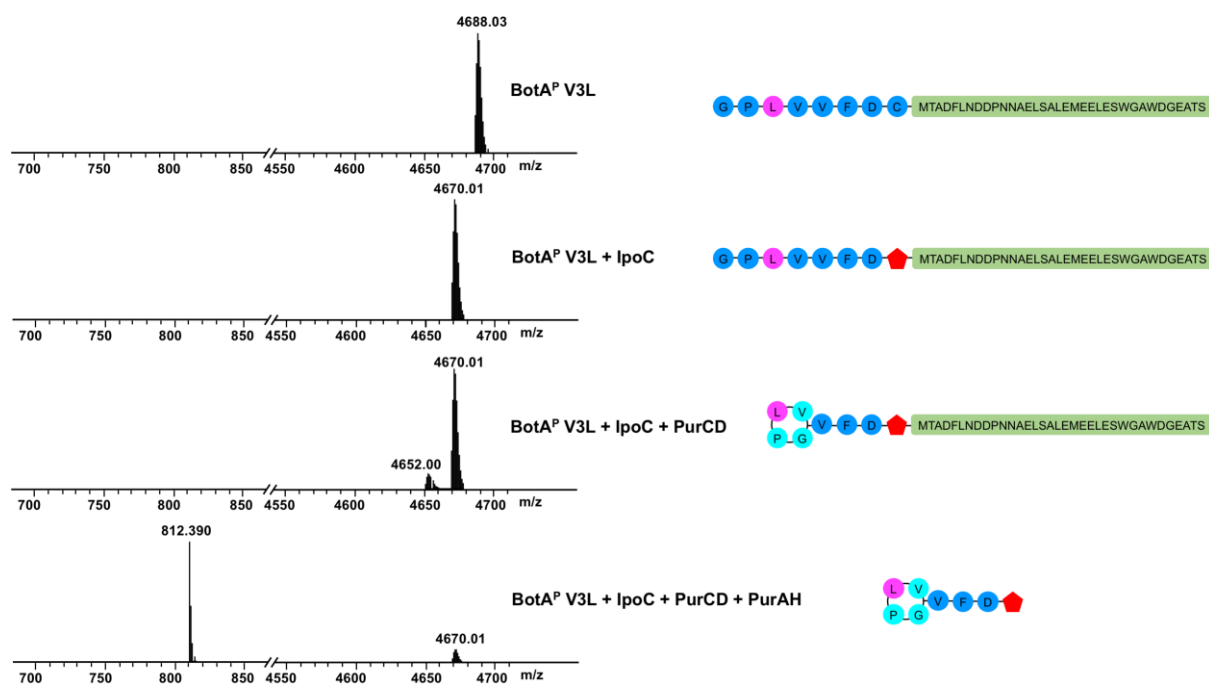


Figure S9: Processing of  $BotA^{P,V3L}$  by  $IpoC$ ,  $PurCD$  and  $PurAH$ . Turnover of heterocyclized  $BotA^{P,V3L}$  by  $PurCD$  is smaller compared to the wild-type  $BotA^{PC}$ . Addition of  $PurAH$  leads to the follower cleavage from  $BotA^{PCCD,V3L}$ , but turnover by  $PurAH$  is not complete. Although turnover of  $V3L$  and  $V4L$  (see S11) by  $PurCD$  are comparable,  $PurAH$  is able to drive the  $BotA^{V4L}$  reaction to completion. Analysis by LC-MS.

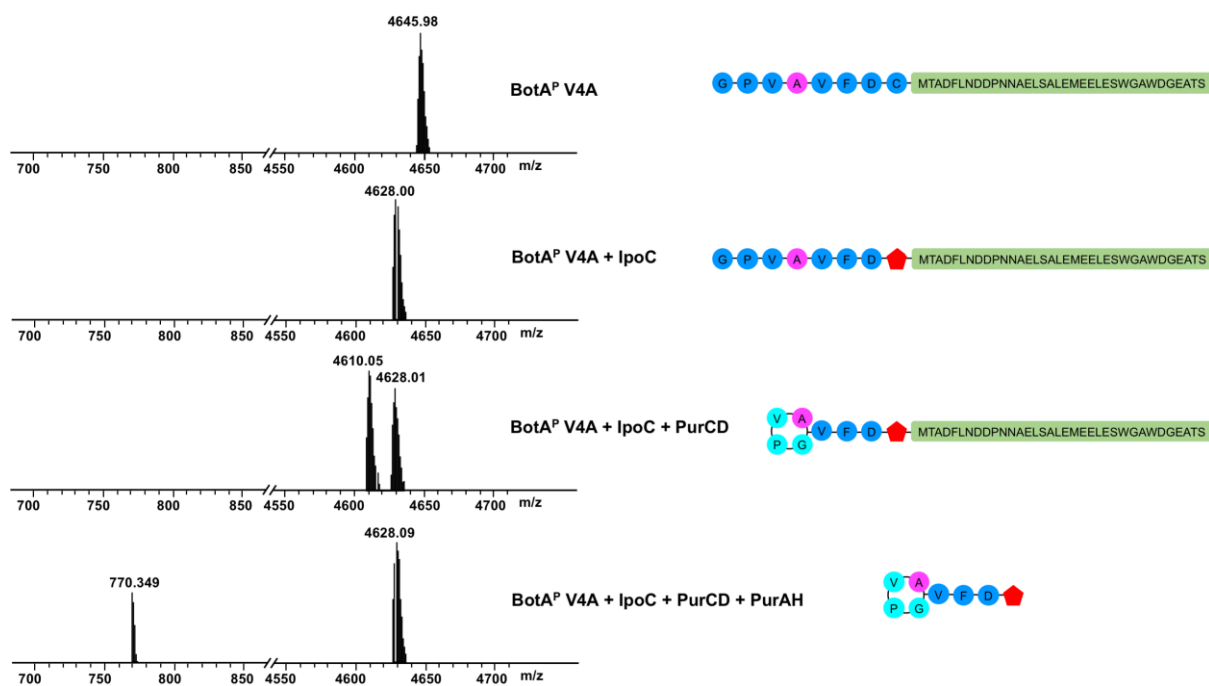


Figure S10: Processing of BotA<sup>P,V4A</sup> by IpoC, PurCD and PurAH. Although turnover by PurCD is reasonable, after addition of PurAH only small amounts of processed peptide without the follower are detected. Altering Val4 to Ala seems to slow down PurAH turnover. Analysis by LC-MS.

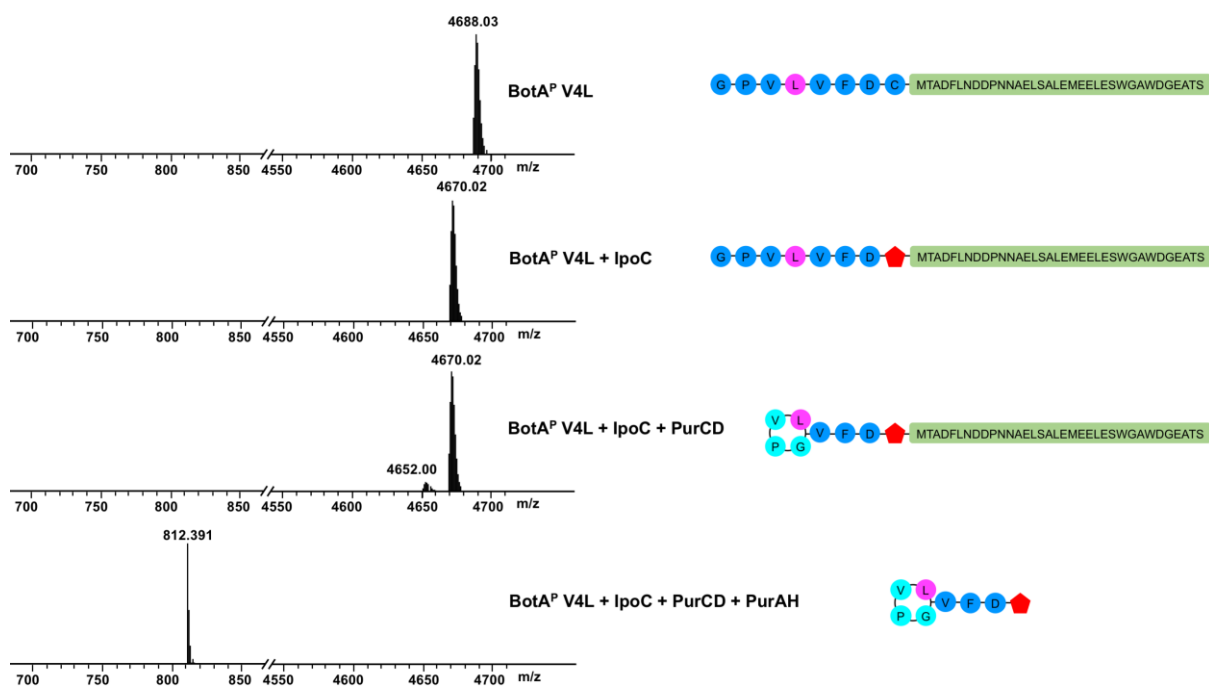


Figure S11: Processing of BotA<sup>P,V4L</sup> by IpoC, PurCD and PurAH. BotA<sup>P,V4L</sup> is a substrate for IpoC, but the turnover of this heterocyclized BotA variant by PurCD is very small compared to the wild type BotA. Addition of PurAH cleaves the follower and drives macroamidine formation to completion. Analysis by LC-MS.

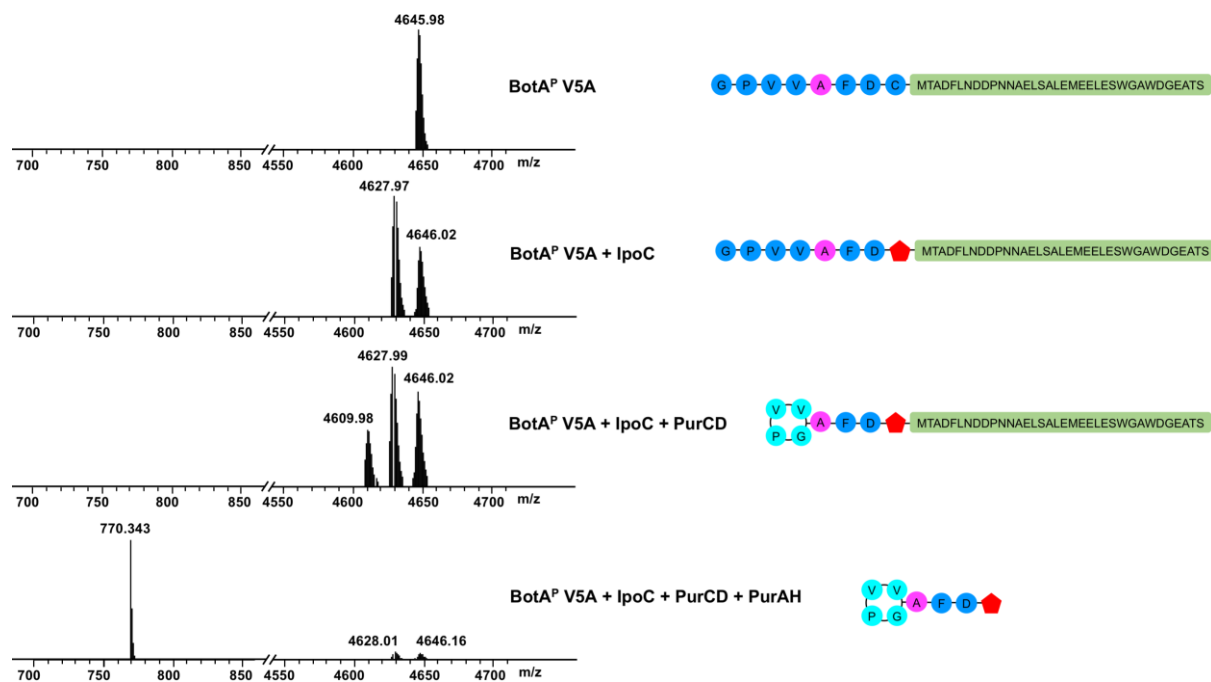


Figure S12: Processing of BotA<sup>P,V5A</sup> by IpoC, PurCD and PurAH. The turnover of BotA<sup>P,V5A</sup> by IpoC is not complete. As seen for the wild type, the PurCD reaction is not unidirectional and an equilibrium between the processed peptide with and without the macroamidase is observed. Addition of PurAH cleaves the follower, but reaction is not complete. Analysis by LC-MS.

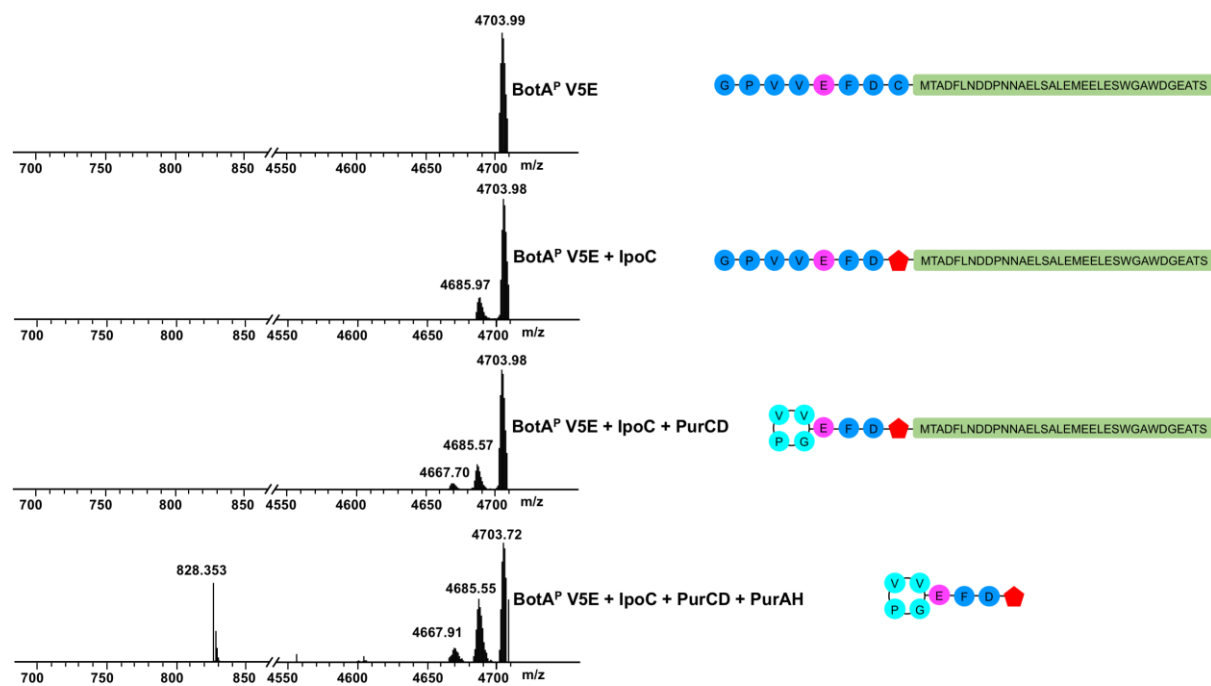


Figure S13: Processing of BotA<sup>P,V5E</sup> by IpoC, PurCD and PurAH. Turnover of BotA<sup>P,V5E</sup> by IpoC is not complete. After addition of PurAH, only very small amounts of follower-cleaved and processed BotA are detected. Analysis by LC-MS.

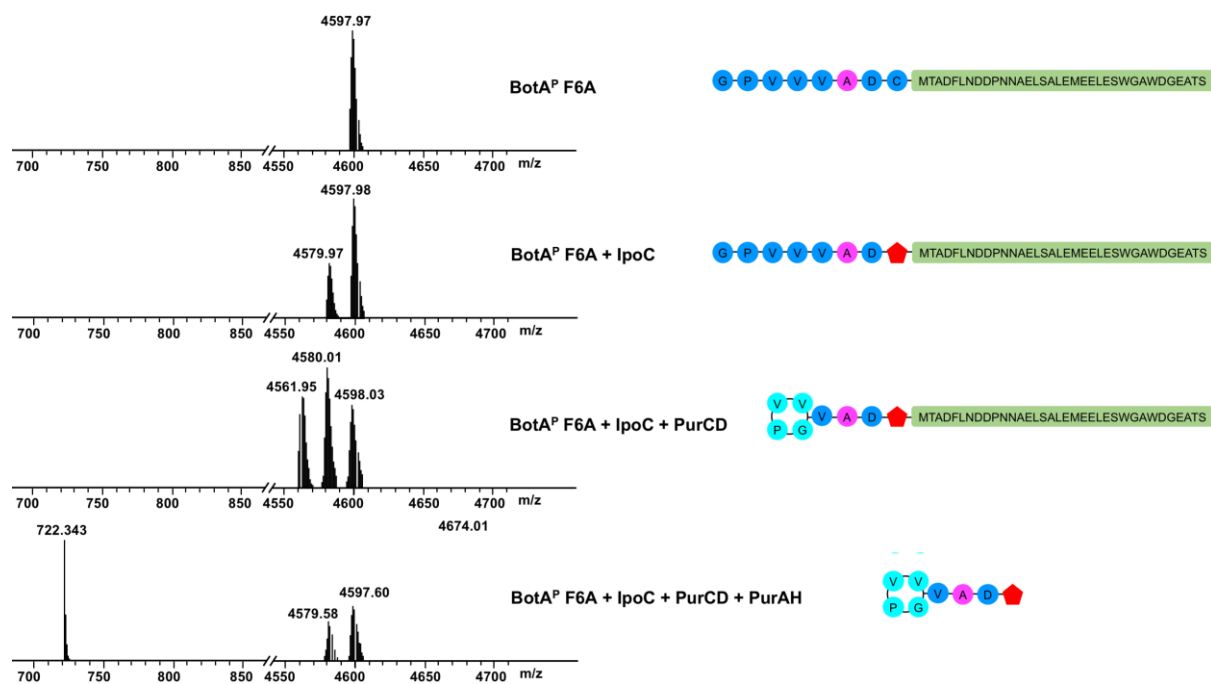


Figure S14: Processing of BotA<sup>P,F6A</sup> by IpoC, PurCD and PurAH. Turnover of BotA<sup>P,F6A</sup> by IpoC is not complete. After addition of PurAH, only small amounts of follower-cleaved and processed BotA are detected. Analysis by LC-MS.

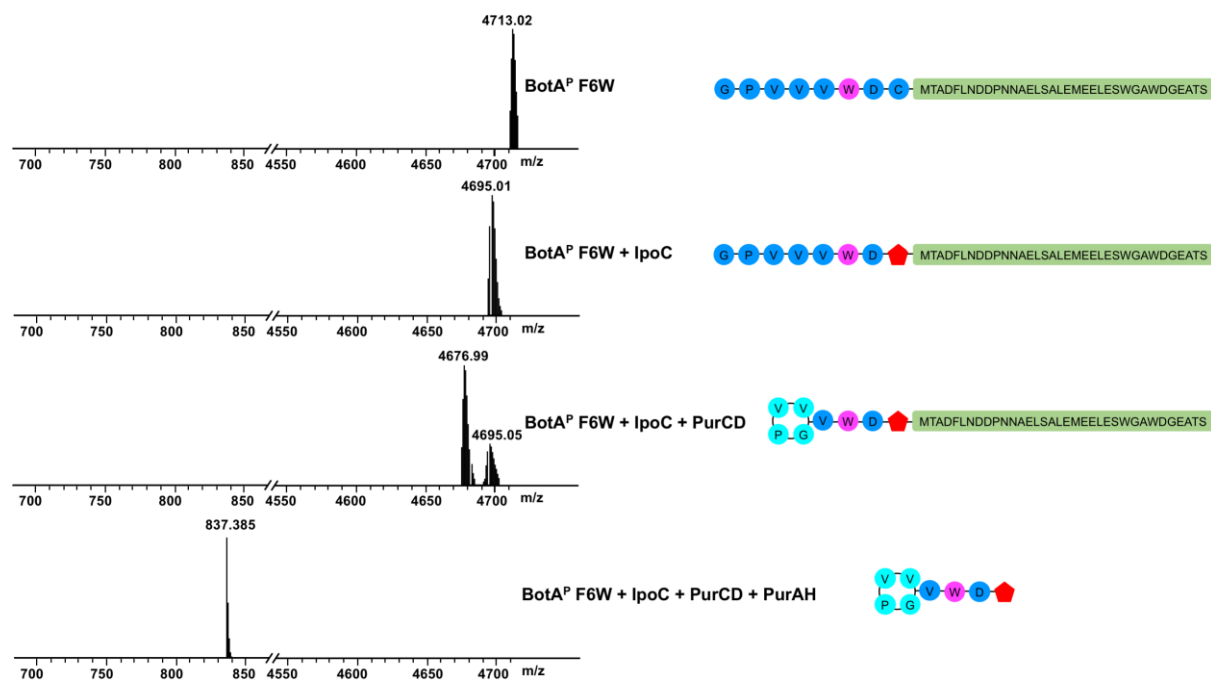


Figure S15: Processing of BotA<sup>P,F6W</sup> by IpoC, PurCD and PurAH. The PurCD reaction is not unidirectional but an equilibrium, in which PurCD catalyzes both, macroamide formation and its reopening. Addition of PurAH cleaves the follower and drives macroamide formation to completion. Analysis by LC-MS.

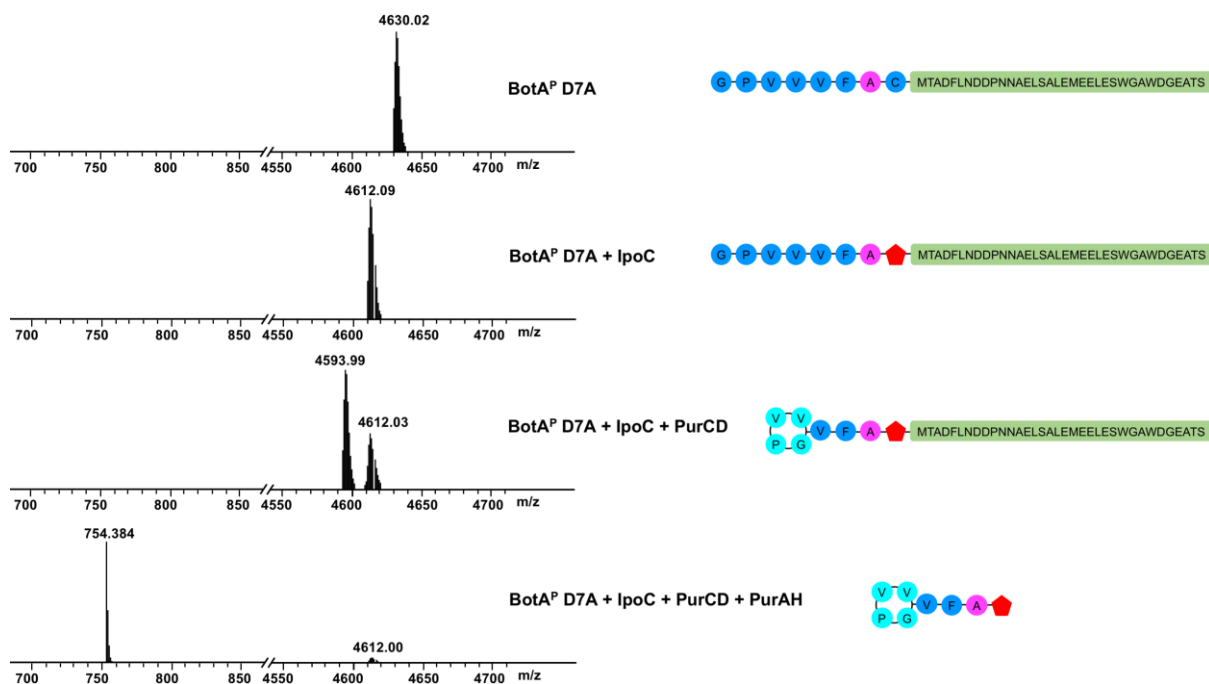


Figure S16: Processing of  $BotA^{P,D7A}$  by IpoC, PurCD and PurAH.  $BotA^{P,D7A}$  is good substrate for IpoC and PurCD (comparable with the wild type  $BotA^P$ ). Addition of PurAH cleaves the follower, but does not drive the macroamidine formation to completion (as seen for the wild type  $BotA^P$ ). The turnover rate of PurAH for the variant D7A seems to be slower than for the wild type. Analysis by LC-MS.

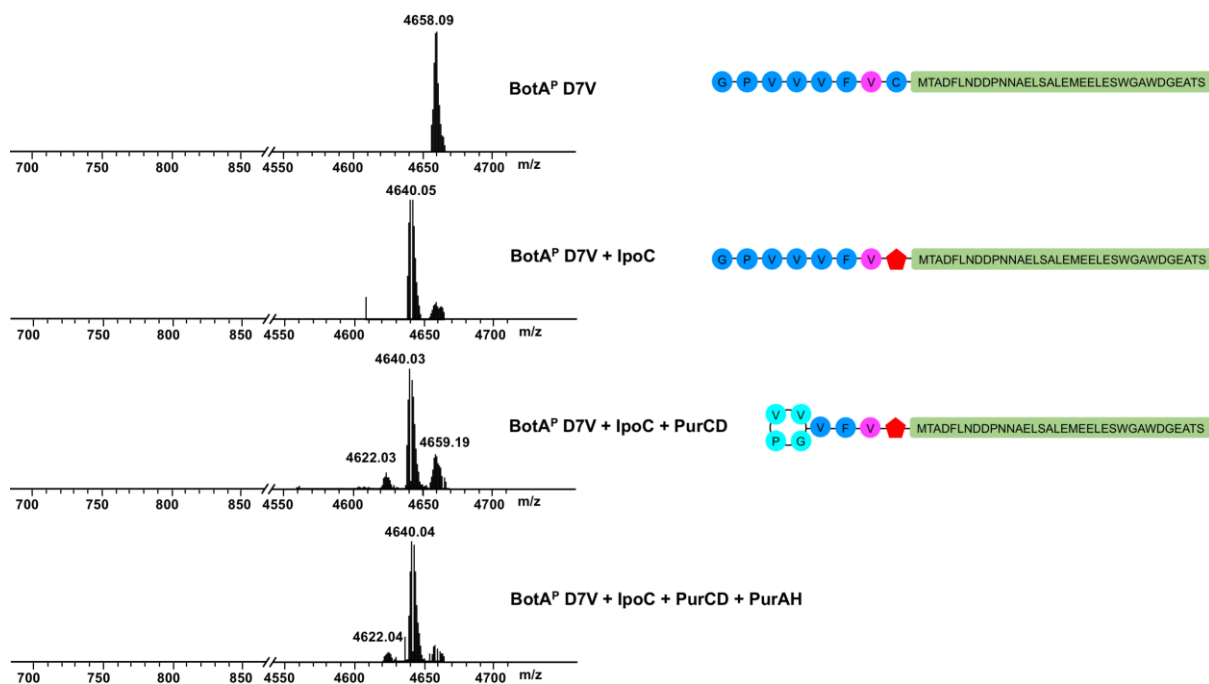


Figure S17: Processing of  $BotA^{P,D7V}$  by IpoC, PurCD and PurAH. Turnover of  $BotA^{P,D7V}$  by IpoC is not complete and the turnover of PurCD is lower than for the wild type peptide. After addition of PurAH, no follower-cleaved product was detected. PurAH seems not to be able to use the D7V variant as substrate. Analysis by LC-MS.

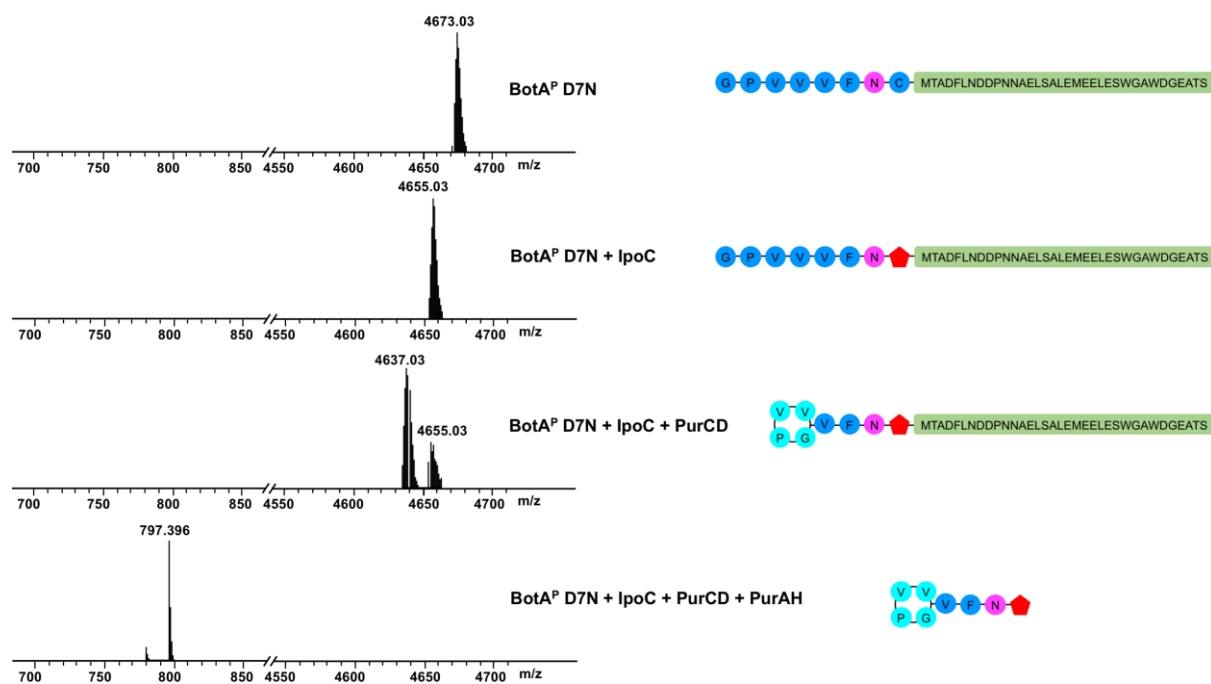


Figure S18: Processing of BotA<sup>P,D7N</sup> by IpoC, PurCD and PurAH. Turnover goes complete when PurAH is added (as seen for the wild-type BotA). Analysis by LC-MS.

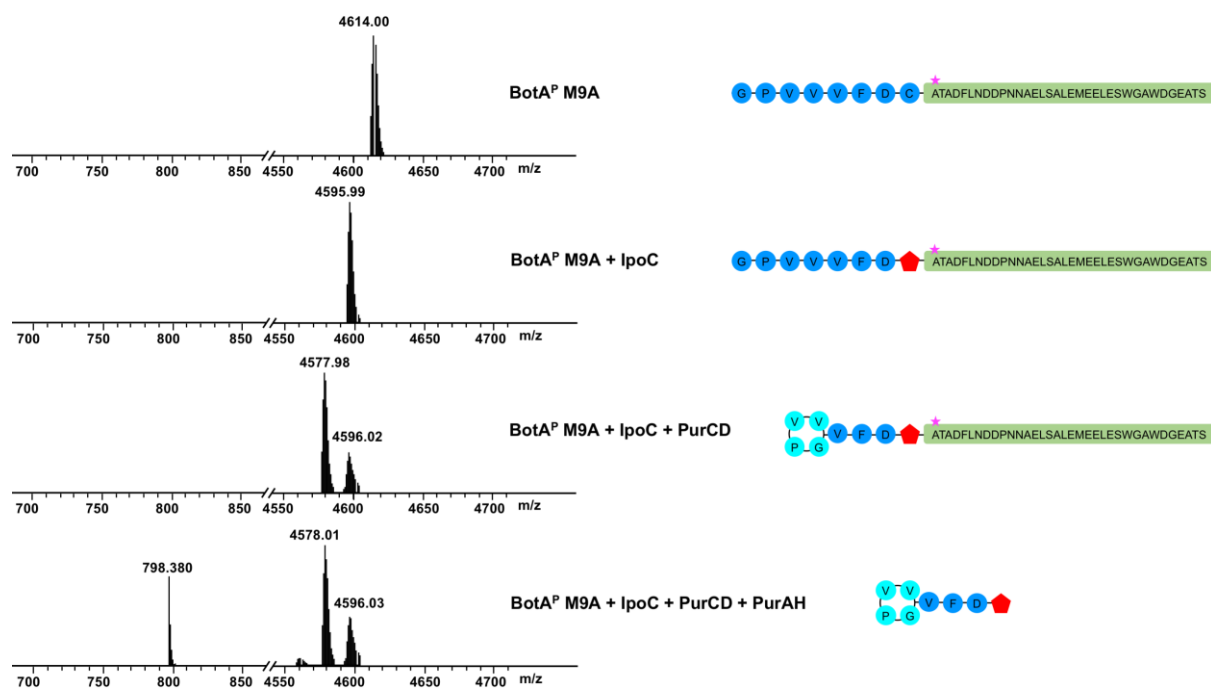


Figure S19: Processing of BotA<sup>P,M9A</sup> by IpoC, PurCD and PurAH. BotA<sup>P,M9A</sup> is a substrate for IpoC and PurCD (comparable to the wild type BotA<sup>P</sup>). When PurAH is added, only small amounts of follower-cleaved peptide are detected. The turnover rate of PurAH for the processed variant is much lower than those for the wild type peptide. Analysis by LC-MS.

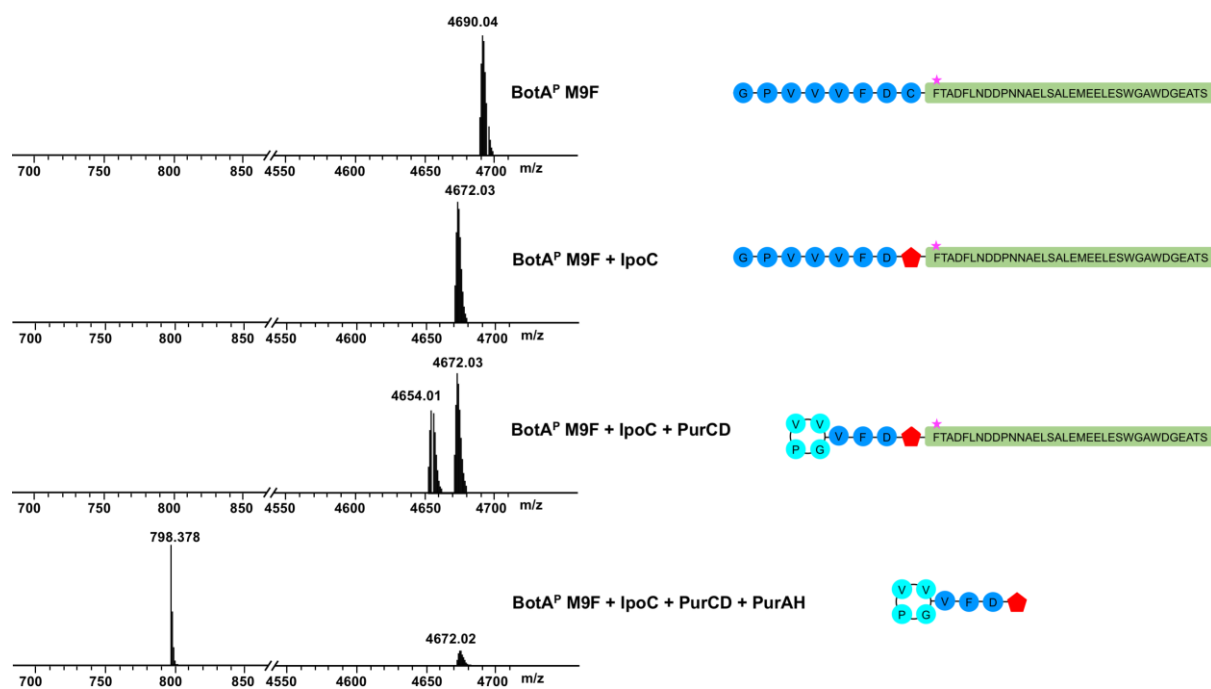


Figure S20: Processing of BotA<sup>P,M9F</sup> by IpoC, PurCD and PurAH. BotA<sup>P,M9F</sup> is a substrate for IpoC and PurCD (comparable to the wild type BotA<sup>P</sup>). Addition of PurAH cleaves the follower and drives macroamidation formation to nearly completion. The turnover rate of PurAH for the processed BotA<sup>P,M9F</sup> is slower than that for the wild type BotA<sup>P</sup>. Analysis by LC-MS.

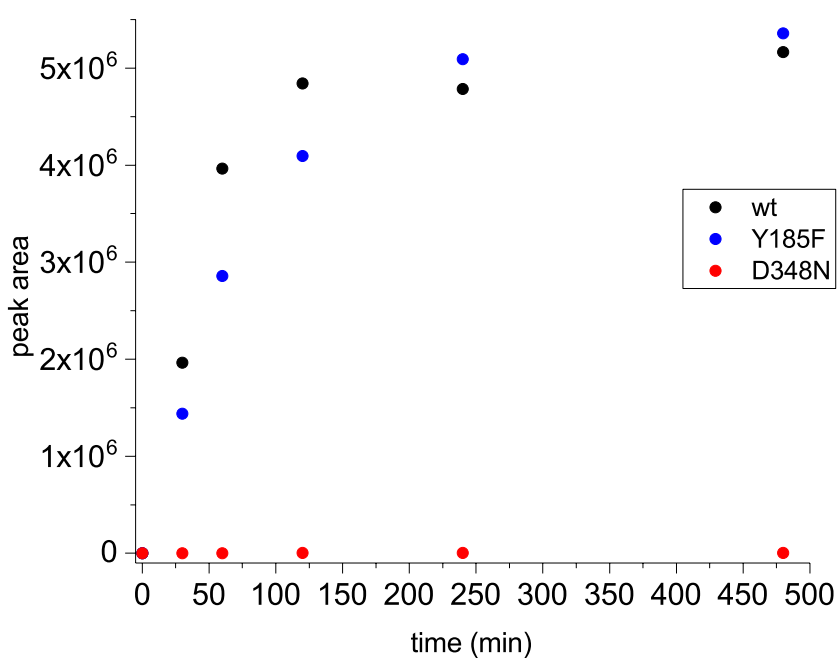


Figure S21: Time dependent formation of 2 (integrated peak areas of [M+H]<sup>+</sup> = 799.3807 ± 0.05 Da) by PurAH wt and mutants (Y185F, D348N) measured by LC-ESI-MS.



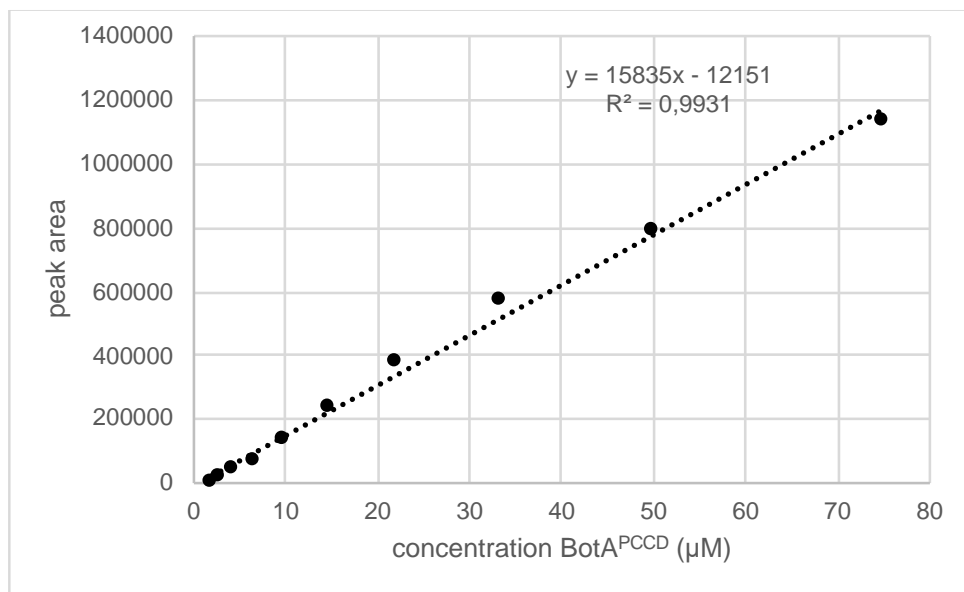


Figure S22: Calibration curve for BotA<sup>PCCD</sup> for the concentration range between 2 and 75 μM. Peak areas for the [M+H]<sup>4+</sup>, [M+H]<sup>5+</sup> and [M+H]<sup>6+</sup> ions (EIC width ±5 ppm) of BotA<sup>PCCD</sup> (see Table S4), measured by LC-ESI-MS, were plotted against the concentration of purified BotA<sup>PCCD</sup> peptide.

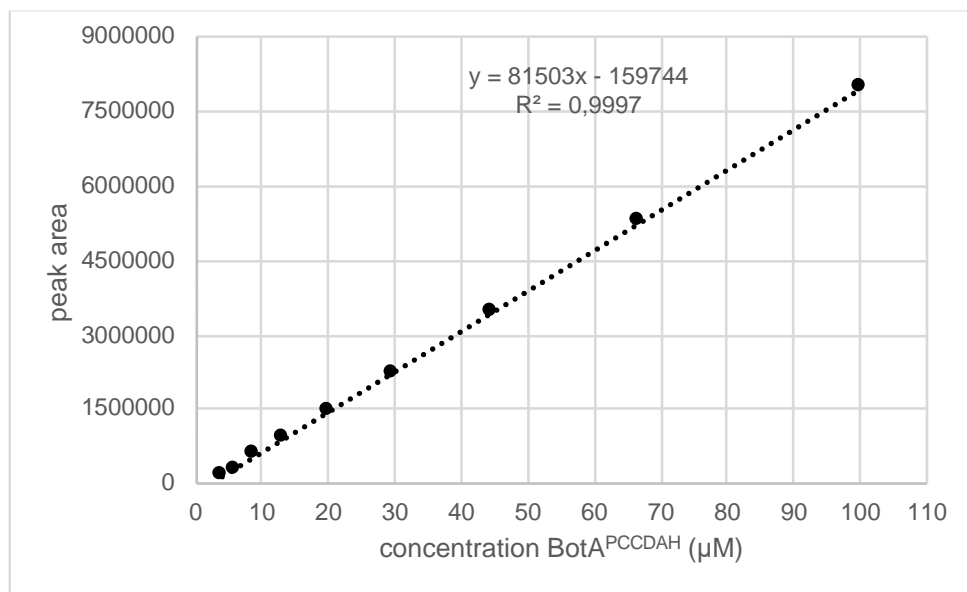


Figure S23: Calibration curve for 2 for the concentration range between 2.6 and 100 μM. The peak area for the [M+H]<sup>+</sup> ion of 2 ([M+H]<sup>+</sup> = 799.3815 ± 0.05 Da), measured by LC-ESI-MS, was plotted against the concentration of purified 2.

Table S1: MS<sup>2</sup> fragmentation data for 1, corresponding to Figure S2.

Seq.	b <sub>n</sub>	Obs. b	Calc. b	y <sub>n</sub>	Obs. y	Calc. y
G	1	-	-	8	817.3898	817.3918
P	2	155.0816	155.0821	7	-	-
V	3	254.1496	254.1605	6	663.3158	663.3176
V	4	353.2178	353.2189	5	564.2475	564.2492
V	5	452.2859	452.2873	4	465.1789	465.1808
F	6	599.3533	599.3557	3	366.1114	366.1124
D <sup>b</sup>	7	-	-	2	219.0432	219.0439
C <sup>b</sup>	8	799.3776	799.3813	1	-	-

b. Cys8 heterocyclization with Asp7 carbonyl

Table S2: MS<sup>2</sup> fragmentation data for 2, corresponding to Figure S2.

Seq.	b <sub>n</sub>	Obs. b	Calc. b	y <sub>n</sub>	Obs. y	Calc. y
G <sup>a</sup>	1	-	-	8	799.3790	799.3807
P	2	155.0816	155.0821	7	-	-
V	3	254.1497	254.1605	6	645.3056	645.3070
V <sup>a</sup>	4	335.2074	335.2083	5	546.2374	546.2386
V	5	434.2757	434.2767	4	465.1798	465.1808
F	6	581.3438	581.3451	3	366.1147	366.1124
D <sup>b</sup>	7	-	-	2	219.0433	219.0439
C <sup>b</sup>	8	781.3688	781.3707	1	-	-

a. Gly1 macrocyclisation with Val4 carbonyl

b. Cys8 heterocyclisation with Asp7 carbonyl

Additionally, fragment a<sub>5</sub> (obs. 406.2809; calc. 406.2818) and b<sub>4</sub><sup>\*</sup> (obs. 363.2023; calc. 363.2032), which is characteristically associated with the bottromycin macrocycle and has been previously observed for bottromycins and macrocyclic peptides, was detected.

Table S3: Data collection and refinement statistics

	PurAH high red	PurAH high res
<b>Data collection</b>		
Space group	C 2 2 21	C 2 2 21
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	100.98 110.93 84.50	100.97 110.73 84.90
$\alpha$ , $\beta$ , $\gamma$ (°)	90.00 90.00 90.00	90.00 90.00 90.00
Resolution (Å)	43.34 - 2.50 (2.60 - 2.50)	74.61 - 1.73 (1.76 - 1.73)
<i>R</i> <sub>merge</sub>	0.129 (0.517)	0.034 (0.693)
<i>I</i> / $\sigma$ <i>I</i>	55.7 (20.2)	18.4 (1.9)
Completeness (%)	99.8 (99.9)	98.1 (99.6)
Redundancy	92.7 (89.8)	4.5 (4.5)
<b>Refinement</b>		
Resolution (Å)		56.04 - 1.73
No. reflections		48843 (4935)
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>		0.1675 / 0.1956
No. atoms		3619
Protein		3403
Ligand/ion		9
Water		207
<i>B</i> -factors		41.04
Protein		40.82
Ligand/ion		51.25
Water		44.25
R.m.s. deviations		
Bond lengths (Å)		0.014
Bond angles (°)		1.10
Molprobit clash score		0.89

\*Values in parentheses are for highest-resolution shell.

Table S4: Calculated masses for the BotA<sup>PC</sup> and BotA<sup>PCCD</sup> (mutant) peptides and [M+H]<sup>4+</sup>, [M+H]<sup>5+</sup> and [M+H]<sup>6+</sup> ions, which were used for the peak area calculation.

peptide	Formula	M <sub>mono.</sub>	[M+H] <sup>4+</sup>	[M+H] <sup>5+</sup>	[M+H] <sup>6+</sup>
BotA <sup>PC</sup> wt	C <sub>202</sub> H <sub>294</sub> N <sub>48</sub> O <sub>73</sub> S <sub>3</sub>	4655.990	1165.006	932.206	777.006
BotA <sup>PCC</sup> wt	C <sub>202</sub> H <sub>292</sub> N <sub>48</sub> O <sub>72</sub> S <sub>3</sub>	4637.983	1160.503	928.604	774.004
BotA <sup>PC</sup> P2A	C <sub>200</sub> H <sub>292</sub> N <sub>48</sub> O <sub>73</sub> S <sub>3</sub>	4629.977	1158.502	927.003	772.670
BotA <sup>PCCD</sup> P2A	C <sub>200</sub> H <sub>290</sub> N <sub>48</sub> O <sub>72</sub> S <sub>3</sub>	4611.967	1156.999	925.801	771.668
BotA <sup>PC</sup> V3L	C <sub>203</sub> H <sub>296</sub> N <sub>48</sub> O <sub>73</sub> S <sub>3</sub>	4670.009	1168.509	935.009	779.342
BotA <sup>PCCD</sup> V3L	C <sub>203</sub> H <sub>294</sub> N <sub>48</sub> O <sub>72</sub> S <sub>3</sub>	4651.998	1164.007	931.407	776.340
BotA <sup>PC</sup> V4L	C <sub>203</sub> H <sub>296</sub> N <sub>48</sub> O <sub>73</sub> S <sub>3</sub>	4670.009	1168.509	935.009	779.342
BotA <sup>PCCD</sup> V4L	C <sub>203</sub> H <sub>294</sub> N <sub>48</sub> O <sub>72</sub> S <sub>3</sub>	4651.998	1164.007	931.407	776.340
BotA <sup>PC</sup> V4A	C <sub>200</sub> H <sub>290</sub> N <sub>48</sub> O <sub>73</sub> S <sub>3</sub>	4627.962	1157.998	926.600	772.334
BotA <sup>PCCD</sup> V4A	C <sub>200</sub> H <sub>288</sub> N <sub>48</sub> O <sub>72</sub> S <sub>3</sub>	4609.951	1153.495	922.998	769.332
BotA <sup>PC</sup> V5A	C <sub>200</sub> H <sub>290</sub> N <sub>48</sub> O <sub>73</sub> S <sub>3</sub>	4627.962	1157.998	926.600	772.334
BotA <sup>PCCD</sup> V5A	C <sub>200</sub> H <sub>288</sub> N <sub>48</sub> O <sub>72</sub> S <sub>3</sub>	4609.951	1153.495	922.998	769.332
BotA <sup>PC</sup> V5E	C <sub>202</sub> H <sub>292</sub> N <sub>48</sub> O <sub>75</sub> S <sub>3</sub>	4685.967	1172.499	938.201	782.002
BotA <sup>PCCD</sup> V5E	C <sub>202</sub> H <sub>290</sub> N <sub>48</sub> O <sub>74</sub> S <sub>3</sub>	4667.957	1167.996	934.599	779.000
BotA <sup>PC</sup> F6W	C <sub>204</sub> H <sub>295</sub> N <sub>49</sub> O <sub>73</sub> S <sub>3</sub>	4695.004	1174.758	940.008	783.508
BotA <sup>PCCD</sup> F6W	C <sub>204</sub> H <sub>293</sub> N <sub>49</sub> O <sub>72</sub> S <sub>3</sub>	4676.993	1170.256	936.406	780.506
BotA <sup>PC</sup> F6A	C <sub>196</sub> H <sub>290</sub> N <sub>48</sub> O <sub>73</sub> S <sub>3</sub>	4579.962	1145.998	917.000	764.334
BotA <sup>PCCD</sup> F6A	C <sub>196</sub> H <sub>288</sub> N <sub>48</sub> O <sub>72</sub> S <sub>3</sub>	4561.951	1141.495	913.398	761.332
BotA <sup>PC</sup> D7A	C <sub>201</sub> H <sub>294</sub> N <sub>48</sub> O <sub>71</sub> S <sub>3</sub>	4612.003	1154.01	923.408	769.674
BotA <sup>PCCD</sup> D7A	C <sub>201</sub> H <sub>292</sub> N <sub>48</sub> O <sub>70</sub> S <sub>3</sub>	4593.993	1149.505	919.806	766.673
BotA <sup>PC</sup> D7N	C <sub>202</sub> H <sub>295</sub> N <sub>49</sub> O <sub>72</sub> S <sub>3</sub>	4655.009	1164.760	932.009	776.842
BotA <sup>PCCD</sup> D7N	C <sub>202</sub> H <sub>293</sub> N <sub>49</sub> O <sub>71</sub> S <sub>3</sub>	4636.999	1160.257	928.407	773.840
BotA <sup>PC</sup> D7V	C <sub>203</sub> H <sub>298</sub> N <sub>48</sub> O <sub>71</sub> S <sub>3</sub>	4640.035	1161.016	929.014	774.346
BotA <sup>PCCD</sup> D7V	C <sub>203</sub> H <sub>296</sub> N <sub>48</sub> O <sub>70</sub> S <sub>3</sub>	4622.024	1156.513	925.412	771.345
BotA <sup>PC</sup> M9A	C <sub>200</sub> H <sub>290</sub> N <sub>48</sub> O <sub>73</sub> S <sub>2</sub>	4595.990	1150.005	920.205	767.006
BotA <sup>PCCD</sup> M9A	C <sub>200</sub> H <sub>288</sub> N <sub>48</sub> O <sub>72</sub> S <sub>2</sub>	4577.979	1145.502	916.603	764.004
BotA <sup>PC</sup> M9F	C <sub>206</sub> H <sub>294</sub> N <sub>48</sub> O <sub>73</sub> S <sub>2</sub>	4672.021	1169.013	935.411	779.677
BotA <sup>PCCD</sup> M9F	C <sub>206</sub> H <sub>292</sub> N <sub>48</sub> O <sub>72</sub> S <sub>2</sub>	4654.010	1164.510	931.809	776.676

Table S5: Nucleotide sequences used for the site-directed mutagenesis (SDM) of PurAH (TS: top strand; BS: bottom strand).

Primer name	Primer sequence (5' to 3')
PurAH Y185F TS	GGCATCAAGGTTTTTCGACGACGTTGAG
PurAH Y185F BS	CTCAACGTCGTCGAAAACCTTGATGCC
PurAH D348N TS	CTGGTCGGTACCAACTGTGGCAACCCG
PurAH D348N BS	CGGGTTGCCACAGTTGGTACCGACCAG

# **Chapter 3: The Post-translational Amino Acid Epimerase BotH Defines a New Group of Atypical alpha/beta- hydrolase-fold Enzymes**

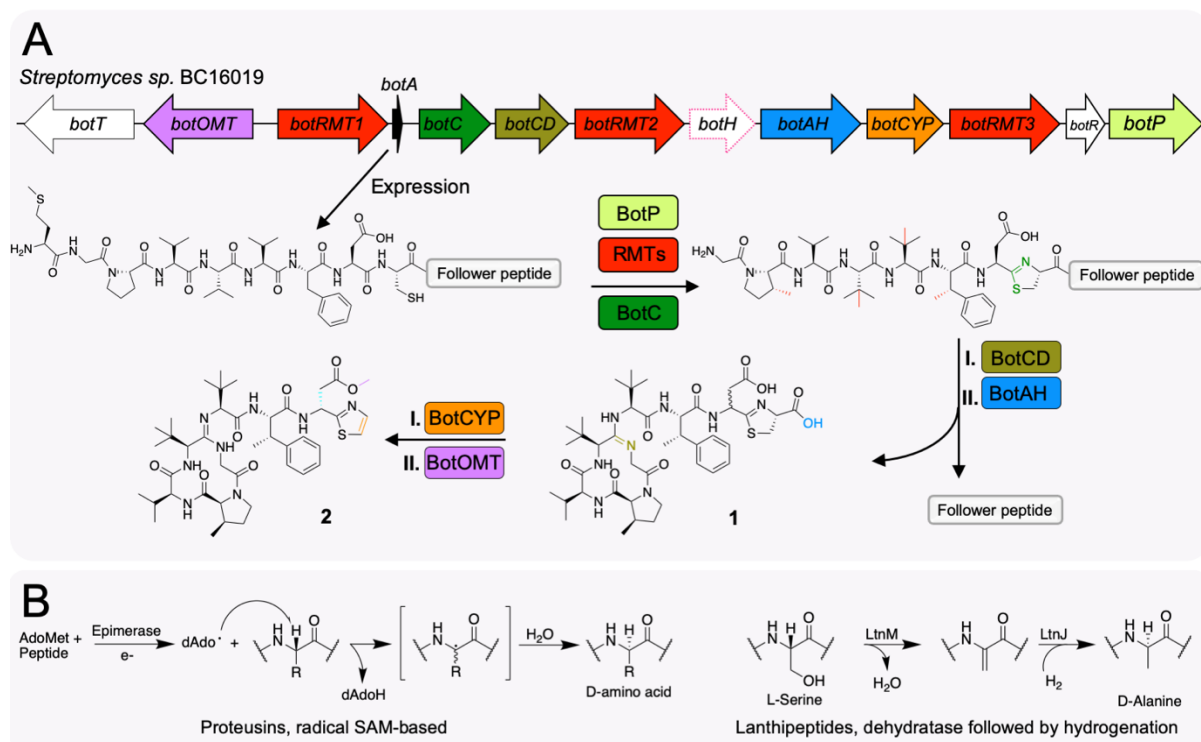
*Asfandyar Sikandar, Laura Franz, Sebastian Adam, Javier Santos-Aberturas, Liliya Horbal,*

*Andriy Luzhetskyy, Andrew W. Truman, Olga V. Kalinina, Jesko Köhnke*

# 1 Introduction

Enzymes belonging to the superfamily of alpha/beta-hydrolase-fold proteins (ABHs) are found in all domains of life<sup>1-3</sup>. Their catalytic roles are highly diverse and they participate in primary and secondary metabolism, where they are usually responsible for the hydrolysis of (thio)ester and peptide bonds<sup>4</sup>. In addition, ABHs have also been reported to function as dehalogenases, epoxide hydrolases, dioxygenases, decarboxylases and haloperoxidases<sup>3,4</sup>. Despite low overall sequence conservation, ABH enzymes share a remarkably conserved core fold<sup>2,3</sup> that has a V-shaped lid domain above the active-site as a frequent addition<sup>5,6</sup>. A majority of ABH family members possesses a canonical Ser, His, Asp catalytic triad at the active-site, but other catalytic residues have also been observed<sup>3,4</sup>. In addition to catalytic roles, a ABH family members have also been reported to fulfil several other important functions, including small-molecule receptors that are involved in signal transduction, cell-cell interaction and channel regulation<sup>5-8</sup>. The ABH fold is thus a prime example for the reappropriation of a conserved core fold during evolution to fulfill a myriad of functions.

An aberrant ABH protein, BotH, is encoded in the biosynthetic gene cluster (BGC) for bottromycins<sup>9-12</sup>. This natural product family was first isolated from the terrestrial bacterium *Streptomyces bottropensis* and originally described as peptidic antibiotics with potent activity against Gram-positive bacteria<sup>13,14</sup>. Bottromycins are active against problematic human pathogens, such as Methicillin-resistant *Staphylococcus aureus*<sup>15,16</sup>, address a novel target (the A-site of the prokaryotic ribosome)<sup>17-19</sup> and belong to the fast growing superfamily of ribosomally synthesized and post-translationally modified peptides (RiPPs)<sup>9-12</sup>. As is typical for RiPPs, their biosynthesis begins with the expression of a small structural gene to yield the precursor peptide (PP)<sup>20</sup>. Uniquely amongst bacterial RiPPs, the bottromycin PP contains an N-terminal core peptide (the eventual natural product) and a C-terminal follower peptide<sup>9-12</sup>, which is important for substrate recognition by several of the biosynthetic enzymes<sup>21,22</sup>. The order of biosynthetic steps (and responsible enzymes) has since been proposed based on an untargeted metabolomics approach using mass spectral networking<sup>23</sup>. Subsequent *in vitro* work has largely corroborated the metabolomics data<sup>21,22,24,25</sup>. In the first phase of bottromycin biosynthesis the N-terminal methionine is removed, proline, valine and phenylalanine residues of the core peptide are C-methylated and a cysteine-derived thiazoline is installed in no particular order<sup>23</sup>. The hallmark macroamidine linkage is then formed in this intermediate<sup>21,22</sup>, which is followed by proteolytic removal of the follower peptide to yield **1** (Figure 1A)<sup>25</sup>.



**Figure 1:** A Bottromycin BGC found in *S. sp.* BC16019. After expression of the precursor peptide *BotA*, its N-terminal methionine is removed by *BotP*, three radical methyl transferases (RMT) perform four C-methylations and *BotC* installs a cysteine-derive thiazoline. This process appears to follow no particular order. Next, *BotCD* catalyzes formation of the macroamidine, after which *BotAH* removes the follower peptide to yield **1**. Oxidative decarboxylation (*BotCYP*) and O-methylation complete the biosynthesis of bottromycin A2 (**2**). Epimerization was proposed to occur spontaneously. Genes, enzymes and modifications have matching colors. White arrows represent genes for which no or a regulatory function have been proposed. The unusual ABH protein encoded in the pathway, *BotH*, is highlighted in pink. **B** Epimerization in RiPP biosynthesis reported to date. In proteusins, including polytheonamides, a radical SAM enzyme epimerizes a range of amino acids. In lanthipeptide biosynthesis, L-Ser can be converted to D-Ala in a two-step process.

To complete the biosynthesis of bottromycin A2 (**2**, Figure 1A), **1** undergoes epimerization, oxidative decarboxylation and O-methylation. While the latter two modifications have been attributed to specific enzymes, epimerization of the L-Asp of **1** was observed to progress spontaneously, albeit very slowly<sup>23</sup>.

Amino acid epimerization in non-ribosomal peptide synthesis is usually catalyzed by epimerization domains embedded within the assembly line that function on carrier protein-bound aminoacyl substrates<sup>26</sup>. Due to their ribosomal origin, RiPPs must undergo post-translational epimerization after the PP has been expressed as an all L-amino acid peptide. To date, only two enzymatic mechanisms for this process have been described in RiPPs, involving either a radical-SAM enzyme<sup>27-29</sup> or a two-step dehydration-hydrogenation process to generate D-alanine from L-serine<sup>30,31</sup>.

Here, we report the identification of *BotH*, an unusual ABH enzyme from the bottromycin BGC, as the epimerase of the biosynthetic pathway. This is the first reported instance of an ABH enzyme catalyzing peptide epimerization and thus expands the catalytic scope of this vast



enzyme family. Biochemical data together with the structure of the BotH–substrate complex allowed us to propose a mechanism for this reaction. Interestingly, BotH is also able to bind bottromycins with high affinity, which hints at additional function(s) in the biosynthetic process. We show that all canonical ABH active-site residues required for hydrolase activity are absent in BotH, and bioinformatic analyses indicate that BotH homologs with comparable non-hydrolytic residues are widespread amongst BGCs and may catalyze similar biosynthetic steps.

## 2 Results and Discussion

### 2.1 BotH is an unusual member of the ABH superfamily

In a search for an enzyme that may catalyze the epimerization in the bottromycin pathway, we noticed that the gene encoding for BotH had been annotated as an ABH, but that the predicted active-site residues (Ser/His/Asp) were not present and there were no prior experimental data on the role of this protein. We therefore expressed, purified and crystallized BotH (see Materials and Methods for details). A high-resolution (1.18 Å) native BotH dataset was collected from a crystal belonging to space group I222, which was phased using seleno-methionine BotH data (data collection and refinement statistics for all structures can be found in Table S1). The crystals contained one protomer in the asymmetric unit and the electron density for residues 10 – 262 was continuous in the refined model. BotH is comprised of the prototypical ABH core structure (Figure S1) and the putative active site was covered by a V-shaped loop consisting of four  $\alpha$ -helices (Figure S1). A search for similar structures using the DALI server<sup>32</sup> revealed 3-oxoadipate-enol-lactonase (PDB 2xua)<sup>33</sup> as the closest structural homolog (Figure S1). A comparison of the two proteins revealed that in BotH, the active-site Ser has been mutated to a Phe, which is part of a Phe-Phe motif that spans a large, hydrophobic plane at the active site (Figure S1). The remaining residues of the catalytic triad are either mutated (His to Ile) or missing (Asp) (Figure S1). In spite of these mutations, the sizeable cavity found in this region of the structure appeared to be large enough for binding **1**.

### 2.2 BotH catalyzes the epimerization of **3b** and **3a**

To confirm this hypothesis, a de-methyl analog of **1**, **3**, was enzymatically produced as reported previously (Figures 2A and S2)<sup>25</sup>. Careful analysis revealed **3** to exist as an epimeric mixture

of **3a** (D-Asp) and **3b** (L-Asp) (Figures 2A and S3 – S5). When 20  $\mu\text{M}$  **3a/b** was incubated with 5  $\mu\text{M}$  BotH and the reactions were analyzed by high-resolution liquid chromatography–mass spectrometry (HR-ESI-LCMS), we observed a change in **3a** : **3b** ratios (Figure 2B), which was BotH-concentration dependent within the time-scale of the experiment (Figure 2C): Increasing the BotH concentration resulted in a shift to **3a** (D-Asp), which is the required epimer to proceed with biosynthesis<sup>34</sup>. To probe if only **3b** or both epimers were substrates, we incubated **3a/b** with BotH in  $\text{D}_2\text{O}$ . This resulted in rapid ( $< 60$  s) deuterium incorporation at the Asp6 position in both peaks, while very little deuterium incorporation was observed in the absence of BotH even after 24 hours (Figures 2D, S4 and S6). In fact, BotH concentrations low enough to leave epimer ratios unchanged still resulted in accelerated deuterium incorporation (Figure S6). Repeating this experiment using the deuterated sample in  $\text{H}_2\text{O}$  showed an equally rapid exchange with solvent protons back to **3a/b** (Figure S6). These data implied that BotH accepts both, **3a** and **3b**, as substrates in a reversible reaction, while favoring D-Asp (**3a**) as the product.

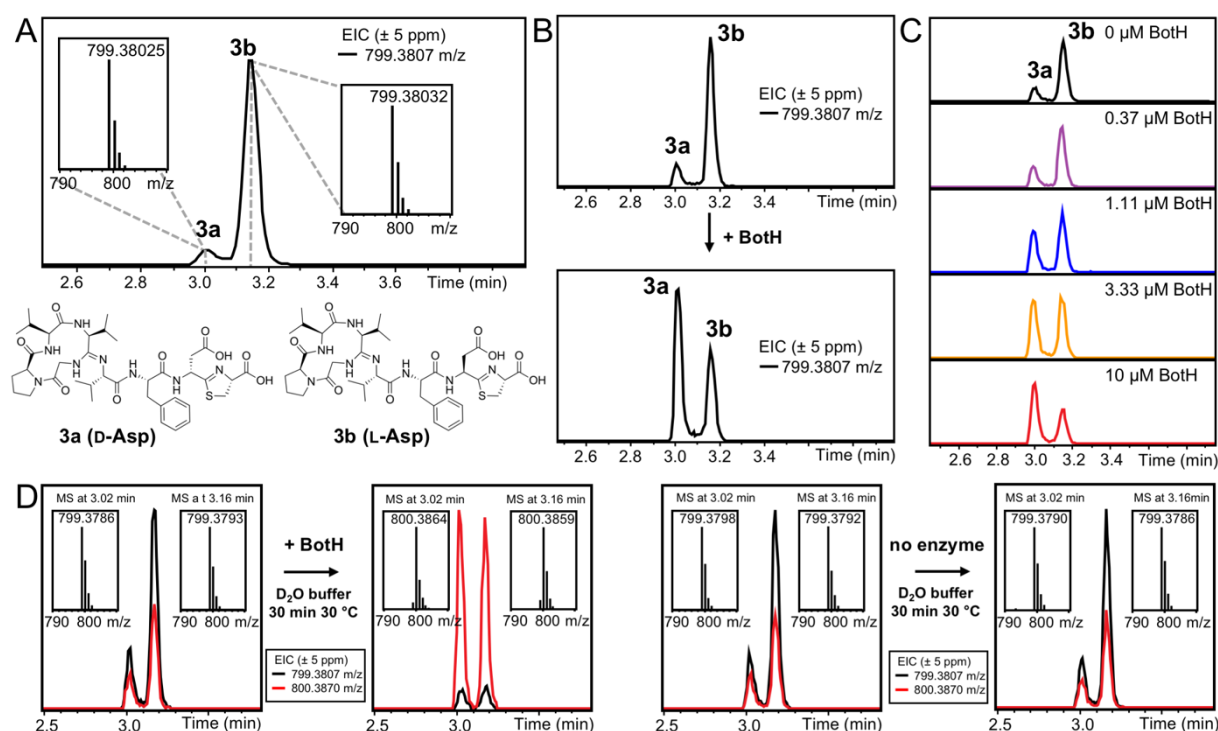


Figure 2: A Spontaneous epimerization of the Asp in position 7 after proteolytic removal of the follower peptide. Marfey's reagent was used to assign the stereochemistry of the Asp for both peaks (Figure S5). Extracted ion chromatograms (EIC) of **3** ( $[M+H]^+$  calc.=799.3807;  $\pm 5$  ppm) and mass spectra at 3.00 and 3.15 min are shown. B Incubation of **3a/b** with BotH results in a change of **3a** : **3b** ratios with **3a** (D-Asp) now the more abundant species. EICs of **3** ( $[M+H]^+$  calc.=799.3807;  $\pm 5$  ppm) are shown. C Incubating 20  $\mu\text{M}$  **3a/b** with increasing concentrations of BotH led to a shift of the equilibrium towards **3a**. Increasing BotH concentrations beyond 10  $\mu\text{M}$  did not lead to a further shift of epimer ratios. EICs of **3** ( $[M+H]^+$  calc.=799.3807;  $\pm 5$  ppm) are shown. D Deuterium incorporation into **3a/b** by BotH in  $\text{D}_2\text{O}$  buffer. EICs for **3** ( $[M+H]^+$  calc.=799.3807;  $\pm 5$  ppm) (black) and deuterium incorporated **3** ( $[M+H]^+$  calc.=800.3870;  $\pm 5$  ppm) (red), as well as mass spectra at the EIC maxima are shown. Representative experiments were repeated independently three times with similar results.

### 2.3 Structure of the BotH-3a complex

To better understand the mechanism of this intriguing enzyme, the high-resolution crystal structure of BotH in complex with its substrate **3a/b** was determined to 1.25 Å resolution. The overall structure of the complex was virtually unchanged when compared to the apo structure ( $C_{\alpha}$  rmsd of 0.12 Å) (Figure S7). The substrate is curled into the active site in a way that places the thiazoline underneath the four amino acid macrocycle and the substrate engages in extensive hydrophobic interactions as well as inter- and intramolecular hydrogen bonds (Figures 3A and S7). Of particular note are hydrogen bonds of the thiazoline carboxy group with the backbone NH of Val41 and Phe110. These two BotH residues are in the position of the oxyanion hole found in many hydrolases. The carboxy group of the substrate's Asp7 is involved in a hydrogen bonding network with ordered water molecules that ultimately link this side-chain to BotH residues (Figure S7). An intramolecular hydrogen bond links the substrate's carbonyl of Val3 with the backbone NH of Asp7. The best fit to the electron density is achieved by choosing the amidine resonance structure that places the double-bond inside the macrocycle and a D-Asp in position 7, which indicates that the ligand observed in the complex structure is **3a** (Figure S7). Since the mechanism would, at least formally, involve proton abstraction and addition, we scanned the complex structure for potential catalytic residues within 4 Å of the  $C_{\alpha}$  hydrogen of Asp7, the site of catalysis, but could not identify any. This left two possibilities: bulk solvent or the side chain of substrate Asp7; its carboxy group is within 2.2 Å of the  $C_{\alpha}$  hydrogen of interest. To probe the importance of the side chain identity in position seven for epimerization, we first tested mutant substrate Asp7Ala and observed no epimerization (Figure S8). Even the most conservative mutation, Asp7Asn, also abolished epimerization (Figure S8). These data imply that the carboxy group of Asp7 is essential for catalysis and that the BotH reaction may be an example of substrate-assisted catalysis.

### 2.4 BotH has relaxed substrate specificity

We had previously reported that the enzymes used to generate **3a/b** possess relaxed substrate specificities for core-peptide residues 2 – 7 (Figure S9)<sup>25</sup>. Since an Asp7 was essential for BotH activity, we wondered if BotH tolerates mutations in the remaining positions. To this end, a series of 13 BotA core-peptide mutants were used to produce the variant BotH substrates enzymatically. These substrates were then incubated with or without BotH and analyzed by LC-MS (Figures 3B and S10 - 11). All but two mutations in positions two to six were processed by

BotH. The positions of Val5 and Phe6 are intimately connected and it appeared that the orientation of the Phe6 side-chain is critical for epimerization. In agreement with this hypothesis, Phe6Ala could not be epimerized by BotH while Phe6Tyr and Phe6Trp were substrates. The side-chain of position 5 (Val) is engaged in hydrophobic interactions with the side-chains of Val4 and Phe6. Accordingly, Val5Thr is not a BotH substrate, while Val5Ala, Val5Leu and Val5Glu can be epimerized by the enzyme. The ability of BotH to process Val5Glu may appear surprising, but modeling of this substrate shows that the C $\beta$  methylene of Glu5 can engage in hydrophobic interactions with Val4 and Phe6, while the C $\gamma$  methylene and terminal carboxy group point towards bulk solvent. Our data demonstrate that BotH is able to process a variety of substrates, which will be invaluable for the production of bottromycin derivatives.

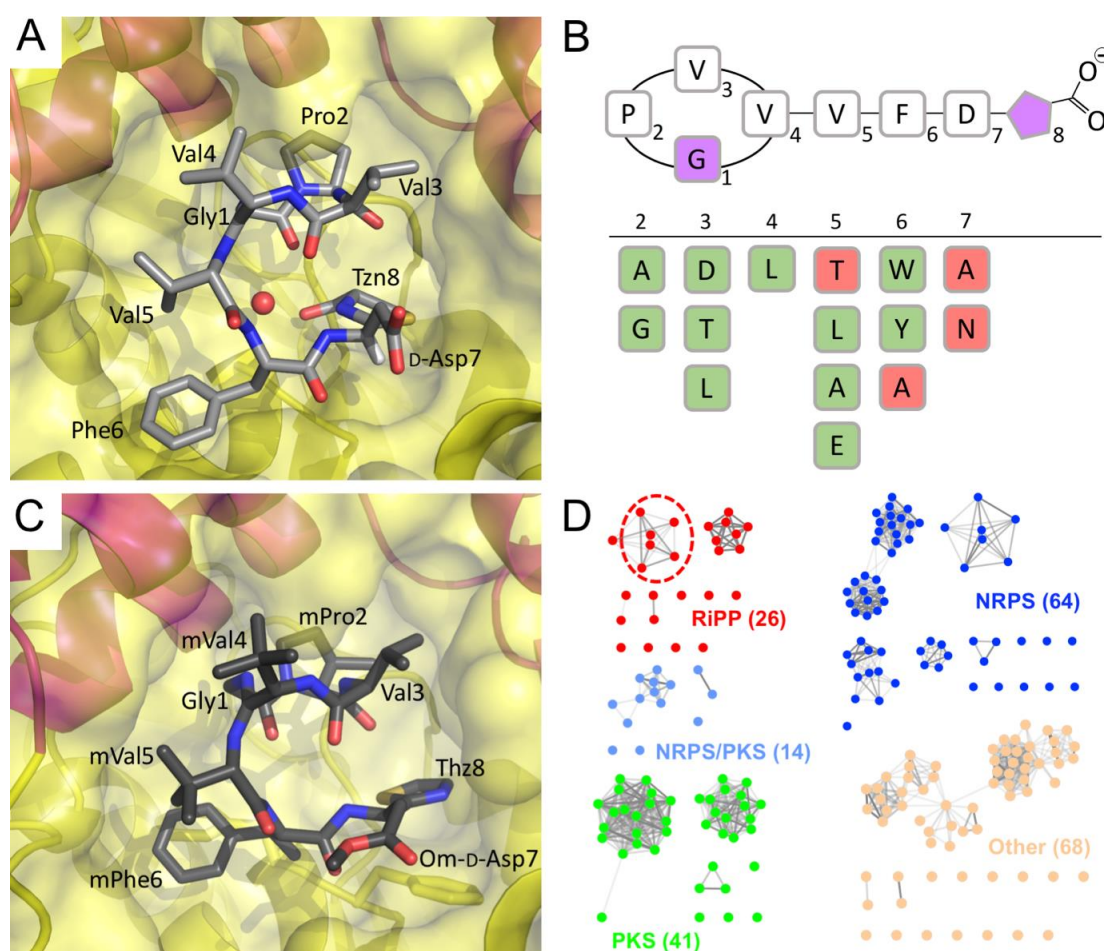


Figure 3: A Close-up of 3a (gray sticks, labeled. Tzn = Thiazoline) bound in the BotH active site (cartoon, yellow/magenta with semi-transparent surface representation). An ordered water molecule trapped between substrate and protein is shown as a red sphere. The Asp7 C $\alpha$  hydrogen is shown as a white stick for clarity. B Summary of BotA point mutants tested as BotH substrates. Purple positions cannot be varied (pentagon represents thiazoline). Accepted mutations are highlighted in green, red mutations cannot be processed by BotH. The accompanying HR-LCMS data can be found in Figures S10 – S11. C Same as A, but bottromycin A2 (dark gray sticks, labeled) bound to the active site of BotH. m = methylated residue, Om = O-methylated residue, Thz = Thiazole. D Sequence similarity network of ABHs homologous to BotH. Of the 213 genes 145 could be assigned to BGCs representing the three large bacterial natural product superfamilies (darker edges represent higher sequence similarity, NRPS = non-ribosomal peptide synthetase; PKS = polyketide synthase).

## 2.5 BotH binds bottromycins, which act as inhibitors of epimerization

Since **3a/b** are close structural homologs of bottromycin A2 (**2**), it seemed feasible that BotH may be able to bind to bottromycins. We thus performed microscale thermophoresis (MST) experiments using heterologously produced **2** and three closely related variants (**4** – **6**, Figure S12). BotH was able to bind all four bottromycins with  $K_D$  values in the high nM to low  $\mu$ M range (Figure S12). To understand the mode of binding, we determined the crystal structures of three bottromycin–BotH complexes to 1.40 (**2**), 1.70 (**5**), and 1.48 (**6**) Å resolution, respectively (Figures 3C and S13 – 14). These also represent the first crystal structures of any bottromycin. As observed for the complex with **3a**, the overall structural change in BotH due to ligand binding was minimal ( $C_\alpha$  RMSD < 0.2 Å) and the bottromycins bound in a similar manner as **3a** (Figures 3A and 3C). Despite the high resolution, it is not obvious which way the thiazole is flipped, since the loss of the carboxy group due to thiazoline oxidation allows a fit of both rotamers without inducing a clash with BotH (Figure S15). The bottromycins themselves are oriented in a twisted fashion that results in several tight, intramolecular hydrogen bonds (Figure S14). Compared to the published NMR structure<sup>35</sup>, bottromycin A2 experiences significant strain as a result of binding to BotH as it is forced to adopt a horseshoe shape with Val5 and Phe6 at its apex and the C-terminal thiazole stacked parallel under the macrocycle (Figure S16). As expected, bottromycin A2 acts as an orthosteric inhibitor of epimerization (Figure S17), which suggests that BotH may be involved in a biosynthetic feedback mechanism to prevent self-poisoning of the producing strain (Figure 4A).

In this model, an increase in the intracellular bottromycin concentrations offers a direct and faster means to reduce bottromycin production than altered gene expression since epimerization of Asp7 is highly important for the activity of the succeeding enzyme BotCYP<sup>34</sup>. In addition, sequestration of mature bottromycin may also aid self-immunity. The *S. scabiei* orthologues of BotH and the putative bottromycin exporter BotT were therefore expressed (independently and in combination) in a strain sensitive to bottromycin and then exposed to increasing concentrations of bottromycin. Expression of BotT resulted in slightly increased resistance and resistance was significantly amplified when both proteins were co-expressed (Figure S18). These data imply that BotH has a secondary role in conferring self-immunity to bottromycin by intracellular sequestration of mature bottromycin. In addition, BotH may directly associate with BotT and aid antibiotic export, although no association was observed in two-hybrid assays (Data not shown).

The ability of BotH to bind both the substrate and the mature natural product with high affinity may serve as a cautionary tale for attempting to identify the binding partners of “non-catalytic” ABHs. When using BotH to capture its ligand/substrate from either supernatant or lysate of a bottromycin producing strain, we were only able to detect bottromycin A2, but not the actual substrate **1** (Figure S19). Since biosynthetic pathway products tend to be present at higher concentrations than pathway intermediates, careful analysis of the biosynthetic pathway supplying the ligand is required to exclude additional, non-canonical catalytic function(s) of the ABH under investigation.

## 2.6 Evolution of BotH and its distribution amongst bacterial BGCs

ABHs comprise one of the largest protein families (~500,000 protein matches in InterPro<sup>36</sup>, ~400,000 of them in Bacteria) and the Ser-His-Asp catalytic triad is surprisingly poorly conserved: in 213,273 bacterial proteins from the InterPro<sup>36</sup> family IPR000073 at least one of the catalytic residues is mutated or missing. In most cases (166,921 sequences), His is missing; and in the majority of these proteins (156,115 sequences) Asp is missing as well. Ser is mutated in only 60,205 sequences, most frequently to an alanine, and completely missing in 4,625 sequences. Using these data, we identified 3,743 proteins that are unlikely to have hydrolase activity (see Materials and Methods section for details), including BotH. Taxonomic and phylogenetic analyses revealed that these non-functional hydrolases are widespread among both Gram-positive and Gram-negative bacteria (Figure S20).

From these sequences, we then selected those that were in or near (less than 1 kb away) BGCs, which resulted in 213 proteins that were used to build a sequence–similarity network (Figures 3D and S21). In this network, all BotH homologs from bottromycin BGCs cluster together. Among the identified BGCs, two other major natural product superfamilies can be identified in addition to RiPPs: NRPS (non-ribosomal peptide synthetase) and PKS (polyketide synthase) clusters.

## 3 Conclusion

We have identified the enzyme BotH as the epimerase of the bottromycin BGC, which is selective for Asp, but promiscuous with regards to mutations at other positions of the substrate. Based on the biochemical and structural data, we propose the following mechanism for epimerization in bottromycin biosynthesis (Figure 4B): Cleavage of the follower peptide converts the 2-thiazoline residue into a 2-thiazoline-4-carboxy moiety at the carboxy terminus of **3b**, which is bound by BotH. Interestingly, the two BotH residues that form hydrogen bonds

with the thiazoline carboxy group are in the position of the canonical oxyanion hole found in most ABHs. By binding, **3b** traps an ordered water molecule within hydrogen-bonding distance of the thiazoline's nitrogen and carboxy group. The side-chain of Asp7 is positioned such that it may serve as a base to abstract the C $\alpha$  proton from itself, which triggers enamine formation and leads to proton transfer from the ordered water molecule to the thiazoline nitrogen. The resulting hydroxide is still hydrogen bonded to the thiazoline's carboxy group and the thiazoline nitrogen proton. Abstraction of this proton by the hydroxide triggers reprotonation of the enamine by the side chain of Asp7 and results in epimerization. The inability of BotH to epimerize substrate with an Asp7Asn mutation suggests an essential role of the carboxylic acid side chain and thus substrate-assisted catalysis. As demonstrated by our hydrogen-deuterium-hydrogen exchange experiments, this reaction is fully reversible. It appears that lowering of the energetic barrier for epimerization by BotH is sufficient to supply the succeeding enzyme with sufficient substrate for complete turnover, even without changing the **3a** : **3b** ratios<sup>34</sup>.

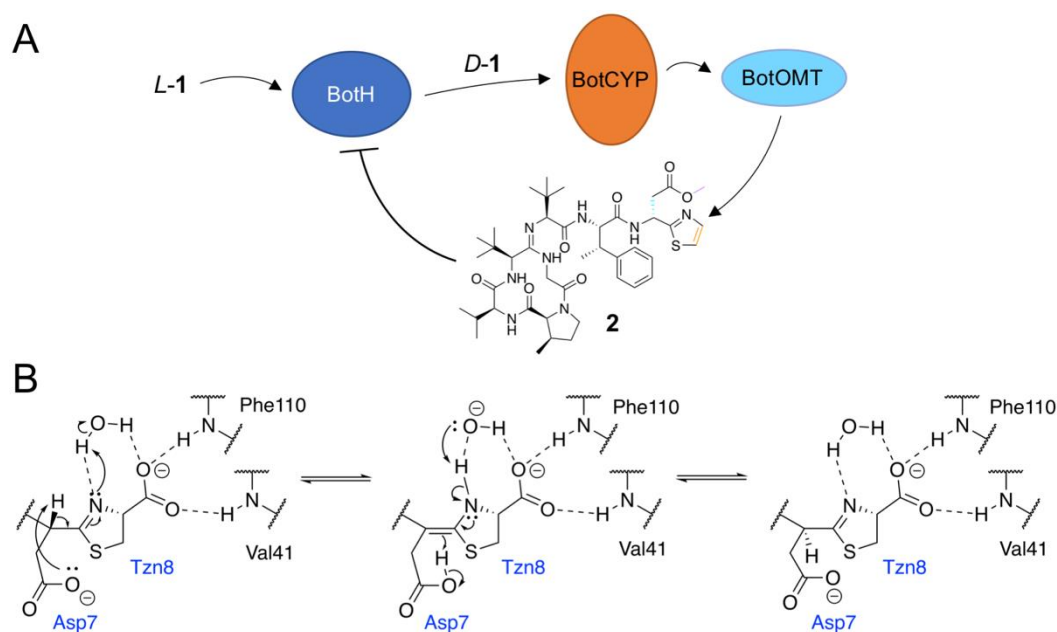


Figure 4: *A Proposed role of BotH in bottromycin biosynthesis. A rise of intracellular bottromycin concentrations leads to an inhibition of BotH epimerase activity, which may in turn prevent self-poisoning of the producing strain and act as an intracellular buffer to store bottromycins. B Proposed mechanism for the epimerization of 3b to yield 3a. BotH residues are labeled in black, substrate residues in blue. Hydrogen bonds are shown as dashed lines.*

It is still unclear why bottromycin A2 contains a D-Asp, as studies with synthetic derivatives have shown that both epimers display the same bioactivity<sup>37</sup>. Selectivity of the bottromycin exporter BotT appears possible, despite a very small fraction of bottromycin A2 in culture supernatant appearing to be the L-Asp epimer (Figure S22). The main benefit of epimerization may be providing increased resistance to proteolytic degradation of bottromycin, since D-amino acid containing peptides have longer half-lives<sup>38,39</sup>. It is of course also possible that the

bottromycin target of organisms in a native setting requires a D-Asp for optimal target binding. Interestingly, the structure elucidation of novel bottromycin analogs via crystallization of their complex with BotH appears to require less compound and be more straightforward than NMR, due to the robust crystallization condition, high affinity and very well-diffracting crystals.

The presence of BotH-like ABH protein encoding genes in over 200 BGCs encoding for the biosynthesis of diverse secondary metabolites places BotH as the founding member of a new subfamily of non-classical ABH enzymes. It will be fascinating to explore the functions of selected homologs, which may be able to epimerize non-Asp stereocenters across families of secondary metabolites via the unprecedented mechanism we have identified for bottromycin.



## 4 Supplementary Information

### 4.1 Methods

#### 4.1.1 Protein expression and purification

The BotH coding sequence was amplified from genomic DNA isolated from *Streptomyces* sp. BC16019 and cloned into the pET-28b plasmid (Novagen). The resulting construct was verified by enzymatic restriction digestion and DNA sequencing before being transformed into *E. coli* BL21(DE3) competent cells.

A single colony was selected and grown in LB liquid medium supplemented with kanamycin (50 µg / mL) to make an overnight culture. This culture was used at a dilution of 1 to 100 to inoculate fresh LB medium containing the appropriate antibiotic, and the culture was grown at 37 °C, 180 rpm. Upon reaching an optical density (OD<sub>600</sub>) of 0.6, the cultures were transferred to a precooled shaker at 16 °C, and protein expression was induced by addition of 1 mM IPTG. The cells were grown at 16 °C and 180 rpm over night before being harvested by centrifugation. Cell pellets were stored at -80 °C until further use.

The cell pellets were resuspended in lysis buffer (20 mM Tris pH 8.0, 500 mM NaCl, 20 mM Imidazole, and 3 mM DTT) supplemented with 0.4 mg DNase per gram of wet cell pellet and cOmplete EDTA-free protease inhibitor tablets (Roche). The cell suspension was lysed via passage through a cell disrupter (Constant Systems) at 30,000 psi, and cell debris was removed by centrifugation (40,000 x g, 4 °C, 15 min). The supernatant was collected and directly loaded onto to a 5 mL Histrap HP column (GE healthcare) preequilibrated with lysis buffer. The column was washed extensively with lysis buffer (30 CV) before the protein was eluted with lysis buffer supplemented with 250 mM imidazole. Fractions containing BotH were directly loaded onto a gel filtration column (HiLoad 16/600 Superdex 200 pg, GE healthcare) preequilibrated in gel filtration buffer A (20 mM HEPES pH 7.4, 200 mM NaCl, 1 mM TCEP). The fractions of the highest purity as judged by SDS-PAGE were pooled and concentrated to 5 mg / mL.

#### 4.1.2 Production of seleno-methionine (SeMet) BotH

L-Selenomethionine-labeled (SeMet) protein was expressed in *E. coli* BL21 cells grown in minimal medium supplemented with glucose-free nutrient mix (Molecular Dimensions), kanamycin (50 µg / mL) and glycerol (5% w / v). After 15 min growth at 37 °C and 180 rpm, 60 mg / mL selenomethionine was added. Upon reaching an OD<sub>600</sub> of 0.6, amino acids (lysine,

phenylalanine, threonine (100 mg / L each) and isoleucine and valine (50 mg / L each)) were added and the culture was grown for an additional 30 min at 37 °C, 180 rpm. Afterwards the cultures were transferred to 16 °C, 180 rpm, and protein expression was induced by addition of 1 mM IPTG. The cells were harvested the next day and protein purified as described above.

### 4.1.3 Crystallization and structure determination

For crystallization, BotH was treated with thermolysin (1 : 100) on ice for 2 h, after which the protein was passed over a HiLoad 16/600 Superdex 200 pg gel filtration column as described above and concentrated to 5 mg / mL. Crystals of apo BotH and BotH in complex with ligands were obtained at 18 °C in 1.2 - 1.8 M ammonium sulfate and 0.1 M Tris-Cl using the hanging drop vapor diffusion method. For the complex crystallization, the thermolysin-treated protein was incubated with excess ligand (1 - 2 mM) on ice overnight before setting up crystallization drops. Crystals appeared after a few days and were allowed to grow for an additional week. The crystals were cryoprotected in mother liquor supplemented with 30% glycerol and 0.5 mM ligand. Data was collected at ESRF (Beamlines ID23-1 and ID23-2), DESY (Beamline P11) and SLS (Beamline X06DA). To solve the apo structure, a single wavelength anomalous dispersion (SAD) data set was collected at the Se K absorption edge. Data were processed using Xia2<sup>40</sup>, the structure was solved using PHENIX *AutoSol*<sup>41</sup> and the initial model used to obtain a molecular replacement solution for the high-resolution native data set using Phaser<sup>42</sup>. The solution was manually rebuilt in COOT<sup>43</sup> and refined using PHENIX *Refine*<sup>41</sup>. This apo structure was then used as a search model for structure determination of the complex crystal structures by molecular replacement (Phaser). Data for all complex crystal structures were processed using XDS<sup>44</sup> and POINTLESS<sup>45</sup>, AIMLESS<sup>46</sup> and Ctruncate<sup>47</sup> implemented in ccp4<sup>48</sup>. All structures were validated using MolProbity. The images presented were created using PyMOL (Schrödinger, LLC) and LigPlot<sup>49</sup>.

### 4.1.4 Microscale Thermophoresis (MST)

Microscale thermophoresis experiments were carried out on a Monolith NT.115 (NanoTemper) in MST buffer (PBS with 0.05% Tween 20) using BotH labelled with the RED-tris-NTA fluorescent dye NT-647 following manufacture's protocol. The concentration of His-labelled BotH was 200 nM. Ligands were diluted in MST buffer to a starting concentration of 62.5 µM and then used in serial dilutions. Instrument settings were 40% excitation power and 40% MST

power. Data fitting and evaluation was performed using *MO.Affinity* analysis software (Nanotemper).

#### 4.1.5 Marfey derivatization

For stereochemical assignment of the aspartate  $\alpha$ -carbon in **3**, Marfey derivatization was performed. For the assay, 300  $\mu\text{g}$  **3** (ratio of **3a** : **3b** of about 1 : 9) was used. Solvent was evaporated at 110 °C and the compound was hydrolyzed by addition of 100  $\mu\text{l}$  6 N HCl and incubation at 110 °C for 35 min in a closed vial with nitrogen. The vial was opened and incubated for further 15 min at 110 °C to dry the contents. Residues were dissolved in 110  $\mu\text{l}$  H<sub>2</sub>O.

For derivatization of the amino acids in the hydrolyzed compound **3** or the amino acid standards (2 mg / mL D- or L- aspartic acid), 50  $\mu\text{L}$  solution were mixed with 20  $\mu\text{l}$  1 N NaHCO<sub>3</sub> and 20  $\mu\text{l}$  of 1% D-FDLA (*N* $\alpha$ -(2,4-dinitro-5-fluorophenyl)-D-leucinylamide) in acetone. The mixture was incubated for 1.5 h at 40 °C, 700 rpm. To stop the reaction, 10  $\mu\text{l}$  of 2 N HCl and 300  $\mu\text{l}$  ACN were added. The derivatized amino acids were separated by RP-HPLC-MS and the stereochemistry of the aspartate in the compounds **3a** and **3b** were assigned by comparing retentions times of the amino acid standards.

## 4.2 Enzymatic reaction of BotH with **3**

To investigate the effect of BotH on **3a/b**, 20  $\mu\text{M}$  **3a/b** were incubated with and without the addition of 5  $\mu\text{M}$  BotH in GF buffer for 30 min. Reactions were stopped by the addition of 2 volumes ACN and were frozen at -80 °C until analysis. Reactions were set up and analyzed in at least triplicates.

To test the effect of different BotH concentrations on the **3a** : **3b** ratio, 20  $\mu\text{M}$  **3a/b** were incubated with 0, 0.37, 1.11, 3.33 or 10  $\mu\text{M}$  BotH at 30 °C for 2 h in GF buffer. Reactions were stopped by the addition of 2 volumes ACN and were frozen at -80 °C until analysis. Higher concentrations than 10  $\mu\text{M}$  or longer incubation times were also tested, but did not lead to a further shift of epimer ratios. Reactions were set up and analyzed in triplicates.

To produce a roughly racemic mixture of **3a** : **3b**, 20  $\mu\text{M}$  **3a/b** were incubated with 4  $\mu\text{M}$  BotH for 30 min at 30 °C in GF buffer. BotH was denatured at 98 °C, 10 min, pelleted by centrifugation (15 min, 15,000 rpm) and the supernatant was lyophilized to remove all solvent.

### 4.3 Bottromycin extraction and purification

For bottromycin production, the *S. sp.* DG2-kmP41hyg strain<sup>50</sup> was cultivated in TSB seed medium at 28 °C. After 2 - 3 days the seed culture was used to inoculated SG production medium (Glucose: 20 g; Peptone: 10 g and CaCO<sub>3</sub>: 2 g per liter) and left to grow for 7 days at 28 °C. To the pooled cultures an equal volume of ethyl acetate was added and mixed for 1 hour. The organic layer was dried and resuspended in methanol before being loaded onto a sephadex LH-20 1-meter long column. The fractions containing bottromycin were pooled together and purified further by RP-HPLC (XBridge® Peptide BEH<sup>TM</sup> CSH C<sub>18</sub> OBD Prep Column, 130 Å, 5 µm, 10 mm x 250 mm, 1 / pkg) using a linear gradient from 95% A (H<sub>2</sub>O, 0.1% formic acid) to 95% B (acetonitrile, 0.1 % formic acid) over 40 min. The fractions containing bottromycin of the highest purity as judged by LC-MS were pooled and dried on a rotary evaporator. The resulting white amorphous solid was resuspend in methanol and dried under nitrogen to yield pure bottromycin A2.

### 4.4 *In vitro* production and purification of **3**

To produce **3**, bottromycin precursor peptide without the N-terminal methionine (BotA<sup>P</sup>) was enzymatically processed by IpoC, PurCD and PurAH, to introduce the thiazoline, the macroamidine and cleave off the follower peptide, respectively (see Figure S2).

BotA<sup>P</sup>, IpoC, PurCD and PurAH were expressed and purified as described previously<sup>21,25</sup>. Large scale IpoC reactions were carried out on a 9 mL scale in GF buffer (150 mM NaCl, 10 mM HEPES, 0.5 mM TCEP, pH 7.4) using the reaction setup 50 µM BotA<sup>P</sup>, 5 µM IpoC, 5 mM ATP as well as 5 mM MgCl<sub>2</sub> and incubating the reaction mixture for 16 h at 37 °C. The reaction mixture was filtered through a 0.22 µm filter and loaded onto a Superdex 30 16/60 size exclusion chromatography column pre-equilibrated in GF buffer. Heterocyclized-peptide-containing fractions were pooled, analyzed by MS and concentrated using a 5 kDa cutoff filter. Next, large scale reactions with PurCD and PurAH were carried out on a 9 mL scale by incubating 50 µM BotA<sup>PC</sup>, 5 µM PurCD, 1 µM PurAH, 10 mM ATP, 10 mM MgCl<sub>2</sub> and 100 µM CoCl<sub>2</sub> for 12 h at 37 °C in reaction buffer (200 mM NaCl, 50 mM Tris, 10% glycerol, pH 8.5). The reaction process was monitored by LC-MS. The analysis under acidic LC conditions, which leads to partial (about 50%) opening of the thiazoline ring, revealed the presence of two peaks (**3a/b**) with identical mass and fragmentation patterns, with a ratio of **3a** : **3b** of ~ 1 : 9.

After completion of the reaction, the mixture was filtered through a 0.22  $\mu\text{m}$  filter and applied to a Superdex 30 16/60 size exclusion chromatography column pre-equilibrated in ultrapure water. Every fraction was analyzed by mass spectrometry and the pure **3a/b**-containing fractions were pooled.

For further purification, **3a/b** was purified in  $\text{H}_2\text{O}$  and ACN containing 0.1% ammonium acetate pH 8.5. Separation was carried out on a Kinetex XB-C18 2.6  $\mu\text{m}$ , 4.6 x 100 mm column (Phenomenex) at 60  $^\circ\text{C}$  using a 8 min gradient from 0 to 80% ACN. Pure fractions were lyophilized and the remaining **3a/b** powder was weighed using a precision scale. A large-scale purification was carried out by Peptide Protein Research Ltd.

#### **4.5 Spontaneous epimerization of 3**

To determine the propensity for spontaneous (non-enzymatic) epimerization of the Asp C $\alpha$  in **3a/b**, a lyophilized, racemic mixture of **3a** : **3b** was dissolved in  $\text{D}_2\text{O}$ . Samples were taken after 2 h, 1 day, 4 days and 6 days and analyzed by LC-MS. Three independent experiments were performed.

#### **4.6 Pull-down of bottromycin A2 and 3a/b using BotH**

*S. sp.* DG2-KMp41hyg strain was cultivated as described above. 50 mL cultures were centrifuged for 10 min at room temperature to separate cells from the medium. The cell pellet was subsequently washed twice with PBS buffer before being resuspended in fresh 50 mL PBS buffer. The cell suspension was then lysed by sonication, centrifuged and the supernatant was decanted. His<sub>6</sub>-BotH (75  $\mu\text{g}$ ) was subsequently incubated with medium (1 mL), supernatant (1 mL) and **3a/b** (50  $\mu\text{g}$  dissolved PBS) for 30 min before being purified using a KingFisher<sup>TM</sup> mL Purification System. The protein was precipitated using ice cold ACN to liberate bound compounds and the supernatant was analyzed by HRLC-MS.

#### **4.7 Incorporation of a deuterium at Asp C $\alpha$ of 3**

Deuterium incorporation at the Asp C $\alpha$  was achieved by performing a BotH reaction in  $\text{D}_2\text{O}$  GF buffer. Lyophilized **3a/b** was dissolved in  $\text{D}_2\text{O}$  GF buffer and incubated for 30 min at 30  $^\circ\text{C}$  with (or without as control) 5  $\mu\text{M}$  BotH. The reactions were stopped by addition of 2 volumes ACN and stored at -80  $^\circ\text{C}$  until analysis. Reactions were set up in triplicates and analyzed by LC-MS and MS-fragmentation.

Back exchange of deuterated **3a/b** with solvent protons was shown by performing a BotH reaction using deuterated **3a/b** in H<sub>2</sub>O GF buffer: A reaction of **3a/b** with 5 μM BotH in D<sub>2</sub>O GF was stopped after 30 min at 30 °C by denaturation of BotH at 98 °C for 10 min. Denatured BotH was pelleted by centrifugation (15 min, 15,000 rpm). The supernatant was lyophilized to remove the D<sub>2</sub>O and re-dissolved in the same volume H<sub>2</sub>O. Samples were split up and incubated 30 min at 30 °C with or without the addition of 5 μM BotH. Reactions were stopped by the addition of 2 volumes ACN and analyzed by LC-MS.

#### 4.8 Reaction of BotH with derivatives of **3**

To generate derivatives of **3**, bottromycin precursor peptides (BotA<sup>P</sup>) with single amino acid changes in the core peptide were cloned, expressed and purified as described previously<sup>21,25</sup>. Shortly, *botA* was mutated by using mutation carrying primers and cloned into a pHisSUMOTEV vector, which was a gift from Dr. Huanting Liu, St. Andrews University<sup>51</sup>. To produce the native N-terminus of the bottromycin precursor peptide BotA<sup>P</sup> (after cleavage of the N-terminal methionine by BotP) (GPVVV....) without cloning artefacts, primers which introduce a lysine residue were used (...GAMAGKGPVVV...). Peptides were expressed in *E. coli* Lemo21(DE3) cells carrying the respective expression vector and purified by Ni<sup>2+</sup>-NTA-chromatography and, after cleavage with Trypsin, by gel filtration<sup>21,25</sup>.

To produce the **3** derivatives, 50 μM BotA<sup>P</sup> derivative was incubated with 5 μM IpoC, 10 mM ATP as well as 5 mM MgCl<sub>2</sub> for 12 h at 37 °C in reaction buffer (200 mM NaCl, 50 mM Tris, 10 % glycerol, pH 8.5). Then, 5 μM PurCD, 1 μM PurAH and 100 μM CoCl<sub>2</sub> were added and incubated for a further 4 h at 37 °C. Each reaction was then divided and incubated with or without the addition of 5 μM BotH for 2 h at 30 °C. Reactions were stopped by the addition of 2 volumes ACN and frozen at -80 °C until analysis. Reactions were set up and analyzed in triplicates.

#### 4.9 Orthosteric inhibition of BotH epimerization by bottromycin A2

0.2 μM BotH in D<sub>2</sub>O GF buffer was pre-incubated with and without 50 μM bottromycin A2 for 30 min at 30 °C. **3a/b** in D<sub>2</sub>O GF buffer was added (final concentration 10 μM) and the solution was incubated for 7 min at 30 °C. The reactions were stopped by the addition of 2 volumes ACN and the samples were frozen at -80 °C until analysis by LC-MS. As a control for non-enzymatic deuterium incorporation, samples without enzyme were set up. Reactions were performed in triplicates and analyzed by LC-MS. For analysis of the LC-MS data, the EICs

( $\pm 5$  ppm) for the 1<sup>st</sup> and the 2<sup>nd</sup> isotope peak of the doubly charged ion of **3** (1<sup>st</sup> isotope peak  $[M+2H]^{2+} = 400.1940$ ; 2<sup>nd</sup> isotope peak  $[M+2H]^{2+} = 400.6955$ ) were generated. The respective areas for **3a** (2.96 - 3.10 min) were calculated using the Bruker Compass DataAnalysis 4.2. software. The ratio of the area of the 1<sup>st</sup> isotope peak to the 2<sup>nd</sup> isotope peak was used to quantify the deuterium incorporation. The significance (*p*-value) of the ratio differences was calculated using an unpaired two-tailed t-test.

#### 4.10 LC-MS and MS<sup>2</sup> analysis

Measurements were performed on a Dionex Ultimate 3000 RSLC system (Thermo Fisher Scientific) using a flow rate of 600  $\mu\text{L min}^{-1}$  and column oven temperature of 45 °C. Standard measurements were performed using a BEH C18, 50 x 2.1 mm, 1.7  $\mu\text{m}$  dp column equipped with a C18 precolumn (Waters). Samples were separated by a gradient from (A) H<sub>2</sub>O + 0.1% formic acid to (B) ACN + 0.1% formic acid. The linear gradient was initiated by a 1 min isocratic step at 5% B, followed by an increase to 95% B in 9 min to end up with a 1.5 min plateau step at 95% B before re-equilibration under the initial conditions. UV spectra were recorded by a DAD in the range from 200 to 600 nm.

For Marfey's analysis, samples were separated on a BEH C18, 100 x 2.1 mm, 1.7  $\mu\text{m}$  dp column equipped with a C18 precolumn (Waters). A multistep gradient using (A) H<sub>2</sub>O + 0.1% formic acid and (B) ACN + 0.1% formic acid was applied for sample preparation. The gradient was initiated by a 1 min isocratic step at 5% B, an increase to 10% B in 1 min, an increase to 35% B in 14 min, followed by an increase to 55% B in 7 min and an increase to 80% B in 3 min to end up with a 1 min plateau step at 80% B before re-equilibration to the initial conditions.

For MS measurements on maXis-4 hr-qTOF mass spectrometer (Bruker Daltonics), the LC flow was split 1:8 before entering the mass spectrometer using the Apollo II ESI source. In the source region, the temperature was set to 200 °C, the capillary voltage was 4000 V, the dry-gas flow was 5.0 L / min and the nebulizer was set to 1 bar. Ion transfer settings were set to Funnel 1 RF 350 Vpp and Multipole RF 400 Vpp, quadrupole settings were set to an ion energy of 5.0 eV and a low mass cut of 120 m / z. The collision cell was set to an energy of 5.0 eV and the pulse storage time was 5  $\mu\text{s}$ . Data were recorded in centroid mode ranging from 150 to 2500 m/z at a 2 Hz scan rate. Calibration of the maXis4G qTOF spectrometer was achieved with sodium formate clusters before every injection to avoid mass drifts. All MS analyses were acquired in the presence of the lock masses C<sub>12</sub>H<sub>19</sub>F<sub>12</sub>N<sub>3</sub>O<sub>6</sub>P<sub>3</sub>, C<sub>18</sub>H<sub>19</sub>O<sub>6</sub>N<sub>3</sub>P<sub>3</sub>F<sub>2</sub> and C<sub>24</sub>H<sub>19</sub>F<sub>36</sub>N<sub>3</sub>O<sub>6</sub>P<sub>3</sub> which generate the  $[M+H]^+$  Ions of 622.028960, 922.009798 and 1221.990638.

LC-MS<sup>2</sup> fragmentation spectra were recorded using a scheduled precursor list (SPL). Separation was achieved using a BEH C18, 100 x 2.1 mm, 1.7 μm dp column equipped with a C18 precolumn (Waters) and a linear gradient from (A) H<sub>2</sub>O + 0.1% formic acid to (B) ACN + 0.1% formic acid. The gradient was initiated by a 1 min isocratic step at 5% B, followed by an increase to 95% B in 18 min to end up with a 2 min plateau step at 95% ACN before re-equilibration to the initial conditions. SPL entries and parameters were set to fragment only the ions of interest. SPL tolerance parameters for precursor ion selection were 0.17 min and 0.05 m / z. The CID energy was ramped from 35 eV for 500 m / z to 45 eV for 1000 m / z. Data were displayed and analyzed using the Bruker Compass DataAnalysis software (Version 4.2). Shown MS spectra are in general single spectra (taken at the EIC maximum), except for Figure S4 where spectra are averaged, as the peaks of **3** and the respective peak for the compound with opened thiazoline do not have the same retention time. Signals in the MS-spectra are labelled with the observed monoisotopic mass. Extracted-ion chromatograms were generated using the calculated monoisotopic mass with a range of 5 ppm. All EICs (except Figures S19 and S22A) and MS-spectra are scaled to a relative intensity of 1.

#### **4.11 Heterologous expression of *btmH* and *btmA* in *Streptomyces coelicolor* M145**

*btmH* was PCR amplified from *S. scabies* genomic DNA using primers *btmH*-start (GAA CCG TTC CAT ATG GTC TCC CGC GA, NdeI restriction site underlined) and *btmH*-end (GTC CTC GGT GGA ATT CAT ACG GCC C, EcoRI restriction site underlined). The resulting PCR product was digested and ligated into the NdeI and EcoRI sites of pIB139-RBS<sup>23</sup> to generate pIB139-RBS-*btmH*. This construct, as well as pIJ10257-*btmA* were introduced by intergeneric conjugation from *E. coli* ET12567/pUZ8002 into *S. coelicolor* M145.<sup>52</sup> Selection was carried out with nalidixic acid (25 μg / mL) and apramycin (50 μg / mL) in SFM for the exconjugants carrying pIB139-RBS-*btmH*, nalidixic acid (25 μg / mL) and hygromycin (50 μg / mL) for the exconjugants carrying pIJ10257-*btmA* and nalidixic acid (25 μg / mL), apramycin (50 μg / mL) and hygromycin (50 μg / mL) for the exconjugants carrying both constructs.

#### **4.12 Bioassays**

To evaluate the effect of *btmA* and *btmH* expression on the sensitivity of *S. coelicolor* M145 to bottromycin A2, 250 μL of spores from a concentrated stock of each strain were added to 4 mL of MQ water and the resulting mixture was agitated manually until an homogeneous mixture



was obtained. Each resulting spore suspension was then overlaid on top of a 100 mm square petri dish containing 35 mL of Difco™ Nutrient Agar, and the liquid layer was dried under a laminar flow hood for 2 hours. Sample diffusion wells were bored into the agar using an inverted 200 µL micropipette tip, and 40 µL of serial dilutions of bottromycin A2 in DMSO were loaded into each well. After this, plates were kept at 4 °C for 2 hours for sample diffusion. Finally, the bioassay plates were incubated at 30 °C for 48 hours.

#### **4.13 Bioinformatics analysis**

We downloaded bacterial sequences of the InterPro<sup>36</sup> family IPR000073 and aligned them with HMMer (<http://hmmer.org/>)<sup>53</sup> using the HMM for the Pfam<sup>54</sup> family PF00561 (Abhydrolase\_1), retaining only those from this Pfam family. Then sequences containing the following mutations at the catalytic site positions were selected with a custom Perl script: Ser mutated to either Val, Ile, Leu, Met, Ala, Phe, Tyr, or Trp; Asp mutated to anything except Asn or Glu or missing; His mutated to any amino acid or missing.

The corresponding genomes were identified by parsing UniProt<sup>55</sup> data files and retrieved from NCBI using the batch download option. Biosynthetic gene clusters (BGCs) were predicted using antiSMASH<sup>56</sup>. All BotH homologs whose genomic coordinates fall within the predicted boundaries of BGCs or lie outside them separated by no more than 1000 nt were considered to be associated with these BGCs.

The distance matrix between all proteins that are associated with BGCs was calculated using the R package seqinr<sup>57</sup>. The sequence similarity networks were visualized in Cytoscape<sup>58</sup>, all edges connecting pairs of proteins that shared less than 15% sequence identity were omitted.

## 5 References

- 1 Ollis, D. L. *et al.* The alpha/beta hydrolase fold. *Protein Eng* **5**, 197-211, doi:10.1093/protein/5.3.197 (1992).
- 2 Nardini, M. & Dijkstra, B. W. Alpha/beta hydrolase fold enzymes: the family keeps growing. *Curr Opin Struct Biol* **9**, 732-737 (1999).
- 3 Mindrebo, J. T., Nartey, C. M., Seto, Y., Burkart, M. D. & Noel, J. P. Unveiling the functional diversity of the alpha/beta hydrolase superfamily in the plant kingdom. *Curr Opin Struct Biol* **41**, 233-246, doi:10.1016/j.sbi.2016.08.005 (2016).
- 4 Rauwerdink, A. & Kazlauskas, R. J. How the Same Core Catalytic Machinery Catalyzes 17 Different Reactions: the Serine-Histidine-Aspartate Catalytic Triad of alpha/beta-Hydrolase Fold Enzymes. *ACS Catal* **5**, 6153-6176, doi:10.1021/acscatal.5b01539 (2015).
- 5 Hamiaux, C. *et al.* DAD2 is an alpha/beta hydrolase likely to be involved in the perception of the plant branching hormone, strigolactone. *Curr Biol* **22**, 2032-2036, doi:10.1016/j.cub.2012.08.007 (2012).
- 6 Guo, Y., Zheng, Z., La Clair, J. J., Chory, J. & Noel, J. P. Smoke-derived karrikin perception by the alpha/beta-hydrolase KAI2 from Arabidopsis. *Proc Natl Acad Sci U S A* **110**, 8284-8289, doi:10.1073/pnas.1306265110 (2013).
- 7 Shimada, A. *et al.* Structural basis for gibberellin recognition by its receptor GID1. *Nature* **456**, 520-523, doi:10.1038/nature07546 (2008).
- 8 Marchot, P. & Chatonnet, A. Enzymatic activity and protein interactions in alpha/beta hydrolase fold proteins: moonlighting versus promiscuity. *Protein Pept Lett* **19**, 132-143, doi:10.2174/092986612799080284 (2012).
- 9 Crone, W. J. K., Leeper, F. J. & Truman, A. W. Identification and characterisation of the gene cluster for the anti-MRSA antibiotic bottromycin: expanding the biosynthetic diversity of ribosomal peptides. *Chemical Science* **3**, 3516-3521, doi:10.1039/C2SC21190D (2012).
- 10 Gomez-Escribano, J. P., Song, L., Bibb, M. J. & Challis, G. L. Posttranslational  $\beta$ -methylation and macrolactamidation in the biosynthesis of the bottromycin complex of ribosomal peptide antibiotics. *Chemical Science* **3**, 3522-3525, doi:10.1039/C2SC21183A (2012).
- 11 Hou, Y. *et al.* Structure and biosynthesis of the antibiotic bottromycin D. *Org Lett* **14**, 5050-5053, doi:10.1021/ol3022758 (2012).
- 12 Huo, L., Rachid, S., Stadler, M., Wenzel, S. C. & Muller, R. Synthetic biotechnology to study and engineer ribosomal bottromycin biosynthesis. *Chem Biol* **19**, 1278-1287, doi:10.1016/j.chembiol.2012.08.013 (2012).
- 13 Waisvisz, J. M. Bottromycin. I. A new sulfurcontaining antibiotic. *J. Am. Chem. Soc.* **79**, 4520-4521 (1957).
- 14 Nakamura, S. Isolation and characterization of bottromycins A and B. *J. Antibiotics, Ser. A* **18**, 47-52 (1965).
- 15 Sowa, S. *et al.* Susceptibility of methicillin-resistant Staphylococcus aureus clinical isolates to various antimicrobial agents. *Hiroshima J Med Sci* **40**, 137-144 (1991).
- 16 Shimamura, H. *et al.* Structure determination and total synthesis of bottromycin A2: a potent antibiotic against MRSA and VRE. *Angew Chem Int Ed Engl* **48**, 914-917, doi:10.1002/anie.200804138 (2009).
- 17 Otaka, T. & Kaji, A. Mode of action of bottromycin A2. Release of aminoacyl- or peptidyl-tRNA from ribosomes. *J Biol Chem* **251**, 2299-2306 (1976).
- 18 Otaka, T. & Kaji, A. Mode of action of bottromycin A2: effect on peptide bond formation. *FEBS Lett* **123**, 173-176 (1981).

- 19 Otaka, T. & Kaji, A. Mode of action of bottromycin A2: effect of bottromycin A2 on polysomes. *FEBS Lett* **153**, 53-59 (1983).
- 20 Arnison, P. G. *et al.* Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* **30**, 108-160, doi:10.1039/c2np20085f (2013).
- 21 Franz, L., Adam, S., Santos-Aberturas, J., Truman, A. W. & Koehnke, J. Macroamidine Formation in Bottromycins Is Catalyzed by a Divergent YcaO Enzyme. *J Am Chem Soc* **139**, 18158-18161, doi:10.1021/jacs.7b09898 (2017).
- 22 Schwalen, C. J. *et al.* In Vitro Biosynthetic Studies of Bottromycin Expand the Enzymatic Capabilities of the YcaO Superfamily. *J Am Chem Soc* **139**, 18154-18157, doi:10.1021/jacs.7b09899 (2017).
- 23 Crone, W. J. *et al.* Dissecting Bottromycin Biosynthesis Using Comparative Untargeted Metabolomics. *Angew Chem Int Ed Engl* **55**, 9639-9643, doi:10.1002/anie.201604304 (2016).
- 24 Mann, G. *et al.* Structure and Substrate Recognition of the Bottromycin Maturation Enzyme BotP. *Chembiochem* **17**, 2286-2292, doi:10.1002/cbic.201600406 (2016).
- 25 Sikandar, A., Franz, L., Melse, O., Antes, I. & Koehnke, J. Thiazoline-Specific Amidohydrolase PurAH Is the Gatekeeper of Bottromycin Biosynthesis. *J Am Chem Soc* **141**, 9748-9752, doi:10.1021/jacs.8b12231 (2019).
- 26 Hur, G. H., Vickery, C. R. & Burkart, M. D. Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat Prod Rep* **29**, 1074-1098, doi:10.1039/c2np20025b (2012).
- 27 Freeman, M. F. *et al.* Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides. *Science* **338**, 387-390, doi:10.1126/science.1226121 (2012).
- 28 Morinaka, B. I. *et al.* Radical S-adenosyl methionine epimerases: regioselective introduction of diverse D-amino acid patterns into peptide natural products. *Angew Chem Int Ed Engl* **53**, 8503-8507, doi:10.1002/anie.201400478 (2014).
- 29 Parent, A. *et al.* Mechanistic Investigations of PoyD, a Radical S-Adenosyl-L-methionine Enzyme Catalyzing Iterative and Directional Epimerizations in Polytheonamide A Biosynthesis. *J Am Chem Soc* **140**, 2469-2477, doi:10.1021/jacs.7b08402 (2018).
- 30 Cotter, P. D. *et al.* Posttranslational conversion of L-serines to D-alanines is vital for optimal production and activity of the lantibiotic lactacin 3147. *Proc Natl Acad Sci U S A* **102**, 18584-18589, doi:10.1073/pnas.0509371102 (2005).
- 31 Yang, X. & van der Donk, W. A. Post-translational Introduction of D-Alanine into Ribosomally Synthesized Peptides by the Dehydroalanine Reductase NpnJ. *J Am Chem Soc* **137**, 12426-12429, doi:10.1021/jacs.5b05207 (2015).
- 32 Holm, L. Benchmarking Fold Detection by DaliLite v.5. *Bioinformatics*, doi:10.1093/bioinformatics/btz536 (2019).
- 33 Bains, J., Kaufman, L., Farnell, B. & Boulanger, M. J. A product analog bound form of 3-oxoadipate-enol-lactonase (PcaD) reveals a multifunctional role for the divergent cap domain. *J Mol Biol* **406**, 649-658, doi:10.1016/j.jmb.2011.01.007 (2011).
- 34 Adam, S., Franz, L., Milhim, M., Bernhardt, R. & Koehnke, J. Characterization of the stereoselective P450 enzyme BotCYP enables the in vitro biosynthesis of the Bottromycin core scaffold. (Submitted).
- 35 Gouda, H. *et al.* Three-dimensional solution structure of bottromycin A2: a potent antibiotic active against methicillin-resistant *Staphylococcus aureus* and vancomycin-resistant Enterococci. *Chem Pharm Bull (Tokyo)* **60**, 169-171, doi:10.1248/cpb.60.169 (2012).

- 36 Mitchell, A. L. *et al.* InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res* **47**, D351-D360, doi:10.1093/nar/gky1100 (2019).
- 37 Yamada, T. *et al.* Synthesis and Evaluation of Antibacterial Activity of Bottromycins. *J Org Chem* **83**, 7135-7149, doi:10.1021/acs.joc.8b00045 (2018).
- 38 Jungheim, L. N. *et al.* Potent human immunodeficiency virus type 1 protease inhibitors that utilize noncoded D-amino acids as P2/P3 ligands. *J Med Chem* **39**, 96-108, doi:10.1021/jm950576c (1996).
- 39 Rink, R. *et al.* To protect peptide pharmaceuticals against peptidases. *J Pharmacol Toxicol Methods* **61**, 210-218, doi:10.1016/j.vascn.2010.02.010 (2010).
- 40 Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* **43**, 186-190, doi:doi:10.1107/S0021889809045701 (2010).
- 41 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).
- 42 McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674, doi:10.1107/S0021889807021206 (2007).
- 43 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).
- 44 Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132, doi:10.1107/S0907444909047337 (2010).
- 45 Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr D Biol Crystallogr* **67**, 282-292, doi:10.1107/S090744491003982X (2011).
- 46 Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204-1214, doi:10.1107/S0907444913000061 (2013).
- 47 French, S. & Wilson, K. On the treatment of negative intensity observations. *Acta Crystallographica Section A* **34**, 517-525, doi:doi:10.1107/S0567739478001114 (1978).
- 48 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242, doi:10.1107/S0907444910045749 (2011).
- 49 Laskowski, R. A. & Swindells, M. B. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model* **51**, 2778-2786, doi:10.1021/ci200227u (2011).
- 50 Horbal, L., Marques, F., Nadmid, S., Mendes, M. V. & Luzhetskyy, A. Secondary metabolites overproduction through transcriptional gene cluster refactoring. *Metab Eng* **49**, 299-315, doi:10.1016/j.ymben.2018.09.010 (2018).
- 51 Liu, H. & Naismith, J. H. A simple and efficient expression and purification system using two newly constructed vectors. *Protein Expr Purif* **63**, 102-111, doi:10.1016/j.pep.2008.09.008 (2009).
- 52 Parish, J. H. Genetic manipulation of streptomyces — A laboratory manual: By D A Hopwood, M J Bibb, K F Chater; T Kieser CJ Bruton, H M Kieser, D J Lydiate, C P Smith, J M Ward and H Schrempf. pp 356. The John Innes Foundation, Norwich, UK and Cold Spring Harbour Laboratory. 1985. \$25 ISBN 0-7084-0336-0. *Biochemical Education* **14**, 196-196, doi:10.1016/0307-4412(86)90228-1 (1986).
- 53 Eddy, S. R. Profile hidden Markov models. *Bioinformatics* **14**, 755-763, doi:10.1093/bioinformatics/14.9.755 (1998).

- 54 El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res* **47**, D427-D432, doi:10.1093/nar/gky995 (2019).
- 55 UniProt, C. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* **47**, D506-D515, doi:10.1093/nar/gky1049 (2019).
- 56 Blin, K. *et al.* antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res* **45**, W36-W41, doi:10.1093/nar/gkx319 (2017).
- 57 Charif, D. & Lobry, J. R. in *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations* (eds Ugo Bastolla, Markus Porto, H. Eduardo Roman, & Michele Vendruscolo) 207-232 (Springer Berlin Heidelberg, 2007).
- 58 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).

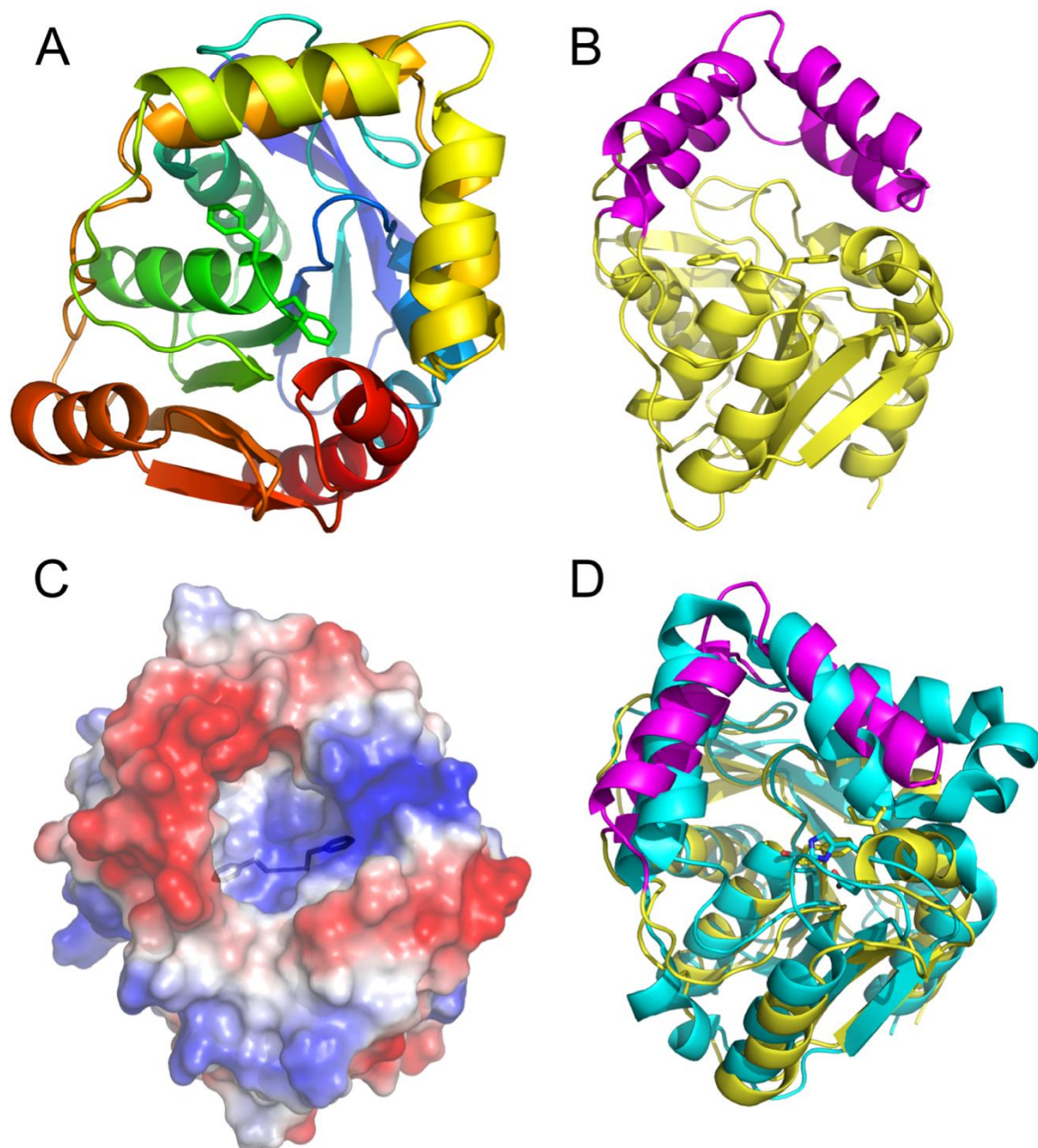


Figure S1: (A) Cartoon representation of apo-BotH. The two Phe residues at the active site are shown as sticks. (B) Cartoon representation of apo-BotH highlighting the V-shaped loop (magenta) positioned above the active site. (C) Electrostatic surface potential of BotH. The Phe-Phe motif at the active-site is shown as black sticks. (D) Superposition of apo-BotH (yellow/magenta) with its closest structural homolog (PDB ID 2xua, cyan) gives a  $C_{\alpha}$ -rmsd of 2.8 Å over 288 residues.

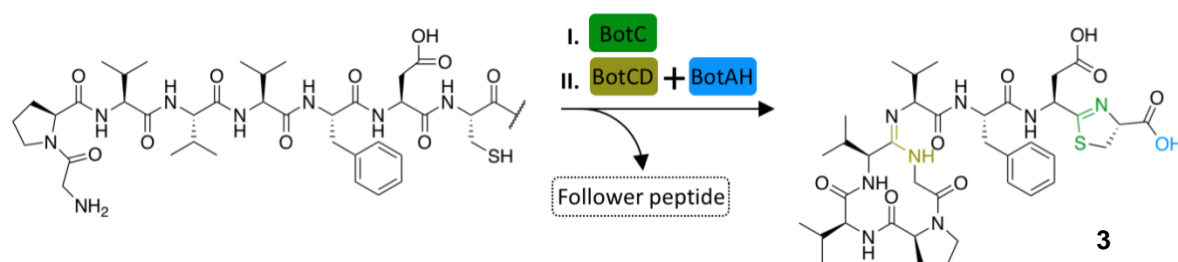


Figure S2: Production of **3** can be accomplished in a two-stage, one-pot reaction with quantitative yields.

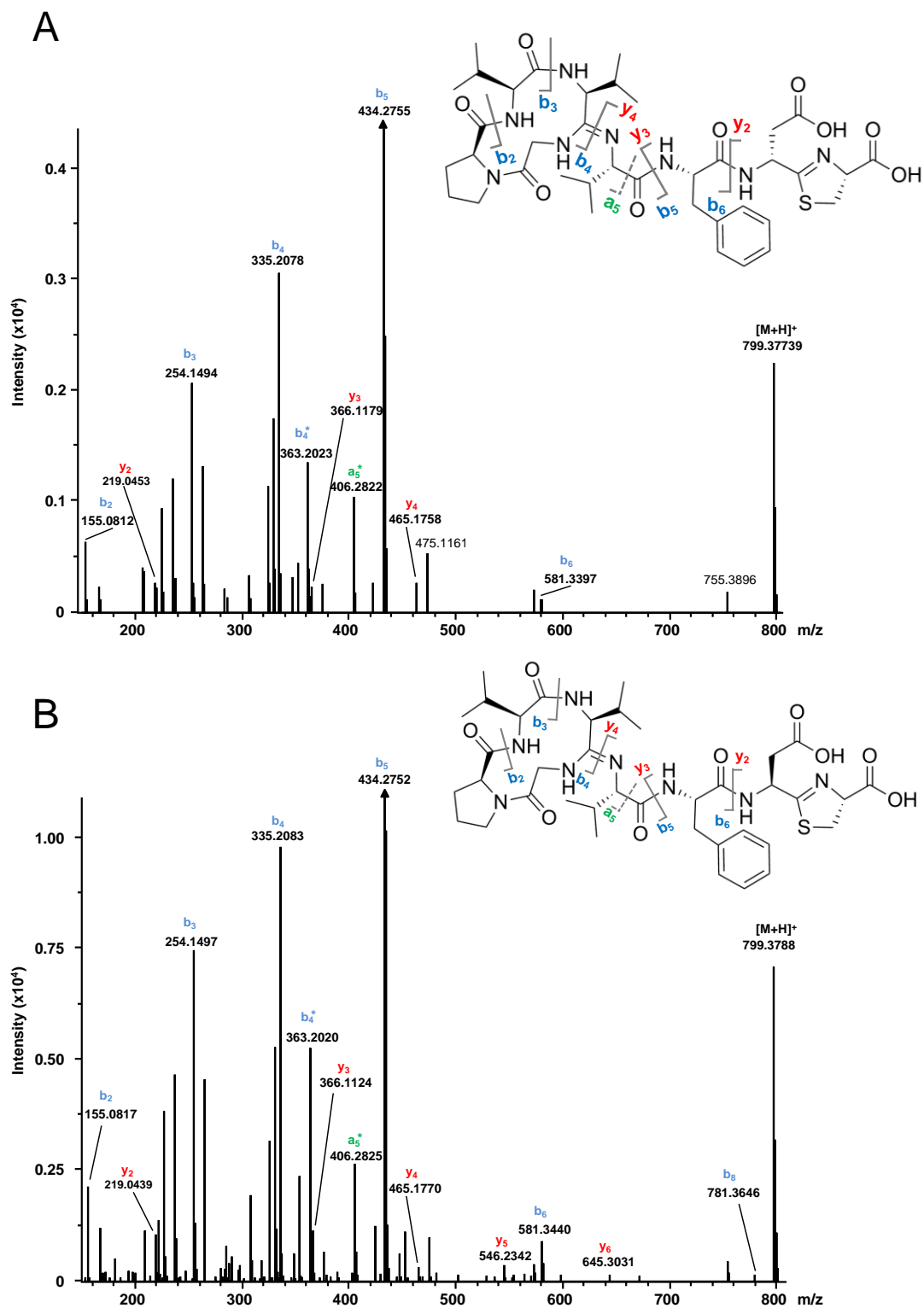


Figure S3: MS<sup>2</sup> fragmentation spectra of **3a** (D-Asp) (A) and **3b** (L-Asp) (B). The associated peak lists can be found in table S2.

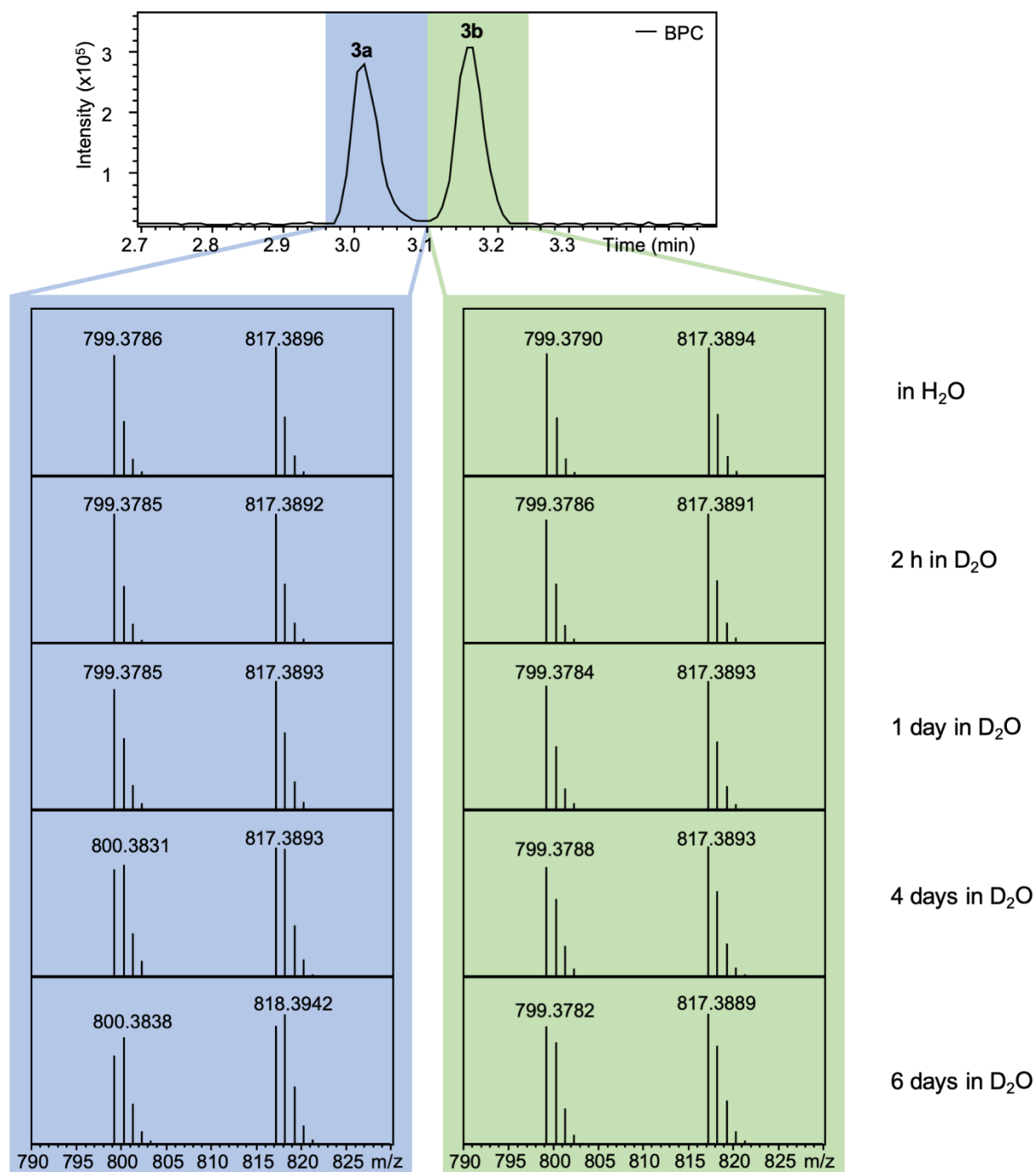


Figure S4: Spontaneous (non-enzymatic) epimerization of the **3a/b** Asp C $\alpha$  in D<sub>2</sub>O. A roughly racemic mixture of **3a/b** was dissolved in D<sub>2</sub>O to analyze spontaneous epimerization. Samples were taken after 2 h, 1 day, 4 days and 6 days. In the acidic (0.1% FA) LC conditions the thiazoline partly (about 50%) re-opens, which leads to the addition of H<sub>2</sub>O (+ 18.015 Da). Shown mass spectra are averaged from 2.96 - 3.10 min and 3.10 - 3.24 min respectively, as the retention times of **3a/b** and the respective compound with re-opened thiazoline slightly differ. Shown is the base peak chromatogram (BPC) in H<sub>2</sub>O. No changes in the shape of the BPCs were observed at the different time points. Representative experiments were repeated independently three times with similar results. BPC = Base Peak Chromatogram.



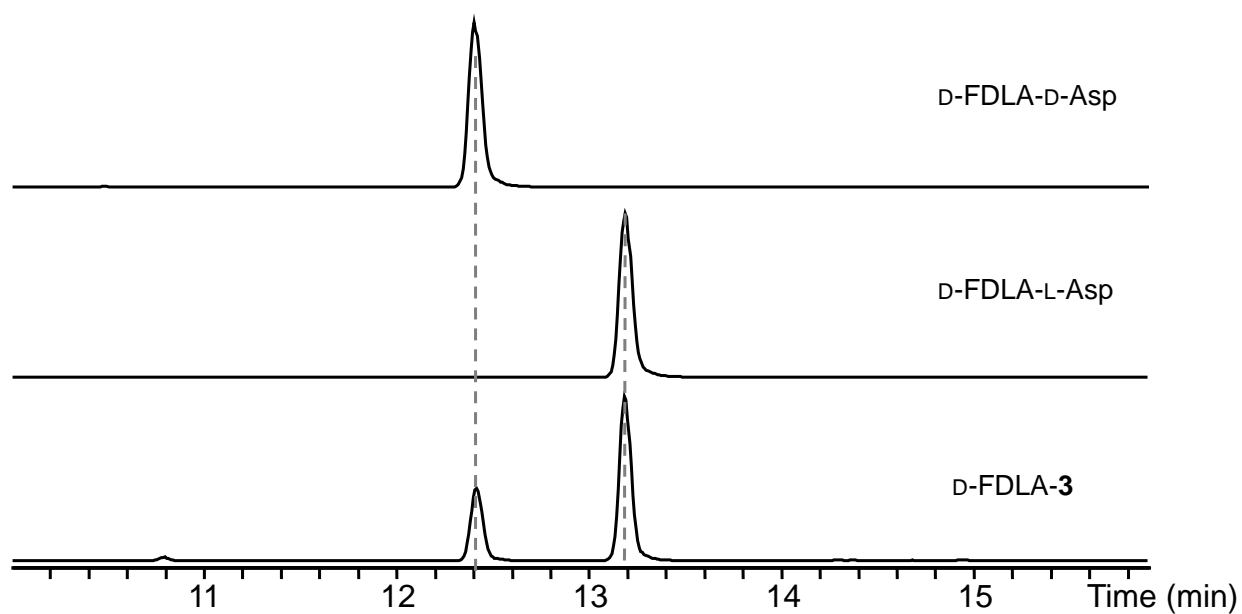


Figure S5: Marfey's analysis of **3a/b** confirmed the presence of D-Asp (**3a**) and L-Asp (**3b**) and allowed us to assign **3a** to the less abundant peak. Shown are the extracted ion chromatograms (EICs  $\pm 5$  ppm) for the FDLA-Asp derivate (428.1412  $m/z$ ).

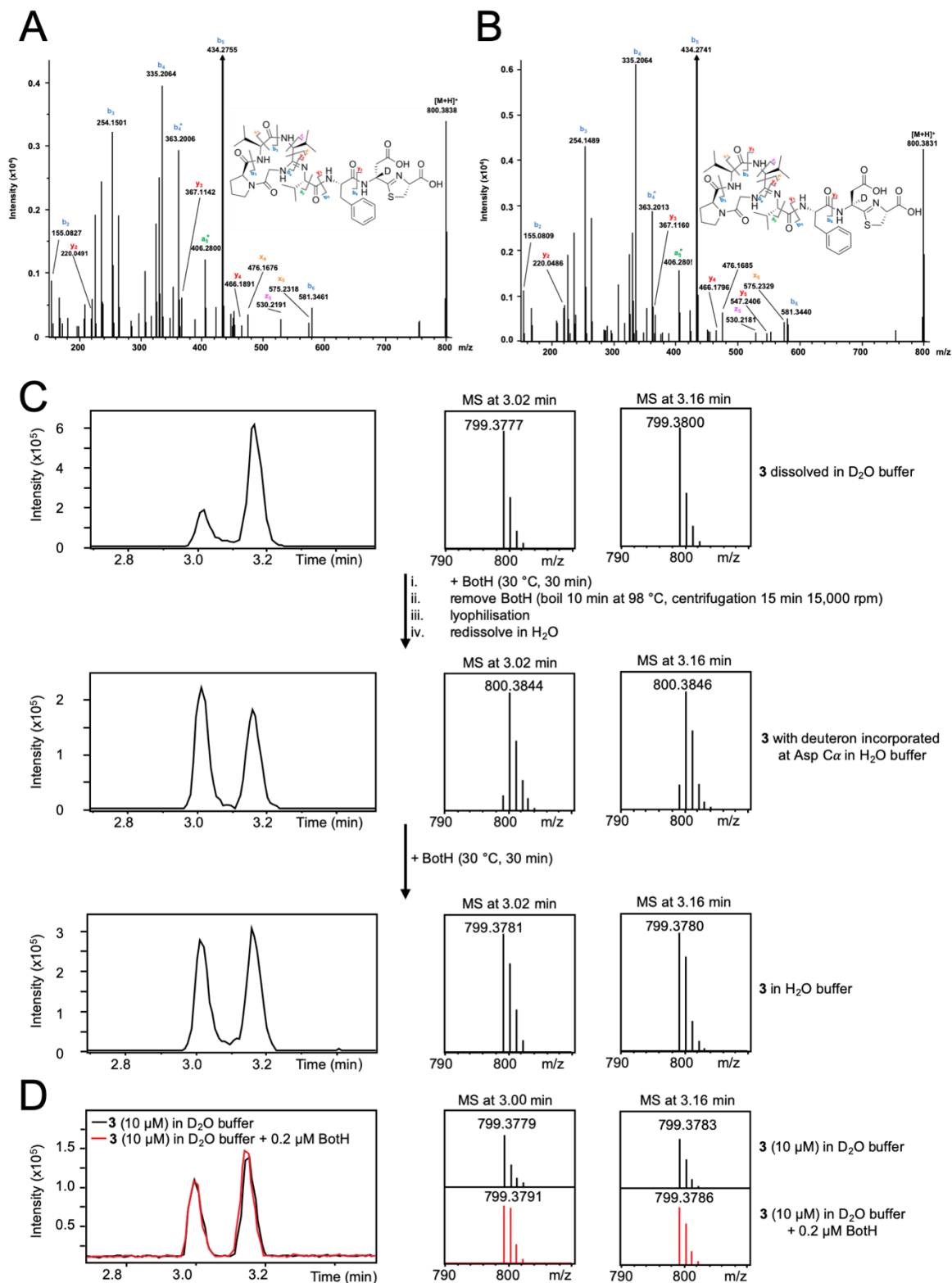


Figure S6: (A) MS<sup>2</sup>-fragmentation of deuterated 3a. The associated peak lists can be found in Table S3A. (B) MS<sup>2</sup>-fragmentation of deuterated 3b. The associated peak lists can be found in Table S3B. (C) Performing a BotH reaction in D<sub>2</sub>O buffer leads to deuterium incorporation at the Asp C $\alpha$ . In reverse, performing a BotH reaction with deuterated 3a/b in H<sub>2</sub>O buffer leads to the exchange of the deuterium with solvent protons. Shown are the base peak chromatograms (BPC). (D) At low BotH concentrations no change of the 3a : 3b ratio is observed within the time-frame of the experiment (7 min), but when the reactions are performed in D<sub>2</sub>O buffer, deuterium is incorporated. Shown are the BPCs. Representative experiments were repeated independently three times with similar results.

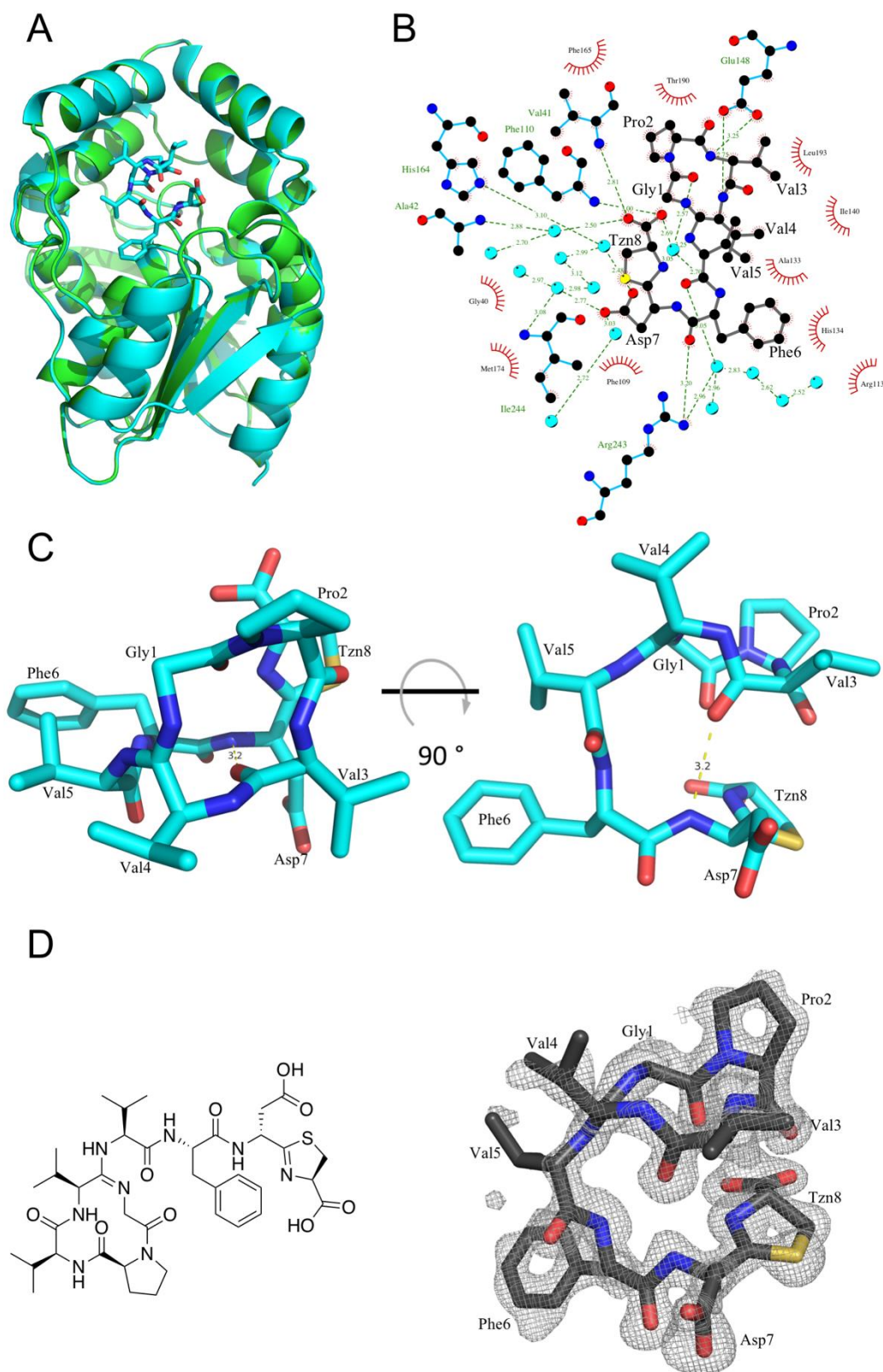


Figure S7: (A) Superposition of the BotH apo structure (green) with the BotH-3a complex structure (cyan). 3a is shown as cyan sticks. (B) LigPlus diagram of the interactions between 3a and BotH. 3a shown with bonds in grey, BotH with bonds in cyan. Water molecules are depicted as cyan spheres, intermolecular hydrogen bonds shown as dashed lines with distances in Å and hydrophobic contacts are shown as red spoked arcs. (C) 3a as observed in the complex structure shown as sticks. The intramolecular hydrogen bond is shown as a dashed line with the distance given in Å. (D) 3a as observed in the complex crystal structure (left) and Polder map of 3a contoured at 3  $\sigma$  (right).

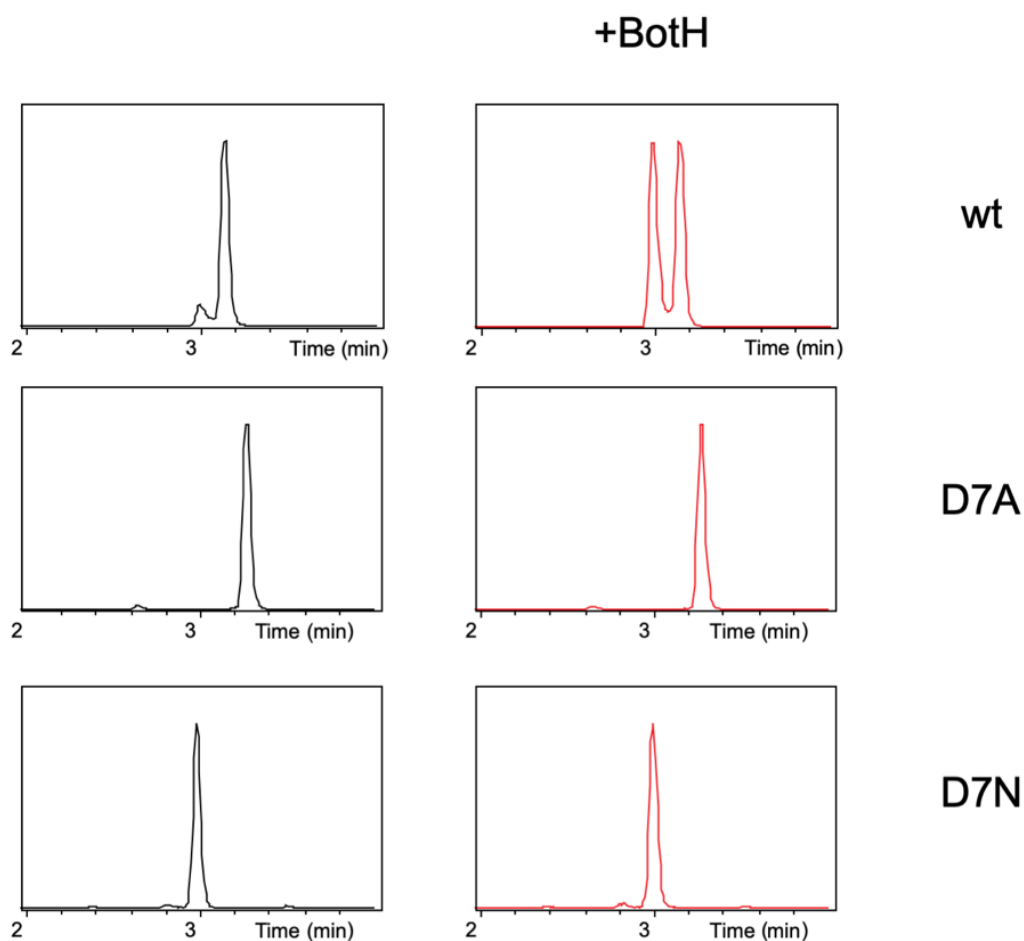


Figure S8: The derivatives of **3**, where Asp was mutated to Ala (D7A) or Asn (D7N) are no substrates of BotH. Graphs show extracted ion chromatograms (EICs, calculated mass see Table S6  $\pm$  5 ppm) of possible BotH substrates with (red) and without (black) addition of BotH. No change of the retention time nor an additional peak could be observed in the respective EICs after addition of BotH. Representative experiments were repeated independently three times with similar results.

<b>G</b>	<b>P</b>	<b>V</b>	<b>V</b>	<b>V</b>	<b>F</b>	<b>D</b>	<b>C</b>
	<b>A</b>	<b>L</b>	<b>A</b>	<b>T</b>	<b>W</b>	<b>A</b>	
	<b>G</b>		<b>L</b>	<b>L</b>	<b>A</b>	<b>N</b>	
				<b>A</b>	<b>Y</b>		
				<b>E</b>			

Figure S9: BotA point mutations that are compatible with the enzymes used to generate the BotH substrate (IpoC/PurCD/PurAH).

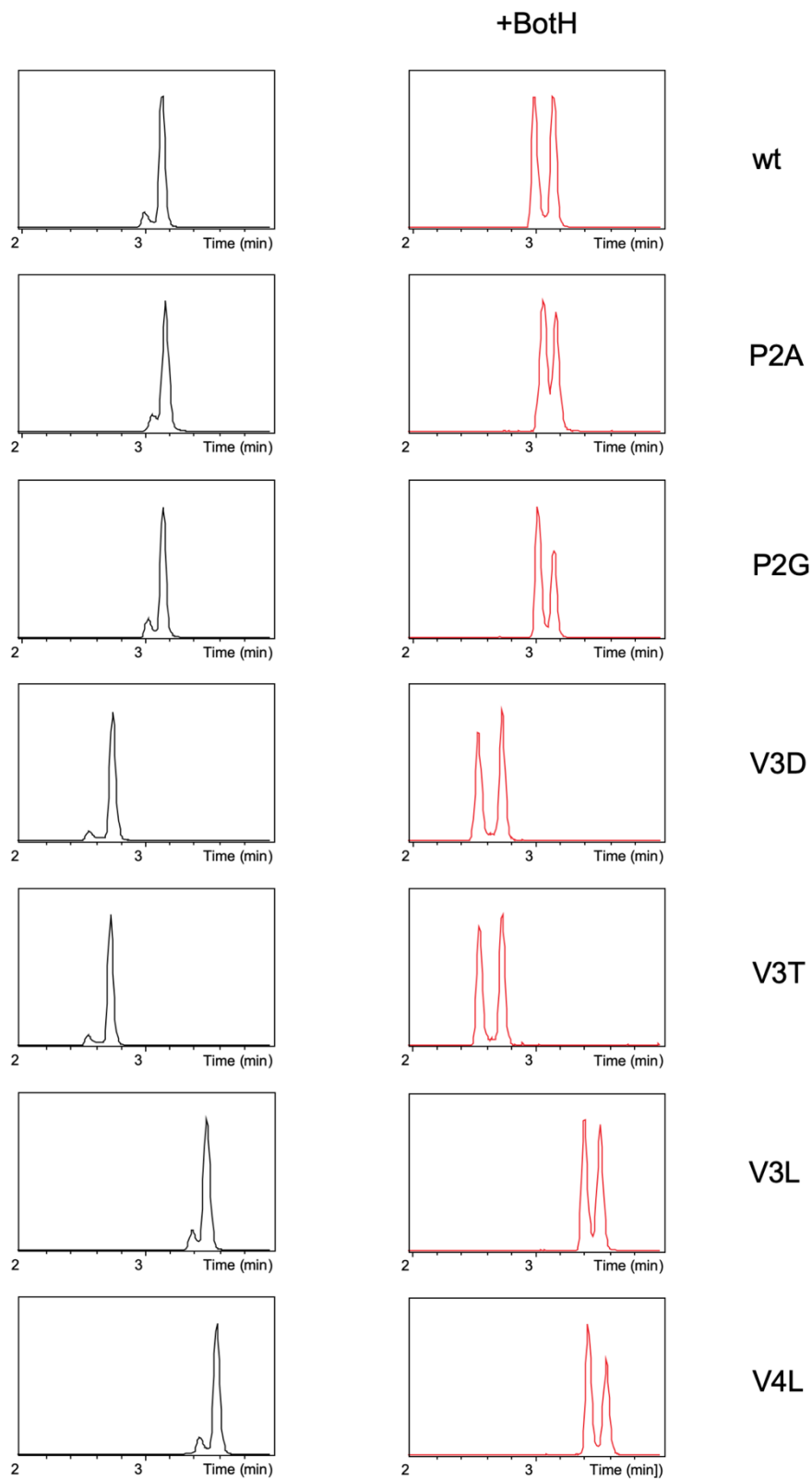
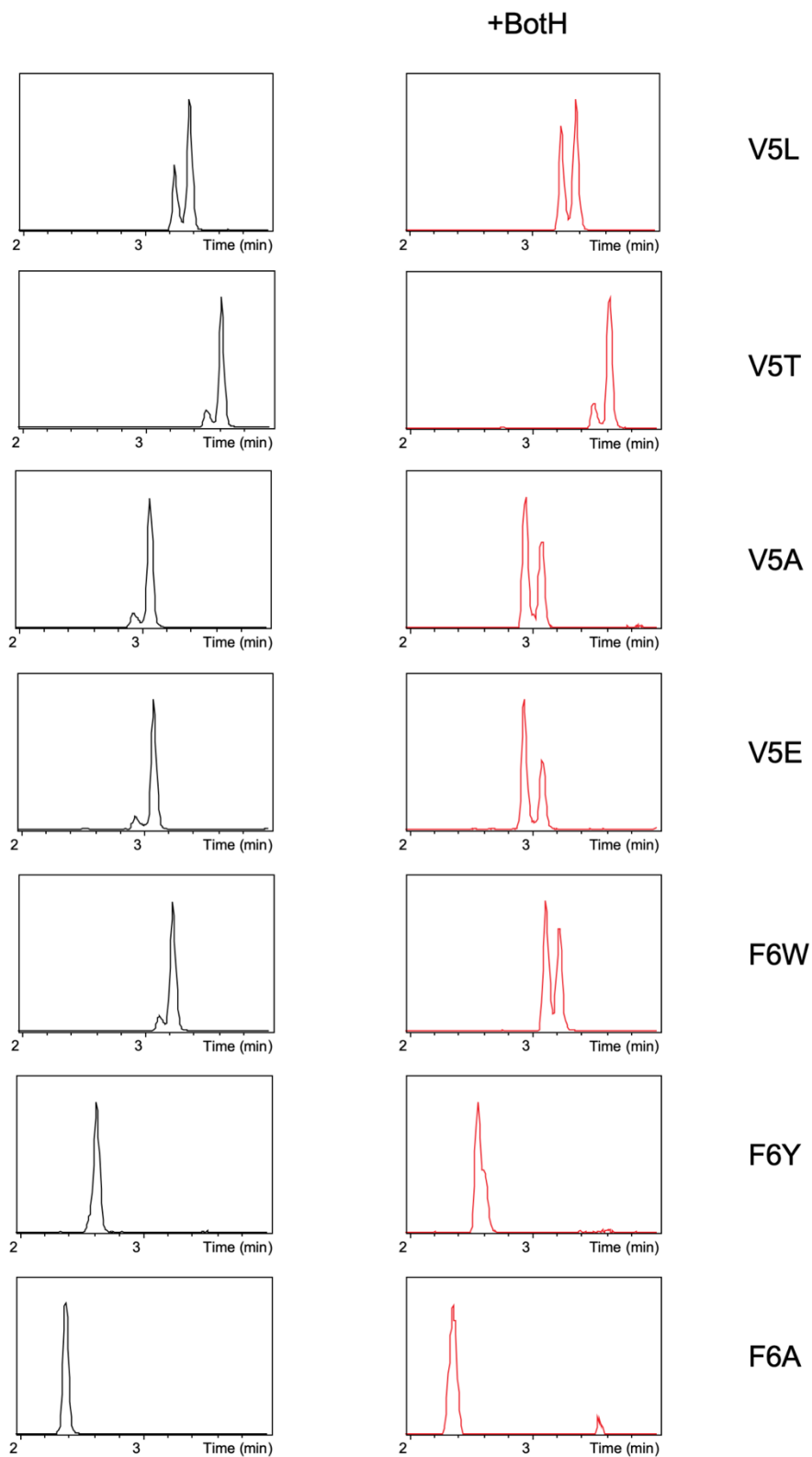


Figure S10: Analysis of the effects of point mutations in the core peptide on BotH activity. Each mutant was analyzed by HRLC-MS in the absence (left column, black) and presence (right column, red) of BotH. Shown are EICs for the monoisotopic  $[M+H]^+$  mass  $\pm 5$  ppm. A complete list of masses is given in Table S6. As can be seen, all mutations in positions 2 – 4 are accepted by the enzyme.



*Figure S11: Analysis of the effects of point mutations in the core peptide on BotH activity. Each mutant was analyzed by HRLC-MS in the absence (left column, black) and presence (right column, red) of BotH. Shown are EICs for the monoisotopic  $[M+H]^+$  mass  $\pm 5$  ppm. A complete list of masses is given in Table S6. Mutations V5T and F6A are not accepted by the enzyme.*

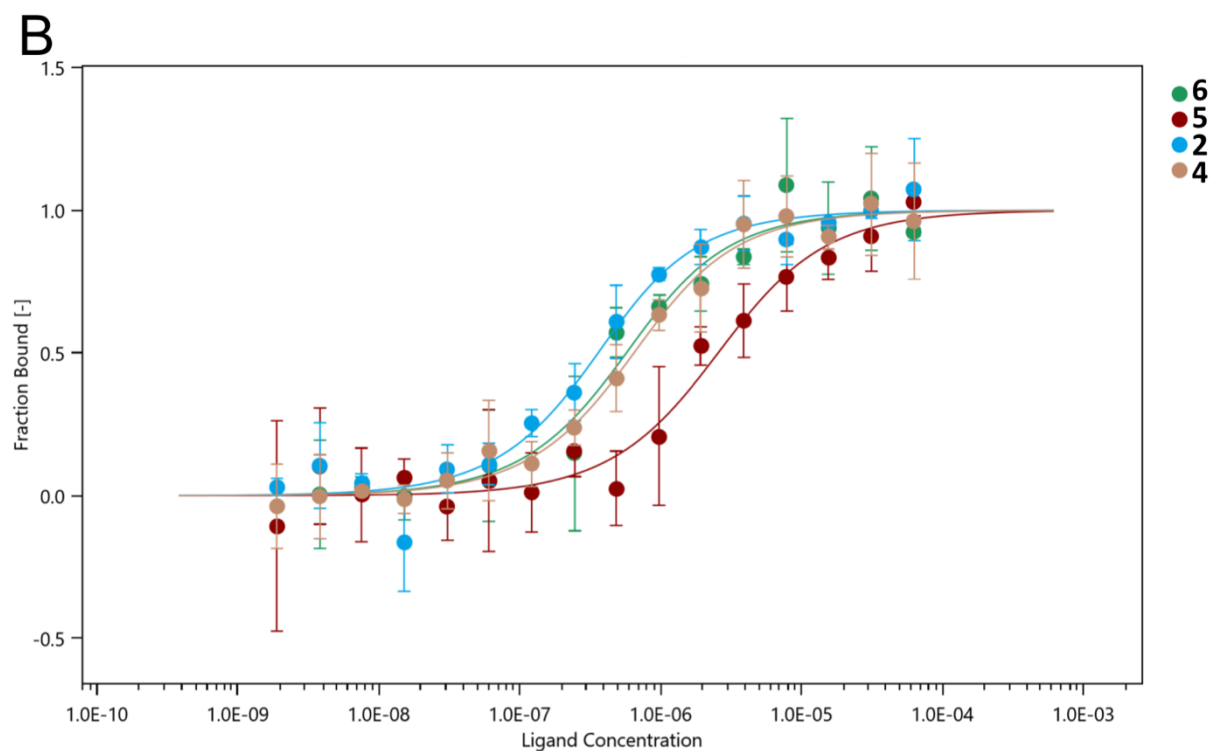
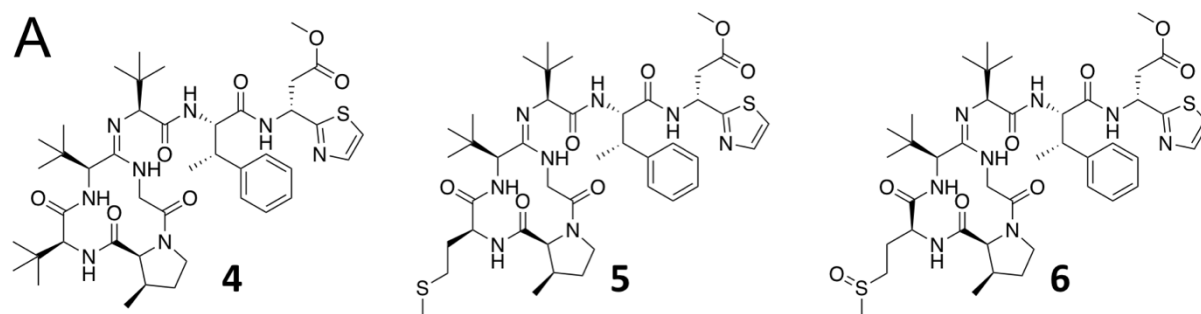


Figure S12: (A) Chemical structures of the three bottromycin derivatives used in this study. (B) MST measurements to determine the affinities of bottromycin A2 (**2**) and the three derivatives shown in A for BotH. The affinity of BotH for **2** is  $232 \pm 81$  nM. The additional methyl group at the  $C_{\beta}$  position of Val3 of (**4**)<sup>50</sup> reduced the affinity to a  $K_D$  of  $472 \pm 62$  nM, while the Val3Met mutation of **5**<sup>50</sup> reduced the affinity by an order of magnitude to a  $K_D$  of  $3.2 \pm 1.9$   $\mu$ M. Oxidation of the methionine sulfur of **5** (**6**) resulted in a  $K_D$  of  $459 \pm 178$  nM. A MS<sup>2</sup> spectrum for **6** can be found in Figure S22. Each curve represents three independent samples and the error bars represent standard deviations.

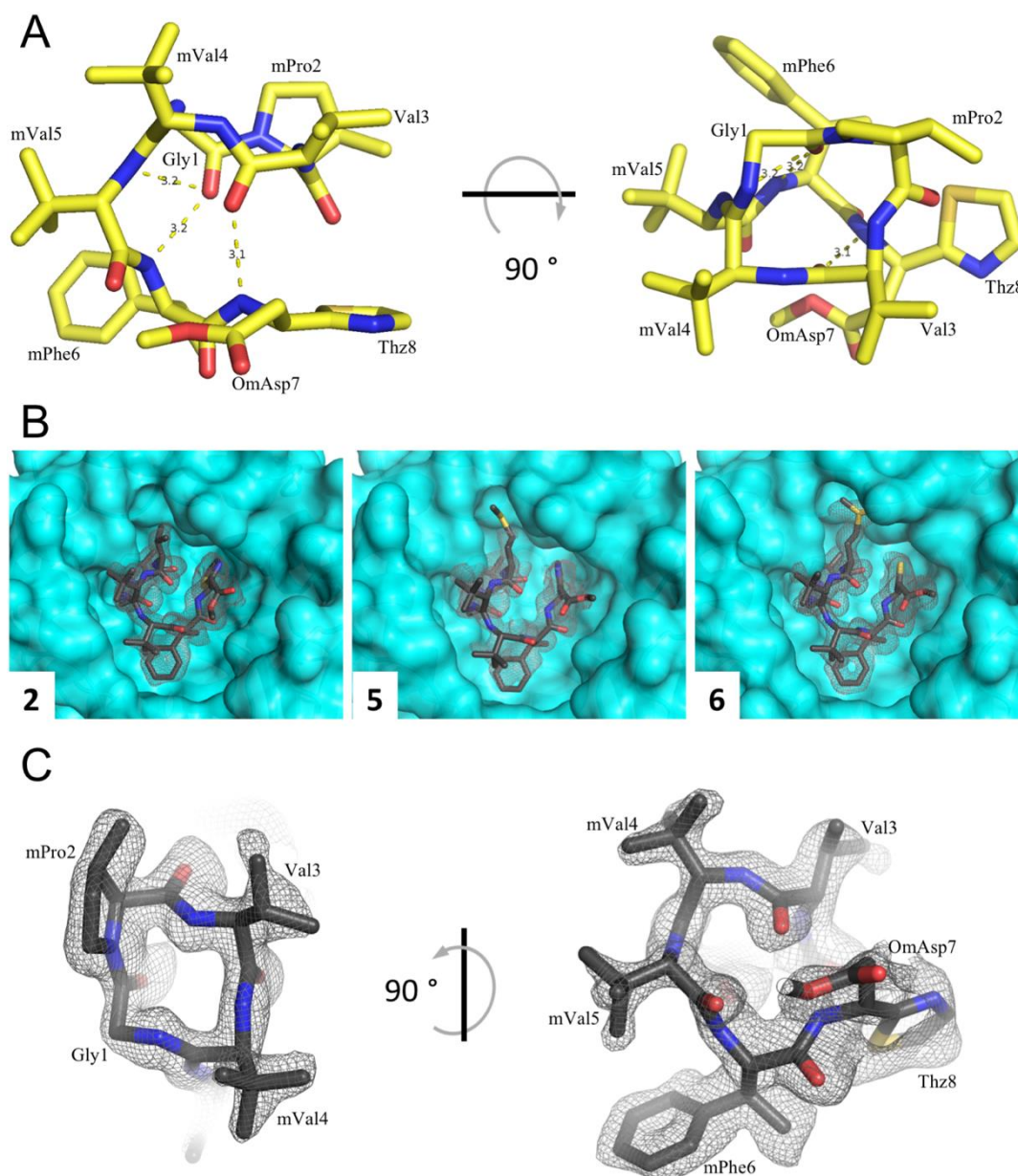


Figure S13: (A) Bottromycin A2 (**2**) as observed in the active-site of BotH. Intramolecular hydrogen bonds are shown as dashed lines and distances given in Å. (B) Polder maps (grey isomesh) for bottromycin A2 (**2**) and bottromycin derivatives **5** and **6** bound at the BotH active-site. BotH is shown as a cyan surface representation, the ligands as sticks. The polder maps were contoured at 3 (**2**), 2.5 (**5**) and 3 (**6**)  $\sigma$ . (C) Close-up polder map of bottromycin A2 from **B** shown as a grey isomesh contoured at 3  $\sigma$ .

The structures also provide insights into the varying affinities of the bottromycin derivatives: Based on the BotH-**2** complex structure, the additional methyl group of **4** clashes with BotH Glu148, which is involved in hydrogen bonding to the compound and BotH residue Tyr160, which results in a 2-fold weaker affinity. The BotH-**5** complex shows that the hydrophobic side-chain of the methionine is forced into a very narrow, highly polar opening lined by Glu148 and His164, which may explain the marked loss in affinity. In the BotH-**6** complex structure, the oxidation of the methionine sulfur triggers a rotation of His164, which allows the protein to easily accommodate the bulky side-chain and enables the formation of a new hydrogen bond between the substrate (methionine sulfoxide oxygen) and the side-chain of BotH Arg168. This rationalizes the increase in affinity when compared to unoxidized **5**.



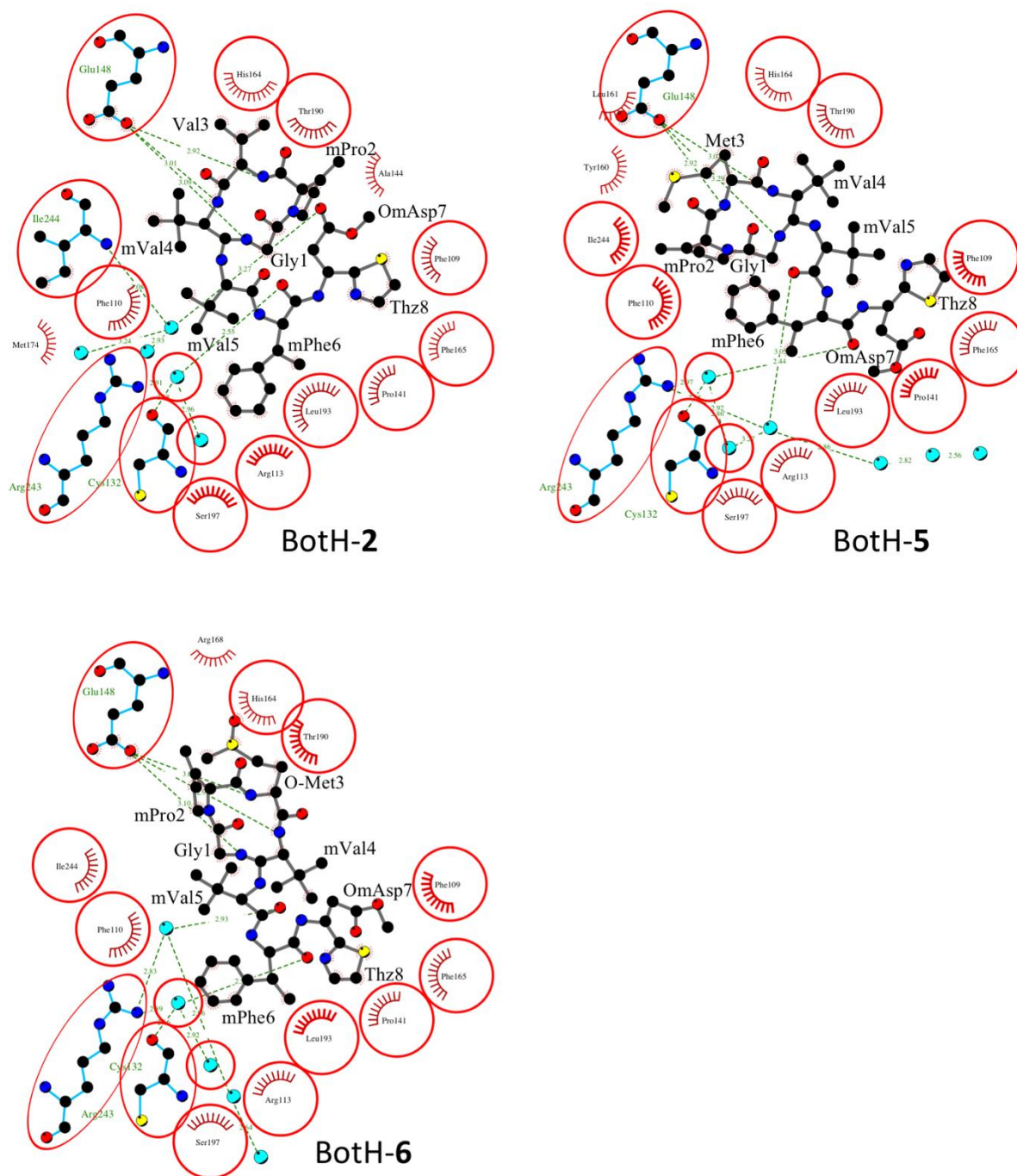
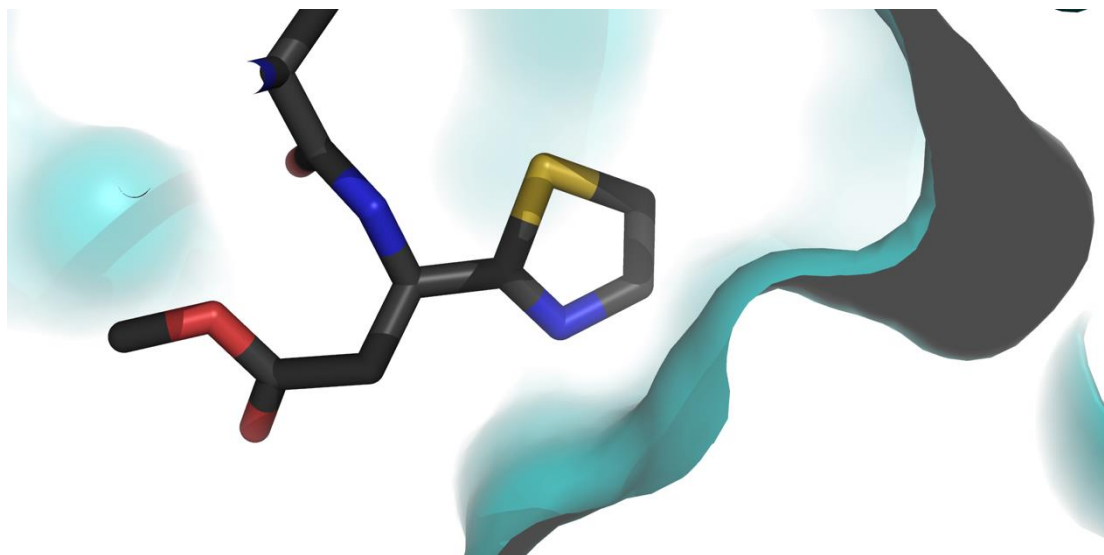


Figure S14: LigPlus diagrams of the three BotH complex structures with bottromycin A2 (**2**) and two analogs (**5** and **6**). Ligands are shown with bonds in grey, BotH with bonds in cyan. Water molecules are depicted as cyan spheres, intermolecular hydrogen bonds are shown as dashed lines with distances in Å and hydrophobic contacts are shown as red spiked arcs. Residues and water molecules engaged in the same interactions in all three structures are circled in red.



*Figure S15: Close-up of the thiazole found in the BotH-bottromycin A2 complex structure. Bottromycin A2 is shown as sticks, BotH as a cyan surface representation. As can be seen, flipping the thiazole 180° such that sulfur and nitrogen swap positions does not lead to a clash and is thus also an allowed orientation.*

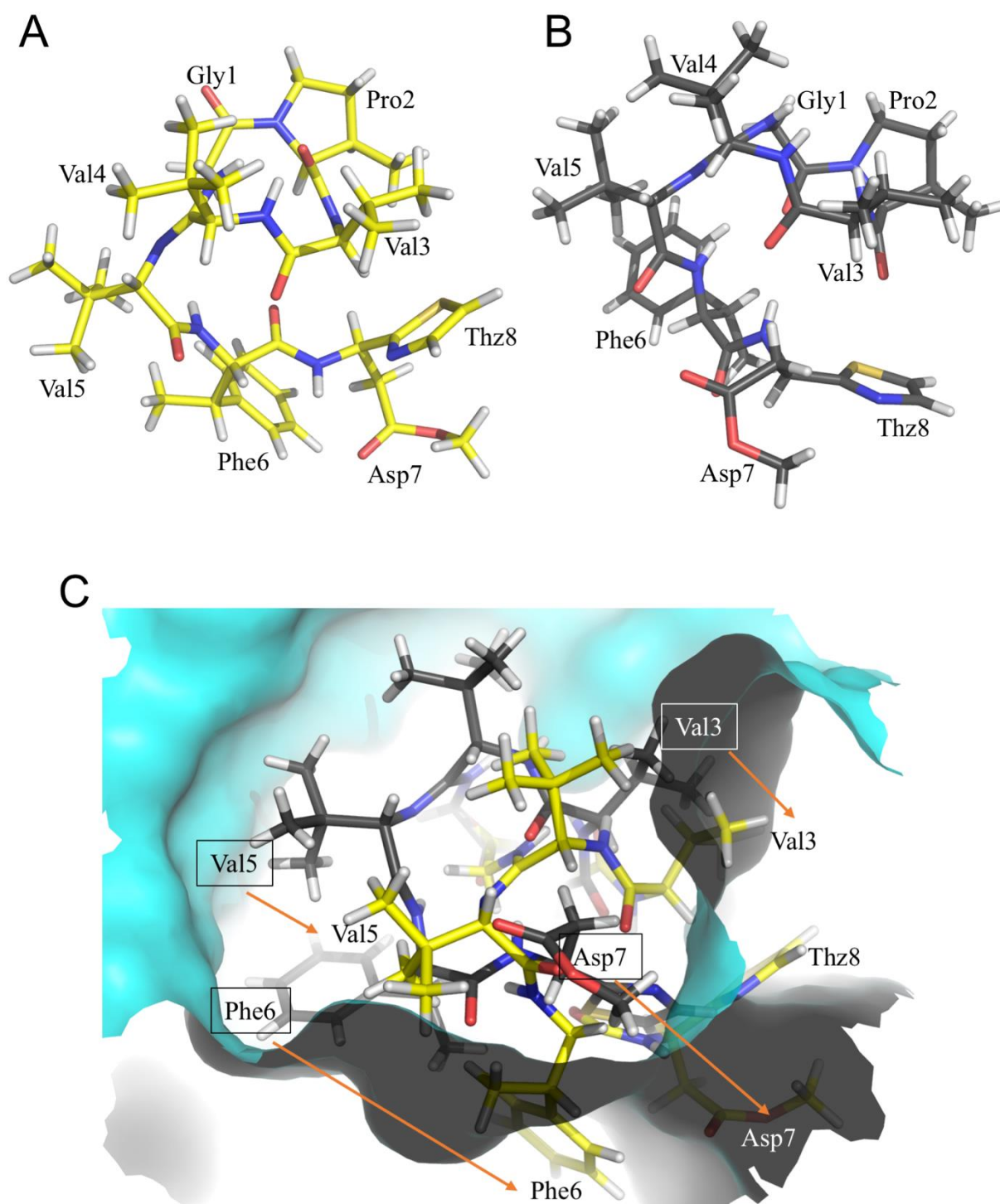


Figure S16: (A) Solution NMR structure of bottromycin A2 shown as yellow sticks. (B) Structure of bottromycin A2 when bound to BotH shown as grey sticks. The same orientation and magnification is used for A and B. Residues are labeled, Thz8 = Thiazole in position 8. (C) Superposition of the solution NMR structure of bottromycin A2 and the bottromycin A2 bound to BotH. Same color scheme as A and B, Pro2 was used as the reference residue. As can be seen, binding to BotH causes a significant conformational change and introduces strain. The more relaxed solution state (yellow) clashes with the protein in positions Val3, Phe6 and Thz8. Boxed labels belong to the complex crystal structure, orange arrows indicate movement required to reach unbound state.

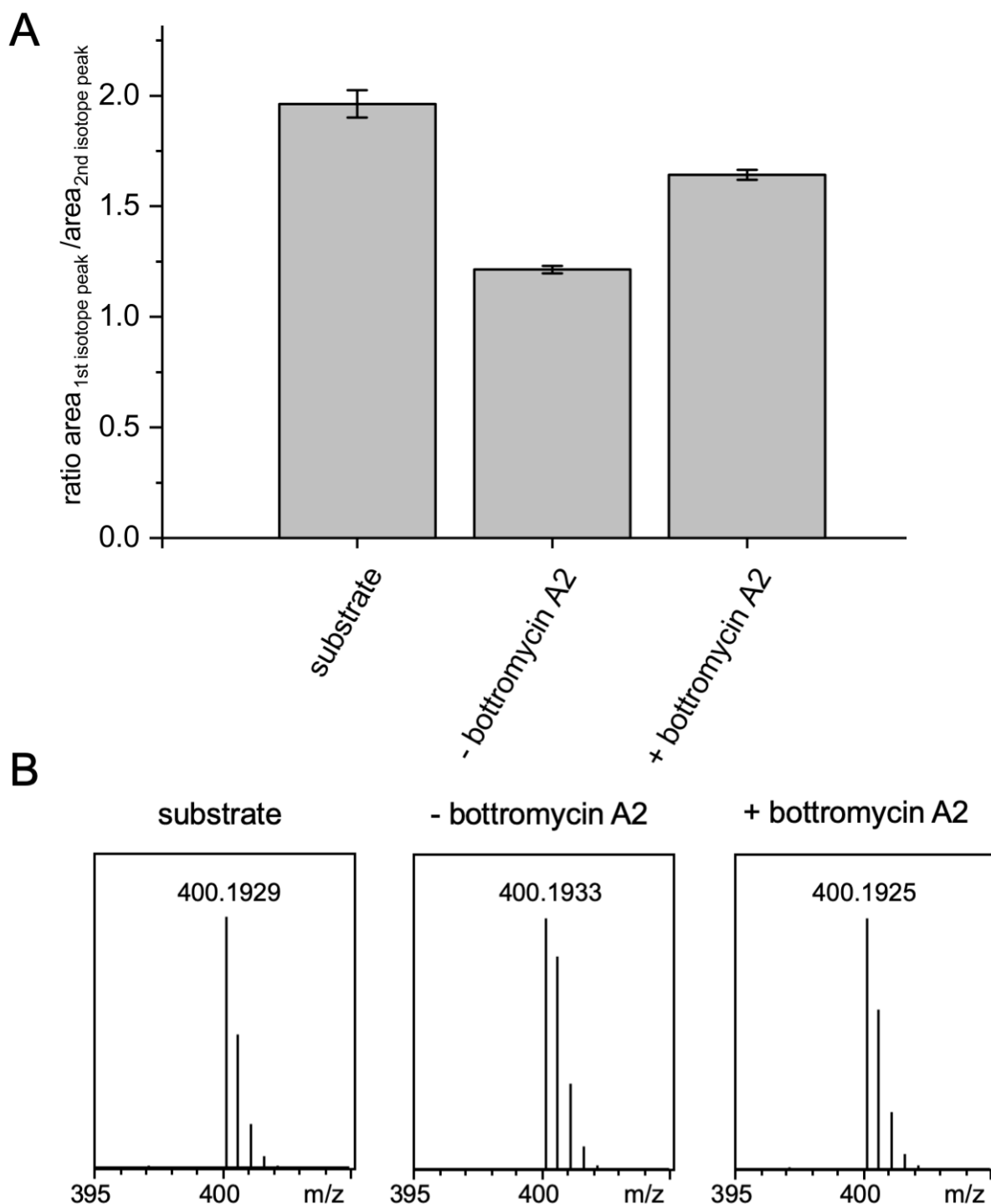
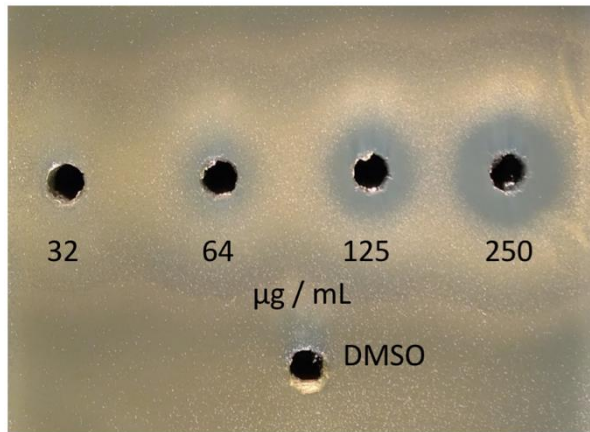
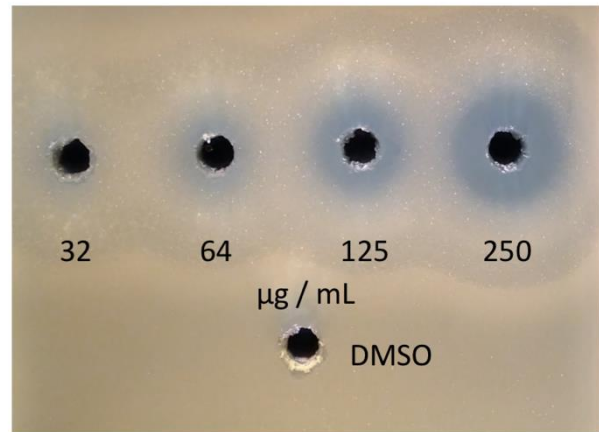


Figure S17: Orthosteric inhibition of BotH epimerization by bottromycin A2. The substrate **3a/b** (10  $\mu$ M in D<sub>2</sub>O buffer) was incubated with BotH (0.2  $\mu$ M) with and without the presence of bottromycin A2 (50  $\mu$ M) for 7 min at 30 °C. BotH activity was measured by the incorporation of a deuterium, which leads to an increase of the 2<sup>nd</sup> isotope peak. In the presence of bottromycin A2, the incorporation of a deuterium in **3a** by BotH is significantly lower. (A) Ratios of the area of the 1<sup>st</sup> to the 2<sup>nd</sup> isotope peak of [**3a**+2H]<sup>2+</sup>. Shown are means  $\pm$  SD (n=3). Differences were calculated to be extremely significant (p-value < 0.0001) (B) Mass spectra (**3a**, 3.02 min) of the substrate (**3a/b** in D<sub>2</sub>O buffer), the substrate after 7 min incubation with BotH and the substrate after 7 min incubation with BotH in presence of bottromycin A2.

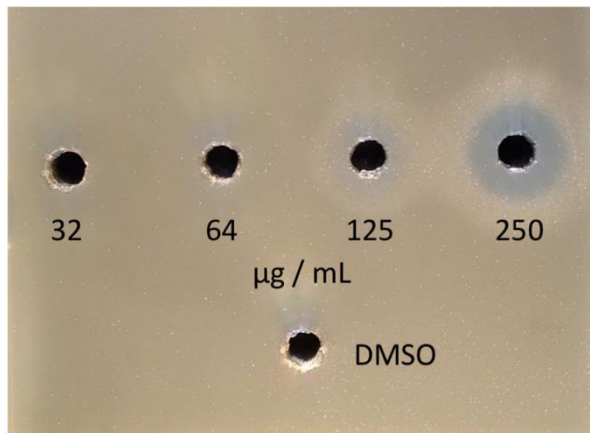
*S. coelicolor* M145



*S. coelicolor* M145 + Epimerase



*S. coelicolor* M145 + Transporter



*S. coelicolor* M145 + Both

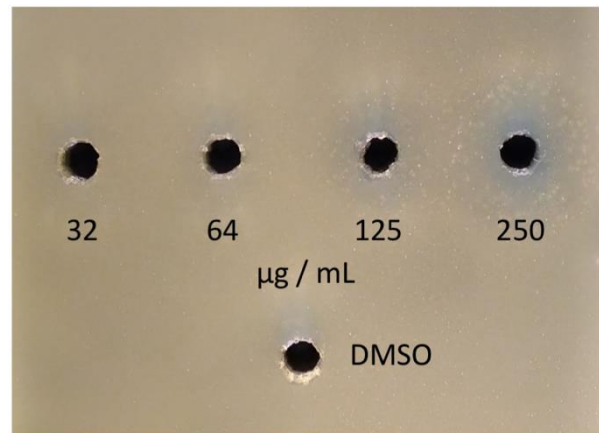


Figure S18: Zone of inhibition assays to demonstrate the effect of increasing concentrations of botromycin A2 on the susceptible strain *S. coelicolor* M145. Experiments were carried out by expressing the very close *BotH* (*BtmH*) and *BotT* (*BtmA*, transporter) homologs from *S. scabies*. Expression of *BtmH* (Epimerase) had no observable effect on resistance, while expression of *BtmA* (Transporter) expectedly increased resistance to botromycin A2. Interestingly, expression of both proteins rendered *S. coelicolor* M145 almost completely resistant within the concentration range tested. Pictures show representative experiments of triplicates.

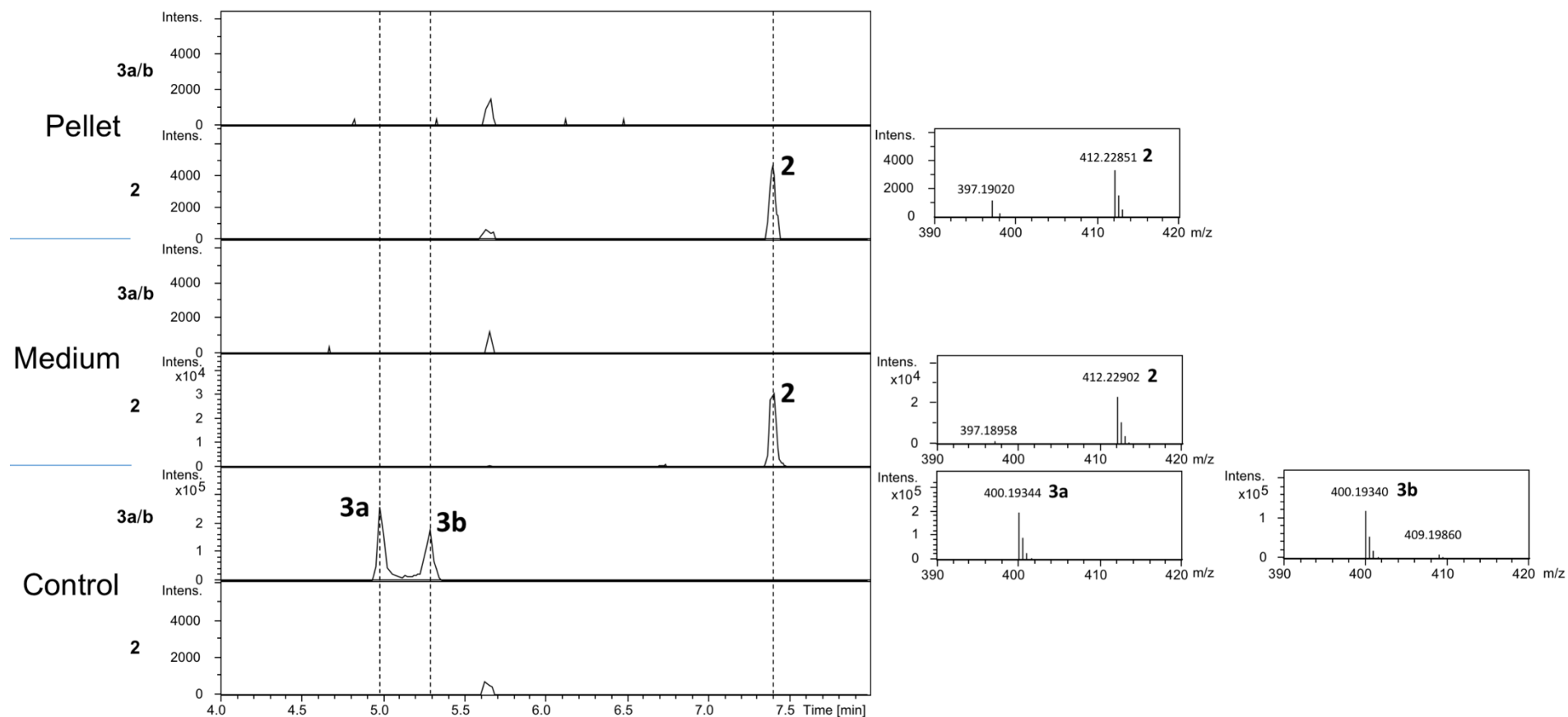


Figure S19: Pull-down data for **2** and **3a/b**. His<sub>6</sub>-tagged BotH was incubated with equal volumes of either lysed cell pellet supernatant (top) or spent culture medium (middle) of the heterologous bottromycin producer *S. sp. DG2-kmP41hyg*<sup>50</sup>. Purified **3a/b** was used as a control (bottom). The BotH was then removed from the mixture using magnetic Ni<sup>2+</sup>-NTA beads, washed thoroughly and eluted from the beads. Denaturing the enzyme with ACN liberated bound ligands. Only bottromycin A2 (**2**) and traces of **4** (data not shown), but not the BotH substrate **3a/b** could be found in the experimental samples. Shown are the EICs (right) and mass spectra for labeled peaks (right) for either **3a/b** or **2** ± 5 ppm using the [M + 2H]<sup>2+</sup> masses.

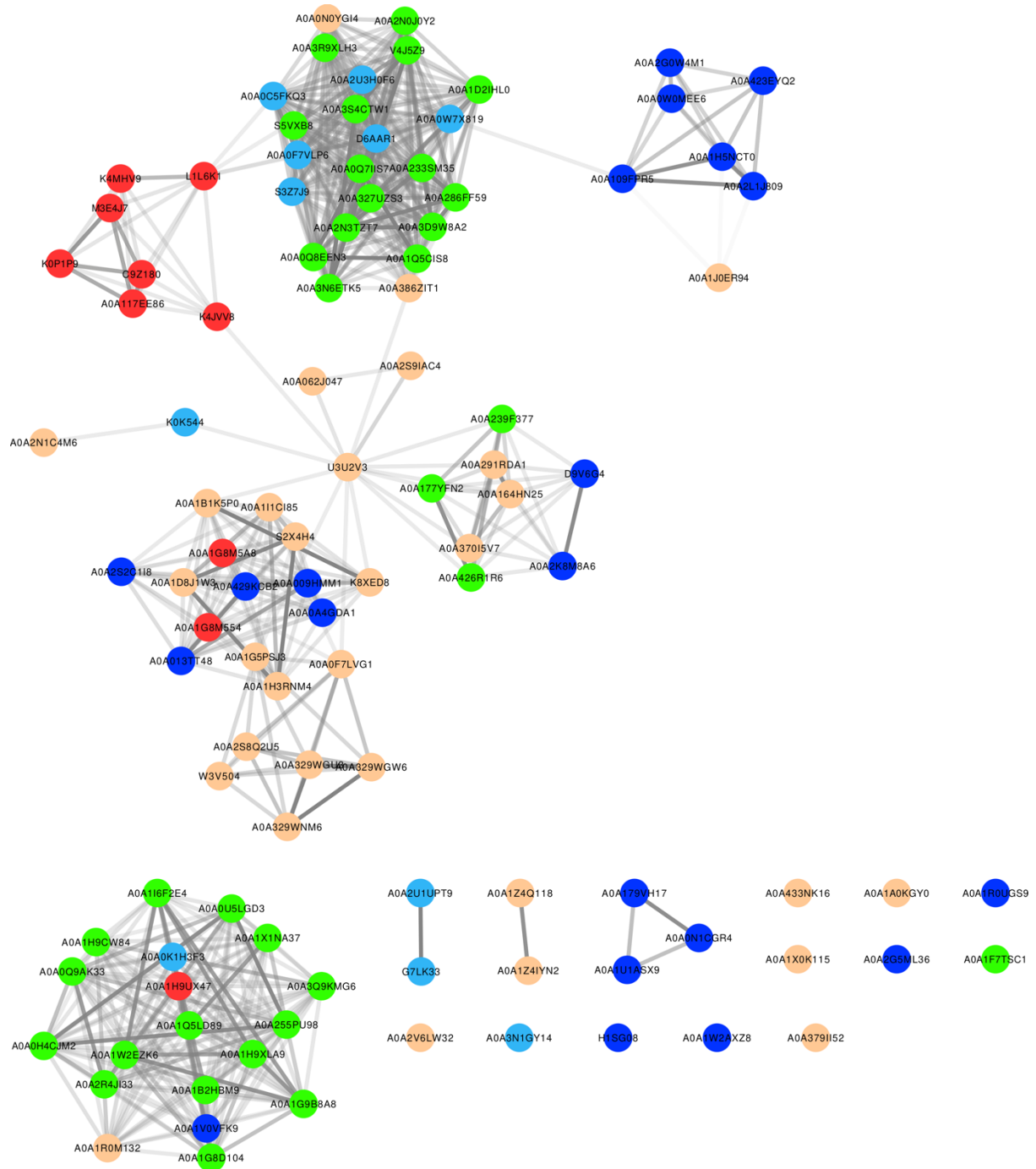


Figure S20: Sequence similarity network for *BotH*-like proteins lying in or near (closer than 1000 nt) biosynthetic gene clusters (BGCs). Nodes represent individual proteins (Accession number given) and colored according to the cluster type (red: RiPP, blue: NRPS, green: PKS, sky: NRPS/PKS, cantaloupe: other), edges represent similarity relationships and are colored according to sequence identity (lighter: less identical, darker: more identical). Edges corresponding to sequence identity below 15% are omitted.

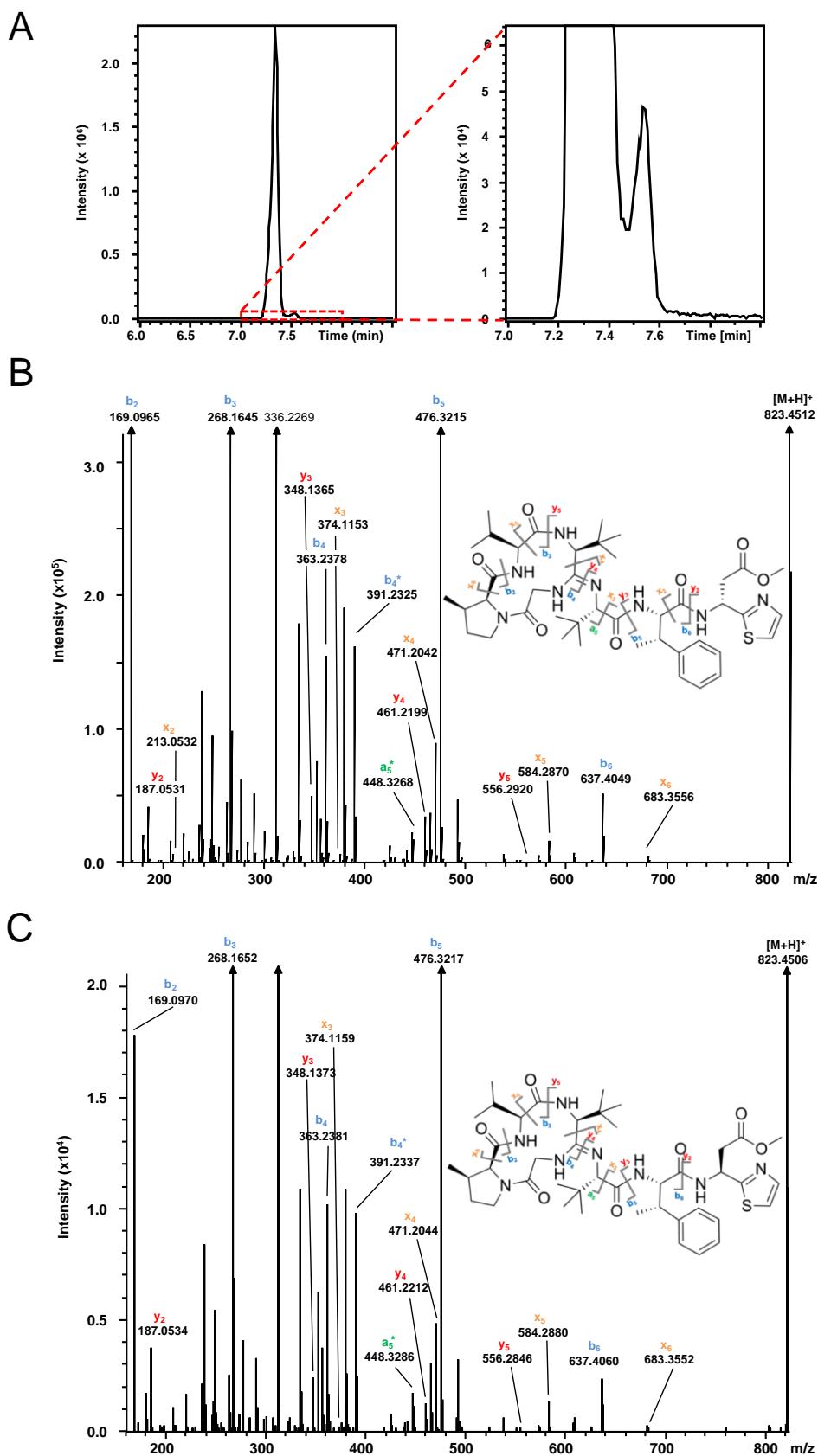


Figure S21: (A) EIC of bottromycin A2 (2) crude extract ( $823.4540 \pm 1$  ppm). (B)  $MS^2$  fragmentation of bottromycin A2 (2) main peak (*o*-Met-*D*-Asp-containing compound; retention time 7.3 min). The associated peak lists can be found in Table S4A. (C)  $MS^2$  fragmentation of minor bottromycin peak (most probably *o*-Met-*L*-Asp-containing compound; retention time 7.55 min). The associated peak lists can be found in Table S4B.



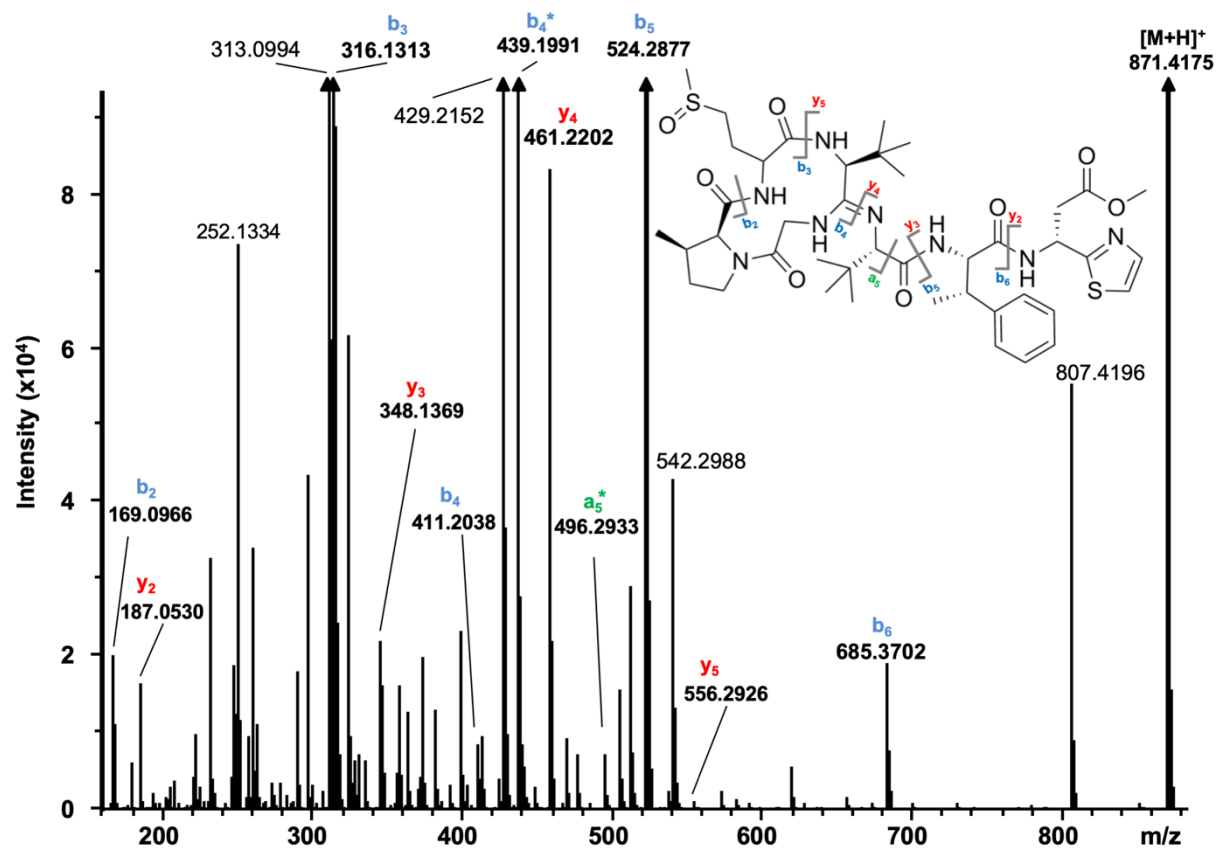


Figure S22: MS<sup>2</sup> fragmentation spectrum of the bottromycin analog **6**. The associated peak lists can be found in Table S5.

**Table S1:** Data collection and refinement statistics

	<i>BotH_SeMet</i>	<i>BotH_apo</i>	<i>BotH-3a</i>	<i>BotH-2</i>	<i>BotH-5</i>	<i>BotH-6</i>
<b>PDB ID</b>	<i>6T6H</i>	<i>6T6H</i>	<i>6T6X</i>	<i>6T6Y</i>	<i>6T6Z</i>	<i>6T70</i>
<b>Data collection</b>						
Space group	<i>I222</i>	<i>I222</i>	<i>I222</i>	<i>I222</i>	<i>I222</i>	<i>I222</i>
Cell dimensions						
<i>a, b, c (Å)</i>	<i>66.7, 80.1, 88.7</i>	<i>66.7, 80.1, 88.7</i>	<i>66.7, 80.1, 89.5</i>	<i>67.0, 80.1, 89.4</i>	<i>66.7, 79.5, 89.0</i>	<i>66.5, 79.4, 88.4</i>
□□□□□□□□□□□□	<i>90.0, 90.0, 90.0</i>	<i>90.0, 90.0, 90.0</i>	<i>90.0, 90.0, 90.0</i>	<i>90.0, 90.0, 90.0</i>	<i>90.0, 90.0, 90.0</i>	<i>90.0, 90.0, 90.0</i>
□□ °)						
Wavelength (Å)	<i>0.97941</i>					
Resolution (Å)	<i>1.70 (1.73-1.70) *</i>	<i>1.18 (1.20-1.18)</i>	<i>1.25 (1.27-1.25)</i>	<i>1.40 (1.42-1.40)</i>	<i>1.70 (1.73-1.70)</i>	<i>1.58 (1.61-1.58)</i>
<i>R<sub>sym</sub> or R<sub>merge</sub></i>	<i>7.2 (50.0)</i>	<i>3.5 (63.4)</i>	<i>7.8 (72.5)</i>	<i>6.9 (80.0)</i>	<i>10.4 (77.8)</i>	<i>6.0 (71.2)</i>
<i>I / σI</i>	<i>47.6 (8.7)</i>	<i>21.4 (2.5)</i>	<i>11.4 (4.2)</i>	<i>14.5 (2.6)</i>	<i>10.5 (2.4)</i>	<i>13.5 (2.0)</i>
Completeness (%)	<i>100 (100)</i>	<i>98.7 (93.5)</i>	<i>99.0 (100)</i>	<i>99.4 (99.5)</i>	<i>99.2 (99.0)</i>	<i>99.8 (99.9)</i>
Redundancy	<i>26.7 (27.8)</i>	<i>6.4 (5.7)</i>	<i>4.3 (4.1)</i>	<i>6.6 (6.7)</i>	<i>5.3 (5.4)</i>	<i>4.3 (4.4)</i>
<b>Refinement</b>						
Resolution (Å)		<i>44.35-1.18</i>	<i>44.77-1.25</i>	<i>44.71-1.40</i>	<i>44.52-1.70</i>	<i>39.72-1.58</i>
No. reflections		<i>77,123</i>	<i>65,761</i>	<i>47,302</i>	<i>26,141</i>	<i>32,386</i>
<i>R<sub>work</sub> / R<sub>free</sub></i>		<i>0.157 / 0.168</i>	<i>0.166 / 0.175</i>	<i>0.159 / 0.182</i>	<i>0.169 / 0.207</i>	<i>0.166 / 0.200</i>
No. atoms		<i>2,267</i>	<i>2,296</i>	<i>2,222</i>	<i>2,244</i>	<i>2,216</i>
Protein		<i>1,929</i>	<i>1,934</i>	<i>1,939</i>	<i>1,939</i>	<i>1,936</i>
Ligand/ion		<i>21</i>	<i>56</i>	<i>58</i>	<i>59</i>	<i>79</i>
Water		<i>317</i>	<i>306</i>	<i>225</i>	<i>246</i>	<i>201</i>
B-factors		<i>20.15</i>	<i>14.84</i>	<i>25.07</i>	<i>22.22</i>	<i>27.14</i>
Protein		<i>17.89</i>	<i>12.60</i>	<i>23.43</i>	<i>20.66</i>	<i>25.68</i>
Ligand/ion		<i>69.91</i>	<i>21.87</i>	<i>39.49</i>	<i>35.40</i>	<i>44.70</i>
Water		<i>30.62</i>	<i>27.69</i>	<i>35.52</i>	<i>31.38</i>	<i>34.24</i>
R.m.s. deviations						
Bond lengths (Å)		<i>0.011</i>	<i>0.008</i>	<i>0.016</i>	<i>0.005</i>	<i>0.012</i>
Bond angles (°)		<i>1.14</i>	<i>1.23</i>	<i>1.43</i>	<i>0.90</i>	<i>1.25</i>

\*I crystal per structure. \*Values in parentheses are for highest-resolution shell.

Table S2: MS<sup>2</sup> fragmentation of **3a** (A) and **3b** (B). Calculated and observed b- and y- ions.

**A**

<i>Seq.</i>	<i>b<sub>n</sub></i>	<i>Obs. b</i>	<i>Calc. b</i>	<i>y<sub>n</sub></i>	<i>Obs. y</i>	<i>Calc. y</i>
<i>G<sup>a</sup></i>	1	-	-	8	799.3793	799.3807
<i>P</i>	2	155.0812	155.0821	7	-	-
<i>V</i>	3	254.1494	254.1605	6	-	645.3070
<i>V<sup>a</sup></i>	4	335.2078	335.2083	5	-	546.2386
<i>V</i>	5	434.2755	434.2767	4	465.1758	465.1808
<i>F</i>	6	581.3397	581.3451	3	366.1180	366.1124
<i>D<sup>b</sup></i>	7	-	-	2	219.0453	219.0439
<i>C<sup>b</sup></i>	8	-	781.3707	1	-	-

- c. Gly1 macrocyclisation with Val4 carbonyl  
d. Cys8 heterocyclisation with Asp7 carbonyl

Additionally, fragment *a<sub>5</sub>* (obs. 406.2822; calc. 406.2818) and *b<sub>4</sub><sup>\*</sup>* (obs. 363.2023; calc. 363.2032), which are characteristically associated with the bottromycin macrocycle and have been observed previously for bottromycins and macrocyclic peptides, were detected.

**B**

<i>Seq.</i>	<i>b<sub>n</sub></i>	<i>Obs. b</i>	<i>Calc. b</i>	<i>y<sub>n</sub></i>	<i>Obs. y</i>	<i>Calc. y</i>
<i>G<sup>a</sup></i>	1	-	-	8	799.3788	799.3807
<i>P</i>	2	155.0817	155.0821	7	-	-
<i>V</i>	3	254.1497	254.1605	6	645.3031	645.3070
<i>V<sup>a</sup></i>	4	335.2076	335.2083	5	546.2343	546.2386
<i>V</i>	5	434.2752	434.2767	4	465.1770	465.1808
<i>F</i>	6	581.3440	581.3451	3	366.1098	366.1124
<i>D<sup>b</sup></i>	7	-	-	2	219.0422	219.0439
<i>C<sup>b</sup></i>	8	781.3646	781.3707	1	-	-

- a. Gly1 macrocyclisation with Val4 carbonyl  
b. Cys8 heterocyclisation with Asp7 carbonyl

Additionally, fragment *a<sub>5</sub>* (obs. 406.2825; calc. 406.2818) and *b<sub>4</sub><sup>\*</sup>* (obs. 363.2020; calc. 363.2032), which are characteristically associated with the bottromycin macrocycle and have been observed previously for bottromycins and macrocyclic peptides, were detected.

Table S3: MS<sup>2</sup> fragmentation of Asp C $\alpha$  deuterated **3a** (A) and **3b** (B). Calculated and observed b- and y- ions.

**A**

Seq.	$b_n$	Obs. $b$	Calc. $b$	$y_n$	Obs. $y$	Calc. $y$	$x_n$	Obs. $x$	Calc. $x$
$G^a$	1	-	-	8	800.3838	800.3870	8	-	-
$P$	2	155.0827	155.0821	7	-	-	7	-	-
$V$	3	254.1501	254.1605	6	-	646.3133	6	-	674.3082
$V^a$	4	335.2064	335.2083	5	-	547.2449	5	575.2318	575.2398
$V$	5	434.2767	434.2767	4	466.1891	466.1871	4	476.1676	476.1714
$F$	6	581.3461	581.3451	3	367.1142	367.1186	3	-	393.0979
$D^{b,c}$	7	-	-	2	220.0491	220.0502	2	-	246.0995
$C^b$	8	-	782.3770	1	-	-	1	-	-

- a. Gly1 macrocyclisation with Val4 carbonyl
- b. Cys8 heterocyclisation with Asp7 carbonyl
- c. Asp7 C $\alpha$ -positon labeled with deuterium

Fragments  $a_5$  (obs. 406.2800; calc. 406.2818) and  $b_4^*$  (obs. 363.2006; calc. 363.2032), which are characteristically associated with the bottromycin macrocycle and have been previously observed for bottromycins and macrocyclic peptides, were detected. Additionally, fragment  $z_5$  (obs. 530.2191; calc. 530.2184) was detected.

**B**

Seq.	$b_n$	Obs. $b$	Calc. $b$	$y_n$	Obs. $y$	Calc. $y$	$x_n$	Obs. $x$	Calc. $x$
$G^a$	1	-	-	8	800.3831	800.3870	8	-	-
$P$	2	155.0809	155.0821	7	-	-	7	-	-
$V$	3	254.1489	254.1605	6	-	646.3133	6	-	674.3082
$V^a$	4	335.2064	335.2083	5	547.2449	547.2449	5	575.2329	575.2398
$V$	5	434.2741	434.2767	4	466.1796	466.1871	4	476.1685	476.1714
$F$	6	581.3440	581.3451	3	367.1160	367.1186	3	-	393.0979
$D^{b,c}$	7	-	-	2	220.0486	220.0502	2	-	246.0995
$C^b$	8	-	782.3770	1	-	-	1	-	-

- a. Gly1 macrocyclisation with Val4 carbonyl
- b. Cys8 heterocyclisation with Asp7 carbonyl
- c. Asp7 C $\alpha$ -positon labeled with deuterium

Fragments  $a_5$  (obs. 406.2801; calc. 406.2818) and  $b_4^*$  (obs. 363.2013; calc. 363.2032), which are characteristically associated with the bottromycin macrocycle and have been previously observed for bottromycins and macrocyclic peptides, were detected. Additionally, fragment  $z_5$  (obs. 530.2181; calc. 530.2184) was detected.

Table S4: MS<sup>2</sup> fragmentation of the bottromycin A2 main (A) and minor (B). Calculated and observed b-, x- and y- ions.

**A**

Seq.	<i>b<sub>n</sub></i>	Obs. <i>b</i>	Calc. <i>b</i>	<i>y<sub>n</sub></i>	Obs. <i>y</i>	Calc. <i>y</i>	<i>x<sub>n</sub></i>	Obs. <i>x</i>	Calc. <i>x</i>
<i>G<sup>a</sup></i>	1	-	-	8	823.4512	823.4540	8	-	-
<i>P</i>	2	169.0965	169.0977	7	-	766.4326	7	-	794.4275
<i>V</i>	3	268.1645	268.1661	6	-	655.3642	6	683.3556	683.3591
<i>V<sup>a</sup></i>	4	363.2378	363.2396	5	556.2920	556.2957	5	584.2870	584.2907
<i>V</i>	5	476.3215	476.3237	4	461.2199	461.2223	4	471.2042	471.2066
<i>F</i>	6	637.4049	637.4077	3	348.1365	348.1382	3	374.1153	374.1175
<i>D<sup>b</sup></i>	7	-	-	2	187.0531	187.0541	2	213.0532	213.0334
<i>C<sup>b</sup></i>	8	-	-	1	-	-	1	-	-

- a. Gly1 macrocyclisation with Val4 carbonyl
- b. Cys8 heterocyclisation with Asp7 carbonyl

Additionally, fragments *a*<sub>5</sub> (obs. 448.3268; calc. 448.3288) and *b*<sub>4</sub>\* (obs. 391.2325; calc. 391.2345), which are characteristically associated with the bottromycin macrocycle and have been previously observed for bottromycins and macrocyclic peptides, were detected.

**B**

Seq.	<i>b<sub>n</sub></i>	Obs. <i>b</i>	Calc. <i>b</i>	<i>y<sub>n</sub></i>	Obs. <i>y</i>	Calc. <i>y</i>	<i>x<sub>n</sub></i>	Obs. <i>x</i>	Calc. <i>x</i>
<i>G<sup>a</sup></i>	1	-	-	8	823.4506	823.4540	8	-	-
<i>P</i>	2	169.0970	169.0977	7	-	766.4326	7	-	794.4275
<i>V</i>	3	268.1652	268.1661	6	-	655.3642	6	683.3552	683.3591
<i>V<sup>a</sup></i>	4	363.2381	363.2396	5	556.2846	556.2957	5	584.2880	584.2907
<i>V</i>	5	476.3217	476.3237	4	461.2212	461.2223	4	471.2044	471.2066
<i>F</i>	6	637.4060	637.4077	3	348.1373	348.1382	3	374.1159	374.1175
<i>D<sup>b</sup></i>	7	-	-	2	187.0534	187.0541	2	-	213.0334
<i>C<sup>b</sup></i>	8	-	-	1	-	-	1	-	-

- a. Gly1 macrocyclisation with Val4 carbonyl
- b. Cys8 heterocyclisation with Asp7 carbonyl

Additionally, fragments *a*<sub>5</sub> (obs. 448.3286; calc. 448.3288) and *b*<sub>4</sub>\* (obs. 391.2337; calc. 391.2345), which are characteristically associated with the bottromycin macrocycle and have been previously observed for bottromycins and macrocyclic peptides, were detected.

Table S5: MS<sup>2</sup> fragmentation of the bottromycin analog **6**. Calculated and observed b-, x- and y- ions.

Seq.	b <sub>n</sub>	Obs. b	Calc. b	y <sub>n</sub>	Obs. y	Calc. y
G <sup>a</sup>	1	-	-	8	871.4175	871.4210
P	2	169.0966	169.0977	7	-	814.3996
M-O	3	316.1313	316.1331	6	-	703.3312
V <sup>a</sup>	4	411.2038	411.2066	5	556.2926	556.2957
V	5	524.2877	524.2907	4	461.2202	461.2223
F	6	685.3702	685.3747	3	348.1369	348.1382
D <sup>b</sup>	7	-	-	2	187.0530	187.0541
C <sup>b</sup>	8	-	-	1	-	-

a. Gly1 macrocyclisation with Val4 carbonyl

b. Cys8 heterocyclisation with Asp7 carbonyl

Additionally, fragments a<sub>5</sub> (obs. 496.2933; calc. 496.2958) and b<sub>4</sub>\* (obs. 439.1991; calc. 439.2015), which are characteristically associated with the bottromycin macrocycle and have been previously observed for bottromycins and macrocyclic peptides, were detected.

Table S6: Sum formulas and calculated masses for the IpoC, PurCD and PurAH treated BotA<sup>P</sup> variants.

BotA <sup>P</sup> variant	sum formula	monoiso. neutral mass	[M+H] <sup>+</sup>
P2A	C <sub>36</sub> H <sub>52</sub> N <sub>8</sub> O <sub>9</sub> S	772.3578	773.3651
P2G	C <sub>35</sub> H <sub>50</sub> N <sub>8</sub> O <sub>9</sub> S	758.3421	759.3494
V3D	C <sub>37</sub> H <sub>50</sub> N <sub>8</sub> O <sub>11</sub> S	814.3320	815.3393
V3T	C <sub>37</sub> H <sub>52</sub> N <sub>8</sub> O <sub>10</sub> S	800.3527	801.3600
V3L	C <sub>39</sub> H <sub>56</sub> N <sub>8</sub> O <sub>9</sub> S	812.3891	813.3964
V4L	C <sub>39</sub> H <sub>56</sub> N <sub>8</sub> O <sub>9</sub> S	812.3891	813.3964
V5T	C <sub>37</sub> H <sub>52</sub> N <sub>8</sub> O <sub>10</sub> S	800.3527	801.3600
V5L	C <sub>39</sub> H <sub>56</sub> N <sub>8</sub> O <sub>9</sub> S	812.3891	813.3964
V5A	C <sub>36</sub> H <sub>50</sub> N <sub>8</sub> O <sub>9</sub> S	770.3421	771.3494
V5E	C <sub>38</sub> H <sub>52</sub> N <sub>8</sub> O <sub>11</sub> S	828.3476	829.3549
F6W	C <sub>40</sub> H <sub>55</sub> N <sub>9</sub> O <sub>9</sub> S	837.3843	838.3916
F6Y	C <sub>38</sub> H <sub>54</sub> N <sub>8</sub> O <sub>10</sub> S	814.3684	815.3756
F6A	C <sub>32</sub> H <sub>50</sub> N <sub>8</sub> O <sub>9</sub> S	722.3421	723.3494
D7A	C <sub>37</sub> H <sub>54</sub> N <sub>8</sub> O <sub>7</sub> S	754.3836	755.3909
D7N	C <sub>38</sub> H <sub>55</sub> N <sub>9</sub> O <sub>8</sub> S	797.3894	798.3967
wt	C <sub>38</sub> H <sub>54</sub> N <sub>8</sub> O <sub>9</sub> S	798.3734	799.3807

## **Chapter 4: Adaptation of a Bacterial Multidrug Resistance System Revealed by the Structure and Function of AlbA**

*Asfandyar Sikandar, Katarina Cirnski, Giambattista Testolin, Carsten Volz, Mark Brönstrup, Olga V. Kalinina, Rolf Müller and Jesko köhnke*

Reproduced with permission from *J. Am. Chem. Soc.* 2018, 140, 48, 16641-16649

Publication Date: November 13, 2018

<https://doi.org/10.1021/jacs.8b08895>

**Copyright © 2018 American Chemical Society**

# 1 Introduction

Natural products continue to be an abundant source of novel and biologically active molecules (e.g.<sup>1</sup>). The majority of drugs currently on the market to treat bacterial infections are natural products and their derivatives – they have become a major pillar of modern medicine and are regarded as essential for human health<sup>2</sup>. With the extensive use of antibiotics, however, the problem of (multi-)resistance development in human pathogens has risen<sup>3</sup>. To keep pace with resistance development and always have effective antibiotics available, constant antibiotic discovery and optimization efforts are required; a challenge that was not met in the past decades<sup>4</sup>. Most problematic with regard to drug development are Gram-negative pathogens, which are becoming a severe threat to public health<sup>5</sup>. Multidrug resistance (MDR) mechanisms pose a particular challenge, since such systems are able to recognize and neutralize structurally and chemically diverse compounds, while still being selective for a certain compound class<sup>6</sup>.

Antibiotic resistance usually results from one of four mechanisms: Modification of the bacterial cell wall to prevent antibiotic entry, expulsion of the antibiotic by general or specific efflux pumps, mutation of the cellular target, or chemical modification of the antibiotic<sup>7,8</sup>. These mechanisms (with the exception of target mutations) are often employed after bacterial antibiotic-specific biosensors bind to an antibiotic and subsequently trigger transcriptional programs<sup>9,10</sup>. One protein class involved in such processes are thiostrepton-induced protein A (TipA) systems<sup>11,12</sup>. TipA belongs to the superfamily of mercuric ion resistance (MerR) -like transcriptional regulators<sup>12-14</sup>, which contain an N-terminal helix-turn-helix (HTH) motif, followed by a coiled-coil region and a C-terminal effector binding domain<sup>15-17</sup>. The *tipA* gene contains two alternative start codons, giving rise to TipA-L and TipA-S protein isoforms (Figure 1A)<sup>17,18</sup>.

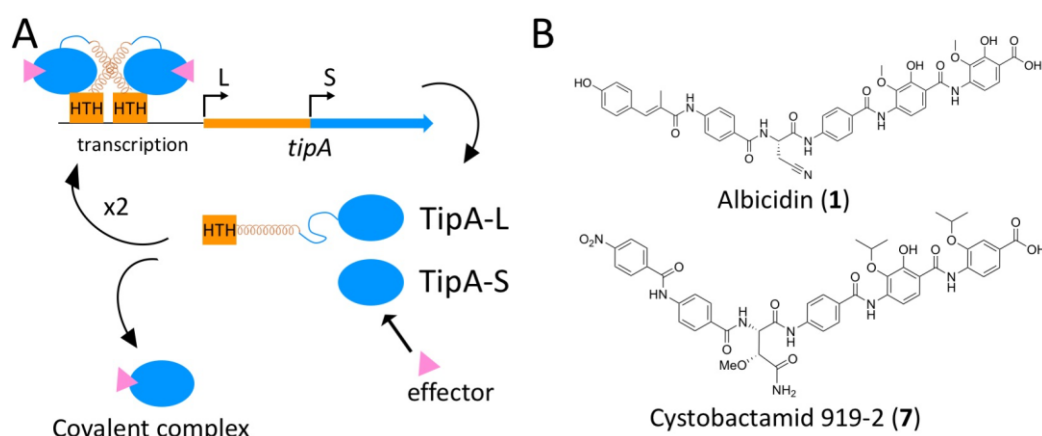


Figure 1. (A) Basic logic of the TipA system. The *tipA* gene contains two alternative start codons giving rise to TipA-L and TipA-S, which are identical in their C-terminal domains. While TipA-S scrubs free antibiotic from the cytoplasm, TipA-L



dimerizes after ligand binding and induces increased expression from the *tipA* gene. (B) Chemical structures of albicidin (1) and cystobactamid 919-2 (7).

While TipA-L follows the basic MerR architecture, TipA-S is the predominant form and consist only of the effector binding domain<sup>11,12,17,19</sup>. This domain is able to covalently bind a wide variety of thiopeptide antibiotics via an active site cysteine<sup>12,15,16</sup> and thus permanently sequester the antibiotics<sup>16</sup>. Upon substrate binding, TipA-L forms a dimer, which then binds to promoters via its HTH domain and induces the transcription of multiple genes, including *tipA*, which can thus be viewed as a minimal autoregulated MDR system<sup>11</sup>. Recently reported NMR-based structural studies of TipA-S allowed the proposition of a model for how this protein is able to neutralize a wide variety of thiopeptide antibiotics<sup>15,16</sup>: Upon thiopeptide binding TipA-S transitions from a partially unstructured to a fully ordered state<sup>16</sup>. In the process, a large substrate binding cleft is formed that contains the active-site cysteine, which reacts with the bound substrate to form a covalent bond<sup>16</sup>.

The natural products albicidin (1) and cystobactamid (7) (Figure 1B) were reported as potent anti-Gram-negative antibiotics with minimal inhibitory concentrations (MICs) in the sub- $\mu\text{g/mL}$  range against *Escherichia coli* and other relevant pathogens<sup>20,21</sup>. Albicidin is produced by the plant pathogen *Xanthomonas albilineans*<sup>22-24</sup>, and was found to be a strong inhibitor of DNA gyrase in Gram-negative bacteria (*E. coli*)<sup>25</sup>. The unusual structure of albicidin was reported in 2015: It consists of *p*-aminobenzoic acid (PABA) building blocks and cyanoalanine as a central unnatural amino acid (Figure 1B)<sup>26</sup>. Concurrently, cystobactamids, produced by myxobacteria, were discovered as a novel class of antibiotics that address the same bacterial target and possess a similar structure (Figure 1B)<sup>21</sup>. Several differences between the two compound classes exist: While the N-terminus of cystobactamids consists of a *p*-nitro-benzoic acid, the equivalent position in albicidin is occupied by 2-methyl-*p*-coumaric acid. Additionally, the two C-terminal building blocks in cystobactamid 919-2 are isopropoxylated in the 3-position, whereas albicidins are methoxylated, and the C-terminal building block of cystobactamid 919-2 lacks hydroxylation in the 2-position. Finally, the most potent cystobactamids have a central  $\beta$ -methoxy-asparagine amino acid instead of the cyanoalanine found in the original albicidin.

Since antibiotics such as albicidin and cystobactamids occur naturally, resistance mechanisms have already been developed. Three proteins have been reported as resistance proteins against albicidin: AlbA, AlbB and AlbD<sup>18,27,28</sup>. AlbA is a distant member of the TipA family and can be found in a variety of *Klebsiellae* strains including the problematic human pathogen *K. pneumoniae*. Earlier studies of AlbA reported the protein to bind albicidin at a single high-

affinity binding site with low nM affinity and that ligand-binding induced major conformational changes<sup>29,30</sup>. Additionally, an alanine scan of AlbA identified several key residues involved in albicidin binding<sup>29</sup> and it has been suggested that AlbA confers albicidin resistance by removing free albicidin from the cell<sup>18,31</sup>. During the evaluation of our manuscript, parallel work by Rostock et al. was published<sup>26</sup>. The very thorough biophysical characterization of the AlbA-albicidin interaction using NMR and fluorescence spectroscopy, as well as the crystal structure of the AlbA-albicidin complex were reported. In addition, AlbA was found to protect the albicidin target DNA gyrase from the antibiotic and several derivatives *in vitro* and bacterial cells in an agar diffusion assay<sup>26</sup>.

Here, we report the crystal structures of both, AlbA and its complex with albicidin. Unexpectedly, we found AlbA to promote a chemical reaction in albicidin slowly, leading to a loss of biological activity that redefines our understanding of AlbA-mediated albicidin resistance. We provide detailed mutational studies, which allowed us to propose a mechanism for this reaction. We used the structural data to conduct a comprehensive analysis of the prevalence of AlbA-like genes in important human pathogens and their evolutionary development. Finally, we demonstrate that AlbA expression is upregulated in *K. pneumoniae* upon treatment with albicidin.

## 2 Results

### 2.1 AlbA is a structural homolog of TipA-S

To understand how AlbA may exert its function, we determined its high-resolution crystal structure. The full-length wild-type AlbA (AlbA<sub>wt</sub>) protein from *K. oxytoca* (see materials and methods for details) formed crystals belonging to space-group P4<sub>3</sub>2<sub>1</sub>2 and the structure was determined at 1.9 Å resolution by single-wavelength anomalous dispersion using selenomethionine (PDB ID 6h95). All data collection and refinement statistics can be found in table S1. The refined model contained one AlbA<sub>wt</sub> molecule in the asymmetric unit and includes residues 1 – 221 (Figure 2A). It is an  $\alpha$ -helical protein with a flattened, oblong shape measuring approximately 60 Å x 40 Å x 25 Å. AlbA<sub>wt</sub> has no sequence homologs in the protein data bank (PDB), and we therefore searched for structural homologs using the DALI server<sup>32</sup>. No protein was found to cover more than 50 % of the AlbA<sub>wt</sub> structure, but upon closer inspection it became apparent that the fully ordered (ligand bound) structure of TipA-S (PDB 2mc0) aligns to the N-terminal 114 residues of AlbA<sub>wt</sub> (C <sub>$\alpha$</sub>  rmsd of 3.7 Å, sequence identity 15 %) and that an

additional copy of the same TipA-S structure also aligns to AlbA<sub>wt</sub> residues 115 – 216 ( $C_{\alpha}$  rmsd of 3.9 Å, sequence identity 16 %) (Figure 2B and S1). Following TipA-S nomenclature<sup>15,16</sup> (TipA-S consists of helices  $\alpha 6 - \alpha 13$ ), AlbA<sub>wt</sub> does not contain  $\alpha 6$ , but all other helices are largely conserved. Helices  $\alpha 7$  and  $\alpha 8$ , which are critical for substrate binding in TipA-S<sup>15,16</sup>, are much shorter in AlbA<sub>wt</sub> and do not form the deep substrate binding cleft observed in TipA-S<sup>15,16</sup>. Best conservation (helix length and relative position) is observed for  $\alpha 9$  and  $\alpha 10$ . Intriguingly, unlike TipA-S,  $\alpha 11$  and  $\alpha 13$  of AlbA<sub>wt</sub> are neither parallel nor in direct contact with each other, but spread apart (Figure 2C). In AlbA<sub>wt</sub>, helix  $\alpha 13$  is significantly longer than in TipA-S and its C-terminal portion aligns with  $\alpha 7$  of the second TipA-S copy (TipA-S') (Figure 2B and S1). The alignment of TipA-S'  $\alpha 8 - \alpha 13$  with AlbA<sub>wt</sub>  $\alpha 8'$  to  $\alpha 13'$  corresponds to the alignment of TipA-S with the N-terminal half of AlbA<sub>wt</sub>. It thus appears that AlbA originated from an internal gene duplication event in a TipA-S-like gene, giving rise to a pseudosymmetric TipA-S dimer. In addition to the change in binding-site architecture, the residues identified as important for substrate binding, including the active-site cysteine, show no conservation. In TipA, the cysteine reacts with dehydroalanine residues when TipA-S/L bind thiopeptide antibiotics to form covalent complexes<sup>12,33</sup>, chemistry which would not be possible with albicidin. Binding of albicidin must therefore follow a different logic, and we identified a putative substrate binding tunnel, which runs across the entire length of the protein (Figure S2). Its formation is a direct result of the movements of helices  $\alpha 11 / \alpha 13$  and  $\alpha 11' / \alpha 13'$  (Figure S3).

In order to understand how albicidin may fit into the putative substrate binding tunnel of AlbA we determined the high-resolution structure of the AlbA<sub>wt</sub>-albicidin complex. The natural product albicidin was synthesized according to a previously published procedure (see supplementary online material)<sup>20</sup>. AlbA<sub>wt</sub> was incubated with an excess of albicidin and subsequently crystallized in space group C222<sub>1</sub>. The structure was determined to 1.55 Å resolution by molecular replacement using the structure of AlbA<sub>wt</sub> as a search model (PDB ID 6h96).

The new crystal form contained two AlbA<sub>wt</sub> monomers in the asymmetric unit and the overall structure of AlbA<sub>wt</sub> is virtually unchanged after binding albicidin (Rmsd of 0.8 Å over all non-hydrogen atoms, Figure S4). It has been reported that the NMR spectra of AlbA show extensive line-broadening in the absence of albicidin, which may be caused by slow internal dynamics of the protein<sup>26</sup>. This observation led to the hypothesis that AlbA may exist in an open and closed conformation, with an ensemble of intermediate states<sup>26</sup>. Since the AlbA structure we

determined corresponds to the AlbA-albicidin complex structure, we suggest that we have crystallized the closed conformation of the protein. We observed clear and unambiguous electron density for albicidin, spanning the entire length of AlbA (Figure 2A and S5), with binding of the extended albicidin molecule shared pseudosymmetrically between the two TipA-S domains. Upon closer inspection of the electron density for the ligand, it became obvious that the central cyanoalanine moiety had cyclized to yield a five-membered ring, in which the nitrile group had been converted to a primary ketimine or succinimide (**2** and **3**, respectively, Figures 2D and S5). Equivalent electron density was also reported by Rostock et al. (PDB ID 6et8)<sup>26</sup>, but an unmodified albicidin molecule was built into the density, which resulted in a poor fit, high B-factor of the nitrile carbon compared to adjacent atoms and very close interatomic distance between the amide proton of NH25 (numbering in Figure 2D) and the triple bond of the nitrile moiety (1.4 Å) (Figure S5). Albicidin is almost completely buried in the AlbA tunnel and engages in extensive hydrophobic interactions. A complete LigPlot interaction diagram can be found in Figure S6. Several residues identified previously as important for albicidin binding through an alanine scan<sup>29</sup> are in fact in contact with albicidin. In addition to the hydrophobic interactions, AlbA and albicidin form a salt-bridge and several hydrogen bonds, all mediated by side-chains. The carbonyl oxygen O13 (numbering in Figure 2D) forms a hydrogen bond with H78, while N75 is a bidentate ligand, forming hydrogen bonds with NH21 (immediately adjacent to the five-membered ring) and O26 (part of the newly formed ring). The ketimine / ketone moiety (NH / O62) is hydrogen bonded to T99, and the final interactions concern the terminal *para*-aminobenzoic acid unit. The side-chain of Y169 is hydrogen-bonded to O56 (methylated *m*-oxygen), while Q205 and R181 form a hydrogen bond and salt bridge, respectively, with the terminal carboxyl group. These extensive interactions explain the high affinity of AlbA<sub>wt</sub> for albicidin.

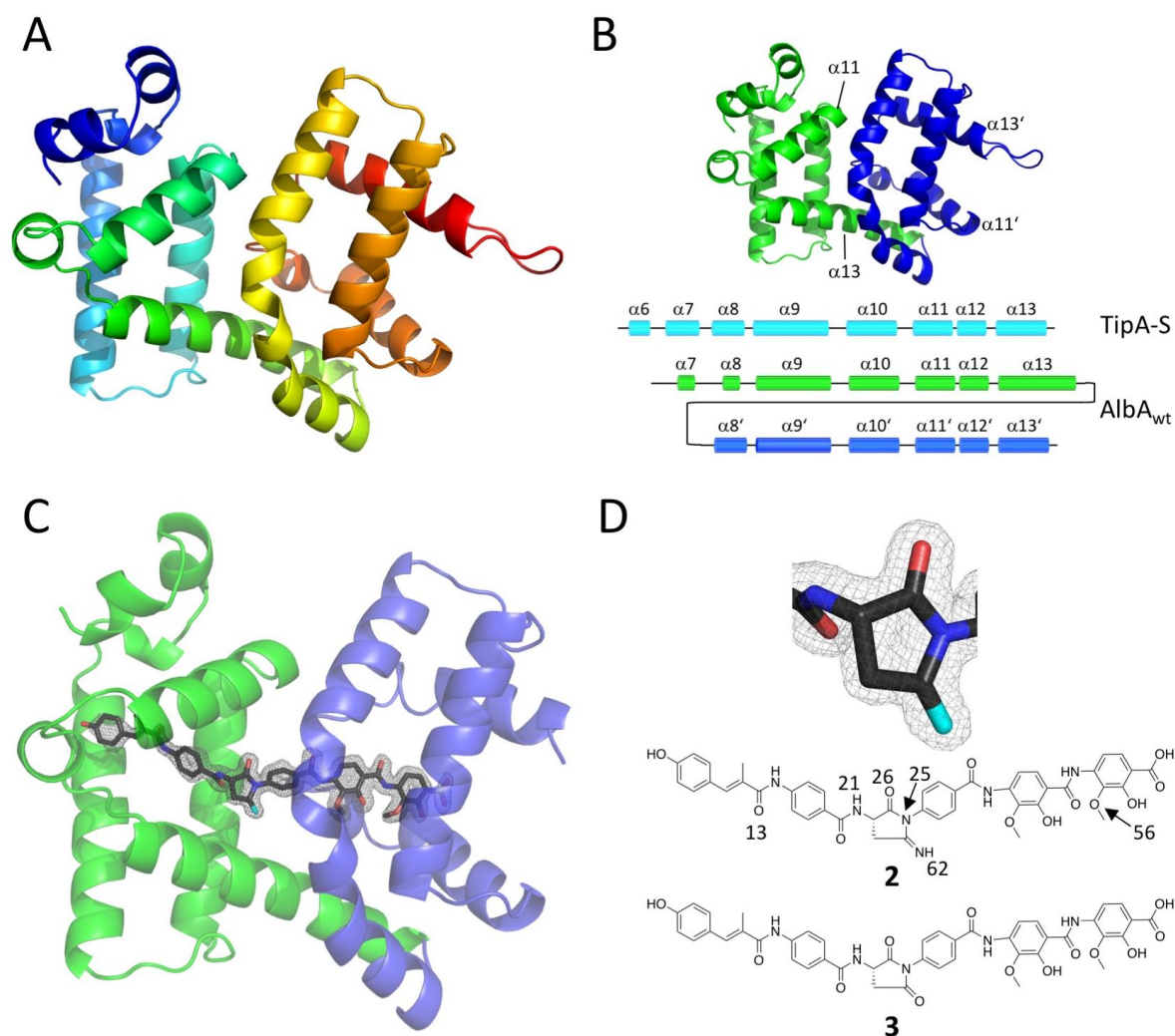


Figure 2. (A) Overall structure of AlbA<sub>wt</sub>. The structure of AlbA<sub>wt</sub> is shown in cartoon representation using the rainbow color scheme (N-terminus blue, C-terminus red). (B) AlbA is a tandem repeat of TipA-S. The AlbA<sub>wt</sub> structure is shown as a cartoon representation with the N-terminal TipA-S repeat in green, and the C-terminal TipA-S repeat in blue. Helices that underwent a drastic change in orientation and relative position in AlbA<sub>wt</sub> (when compared to TipA-S) are labeled. A secondary structure diagram shows the conservation of helix length and position between the two proteins. (C) AlbA<sub>wt</sub>-albicidin complex structure. Protein representation and color scheme as in B. Albicidin is represented as sticks. Carbon atoms black, oxygen atoms red, nitrogen atoms blue, unassigned atom cyan. The difference electron density ( $F_o - F_c$  contoured at  $3\sigma$  with phases calculated from a model which was refined with no albicidin present) is shown as a grey isomesh. (D) Chemical structures of the two possible albicidin conversion products, the primary ketimine 2 and the succinimide 3.

## 2.2 AlbA<sub>wt</sub> promotes the cyclization of albicidin

We wondered whether the observed five-membered ring was an artefact or an AlbA-mediated modification of albicidin. When equimolar amounts of AlbA<sub>wt</sub> and albicidin (100  $\mu$ M each) were incubated at 37 °C for 24 h prior to analysis by high-resolution liquid chromatography-mass spectrometry (HR-LCMS), we discovered that the AlbA<sub>wt</sub>-albicidin sample contained four additional peaks (Figure 3A). In contrast, control samples set up in the reaction buffer with BSA (Figure S7) or without protein (Figure 3A) showed no conversion of albicidin. In fact, only traces of converted albicidin could be detected in reaction buffer without protein after 14 days at 37 °C (Figure S7). This reaction was temperature-dependent – after 24 h, no conversion was

observed at 4 °C, while little conversion was detected at 20 °C. At 37 °C, more than half of the albicidin was converted (Figures 3A and S8).

In the AlbA<sub>wt</sub>-containing sample, the peak representing albicidin (m/z calc: 843.2620, observed: 843.2621,  $\Delta$ ppm 0.1) had decreased significantly, while two new main peaks appeared. The most abundant species was an isotopomer of albicidin with a different retention time, indicating **2** (m/z calc: 843.2620, observed: 843.2632,  $\Delta$ ppm 1.4), which was supported by tandem mass spectrometry (MS<sup>2</sup>, Figure S9). Since primary ketimines are unstable in aqueous solution, we expected the third main peak to be the hydrolysis product of **2**, the succinimide **3** (m/z calc: 844.2461, observed: 844.2466,  $\Delta$ ppm 0.6), and the identity of **3** was in agreement with MS<sup>2</sup> data (Figure S10). The two minor peaks had near identical masses, and we assigned them as two diastereomers of the intermediate between **2** and **3**, corresponding to a hemiaminal **4** (m/z calc: 861.2726, observed: **4a** 861.2740,  $\Delta$ ppm 1.6 and **4b** 861.2738,  $\Delta$ ppm 1.4). While it is not possible to determine which peak corresponds to which diastereomer, MS<sup>2</sup> data corroborated the assigned identity of the compounds (Figures S11 and S12). To ascertain whether **2** or **3** was found in the AlbA<sub>wt</sub>-albicidin complex structure, we harvested multiple AlbA-albicidin crystals, washed them thoroughly and dissolved them in acetonitrile prior to MS analysis. The predominant species was **2**, with a small fraction of **3** that is likely the result of hydrolysis during analysis of the sample and traces of **1** (Figure S8). Therefore, rather than just neutralizing albicidin by binding to it, AlbA<sub>wt</sub> fosters the conversion of albicidin to **3**, with **2** representing the protein-bound (and solvent-protected) intermediate.

To better understand the albicidin conversion, we set up a time course experiment and observed a nearly completed process after 64 h when using a 1 : 1 molar ratio of protein and compound (Figure S13). This very slow progression would imply that despite the non-covalent nature of their interaction, cells would still have to produce stoichiometric quantities of AlbA<sub>wt</sub> for protection, mirroring the situation for TipA. When sub-stoichiometric quantities of AlbA<sub>wt</sub> were used, the conversion progressed very slowly, but one AlbA<sub>wt</sub> molecule was able to promote cyclization of multiple albicidin molecules (Figure S13).

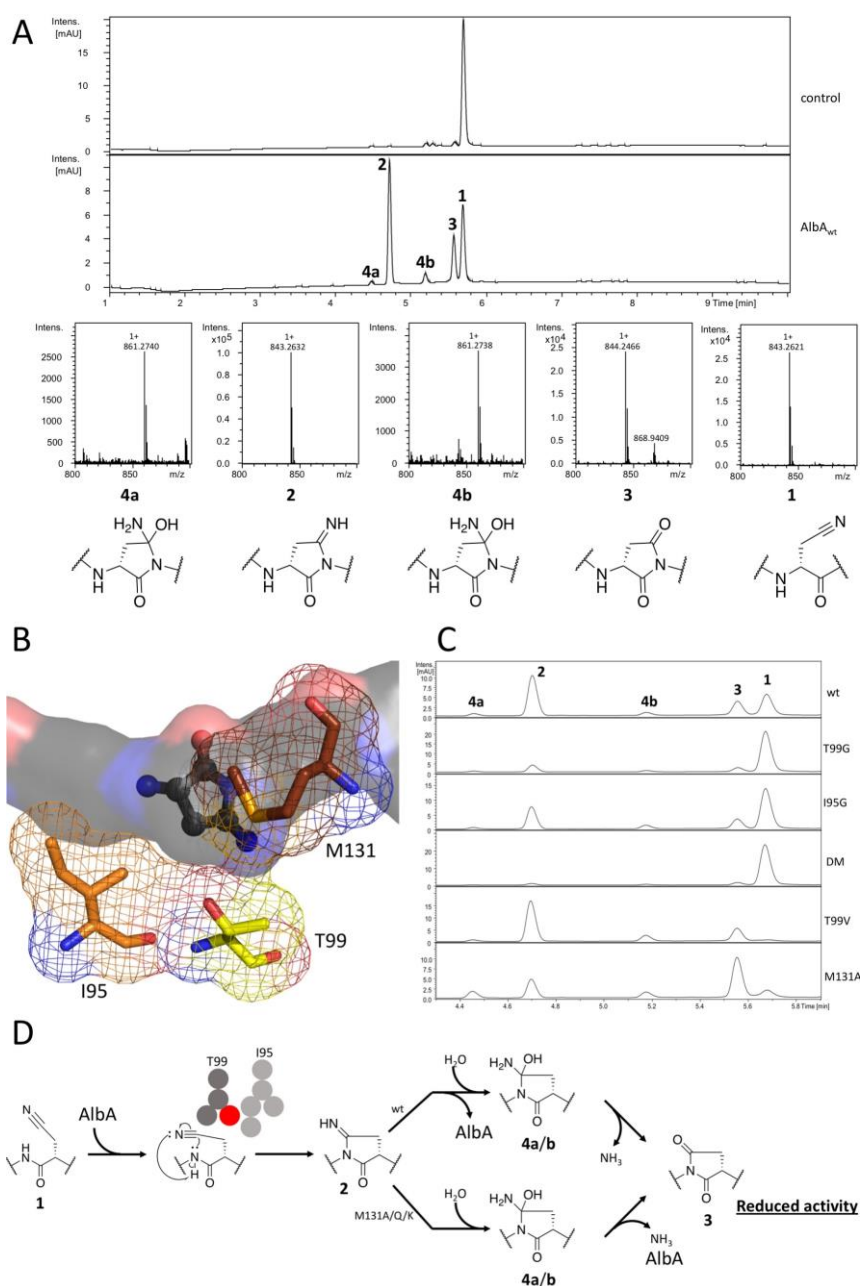


Figure 3. (A) Incubation of albicidin with AlbA<sub>wt</sub> leads to the cyclization of albicidin at the cyanoalanine moiety. The four product peaks were assigned to their chemical structures by MS<sup>2</sup> (see SOI). (B) Three residues are in direct contact with the newly formed ketimine: I95 (orange), T99 (yellow) and M131 (brown). Their surface is represented as an isomesh with corresponding colors. The solid surface of albicidin is colored as carbon atoms black, nitrogen atoms blue, oxygen atoms red. The ketimine moiety is shown as a ball and stick model using the same color scheme. (C) The effects of four point- and one double-mutant (DM, I95G/T99G) on albicidin. The 3 : 2 ratio is inverted for AlbA<sup>M131A</sup> when compared to the other mutants or wt protein. (D) Proposed mechanism for albicidin conversion. While the 2 has to be released from the protein before it can hydrolyze to 3, mutation of M131 to Ala or a hydrophilic residue allows hydrolyzation while 2 is still bound to the protein

Interestingly, we also detected varying amounts of the hydrolysis product of 3 in samples containing substoichiometric AlbA concentrations, in which the central linker is either an aspartate or iso-aspartate (m/z calc: 862.2566, observed: 6 862.2569, Δppm 0.3, Figure S13). Formation of 6 is possibly the result of the extensive incubation time required and this variant

of albicidin has been reported to be inactive<sup>34</sup>. We speculated that the relative affinities of **1** and **2** may be a reason for slow cyclization. When we analyzed the interaction of AlbA<sub>wt</sub> with albicidin by surface plasmon resonance (SPR), the  $K_D$  for this interaction was calculated as 2.1 nM (Figure S14), which is on par with the  $K_D$  reported by Rostock et al.<sup>26</sup>. The main contributor to this high affinity was the very slow off-rate of just  $1.1 \times 10^{-4} \text{ s}^{-1}$ , which may explain the slow progression of cyclization observed at sub-stoichiometric protein concentrations.

The affinity of **3** was slightly weaker (11.6 nM) and, crucially, **3** retained a very slow albeit slightly faster off-rate of  $2.8 \times 10^{-4} \text{ s}^{-1}$  (Figure S14). This implies very slow release of **2/3** may be the main contributor the slow progression of cyclization. Overall these data therefore argue that AlbA, in addition to being a very strong albicidin binder, aids in the cyclization of albicidin. We are unaware of comparable processes in natural products.

### 2.3 Rationalizing AlbA activity

We expected cyclization of albicidin to **2** to follow standard chemistry – deprotonation of the amide nitrogen (NH25), followed by nucleophilic attack on the nitrile carbon and subsequent protonation of the intermediate to yield the primary ketimine function. Since the amide nitrogen would be a weak nucleophile, we suspected that the reaction may be aided by protonation of the nitrile nitrogen prior to nucleophilic attack, but no aspartate or glutamate residues were found in close proximity. To shed light on the reaction mechanism, we first investigated the effects of general acid and / or base. We incubated AlbA<sub>wt</sub> with an equimolar amount of albicidin at different pH values from 5.5 to 9.0 in increments of 0.5 (Figure S15). At low pH, we observed little cyclization, but as the pH increased, so did cyclization. No cyclization was observed in control reactions without AlbA<sub>wt</sub> unless Tris buffer was used, in which case we observed slow, AlbA<sub>wt</sub>-independent cyclization of albicidin at pH 8.5 and 9.0 (Figure S15). While a normal peptide amide would not be deprotonated under such mild conditions, the amide in question can be viewed as an aniline-derivative with a carbonyl group in the *p*-position, which presumably significantly lowers the  $pK_a$  of this amide. These observations are suggestive of the involvement of a general base.

A detailed analysis of the residues in direct contact with the newly formed ketimine – I95, T99 and M131 – was performed (Figure 3B). The side-chain of T99 packs against the side-chain of I95 and is hydrogen-bonded to the ketimine. We thus wondered whether T99 played a role in cyclization and produced AlbA<sup>T99V</sup> and AlbA<sup>T99G</sup>. All mutants reported expressed like wild-type protein and showed highly similar elution profiles in size-exclusion chromatography



(Figure S16). The overall structure of apo-AlbA<sup>T99V</sup> (PDB ID 6h97), determined at 2.7 Å resolution (Table S1), was unchanged and the mutant side-chain had the same orientation and position as T99 in the AlbA<sub>wt</sub> structure (Figure S17). When we tested the effect of AlbA<sup>T99V</sup> on **1**, we found cyclization to be accelerated ~2.5 x when compared to the wt protein (Figures 3C and S18). In contrast, AlbA<sup>T99G</sup> promoted little cyclization, even after 24 h (Figure 3C). We thus propose that the side-chain of T99 creates a bottleneck in the albicidin binding tunnel, that forces the nitrile group of albicidin into a position that favors nucleophilic attack of amide N25 on nitrile C28, leading to the formation of **2** (Figure 3D). By exchanging the threonine to a slightly bulkier valine, we strengthened the hydrophobic interactions between I95 and the amino acid in position 99 (now V). This could lead to a more stable barrier and thus promote faster cyclization of albicidin. Accordingly, with the barrier largely removed in AlbA<sup>T99G</sup>, albicidin cyclization was almost abolished. In agreement with this hypothesis, AlbA<sup>I95G</sup> showed reduced but still appreciable effects on albicidin (Figure 3C), since now merely the back-stop of the actual barrier was removed. When incubated with the double mutant, AlbA<sup>I95G/T99G</sup>, albicidin was virtually unaffected (Figure 3C).

The position of the side-chain of M131 appeared to protect the ketimine function of **2** from bulk solvent and to prevent the formation of the hemiaminal **4a/b** during hydrolysis of **2**. In consequence, one would expect a mutation of the methionine to a residue with a less bulky side-chain to promote the formation of **3**. Accordingly, AlbA<sup>M131A</sup> incubated with albicidin, led to a much larger percentage of **3** than the wt protein or any previously analyzed mutant (Figure 3C). We determined the complex structure of AlbA<sup>M131A</sup> with albicidin at 2.2 Å (PDB ID 6hai) (Figure S19 and Table S1). In the structure, the small alanine side-chain affords ample room to form the hemiaminal intermediate (Figure S20). It should be noted that in the complex structure of AlbA<sub>wt</sub> with **2**, two ordered water molecules (HOH6 and 102) are in a position to attack the ketimine carbon (Figure S20) and may promote hydrolysis.

## 2.4 Cyclization decreases albicidin activity, but AlbA is insufficient to protect cells from albicidin

An important question with respect to the resistance mechanism exerted by AlbA was whether the protein-mediated modification altered albicidin activity. We attempted to purify **2** and **3** in sufficient quantities for biological testing by setting up large-scale conversion reactions. As expected, it was not possible to purify **2**, as the peak converted to **3** during purification. Compound **3**, on the other hand, was stable and we tested its activity against *E. coli*, *Staphylococcus aureus*, *Bacillus subtilis* and *Micrococcus luteus* (Table 1). The most

pronounced effects were observed for *E. coli* and *S. aureus*, where **3** showed a decrease of activity by 300 x and 222 x, respectively. For the other two strains, activity was reduced by approximately two orders of magnitude. We have therefore been able to demonstrate that the modification of albicidin to **3** results in a significant decrease of albicidin activity.

Table 1: Activity of **3** and **1** on selected bacterial strains. Cyclization of cyanoalanine to succinimide leads to a significant decrease in activity.

		<i>E. coli</i>	<i>S. aureus</i>	<i>B. subtilis</i>	<i>M. luteus</i>
		DSM1116	Newman	DSM10	DSM1790
MIC	<b>3</b>	1.8	28.8	5.4	43.5
( $\mu\text{g/mL}$ )	<b>1</b>	0.006	0.13	0.03	0.5
MIC shift	<b>3/1</b>	300	222	108	87

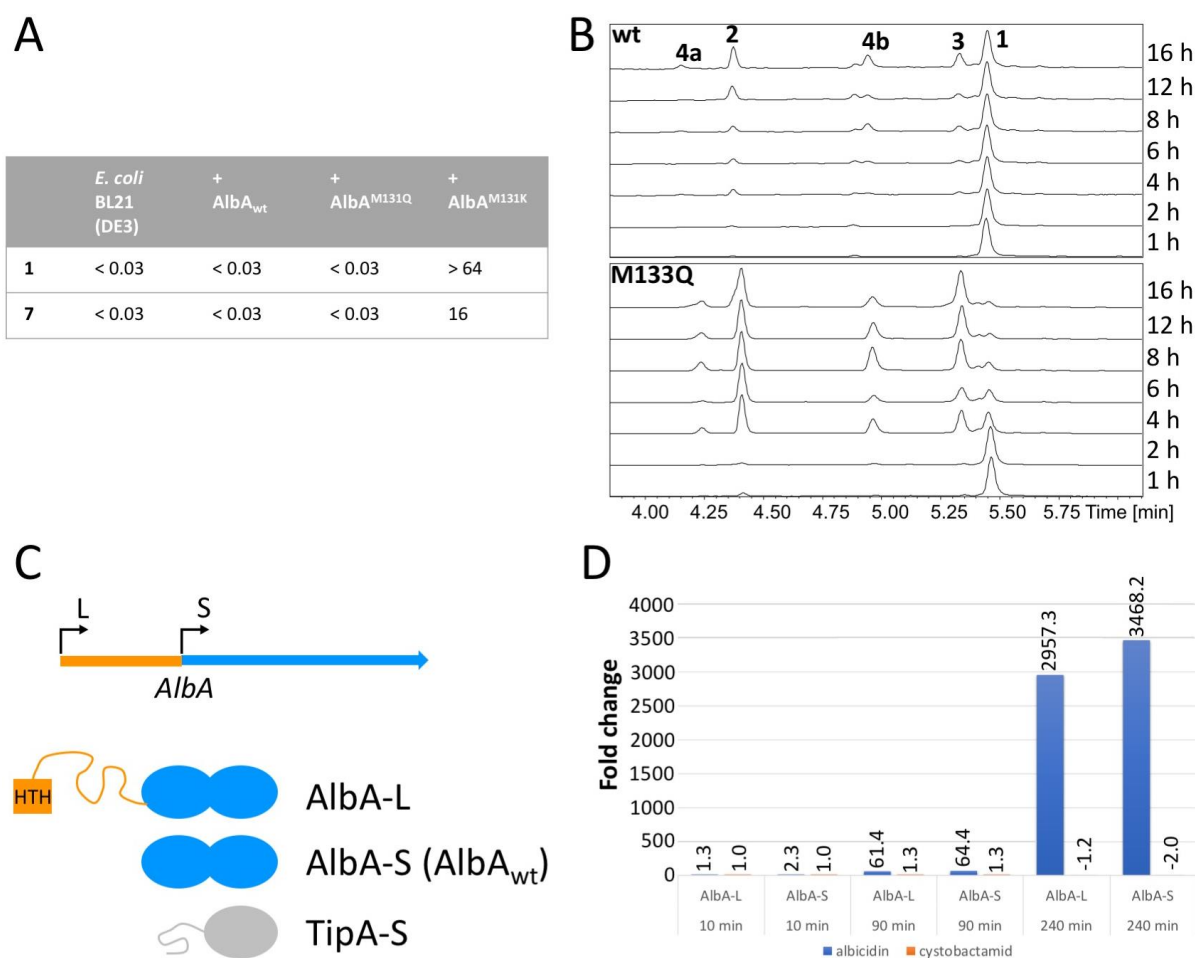


Figure 4. (A) MICs demonstrating that AlbA<sub>wt</sub> is insufficient to protect *E. coli* BL21 (DE3) from **1** or **7**. Naturally occurring point-mutation M131K on the other hand renders these cells albicidin resistant and protects the cells largely from **7**. MICs are given as compound concentrations ( $\mu\text{g/mL}$ ) (B) Time-course for the cyclization of albicidin when incubated with AlbA<sub>wt</sub> (top) and AlbA<sup>M133Q</sup> (bottom) using **1** : **1** protein : compound ratios. The M131Q point mutation leads to significantly accelerated cyclization. (C) Reevaluation of the *alba* sequence reveals an alternative start codon upstream of the reported start. Translation from the alternative start codon would yield an AlbA-L protein that also contains an N-terminal HTH motif followed by a random coil, analogous to TipA. (D) RT-qPCR results demonstrating that exposure of *K. pneumoniae* to **1**, but not **7** induces *alba* transcription.

It has been demonstrated in agar diffusion assays that when AlbA is pre-incubated with an equimolar amount of albicidin, the remaining concentration of free albicidin is too small to inhibit bacterial growth<sup>26</sup>. The same observation was made for the activity of DNA gyrase, the albicidin target, *in vitro*<sup>26</sup>. Since the read-out of these experiments is compound-protein affinity, we wondered if *E.coli* overexpressing AlbA would be resistant to albicidin. When we tested *E.coli* BL21 over-expressing AlbA, we found the cells to be susceptible to **1** (Figure 4A). Since *Klebsiellae* containing the *albA* gene are fully resistant to albicidin<sup>20,21</sup>, it thus appears highly probable that additional factors are required for albicidin resistance.

## 2.5 Evolution of AlbA and evidence for a more efficient homolog in *K. pneumoniae*

We aimed to understand how AlbA-mediated albicidin resistance had evolved and how widely proteins that could potentially bind albicidin and related compounds are distributed amongst sequenced proteobacteria. A search of the non-redundant database for AlbA homologs returned 1,906 sequences from beta-, gamma-, and delta-proteobacteria, all of which were annotated as mercuric ion resistance (MerR) family transcription regulators. Next, we used the AlbA-albicidin complex structure to define the residues in direct contact with albicidin and how each of these residues could be varied to still allow albicidin binding. This contact pattern (Figure S21) was then used to filter the obtained homologs. Phylogenetic analysis reveals that this pattern can occasionally be found in *Klebsiella* and *Enterobacter* species, as well as in several closely related sequences from *Raoultella*, *Kosakonia*, *Kluyvera*, *Escherichia*, *Pseudoescherichia*, and *Leclercia* spp.. The distribution of the relevant sequences in the phylogenetic tree (Figure S22) suggests that in *Klebsiella* and *Enterobacter* the pattern evolved on several independent occasions, whereas in *Raoultella*, *Kosakonia*, *Kluyvera*, *Escherichia*, *Pseudoescherichia*, and *Leclercia*, close homology coupled with the virtual omnipresence of sequences containing the contact pattern suggests introduction of the protein into these species via horizontal gene transfer as a means of resistance against albicidin. Additionally, the *Raoultella* branch is embedded into the *Klebsiella* branch, which suggests *Klebsiella* as the source of the horizontal gene transfer into *Raoultella*. For other species, the phylogeny is not well resolved at the branching point, so the source of the transfer is more difficult to pinpoint. Since AlbA was originally discovered in *K. oxytoca*, we had a closer look at the evolution of AlbA in *Klebsiellae* as a whole. When searching for AlbA homologs in publicly available whole genome sequencing data and filtering identical sequences, two clades of AlbA were found: One

group, consisting mostly of *K. oxytoca* and *K. michiganensis* strains, is currently comprised of 36 sequences with no less than 90 % homology to AlbA. These sequences contain no variations in the residues making contact with albicidin. The second group contains 129 sequences that are all  $77 \pm 2$  % identical to AlbA and is found mainly in *K. pneumoniae*, *K. variicola*, and *K. quasipneumoniae*. These sequences contain a single point mutation in the residues contacting albicidin. Here, the M131 residue, which has been identified by us as critical to shield the intermediate **2** from water, is mutated to either Q (most cases) or K (some instances). The exchange of M131 to a hydrophilic residue opened the possibility that these AlbA homologs would behave in a manner similar to AlbA<sup>M131A</sup>. We therefore introduced the point mutations M131Q and M131K into AlbA<sub>wt</sub> and analyzed the effects of these mutations on albicidin cyclization. Both mutants promoted much faster cyclization of albicidin than the wild-type protein (Figure 4B), with the lysin mutant being faster than the glutamine mutant. When the more prevalent mutant M131Q was used at sub-stoichiometric concentrations it showed faster cyclization than AlbA<sub>wt</sub> (Figure S23). Interestingly, we did not observe the same effect of pH on AlbA<sup>M131Q</sup> as AlbA<sub>wt</sub>, which may suggest that the glutamine now acts as a specific base to promote the reaction (Figure S23). We were curious whether these mutations altered the behavior of AlbA enough to now allow AlbA to protect cells from albicidin. When we tested *E. coli* over-expressing AlbA<sup>M131Q</sup> or AlbA<sup>M131K</sup>, we found the M131Q mutation to have no effect on the minimum inhibitory concentration (MIC) of albicidin. In stark contrast, cells overexpressing the M131K mutant were fully resistant to albicidin (Figure 4A). Whether this is a direct function of cyclization or involves additional factors will require further study. It thus appears as if a version of AlbA that is more efficient at protecting cells from albicidin already preexists in the microbial community and may be passed on to critical human pathogens via horizontal gene transfer. In fact, several instances of this point mutation can be found in AlbA homologs present in *E. coli*.

## 2.6 AlbA defines a new TipA-like family

One characteristic of the AlbA homologs from both clades described above was prominent: Virtually all possess an N-terminal extension when compared to AlbA. This extension has homology to known HTH DNA-binding domains of transcription factors and is followed by a coiled-coil region, after which homology to AlbA begins. The basic architecture of these AlbA homologs therefore resembles that of the autoregulatory TipA system, with a substrate binding domain that has been extended through internal gene duplication. A closer look at AlbA reveals

the existence of an alternative in-frame start codon that would lead to an N-terminal extension by a HTH-domain, followed by a coiled-coil region (Figure 4C). This observation was also noted by Rostock and colleagues<sup>26</sup> and implies that the original annotation of the AlbA gene is incomplete. Analogously to TipA, two versions of AlbA may exist in cells: AlbA-L, capable of driving transcriptional events after binding to albicidin and AlbA-S/AlbA<sub>wt</sub>, which removes free albicidin from cells (Figure 4C). In the TipA system, the S version is expressed at a > 20-fold excess when compared to the L version, presumably because far fewer DNA-binding protein copies (TipA-L) are required than those simply neutralizing the thiopeptide antibiotics (TipA-S)<sup>17</sup>. Interestingly, in the work describing the discovery of AlbA, it was noted that the minimal DNA fragment providing albicidin resistance contained two gene products, which appeared to be under the control of one promoter<sup>18</sup>. They were believed to be two different proteins, but their relative abundance roughly reflected the 1 : 20 ratio observed for TipA-L and TipA-S and their molecular weights are in agreement with the predicted molecular weight of AlbA-L and the observed molecular weight of AlbA-S<sup>18</sup>. If this hypothesis were correct, exposure of *albA*-positive bacteria to albicidin would be expected to induce transcription of AlbA. We thus treated *albA*-positive *K. pneumoniae*, which are fully resistant to albicidin and harbor the M131Q mutation, with albicidin and analyzed the transcription levels of AlbA-L and AlbA-S by RT-qPCR at different time-points (Figure 4D). 10 min after exposure, virtually no change in AlbA transcription levels was detected, while transcription was upregulated ~ 60-fold 90 min after the addition of albicidin. After 4 h, transcription levels were upregulated ~3000-fold, indicating a very strong response to albicidin. These data suggest that AlbA acts in a manner similar to TipA, where binding of the effector molecule triggers protein expression of the effector-binding protein.

Since TipA is able to bind structurally diverse thiopeptides, we wondered if AlbA<sub>wt</sub> is also able to bind other compounds, in particular cystobactamid. The  $K_D$  of AlbA<sub>wt</sub> for an albicidin derivative with cystobactamid-like features (**5**, Figure S24) was reported as 14 nM based on a fluorescence quenching assay<sup>26</sup>. Yet despite this tight interaction, the antibacterial effects of **5** were not neutralized by AlbA in agar diffusion assays and **5** did not stabilize AlbA in NMR experiments<sup>26</sup>. SPR experiments using **7** and AlbA<sub>wt</sub> suggest a  $K_D$  in the low  $\mu$ M range (Figure S25), which may reconcile affinity with NMR data and agar diffusion assay results. To probe the effect of AlbA<sub>wt</sub> on **7**, we incubated it with protein to investigate if AlbA<sub>wt</sub>, AlbA<sup>T99V</sup>, AlbA<sup>M131A</sup> or AlbA<sup>M131Q</sup> were able to also promote the cyclization of cystobactamid (convert the central  $\beta$ -methoxy-asparagine into an aspartimide). Even after extensive incubation times at 37 °C, we did not observe any modification of cystobactamid (Figure S25). When comparing

the MICs of **1** and **7** for *E.coli* BL21 and *E.coli* BL21 overexpressing AlbA<sub>wt</sub>, we found no difference, perhaps because the cells were extremely sensitive to both compounds (Figure 4A). The M131K mutant also provided protection against **7**, but not to the same extent as against **1** (Figure 4A). Since *alba*-containing *K. pneumoniae* have an MIC against both compounds of > 64 µg mL<sup>-1</sup>, we investigated bacterial growth after addition of either **1** or **7**. While cultures continued to grow, only **7**, but not **1**, had a statistically significant negative effect on bacterial growth (Figure S25). Finally, we tested whether **7** was also able to induce an increase in AlbA copy numbers in *K. pneumoniae* by RT-qPCR. In striking contrast to **1**, cystobactamid 919-2 was unable to increase *alba* RNA copy numbers, even 4 h after the addition of compound (Figure 4D). These data hint that binding of **7** is insufficient to cause the dimerization of AlbA-L and thus drive transcription of the *alba* gene. We modelled the interaction of **7** and AlbA<sub>wt</sub> to rationalize the differences in binding between the two compounds (Figure S24). Which AlbA-independent resistance mechanism is employed by *K. pneumoniae* to protect against cystobactamid will require further study.

The internal gene duplication of the TipA-S-like antibiotic binding domain changes the architecture of the protein to allow binding of extended, hydrophobic antibiotics in a binding tunnel. This raises the question how wide-spread this new architecture actually is. When we analyzed all > 3000 UniProt sequences listed in Pfam as containing TipA, we found that the duplication of the TipA-S region has occurred exactly twice in the evolution of this family (Figure S26). One branch contains AlbA and related sequences, which are a colorful mixture of beta- and gamma-proteobacteria and enterobacteria. Overall, this branch contains an assortment of TipA sequences with and without the internal gene duplication. Curiously, the second branch is very compact and contains exclusively TipA-like proteins with a duplication of TipA-S. All of these sequences are from *Clostridiaceae*, including *C. botulinum*, and part of a larger *Clostridium* branch (Figure S26). They belong to the larger family of MerR-type transcriptional regulators and also encode an N-terminal HTH domain, followed by a coiled-coil region, reflecting the TipA and AlbA architecture. It is completely unclear which extended, hydrophobic natural product is bound by the proteins belonging to the second branch, or which transcriptional events are triggered in *Clostridiaceae* upon exposure.

### 3 Discussion

The vast majority of currently used antibiotics have been isolated from microorganisms and modified for clinical application. From an evolutionary standpoint it is intuitive to search in this

space – microbes have evolved to secure and defend their ecological niche from other microbes, often using antibiotics as chemical weapons. But this also poses an inherent problem: if an antibiotic has been used for eons, it is only logical that defense strategies – antibiotic resistance – have also evolved. These resistance mechanisms then preexist in the microbial population and can be disseminated if a particular antibiotic finds broad use, for example in human antibiotic therapy. To circumvent this problem, rational compound design is required, which relies on a detailed understanding of these mechanisms of resistance. In determining the structure of AlbA, we sought to enable the design of compounds circumventing AlbA-mediated resistance. Unexpectedly, the structure revealed AlbA to be a member of the TipA family of multi-drug resistance autoregulatory systems. Consequently, the AlbA gene was reevaluated and an alternative upstream start codon found, which may give rise to AlbA-L and AlbA-S analogously to TipA. What transcriptional events may be controlled by AlbA-L will require further study. It is unclear if AlbA itself is a multi-drug resistance protein or merely evolved from such a system, since the AlbA system is not induced by the related cystobactamid. However, the observation that the internal gene duplication of the effector binding site leads to a complete change in binding-site architecture and provides a new antibiotic binding scaffold highlights the adaptability of antibiotic resistance mechanisms.

TipA is able to form a covalent complex with diverse thiopeptide antibiotics, which sequesters them from the cytoplasm and thus inactivates them. AlbA is unable to form a covalent bond with albicidin, but forms an exceptionally stable complex<sup>31</sup> from which albicidin dissociates very slowly. In addition, we discovered that AlbA promotes the cyclization of albicidin, which leads to a loss of activity and decreased affinity. The mechanism of very tight binding combined with slow chemical modification may also be beneficial for transcriptional control, since it allows elevated transcription to cede once exposure to albicidin is stopped. In AlbA, the combination of high-affinity binding with chemical modification may be a way to approximate the beneficial effect of the covalent TipA-thiopeptide bond. It also opens the possibility of acquired AlbA mutations, which lead to faster cyclization of albicidin and in turn a lower energetic burden on cells, since only sub-stoichiometric quantities of AlbA-S would then be required. A first step in that direction may be the M131K mutation observed in *K. pneumoniae*. This underscores the importance of investigating resistance mechanisms in detail to accelerate the developments of new antibiotics circumventing pre-existing resistance mechanisms.





## 4 Supplementary Information

### 4.1 Methods

#### 4.1.1 Cloning and mutagenesis

AlbA<sub>wt</sub> was cloned from genomic DNA (*K. oxytoca*) into pHisSUMOTEV, which was a gift from Dr. Huanting Liu (University of St. Andrews). Point mutations were designed using the overlap extension method after which the PCR products were cloned into the same vector as AlbA<sub>wt</sub>. The resulting protein expression plasmids were verified by enzyme restriction digestion and DNA sequencing before being transformed into *E. coli* BL21 (DE3) or Rosetta™ (DE3).

#### 4.1.2 Protein expression and purification

A single colony was picked into LB liquid medium containing the appropriate antibiotics (50 µg / mL Kanamycin and / or 34 µg / mL Chloramphenicol) to make an overnight culture. The overnight culture was inoculated 1 to 100 into fresh LB medium supplemented with antibiotics and was grown at 37 °C until the optical density (OD<sub>600</sub>) reached 0.6; protein expression was then induced by the addition of 1 mM IPTG and the cells grown at 16 °C overnight. L-Selenomethionine-labelled (SeMet) protein was expressed *E. coli* BL21 (DE3) cells grown in minimal medium supplemented with glucose-free nutrient mix (Molecular Dimensions), 50 µg / mL Kanamycin and 5 % glycerol. This medium was inoculated with overnight culture grown in LB medium, which was washed three times in minimal medium. After 15 min growth at 37 °C, 60 mg / mL of L-Selenomethionine was added and the culture was grown until the OD<sub>600</sub> reached 0.6, when 100 mg / mL each of lysine, phenylalanine, threonine and 50 mg / mL each of isoleucine and valine were added. After addition of amino acids, the culture was grown for an additional 20 min 37 °C, after which expression was induced by addition of 1mM IPTG and the cells were grown at 16 °C for 24 h. Cells were harvested by centrifuging the culture at 6,000 x g at 4 °C for 15 min and the cell pellets frozen at -80 °C.

The cell pellets were resuspended in lysis buffer (20 mM Tris pH 8.0, 200 mM NaCl, 20 mM imidazole, 10 % glycerol (w/v) and 3 mM DTT) supplemented with 0.4 mg DNase per gram of wet cell pellet and cOmplete EDTA-free protease-inhibitor tablets (Roche). The cells were lysed by sonication and cell debris removed by centrifugation (40,000 x g, 4 °C, 20 min). The supernatant was loaded onto a pre-equilibrated (lysis buffer) 5 mL Histrap HP column (GE healthcare) and was washed with 20 column volumes of lysis buffer. The protein was eluted

from the column with 250 mM imidazole and then passed over a desalting column (Desalt 16/10, GE healthcare) to change the buffer back to lysis buffer. Tobacco etch virus (TEV) protease was added to the protein at a mass-to-mass ratio of 1:10 and incubated at 4 °C for 12 h to cleave the His-Sumo tag. Digested protein was then passed over second Histrap HP column, and the flow through was collected and loaded onto a gel filtration column (HiLoad 16/600 Superdex 200 pg, GE healthcare) preequilibrated in 10 mM HEPES pH 7.5. The fractions of the highest purity were pooled together and concentrated to 10 mg mL<sup>-1</sup>.

#### 4.1.3 Crystallization and structure determination

Crystals of SeMet-AlbA<sub>wt</sub>, AlbA<sup>T99V</sup>, AlbA-albicidin and AlbA<sup>M131A</sup>-albicidin complex were obtained at 18 °C in 0.2 - 0.4 M ammonium sulfate, 0.8 - 1.2 M lithium sulfate and 0.1 M sodium citrate tribasic. The protein concentration used for crystallization was 10 mg mL<sup>-1</sup> and the final pH of the crystallization condition was 7.0. For complex crystallization, AlbA was incubated with excess albicidin (1mM) on ice overnight. Crystals appeared after four days and were allowed to grow for six additional days. The crystals were cryoprotected in mother liquor supplemented with 35 % sucrose and flash cooled in liquid nitrogen. Data was collected at ESRF (Beamline: ID29 and ID23-1). To solve the AlbA<sub>wt</sub> structure, a single-wavelength anomalous dispersion (SAD) data set was collected at the Se K absorption edge. Data were processed using Xia2 and the structures were solved using *AutoSol* from the PHENIX<sup>35</sup> crystallography suite. The models were manually rebuilt in COOT<sup>36</sup> and refined using PHENIX<sup>35</sup> and Refmac5<sup>37</sup>. The structures were validated using *MolProbity*<sup>38</sup>, and all images presented were created using PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.). Interaction diagrams were created using Ligplot<sup>39</sup>.

#### 4.1.4 Affinity measurements

All surface plasmon resonance (SPR) experiments were performed on a Biacore X100 system (GE Healthcare). AlbA<sub>wt</sub> with a His-SUMO-tag was coupled on CM5 sensor chips (GE Healthcare) by the amine coupling method using the amine coupling kit from GE Healthcare. The optimal coupling conditions were determined using the pH scouting procedure. AlbA<sub>wt</sub> in 10 mM NaAc, pH 4.0, at a concentration of 25 µg / mL was used for coupling and the contact time was adjusted to achieve roughly ~ 8000 relative response units (RU) coupled protein. The SPR running conditions were: 1 x HBS-EP buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA and 0.005 % v / v Tween 20) with an association and dissociation time of 180 s and

600 s, respectively. The stock solutions of albicidin and compound **3** were prepared in the HBS-EP buffer and diluted to different concentrations by two-fold serial dilutions. Data analysis and curve fitting was performed using the Biacore evaluation software (Version 2.0.1, GE) using the kinetic model and 1:1 binding.

#### 4.1.5 MS and MS<sup>2</sup> analyses

LC-MS was performed on a Dionex Ultimate 3000 RSLC system using BEH C18 column (100 mm x 2.1 mm, 1.7  $\mu$ m) equipped with a C18 precolumn (Waters). Solvent A was H<sub>2</sub>O containing 0.1% formic acid, and solvent B was acetonitrile containing 0.1 % formic acid. Gradient: 0-2.5 min, 5-35 % B; 2.5-5.5 min, 35-42.5 % B; 42.5-95 % B, 5.5-6.0 min or 0-0.5 min, 5% B; 0.5 – 18.5 min, 5 – 95% B; 18.5 – 20.5 min, 95% B; 20.5 – 21 min, 95 – 5% B; 21-22.5 min, 5% B. After a 2 min step at 95 % B the system was re-equilibrated to the initial conditions (5 % B). UV spectrum was recorded by a DAD in the range from 200 to 600 nm.

MS was performed using either an amaZon speed or maXis-2 UHR-TOF mass spectrometer (both Bruker Daltonics). For amaZon, the LC flow was split 1:8 before entering the mass spectrometer using the Apollo ESI source. The following conditions were used: capillary voltage 4500 V, temperature 300 °C, dry-gas flow rate 10 L/min and nebulizer 30 psi. Data was recorded in the mass range from 250 to 2500 m/z.

For maXis, the LC flow was split 1:8 before entering the mass spectrometer using the Apollo ESI source. The following conditions were used: capillary voltage 4000 V, temperature 200 °C, dry-gas flow rate 5 L/min and nebulizer 14.5 psi. The generated ions were trapped in the collision cell for 100 ms and then transferred to within 10 ms through the hexapole into the ICR cell. Data was recorded in the mass range from 150 to 2500 m/z.

## 4.2 Biochemical assays

For all biochemical assays AlbA<sub>wt</sub> and mutants were used at 100  $\mu$ M and an equimolar amount of albicidin or cystobactamid 919-2 was used unless otherwise indicated. Incubations were carried out in PBS at 37 °C with a final DMSO concentration of 0.4 %. For large scale conversion of albicidin for biological testing of **3**, albicidin (400  $\mu$ g) was incubated with an equimolar amount of AlbA<sup>T101V</sup> or AlbA<sup>M133Q</sup> at 37 °C for 3 days. The protein was precipitated by acetonitrile and the succinimide **3** was purified by RP-HPLC (XBridge® Peptide BEH<sup>TM</sup> CSH C<sub>18</sub> OBD Prep Column, 130 Å, 5  $\mu$ m, 10 mm x 250 mm, 1/pkg) using a linear gradient from 95 % A (H<sub>2</sub>O, 0.1 % formic acid) to 95 % B (acetonitrile, 0.1 % formic acid) over 40 min.

### **4.3 Testing of biological activity**

MIC values of albicidin and **3** against *E. coli* (DSM1116), *S. aureus* (Newman), *B. subtilis* (DSM10) and *M.leuteus* (DSM1790) were determined following the European Society of Clinical Microbiology and Infectious Diseases (EUCAST) guidelines.

### **4.4 Determining MIC values for AlbA-overexpressing *E.coli***

Overnight cultures of *E. coli* BL21 (DE3) transformed with expression plasmids for AlbA<sub>wt</sub>, AlbA<sup>M131K</sup>, AlbA<sup>I31Q</sup> (see protein expression and purification) and an empty pHisTEV-SUMO plasmid were inoculated 1 to 100 into fresh LB medium supplemented with antibiotic and grown at 37° C until the optical density at a wavelength of 600 nm (OD<sub>600</sub>) reached 0.6. Protein expression was then induced by the addition of 1 mM IPTG and the cells were grown at 28 °C for 3 hours. MIC values were then determined in medium supplemented with Kanamycin (50 µg/mL) and IPTG (1 mM) following the European Society of Clinical Microbiology and Infectious Diseases (EUCAST) guidelines.

### **4.5 Growth curves**

*K. pneumoniae* DSM30104 culture was prepared in Mueller Hinton Broth medium and incubated at 37 °C with shaking (180 rpm) for 16 h. Growth was determined in presence of cystobactamid 919-2 (64 µg/mL), albicidin (64 µg/mL) and ciprofloxacin (6.4 µg/mL). Start OD<sub>600</sub> was adjusted to 0.05 and measurements were done at designated timepoints.

### **4.6 RT-qPCR**

#### **4.6.1 RNA extraction**

An overnight culture of *K. pneumoniae* DSM30104 was incubated at 37 °C on a rotary shaker. The following day the culture was split in three parts: (i) non-treated culture, (ii) exposure to 10 µg/mL albicidin and (iii) exposure to 10 µg/mL cystobactamid 919-2. Sampling for RNA extraction was done at 0, 90 and 240 min after exposure for all three. Total RNA was isolated from 5 mL cultures. Liquid cultures were centrifuged at 8000 rpm for 10 min. Supernatant was discarded and 4 mL of hot NAE-phenol was added to the pellet followed by a 5 min incubation at 60 °C. Next 4 mL of hot NAES buffer was added followed by a 5 min incubation at 60 °C

and 5 min incubation on ice. The mixture was centrifuged at 4 °C and 8000 rpm for 10 min, water phase was transferred to Phase Lock Gel Light 15-mL tubes (5Prime), 4 mL of phenol:chloroform (6:1) was added and the tubes were inverted for 2 min followed by centrifugation at 4 °C and 8000 rpm for 10 min. This step was repeated twice. RNA precipitation was done with 4 mL ice cold isopropanol and 400 µL 3 M Na-Acetate (pH 5.1) overnight at -20°. Next morning the mixture was centrifuged at 4 °C and 8000 rpm for 40 min, supernatant was discarded and the pellet was washed with 70 % EtOH followed by a drying step in the exsiccator. Finally, the pellet was resuspended in 500 µL of DEPC water. Resuspended RNA was subjected to DNase I digestion according to manufacturer's protocol (NEB). To obtain higher purity of the RNA samples, we used the miRNAeasy Mini Kit (Qiagen) according to manufacturer's protocol. Concentration and purity of RNA samples was determined with a NanoDrop spectrophotometer ND-1000 (Thermo Scientific).

#### **4.6.2 cDNA synthesis**

1 µg of RNA was used for synthesis of cDNA with Revert Aid Premium (Thermo Scientific) according to manufacturer's protocol.

#### **4.6.3 Quantitative real-time PCR**

Amplification of 23s RNA, AlbA-L and AlbA-S was carried out using GoTaq qPCR MasterMix (Promega) according to manufacturer's protocol in triplicate runs with primers 523s, 323s, 5L, 3L, 5S and 3S:

5L	GATAGATACGCTCAACGCC
3L	TCCAGATCGGGTTCATTGC
5S	GCAATCATCGCCTATATCACC
3S	CTGTAGCCGGTCAAAATAGTG
523s	TTACGCTTTGGGAGGAGAC
323s	AGCCAACCTTGAAATACCAC

PeqStar 96Q was used for quantification of cDNA and 23s RNA (housekeeping) gene was used to normalize the data. Cycling conditions were as follows: (i) hold stage 2 min at 95 °C; (ii) PCR stage 40 cycles of 15 s at 95 °C followed by 1 min at 50 °C; (iii) melting stage 15 s at 95 °C, 1 min at 60 °C and 15 s at 95 °C. Control reactions without cDNA were included to serve as no-template control. Comparative threshold cycle ( $\Delta\Delta C_t$ ) method was used to determine relative mRNA quantity with 23S rRNA normalization.

## 4.7 Phylogenetic analysis

Homologs of AlbA were collected using BLAST<sup>40</sup> with default parameters searching in the non-redundant protein database, and additionally in the translated databases of whole genome sequencing experiments for eight bacterial genera that contain human pathogens: *Acinetobacter*, *Bacillus*, *Burkholderia*, *Cronobacter*, *Enterobacter*, *Escherichia*, *Pseudomonas*, and *Streptococcus*. Sequences were aligned with MAFFT<sup>41</sup>, and phylogenetic trees were reconstructed using RAxML<sup>42</sup> under the GTR+Gamma substitution model.

## 4.8 Synthesis of albicidin

Commercially available reagents and solvents were used as supplied. All reactions were performed in oven-dried glassware under an atmosphere of nitrogen gas unless otherwise stated. NMR spectra were recorded using a Bruker Advance-III HD 500 MHz or Bruker Advance-III HD 700 MHz spectrometer. Multiplicities are described using the following abbreviations: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad signal. Chemical shift values of <sup>1</sup>H and <sup>13</sup>C NMR spectra are commonly reported as values in ppm relative to residual solvent signal as internal standard.

High resolution mass spectra were recorded on a Bruker maXis HD spectrometer using negative electrospray ionization (ESI).

LCMS measurements were performed using a Agilent technologies 1200 series (LC) coupled to Bruker amaZon SL (ion trap MS) using a Gemini-NX 3u C18 110A 50x2.0 mm column (for Marfey's method), or on Agilent technologies 1260 Infinity II (LC) coupled to Agilent technologies 6130 a (quadrupol MS) using an Agilent poroshell 120 SB-C18 2.7 μm 2.1x30 mm column (for reaction monitoring).

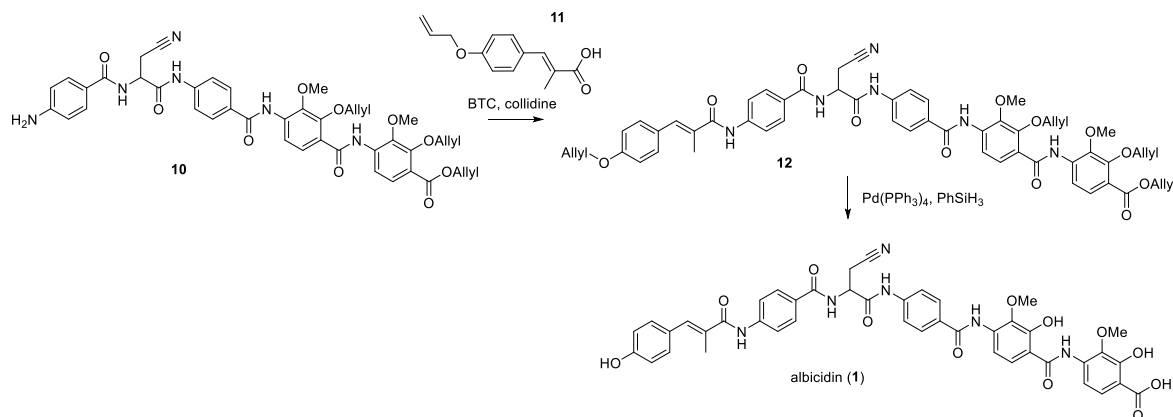
Analytical thin-layer chromatography was performed using pre-coated silica gel 60 F<sub>254</sub> plates (Merck, Darmstadt) and the spots were visualized with UV light at 254 nm or alternatively by staining with potassium permanganate or cerium sulfate.

Chromatographic separations were performed by automated flash chromatography using Grace Reveleris® X2 flash chromatography system or via flash chromatography using silica gel 60M MACHEREY-NAGEL (0.040-0.063 mm; 230–400 mesh).

Preparative reversed phase high performance liquid chromatography (HPLC) was carried out with a Thermo Scientific Dionex (UltiMate 3000 HPLC system) with a Phenomenex Gemini, 5u, C18, 110A, 250x21.20 mm, 5 μm. The collected fractions were lyophilized after their

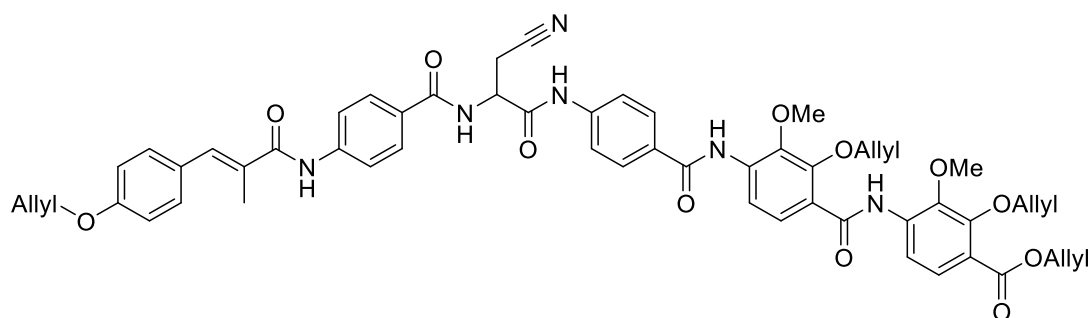
identity and purity was verified by LCMS. Freezedrying was done using LYO Christ alpha 1-4 coupled to high vacuum oil pump.

Albicidin was synthesized using the reported experimental procedures<sup>20</sup>. The analytical characterization of the last intermediates of the synthesis is given:



Scheme 3: last synthetic steps for albicidin.

### Allyl protected albicidin (**12**)



Chemical Formula:  $\text{C}_{56}\text{H}_{54}\text{N}_6\text{O}_{12}$

Exact Mass: 1002,38

Carboxylic acid **11** (205 mg, 0.94 mmol) and BTC (91 mg, 0.31 mmol) were dissolved in THF (16 mL) under a  $\text{N}_2$  atmosphere and cooled to  $0^\circ\text{C}$ , followed by drop wise addition of 2,4,6-collidine (0.285 mL, 2.16 mmol). The reaction was stirred for 20 min and then added to a solution of amine **10** (215 mg, 0.27 mmol) and DiPEA (0.47 mL, 2.70 mmol) in THF (16 mL) cooled to  $0^\circ\text{C}$ . The reaction was stirred for 3 hours and quenched with water. The solvent was partially reduced under vacuum and then diluted with EtOAc (100 mL) and HCl 1 N (100 mL), the organic phase was washed with  $\text{NaHCO}_3$  saturated solution (100 mL) and brine (100 mL), was dried over sodium sulphate and reduced under vacuum. The reaction crude was purified on silica gel with a gradient 0-2% MeOH in DCM, to give 250 mg of a pale yellow solid (0.25 mmol,  $y = 93\%$ ).

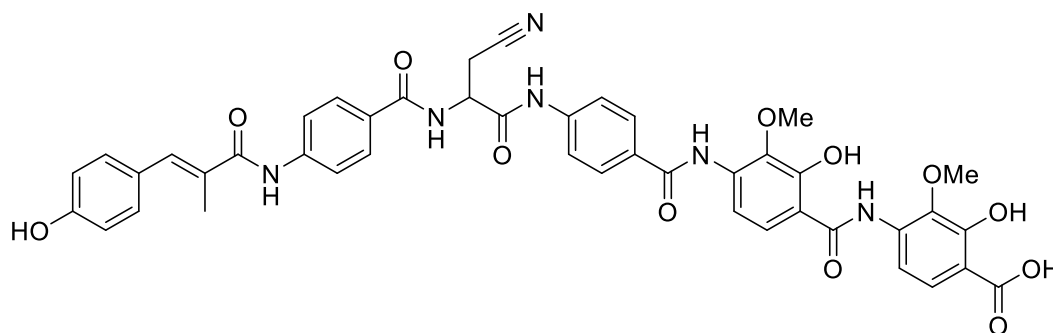
$^1\text{H}$  NMR (500 MHz, DMSO)  $\delta$  10.66 (s, 1H), 10.58 (s, 1H), 10.15 (s, 1H), 9.69 (s, 1H), 9.03 (d,  $J = 7.7$  Hz, 1H), 8.33 (d,  $J = 8.8$  Hz, 1H), 8.00 (d,  $J = 8.6$  Hz, 2H), 7.93 (dd,  $J = 8.7, 3.4$  Hz,

3H), 7.86 (d,  $J = 8.7$  Hz, 2H), 7.81 (dd,  $J = 8.7, 3.4$  Hz, 3H), 7.57 (d,  $J = 8.8$  Hz, 1H), 7.46 (d,  $J = 8.7$  Hz, 2H), 7.31 (s, 1H), 7.04 (d,  $J = 8.7$  Hz, 2H), 6.19 – 5.97 (m, 4H), 5.47 – 5.34 (m, 4H), 5.34 – 5.20 (m, 4H), 4.99 (dd,  $J = 13.7, 8.0$  Hz, 1H), 4.79 (dd,  $J = 15.3, 5.7$  Hz, 4H), 4.58 (dd,  $J = 37.4, 5.4$  Hz, 4H), 3.93 (s, 3H), 3.92 (s, 3H), 3.12 (ddd,  $J = 25.7, 16.9, 7.1$  Hz, 2H), 2.13 (s, 3H).

$^{13}\text{C}$  NMR (126 MHz, DMSO)  $\delta$  169.23, 168.69, 166.63, 165.32, 164.94, 162.89, 158.48, 151.58, 150.06, 144.85, 143.12, 142.97, 142.42, 137.03, 136.99, 134.45, 134.04, 133.89, 133.17, 133.12, 131.58, 131.09, 129.29, 128.85, 128.74, 128.21, 126.77, 125.96, 123.30, 120.75, 120.65, 119.63, 119.42, 119.22, 118.62, 118.33, 118.11, 115.31, 115.16, 75.61, 75.03, 68.69, 65.58, 61.50, 61.38, 51.08, 20.46, 14.98.

HRMS (ESI-):  $m/z$  for  $\text{C}_{56}\text{H}_{53}\text{N}_6\text{O}_{12}$  [M-H] $^-$ : calculated: 1001.3727, found: 1001.3736.

### Albicidin (1)



Chemical Formula:  $\text{C}_{44}\text{H}_{38}\text{N}_6\text{O}_{12}$   
Exact Mass: 842,25

Allyl protected albicidin (**12**) (110 mg, 0.11 mmol) was dissolved in THF (11 mL) under a  $\text{N}_2$  atmosphere, to this solution phenyl silane (122  $\mu\text{L}$ , 0.99 mmol) and palladium tetrakis (63 mg, 0.055 mmol) were added and the reaction was stirred at room temperature for 23 hours.

The solvent was reduced under vacuum, the crude was dissolved in DMSO (12 mL) and acetic acid (3 mL). The product was purified in several runs by preparative reversed-phase HPLC using a gradient 40-95% of  $\text{CH}_3\text{CN}$  +0.1% TFA in water +0.1% TFA to afford 43 mg of desired compound (0.25 mmol,  $y = 46\%$ ).

$^1\text{H}$  NMR (700 MHz, DMSO)  $\delta$  13.92 (br, 1H), 11.59 (br, 1H), 11.52 (s, 1H), 11.17 (s, 1H), 10.57 (s, 1H), 10.10 (s, 1H), 9.76 (s, 1H), 9.69 (s, 1H), 9.01 (d,  $J = 7.7$  Hz, 1H), 8.05 (d,  $J = 8.9$  Hz, 1H), 7.99 (d,  $J = 8.8$  Hz, 2H), 7.92 (d,  $J = 8.8$  Hz, 2H), 7.87 – 7.83 (m, 2H), 7.80 (dd,  $J = 12.2, 8.8$  Hz, 3H), 7.58 (dd,  $J = 13.1, 8.8$  Hz, 2H), 7.35 (d,  $J = 8.7$  Hz, 2H), 7.27 (s, 1H), 6.86 – 6.82 (m, 2H), 4.99 (td,  $J = 8.2, 5.5$  Hz, 1H), 3.92 (s, 3H), 3.78 (s, 3H), 3.15 (dd,  $J = 16.9, 5.4$  Hz, 1H), 3.07 (dd,  $J = 16.9, 8.9$  Hz, 1H), 2.12 (d,  $J = 1.3$  Hz, 3H).

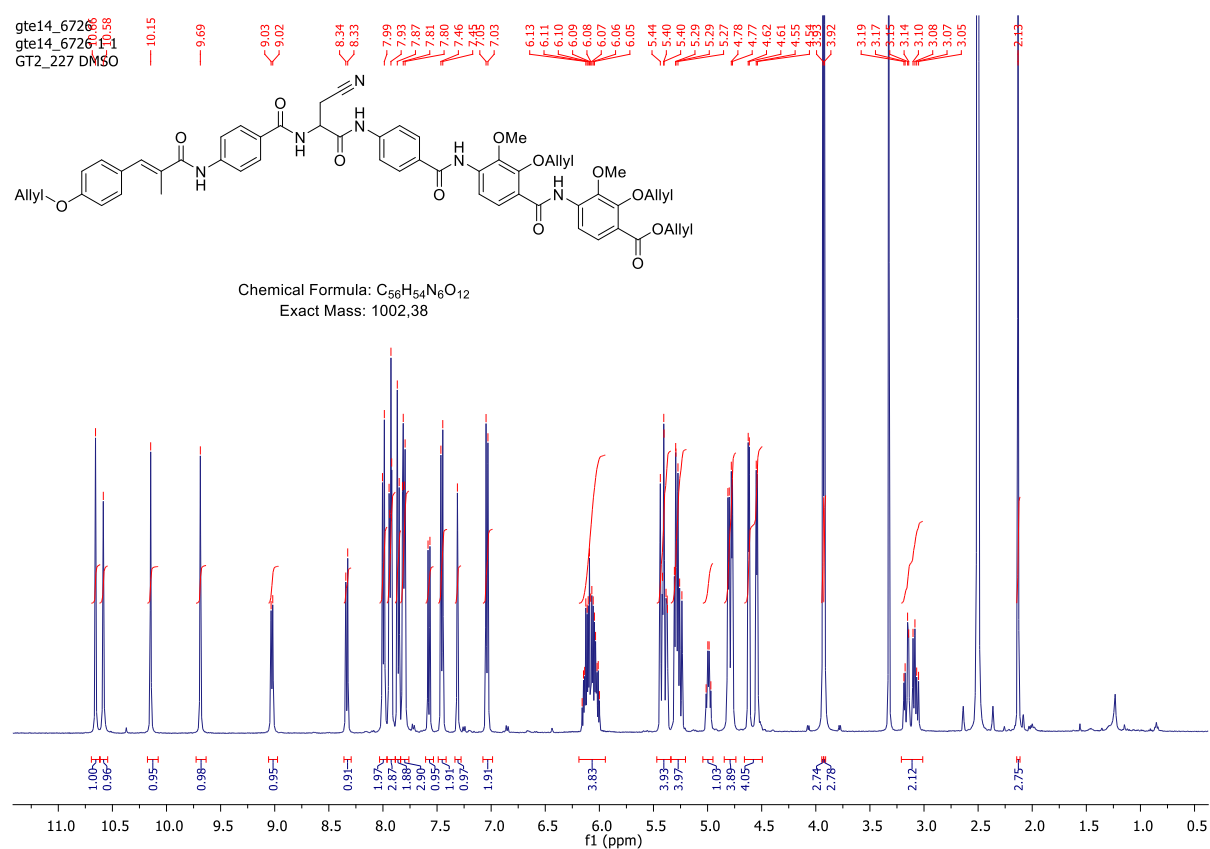


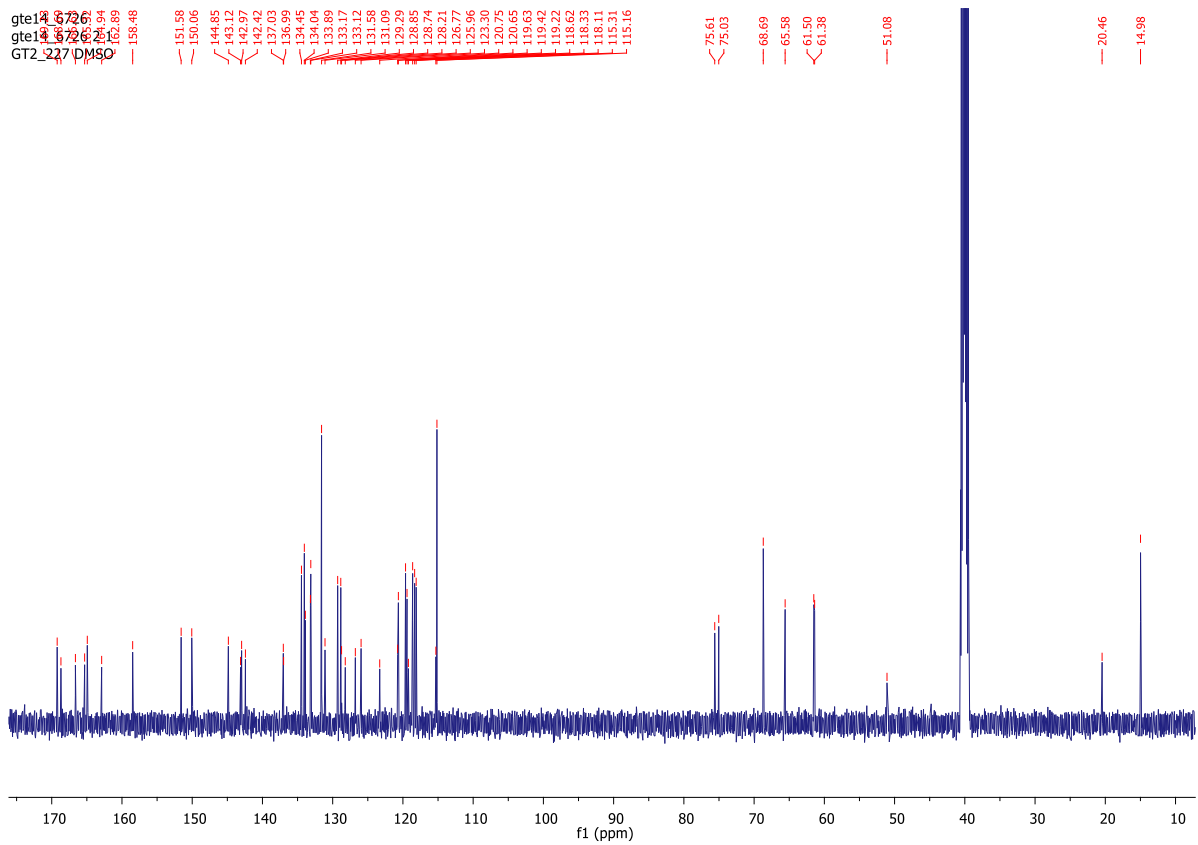
$^{13}\text{C}$  NMR (176 MHz, DMSO)  $\delta$  172.01, 168.86, 168.19, 166.14, 164.84, 163.30, 157.52, 154.36, 149.69, 142.70, 141.85, 140.24, 137.88, 136.10, 135.92, 133.89, 131.28, 129.50, 128.98, 128.78, 128.34, 127.65, 126.53, 125.69, 125.47, 119.10, 118.92, 118.19, 116.22, 115.37, 114.92, 110.27, 108.95, 60.52, 60.20, 50.58, 19.97, 14.51.

HRMS (ESI-):  $m/z$  for  $\text{C}_{44}\text{H}_{37}\text{N}_6\text{O}_{12}$   $[\text{M}-\text{H}]^-$ : calculated: 841.2475, found: 841.2475.

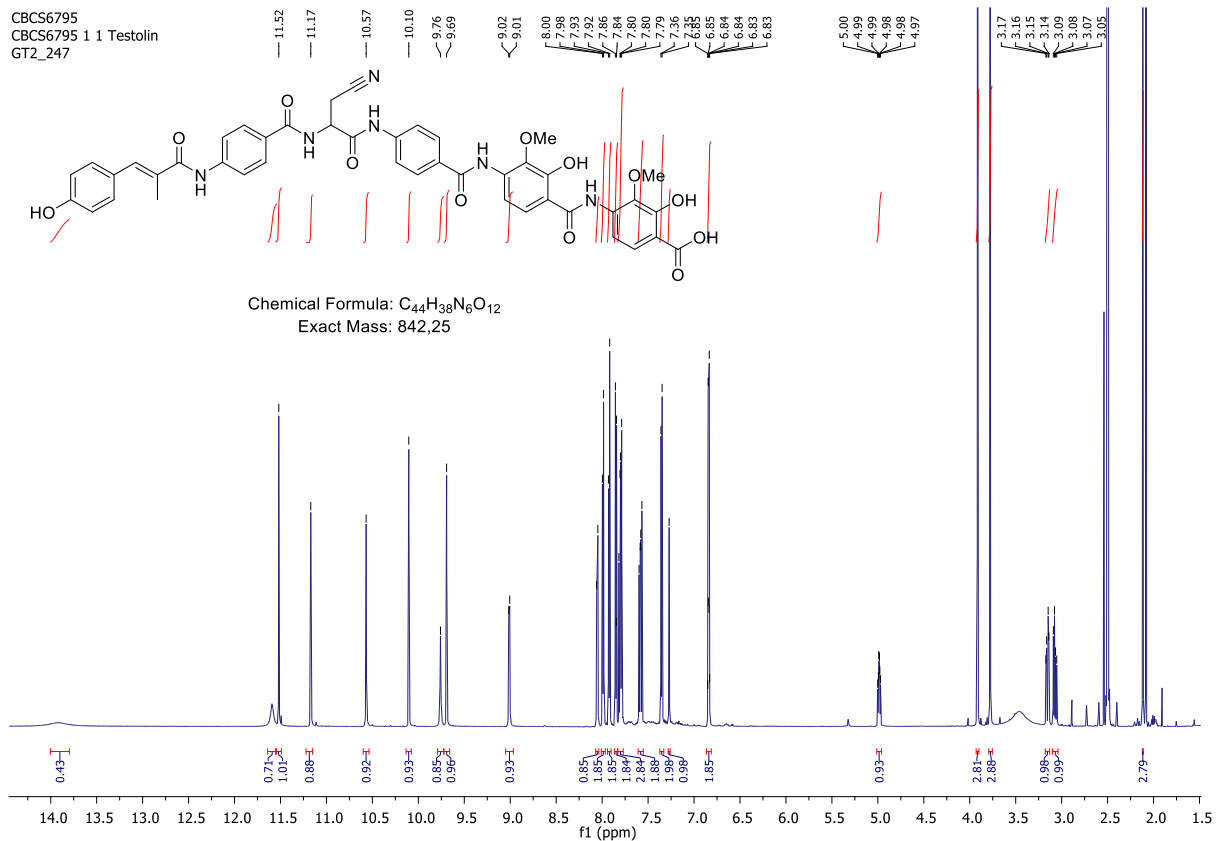
## 4.9 NMR spectra

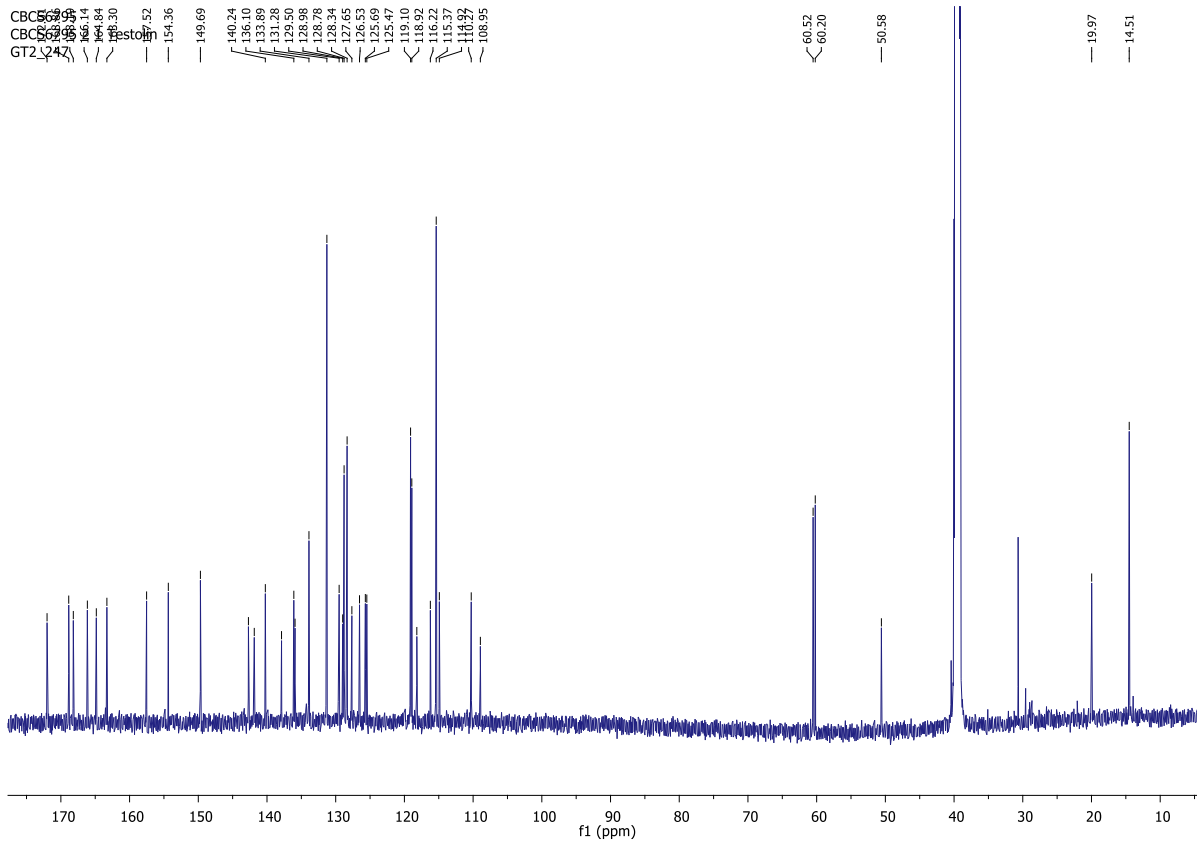
### 4.9.1 Allyl protected albicidin





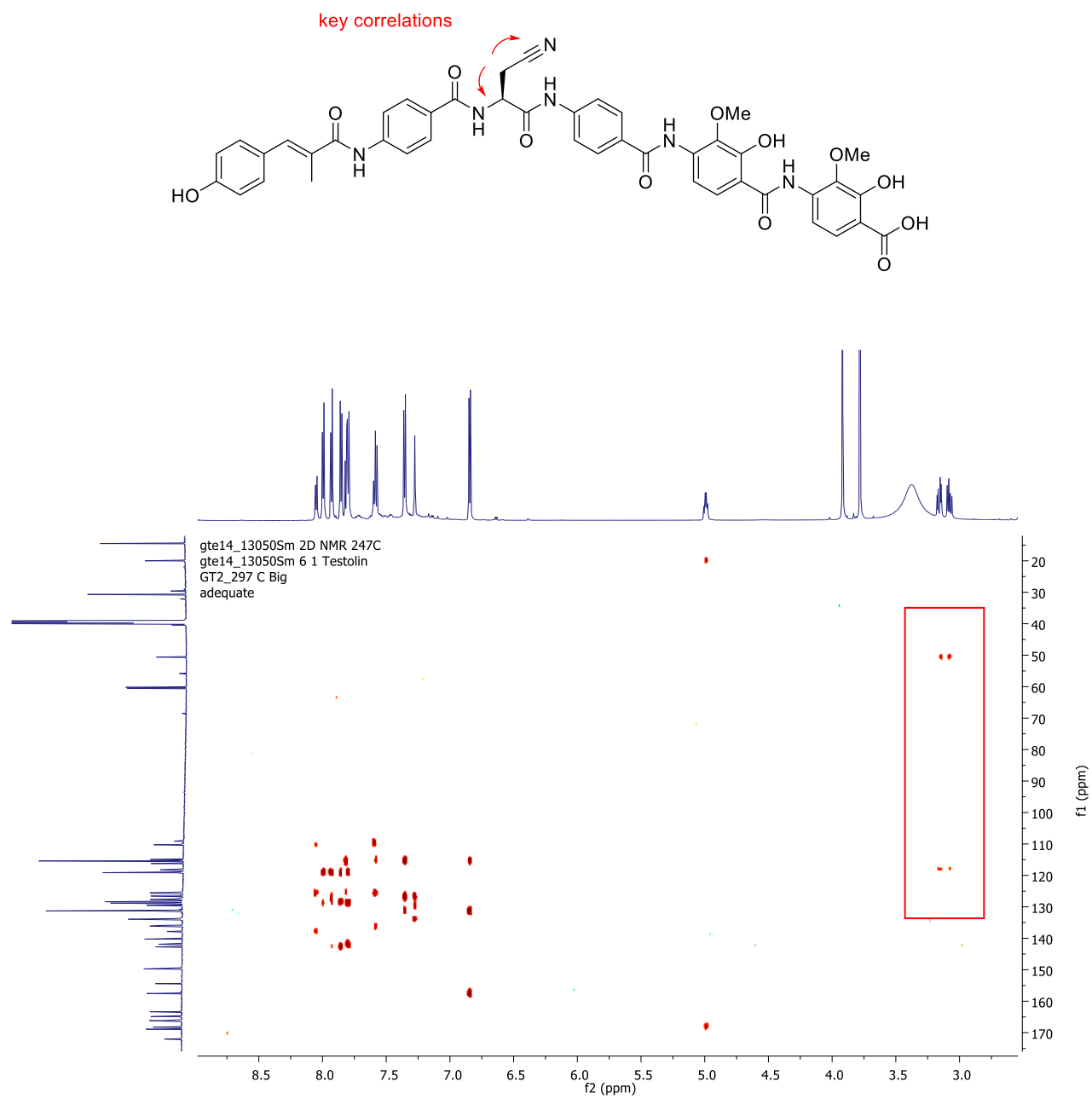
## 4.9.2 Albicidin





### 4.9.3 1, 1-adequate

evidence of  $\beta$ -cyanoalanine



## 5 References

- 1 Cragg, G. M. & Newman, D. J. Natural products: a continuing source of novel drug leads. *Biochim Biophys Acta* **1830**, 3670-3695, doi:10.1016/j.bbagen.2013.02.008 (2013).
- 2 Dias, D. A., Urban, S. & Roessner, U. A historical overview of natural products in drug discovery. *Metabolites* **2**, 303-336, doi:10.3390/metabo2020303 (2012).
- 3 Ventola, C. L. The antibiotic resistance crisis: part 1: causes and threats. *P T* **40**, 277-283 (2015).
- 4 Fair, R. J. & Tor, Y. Antibiotics and bacterial resistance in the 21st century. *Perspect Medicin Chem* **6**, 25-64, doi:10.4137/PMC.S14459 (2014).
- 5 Exner, M. *et al.* Antibiotic resistance: What is so special about multidrug-resistant Gram-negative bacteria? *GMS Hyg Infect Control* **12**, Doc05, doi:10.3205/dgkh000290 (2017).
- 6 Nikaido, H. Multidrug resistance in bacteria. *Annu Rev Biochem* **78**, 119-146, doi:10.1146/annurev.biochem.78.082907.145923 (2009).
- 7 Dever, L. A. & Dermody, T. S. Mechanisms of bacterial resistance to antibiotics. *Arch Intern Med* **151**, 886-895 (1991).
- 8 Munita, J. M. & Arias, C. A. Mechanisms of Antibiotic Resistance. *Microbiol Spectr* **4**, doi:10.1128/microbiolspec.VMBF-0016-2015 (2016).
- 9 Goh, E. B. *et al.* Transcriptional modulation of bacterial gene expression by subinhibitory concentrations of antibiotics. *Proc Natl Acad Sci U S A* **99**, 17025-17030, doi:10.1073/pnas.252607699 (2002).
- 10 Davies, J., Spiegelman, G. B. & Yim, G. The world of subinhibitory antibiotic concentrations. *Curr Opin Microbiol* **9**, 445-453, doi:10.1016/j.mib.2006.08.006 (2006).
- 11 Murakami, T., Holt, T. G. & Thompson, C. J. Thiostrepton-induced gene expression in *Streptomyces lividans*. *J Bacteriol* **171**, 1459-1466 (1989).
- 12 Chiu, M. L. *et al.* Characterization of the covalent binding of thiostrepton to a thiostrepton-induced protein from *Streptomyces lividans*. *Biochemistry* **35**, 2332-2341, doi:10.1021/bi952073e (1996).
- 13 Summers, A. O. Untwist and shout: a heavy metal-responsive transcriptional regulator. *J Bacteriol* **174**, 3097-3101 (1992).
- 14 Barrineau, P. *et al.* The DNA sequence of the mercury resistance operon of the IncFII plasmid NR1. *J Mol Appl Genet* **2**, 601-619 (1984).
- 15 Kahmann, J. D. *et al.* Structural basis for antibiotic recognition by the TipA class of multidrug-resistance transcriptional regulators. *EMBO J* **22**, 1824-1834, doi:10.1093/emboj/cdg181 (2003).
- 16 Habazettl, J. *et al.* Structural basis and dynamics of multidrug recognition in a minimal bacterial multidrug resistance system. *Proc Natl Acad Sci U S A* **111**, E5498-5507, doi:10.1073/pnas.1412070111 (2014).
- 17 Holmes, D. J., Caso, J. L. & Thompson, C. J. Autogenous transcriptional activation of a thiostrepton-induced gene in *Streptomyces lividans*. *EMBO J* **12**, 3183-3191 (1993).
- 18 Walker, M. J., Birch, R. G. & Pemberton, J. M. Cloning and characterization of an albicidin resistance gene from *Klebsiella oxytoca*. *Mol Microbiol* **2**, 443-454 (1988).
- 19 Chiu, M. L., Viollier, P. H., Katoh, T., Ramsden, J. J. & Thompson, C. J. Ligand-induced changes in the *Streptomyces lividans* TipAL protein imply an alternative mechanism of transcriptional activation for MerR-like proteins. *Biochemistry* **40**, 12950-12958 (2001).

- 20 Kretz, J. *et al.* Total synthesis of albicidin: a lead structure from *Xanthomonas albilineans* for potent antibacterial gyrase inhibitors. *Angew Chem Int Ed Engl* **54**, 1969-1973, doi:10.1002/anie.201409584 (2015).
- 21 Baumann, S. *et al.* Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity. *Angew Chem Int Ed Engl* **53**, 14605-14609, doi:10.1002/anie.201409964 (2014).
- 22 Birch, R. G. & Patil, S. S. Preliminary characterization of an antibiotic produced by *Xanthomonas albilineans* which inhibits DNA synthesis in *Escherichia coli*. *J Gen Microbiol* **131**, 1069-1075, doi:10.1099/00221287-131-5-1069 (1985).
- 23 Birch, R. G. *Xanthomonas albilineans* and the antipathogenesis approach to disease control. *Mol Plant Pathol* **2**, 1-11 (2001).
- 24 Rott, P. C., Costet, L., Davis, M. J., Frutos, R. & Gabriel, D. W. At least two separate gene clusters are involved in albicidin production by *Xanthomonas albilineans*. *J Bacteriol* **178**, 4590-4596 (1996).
- 25 Hashimi, S. M., Wall, M. K., Smith, A. B., Maxwell, A. & Birch, R. G. The phytotoxin albicidin is a novel inhibitor of DNA gyrase. *Antimicrob Agents Chemother* **51**, 181-187, doi:10.1128/AAC.00918-06 (2007).
- 26 Rostock, L. *et al.* Molecular insights into antibiotic resistance - how a binding protein traps albicidin. *Nat Commun* **9**, 3095, doi:10.1038/s41467-018-05551-4 (2018).
- 27 Basnayake, W. V. & Birch, R. G. A gene from *Alcaligenes denitrificans* that confers albicidin resistance by reversible antibiotic binding. *Microbiology* **141** ( Pt 3), 551-560, doi:10.1099/13500872-141-3-551 (1995).
- 28 Zhang, L. & Birch, R. G. The gene for albicidin detoxification from *Pantoea dispersa* encodes an esterase and attenuates pathogenicity of *Xanthomonas albilineans* to sugarcane. *Proc Natl Acad Sci U S A* **94**, 9984-9989 (1997).
- 29 Weng, L. X. *et al.* Molecular and conformational basis of a specific and high-affinity interaction between AlbA and albicidin phytotoxin. *Appl Environ Microbiol* **71**, 1445-1452, doi:10.1128/AEM.71.3.1445-1452.2005 (2005).
- 30 Weng, L. X., Xu, J. L., Li, Q., Birch, R. G. & Zhang, L. H. Identification of the essential histidine residue for high-affinity binding of AlbA protein to albicidin antibiotics. *Microbiology* **149**, 451-457, doi:10.1099/mic.0.25942-0 (2003).
- 31 Zhang, L., Xu, J. & Birch, R. G. High affinity binding of albicidin phytotoxins by the AlbA protein from *Klebsiella oxytoca*. *Microbiology* **144** ( Pt 2), 555-559, doi:10.1099/00221287-144-2-555 (1998).
- 32 Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**, W545-549, doi:10.1093/nar/gkq366 (2010).
- 33 Chiu, M. L. *et al.* Broad spectrum thiopeptide recognition specificity of the *Streptomyces lividans* TipAL protein and its role in regulating gene expression. *J Biol Chem* **274**, 20578-20586 (1999).
- 34 Gratz, S. *et al.* Synthesis and Antimicrobial Activity of Albicidin Derivatives with Variations of the Central Cyanoalanine Building Block. *ChemMedChem* **11**, 1499-1502, doi:10.1002/cmde.201600163 (2016).
- 35 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).
- 36 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).
- 37 Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355-367, doi:10.1107/S0907444911001314 (2011).

- 38 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21, doi:10.1107/S09074444909042073 (2010).
- 39 Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng* **8**, 127-134 (1995).
- 40 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403-410, doi:10.1016/S0022-2836(05)80360-2 (1990).
- 41 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).
- 42 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313, doi:10.1093/bioinformatics/btu033 (2014).

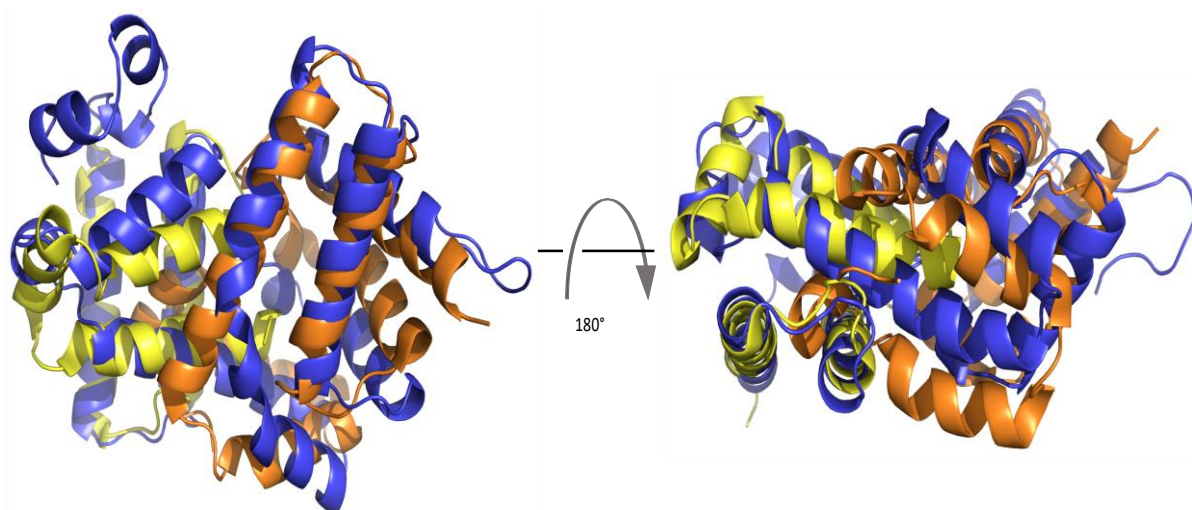


Figure S1: Cartoon representation of  $AlbA_{wt}$  (blue) aligned with two copies of ligand-bound TipA-S (PDB ID 2mc0, yellow and orange). The  $C_{\alpha}$  RMSD between  $AlbA$  and the yellow copy of TipA-S was 3.7 Å, while the  $C_{\alpha}$  RMSD between  $AlbA$  and the orange copy of TipA-S was 3.9 Å.

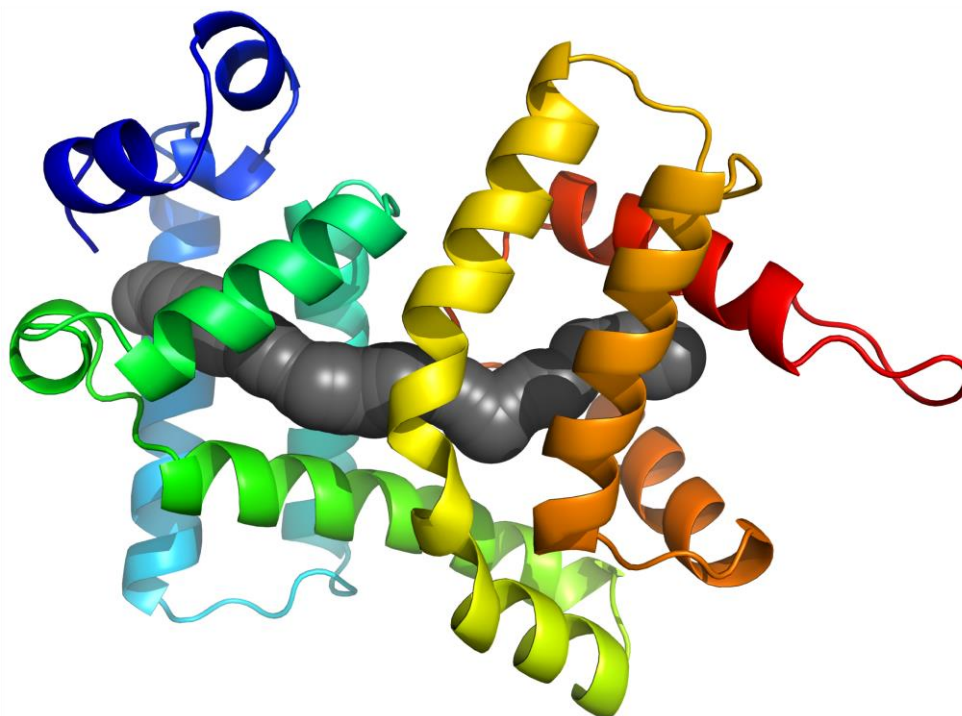


Figure S2: Cartoon representation of  $AlbA_{wt}$  using the rainbow color scheme (N-terminus blue, C-terminus red). The putative substrate binding tunnel (grey) was identified using the Cover 3.0.1 plug-in for PyMol. The tunnel volume was calculated as approximately 700 Å<sup>3</sup>.



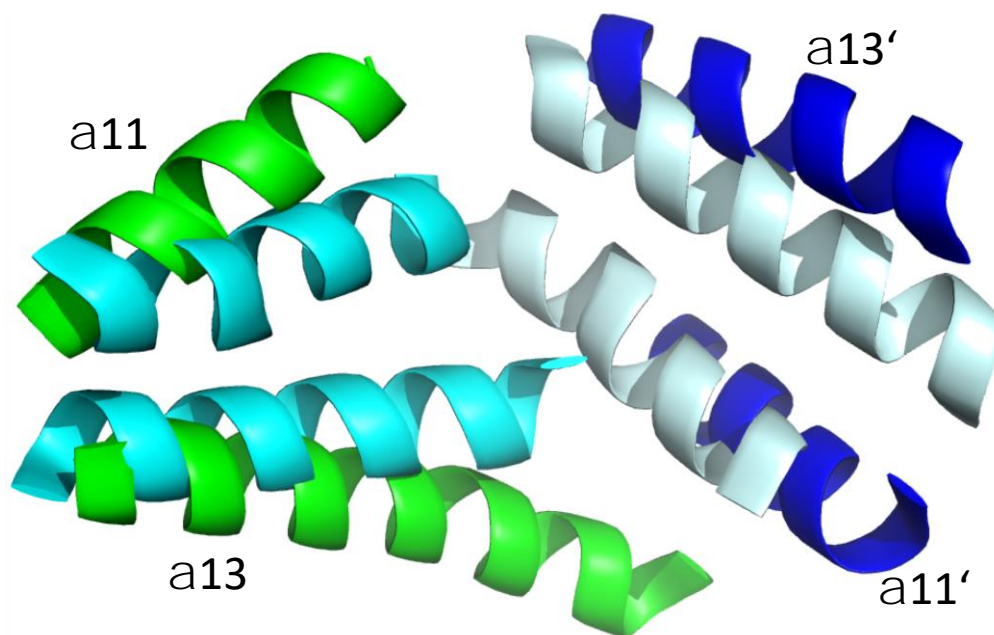


Figure S3: Cartoon representation of  $\alpha$ -helices 11 and 13 as well as 11' and 13' from  $AlbA_{wt}$  (green and blue) with the corresponding helices from  $TipA-S$  (PDB ID 2mc0, cyan and light blue). In  $AlbA_{wt}$ , the helices are no longer in close proximity and parallel, but spread apart to create the substrate binding tunnel.

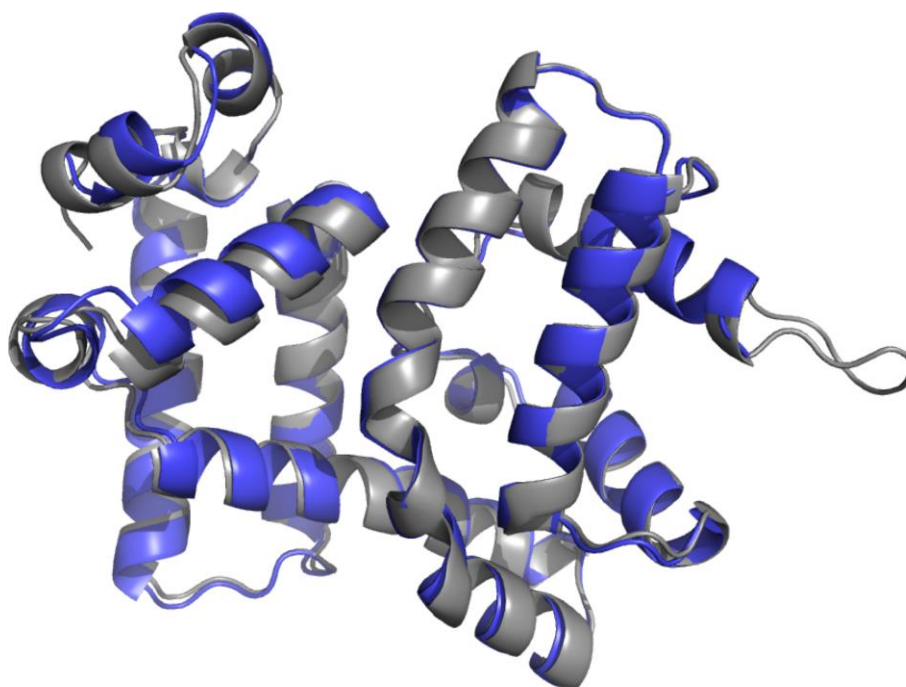


Figure S4: Superposition of the  $AlbA_{wt}$  structure with (blue) and without (grey) albicidin bound. Changes in the overall structure of the protein are minimal (Rmsd of 0.8 Å over all non-hydrogen atoms).

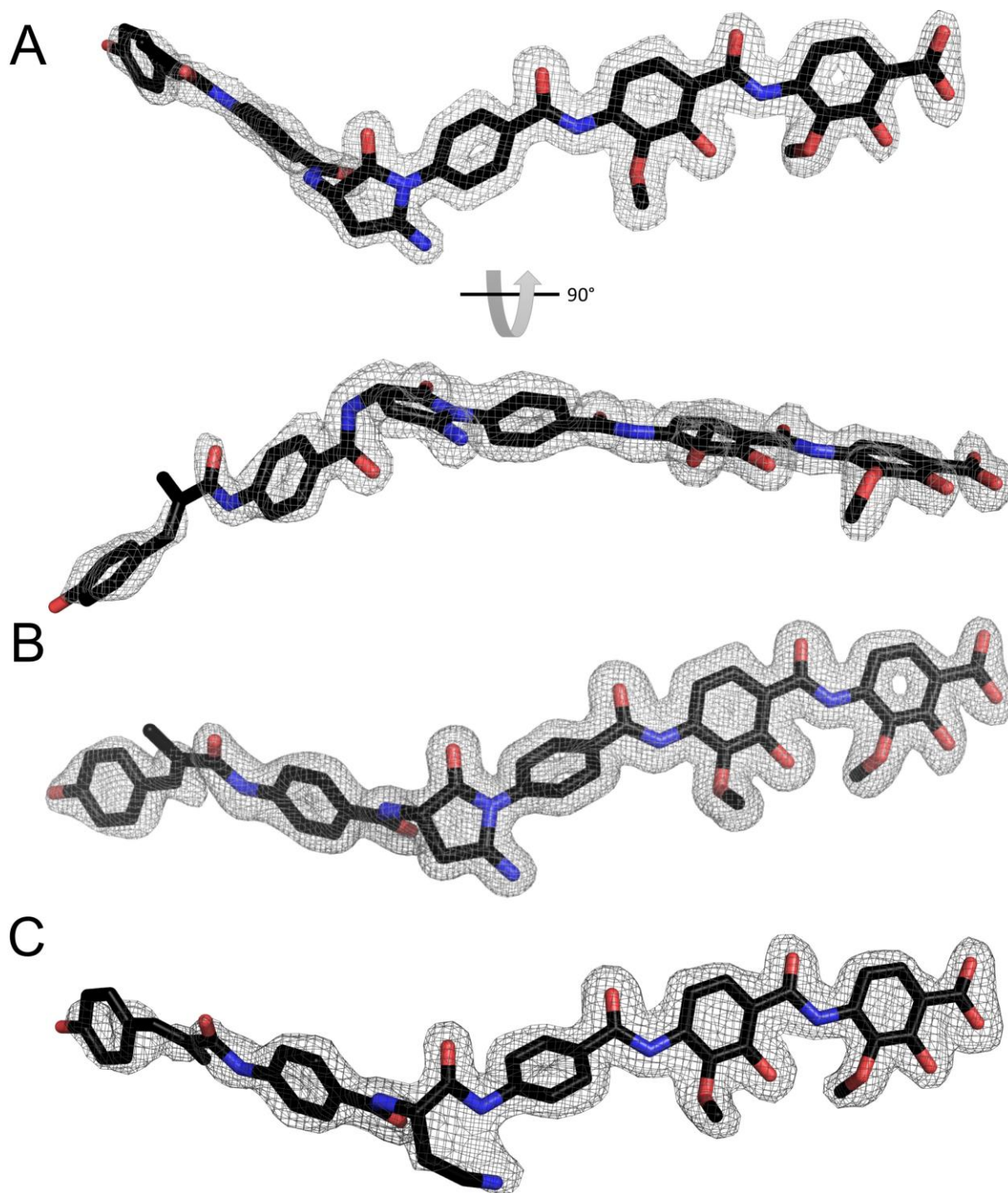


Figure S5: (A) Stick representation of 2 (Atom colors: carbon black, nitrogen blue, oxygen red) and the corresponding difference electron density map (grey isomesh). The difference electron density ( $F_o - F_c$ ) was contoured at  $3\sigma$  with phases calculated from a model which was refined with no 2 present. (B) Same as A, but polder map of the ligand contoured at  $5\sigma$ . (C) Same as A, but the data published by Rostock et al. was used (PDB 6et8). As can be seen, 1 does not fit the electron density.

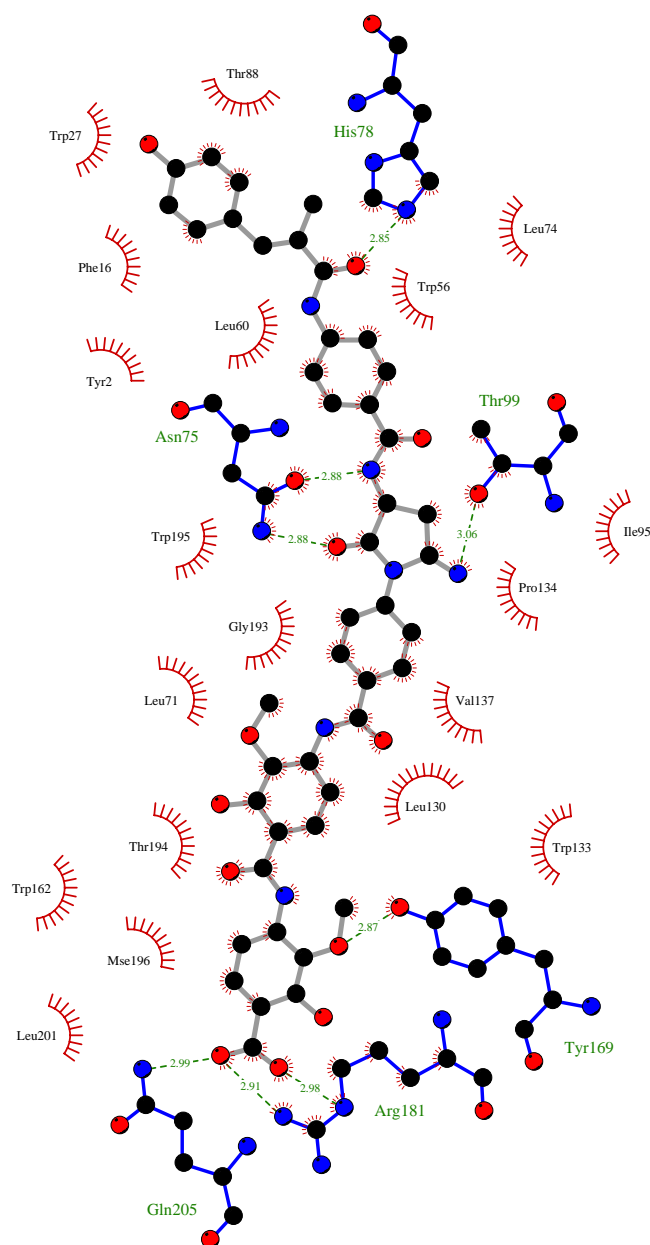


Figure S6: LigPlot diagram of the interactions between *AlbA<sub>wt</sub>* and albicidin. Albicidin is almost completely buried in the *AlbA* tunnel and engages in extensive hydrophobic interactions. Several residues identified previously as important for albicidin binding through an alanine scan are in fact in contact with albicidin, but they are also part of the hydrophobic core of *AlbA*. It is thus unclear how specific the observed effects were. In addition to the hydrophobic interactions, *AlbA* and albicidin form a salt-bridge and several hydrogen bonds, all mediated by side-chains. The carbonyl oxygen O13 (numbering in Figure 2B) forms a hydrogen bond with H78, while N75 is a bidentate ligand, forming hydrogen bonds with NH21 (immediately adjacent to the five-membered ring) and O26 (part of the newly formed ring). The ketimine / ketone moiety (NH / O62) is hydrogen bonded to T99, and the final interactions concern the terminal para-aminobenzoic acid unit. The side-chain of Y169 is hydrogen-bonded to O56 (methylated *m*-oxygen), while Q205 and R181 form a hydrogen bond and salt bridge, respectively, with the terminal carboxyl group. Albicidin bonds are grey, protein residue bonds blue. Atoms are represented as circles (carbon black, oxygen red, nitrogen blue). Salt-bridges and hydrogen-bonds are shown as dashed green lines with distances given. Hydrophobic interactions are depicted as red spoked arcs.

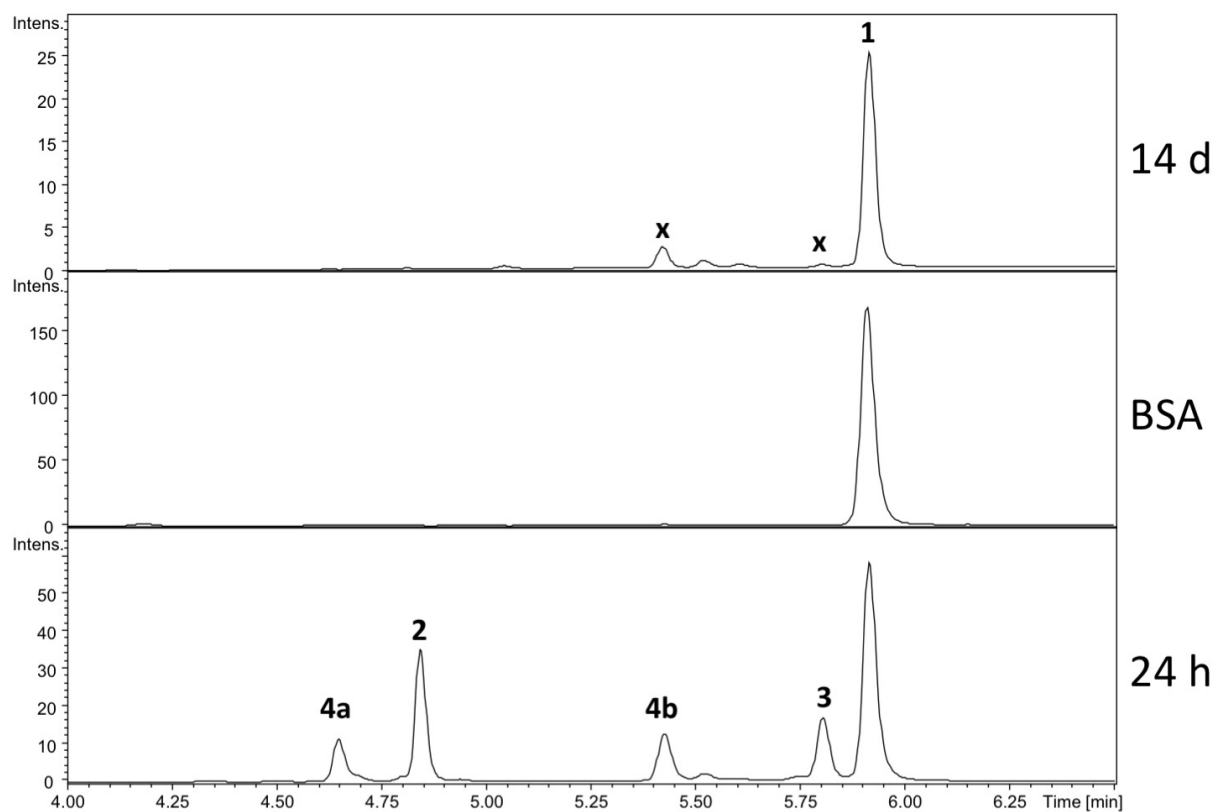


Figure S7: LC-MS traces of albicidin incubated in reaction buffer (PBS) for 14 days at 37 °C (top), for 24 h with BSA (middle) and with AlbA<sub>wt</sub> for 24 h at 37 °C. x denotes peaks that are unrelated to albicidin.

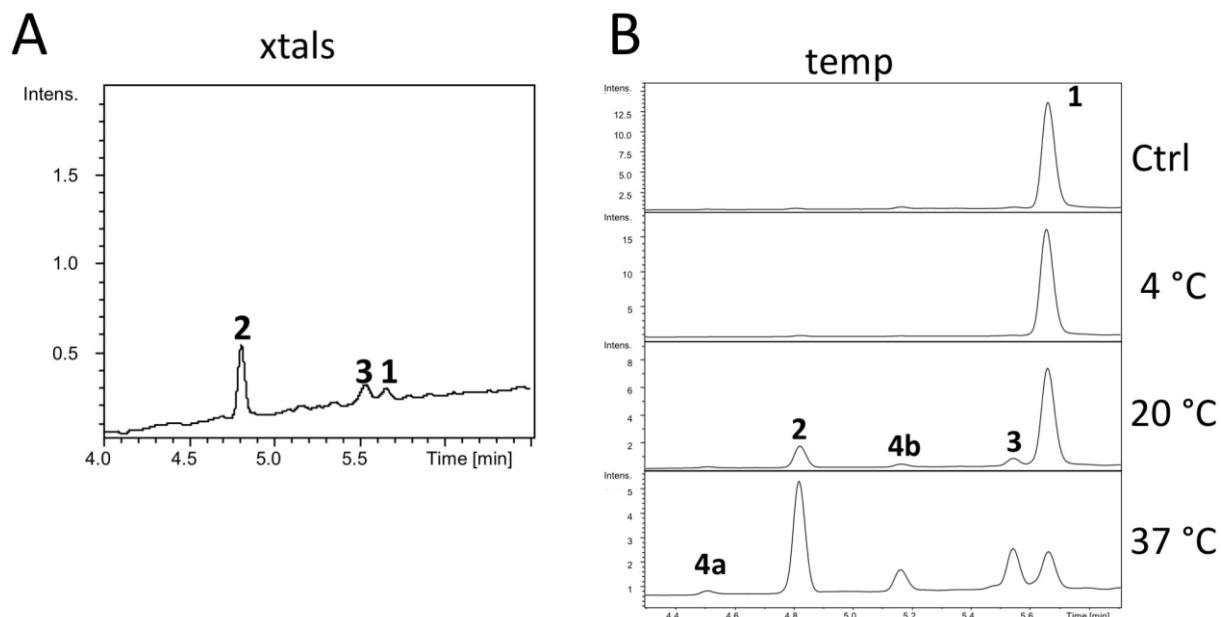


Figure S8: (A) LC-MS trace of AlbA<sub>wt</sub>-albicidin complex crystals harvested after 7 days. The predominant species is **2**, which was consequently used in refinement. Traces of **3** are likely the result of sample processing. We observed minute traces of **1**, which may be due to the short incubation time, compound adsorbed to the crystals or enclosed in the crystal lattice. (B) Temperature-dependence of the AlbA-dependent albicidin (1 : 1) cyclization. Samples were taken after 24 h incubation at the indicated temperature.

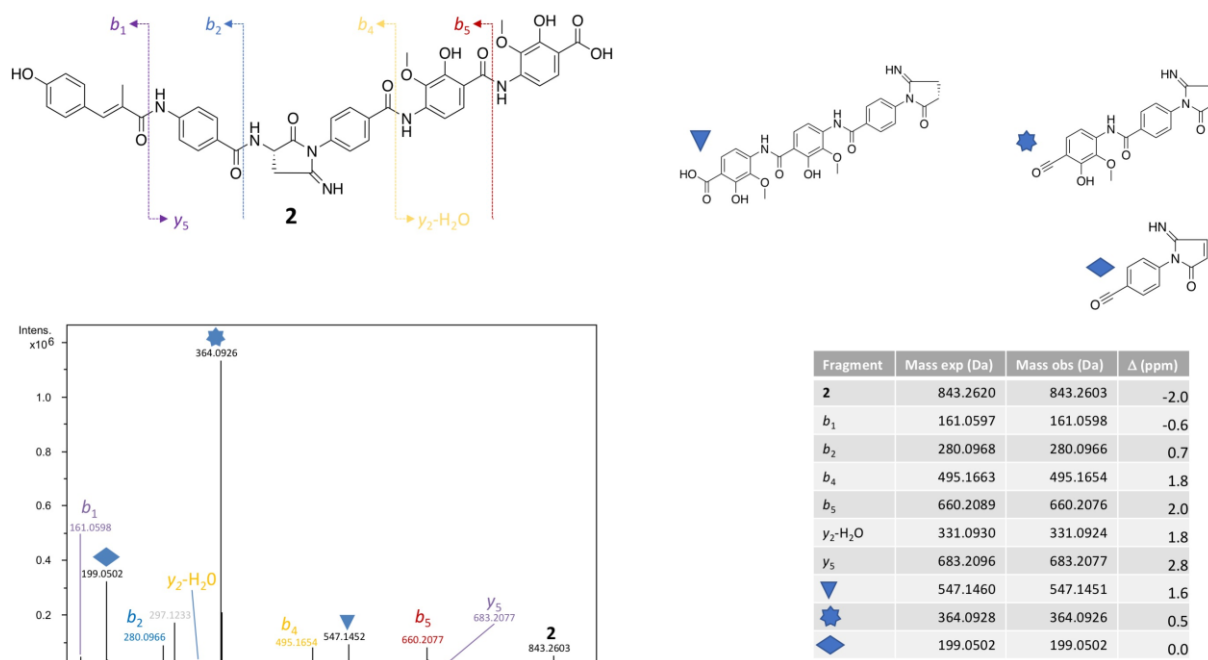


Figure S9: MS<sup>2</sup> of **2** with assigned fragments color coded. Masses and corresponding errors are given in the table.

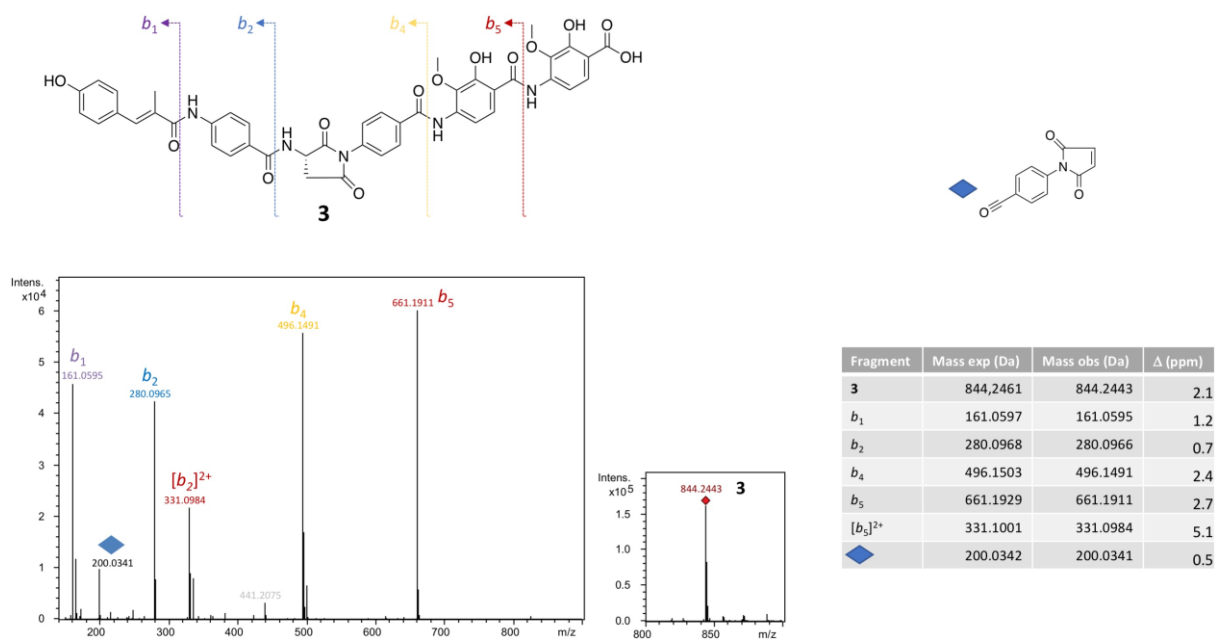


Figure S10:  $MS^2$  of **3** with assigned fragments color coded. Masses and corresponding errors are given in the table.

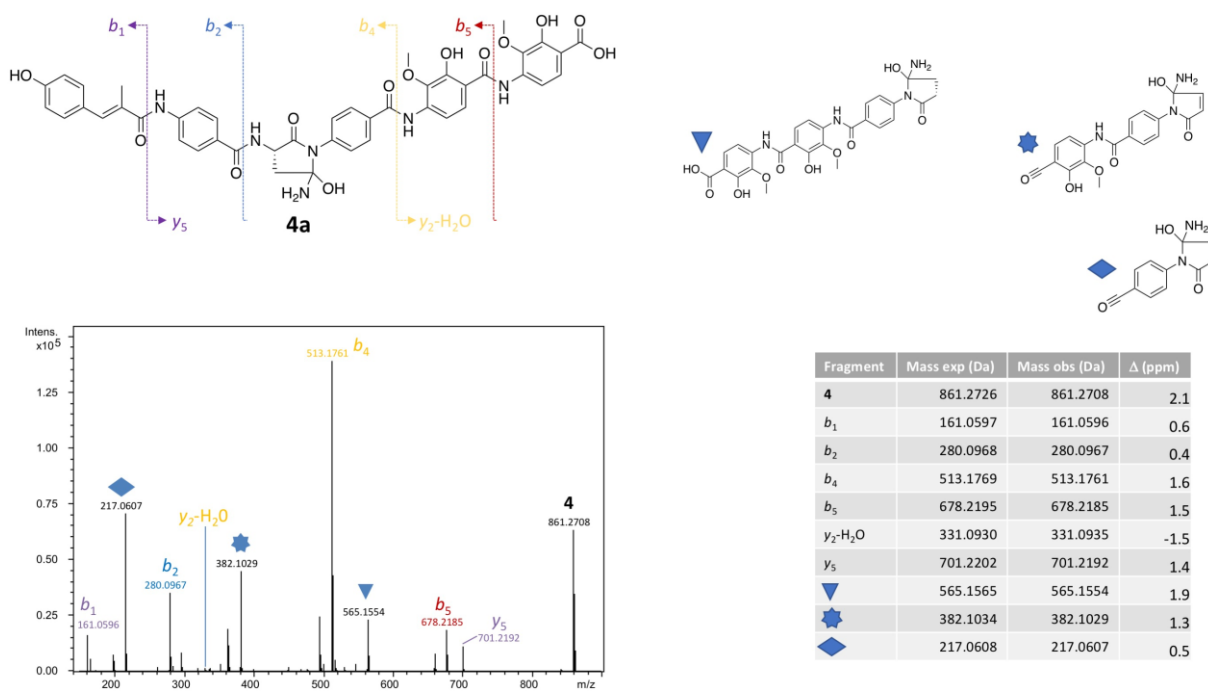
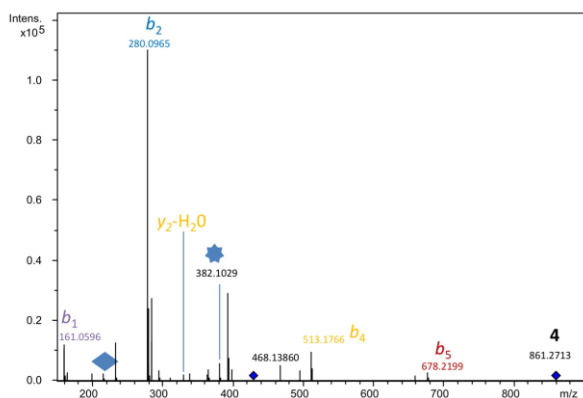
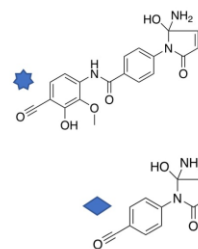
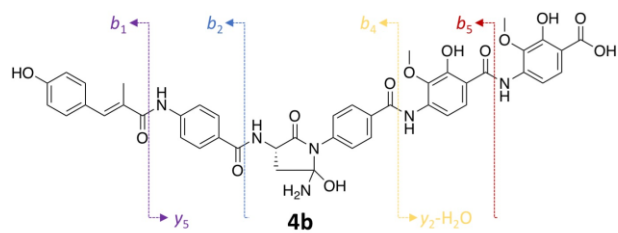


Figure S11:  $MS^2$  of **4a** with assigned fragments color coded. Masses and corresponding errors are given in the table.



Fragment	Mass exp (Da)	Mass obs (Da)	$\Delta$ (ppm)
<b>4</b>	861.2726	861.2713	-1.5
b <sub>1</sub>	161.0597	161.0596	0.6
b <sub>2</sub>	280.0968	280.0965	-1.1
b <sub>4</sub>	513.1769	513.1766	-0.6
b <sub>5</sub>	678.2195	678.2199	0.6
y <sub>2</sub> -H <sub>2</sub> O	331.0930	331.0940	3.0
Star	382.1034	382.1029	1.3
Diamond	217.0608	217.0607	0.5

Figure S12: MS<sup>2</sup> of 4b with assigned fragments color coded. Masses and corresponding errors are given in the table.

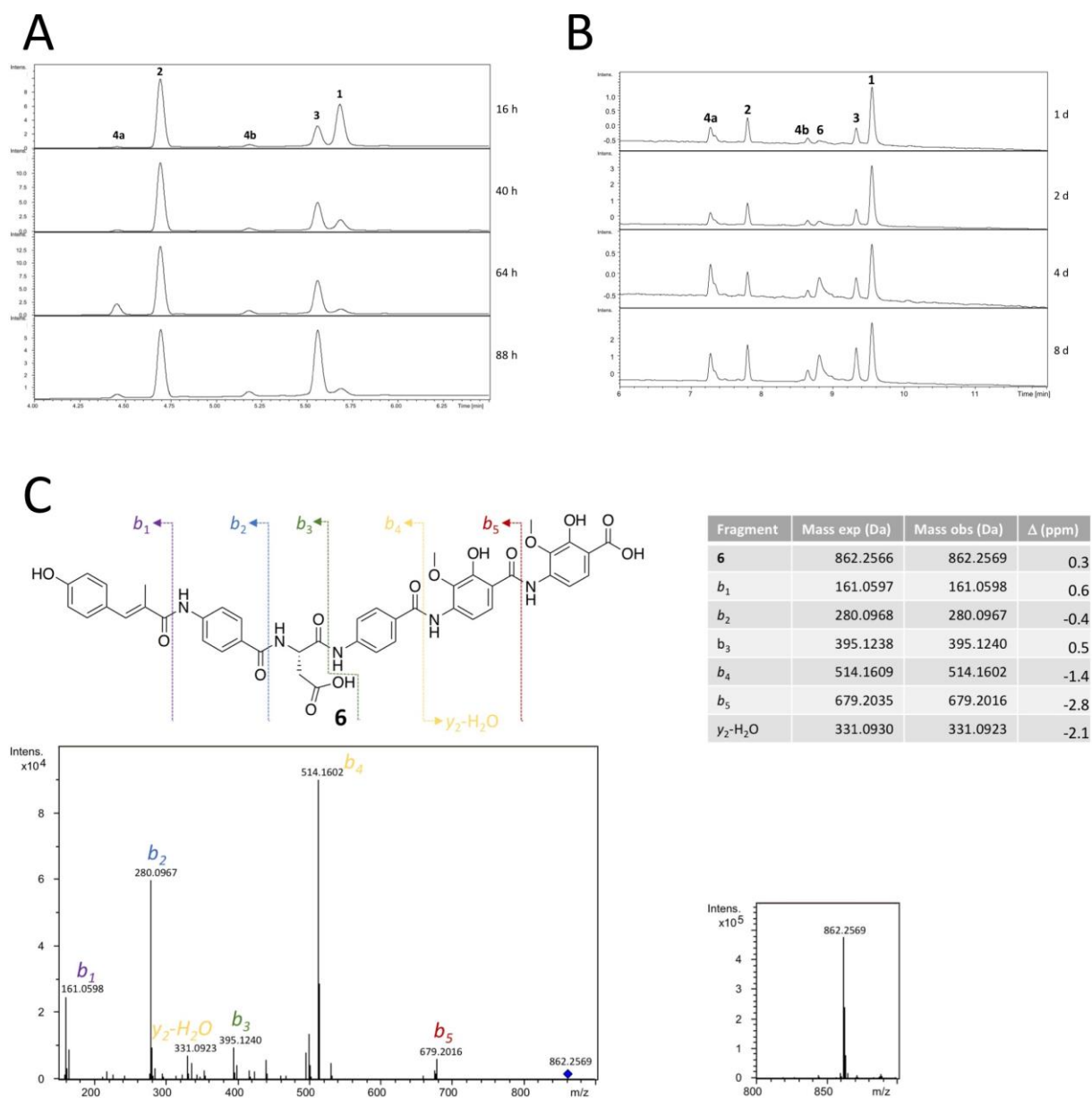
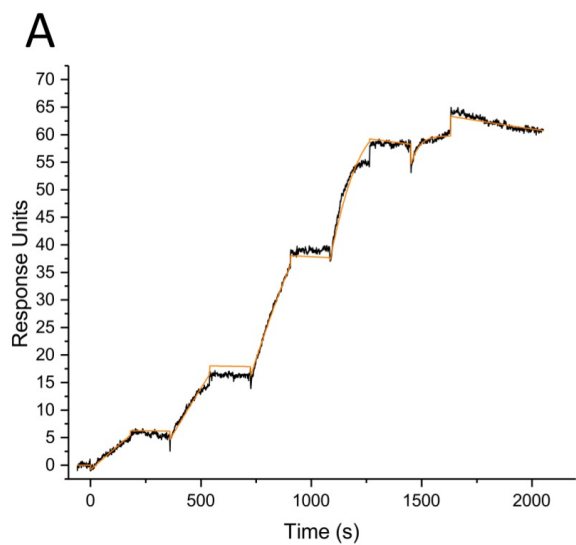


Figure S13: (A) Time course of  $\text{AlbA}_{wt}$  incubated with albicidin at  $37^\circ\text{C}$  at a 1 : 1 ratio. Time points are indicated and peaks assigned by LC-MS. The initial product is 2, which then slowly converts to 3. After 64 h albicidin is consumed, while 2 is continuously converted to 3. (B) Same as A, but using a 1/10  $\text{AlbA}_{wt}$  concentration. Naturally, much longer incubation times were required. (C)  $\text{MS}^2$  analysis of 6.

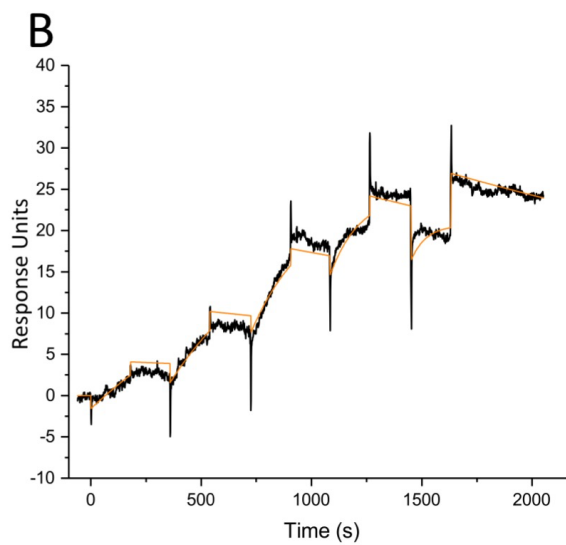




$$k_a = 4.982 \times 10^4 \text{ M}^{-1}\text{s}^{-1} \pm 7.0 \times 10^2$$

$$k_d = 1.059 \times 10^{-4} \text{ s}^{-1} \pm 6.0 \times 10^{-6}$$

$$\underline{K_D = 2.1 \text{ nM}}$$



$$k_a = 2.415 \times 10^4 \text{ M}^{-1}\text{s}^{-1} \pm 5.1 \times 10^2$$

$$k_d = 2.798 \times 10^{-4} \text{ s}^{-1} \pm 5.5 \times 10^{-5}$$

$$\underline{K_D = 11.6 \text{ nM}}$$

Figure S14: SPR of AlbA<sub>wt</sub> with 1 (A) and 3 (B). The affinity of the product 3 is six-fold weaker than that of the substrate, and the off-rate three times faster. These comparably small differences may be the main reason conversion of albicidin is slow.

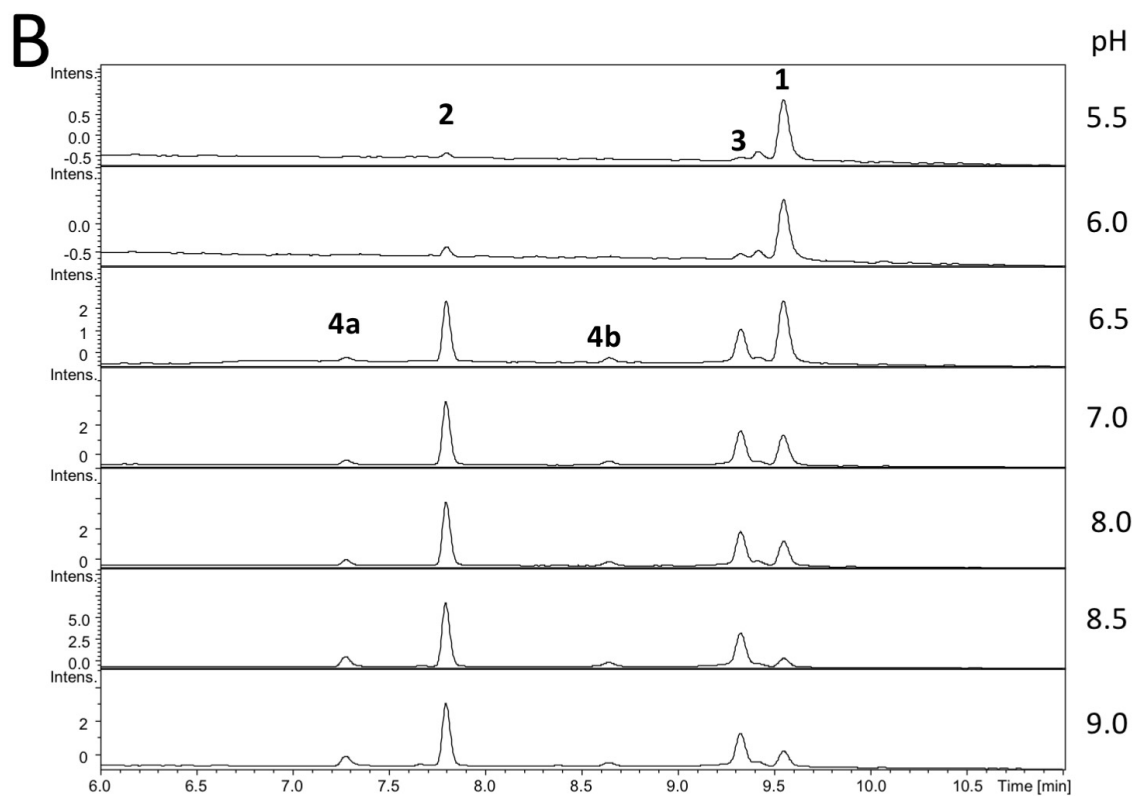
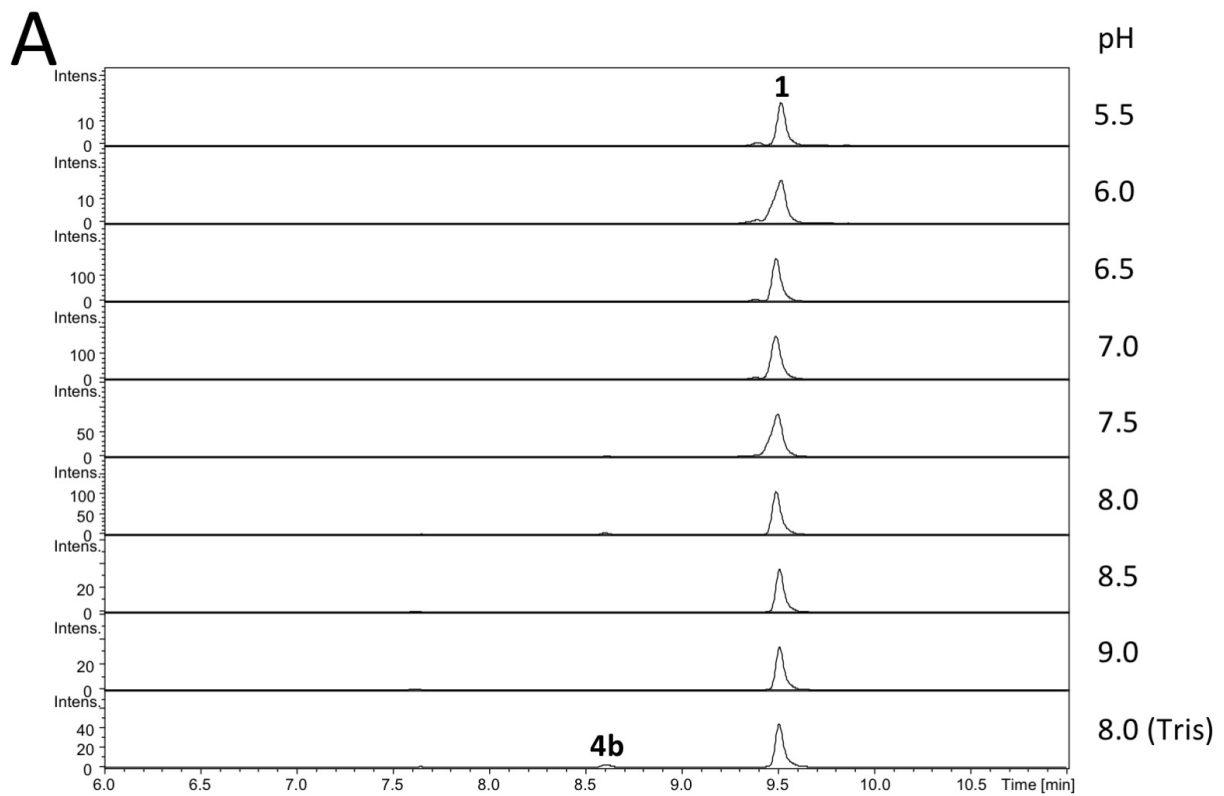


Figure S15: (A) Controls containing albicidin but no  $AlbA_{wt}$  were incubated at different pH at 37 °C for 24 h. pH 5.5 – 6.5 MES, pH 7.0 – 7.5 HEPES, pH 8.0 – 9.0 Bicine. (B) Same as A, but with  $AlbA_{wt}$ . pH 7.5 was omitted since all other incubations were carried out in PBS (pH 7.4).

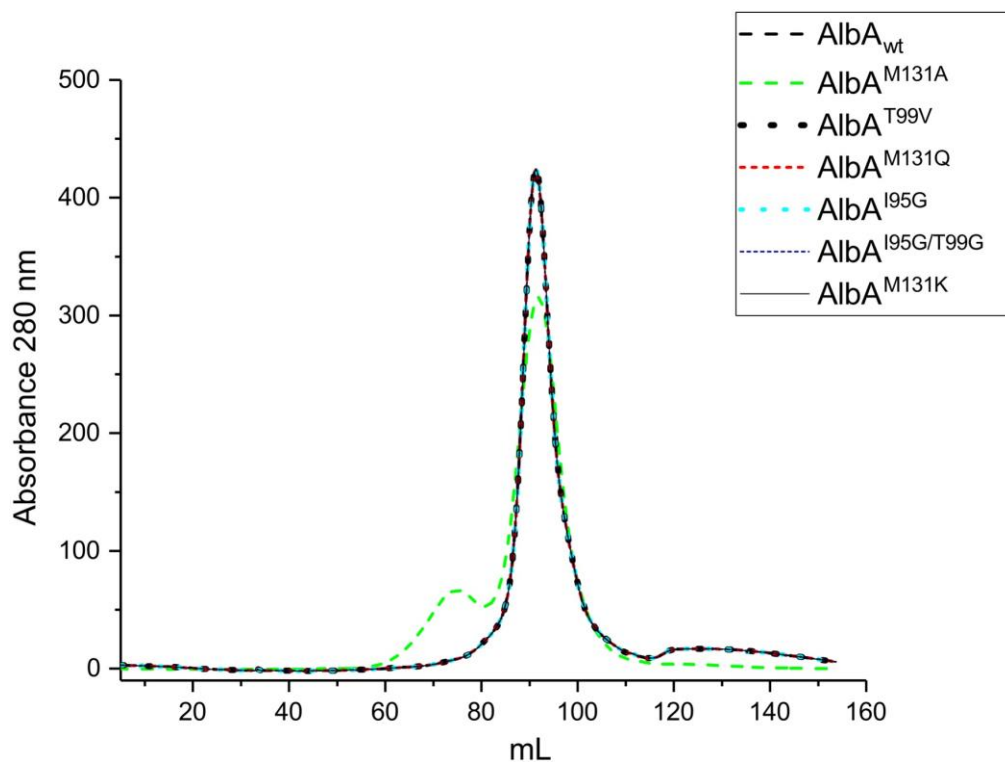


Figure S16: Superposition of size-exclusion chromatography traces of  $\text{AlbA}_{\text{wt}}$  and all mutants. The proteins show near-identical elution volumes.

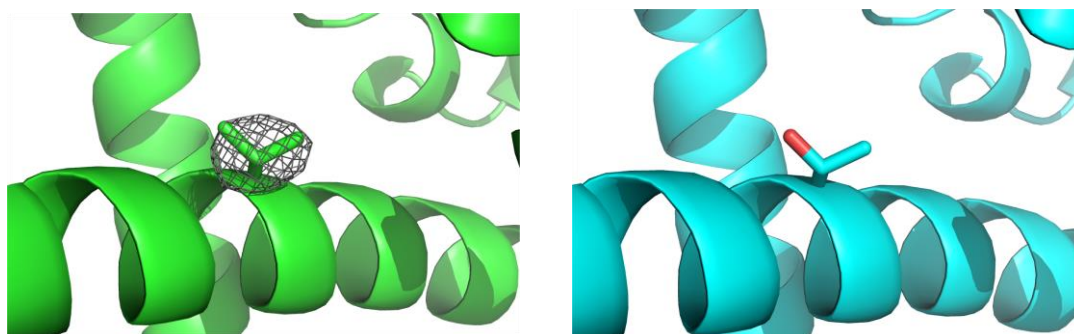


Figure S17: Comparison of the position and orientation of the side-chains  $\text{AlbA}^{\text{T99V}}$  (green, left) with  $\text{AlbA}_{\text{wt}}$ . The difference electron density ( $F_o - F_c$ ) for the V99 residue (grey isomesh) was contoured at  $3\sigma$  with phases calculated from a model which was refined with a glycine in position 99.

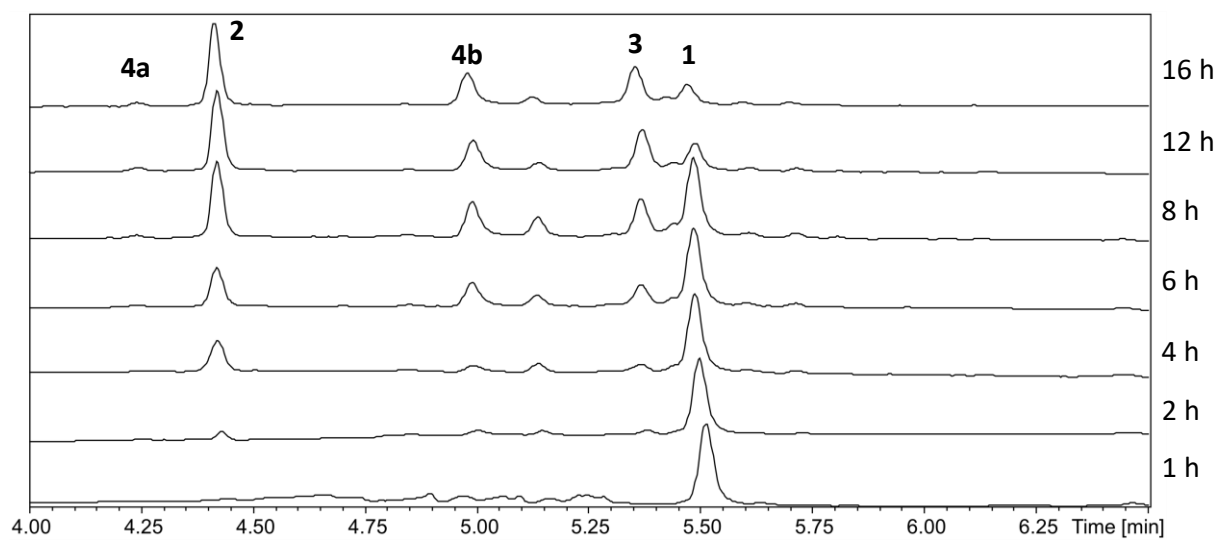


Figure S18: Time-course of the  $AlbA^{T99V}$ -dependent cyclization of albicidin. LC-MS sampling points are indicated on the right. After 16 h, almost all albicidin is consumed.

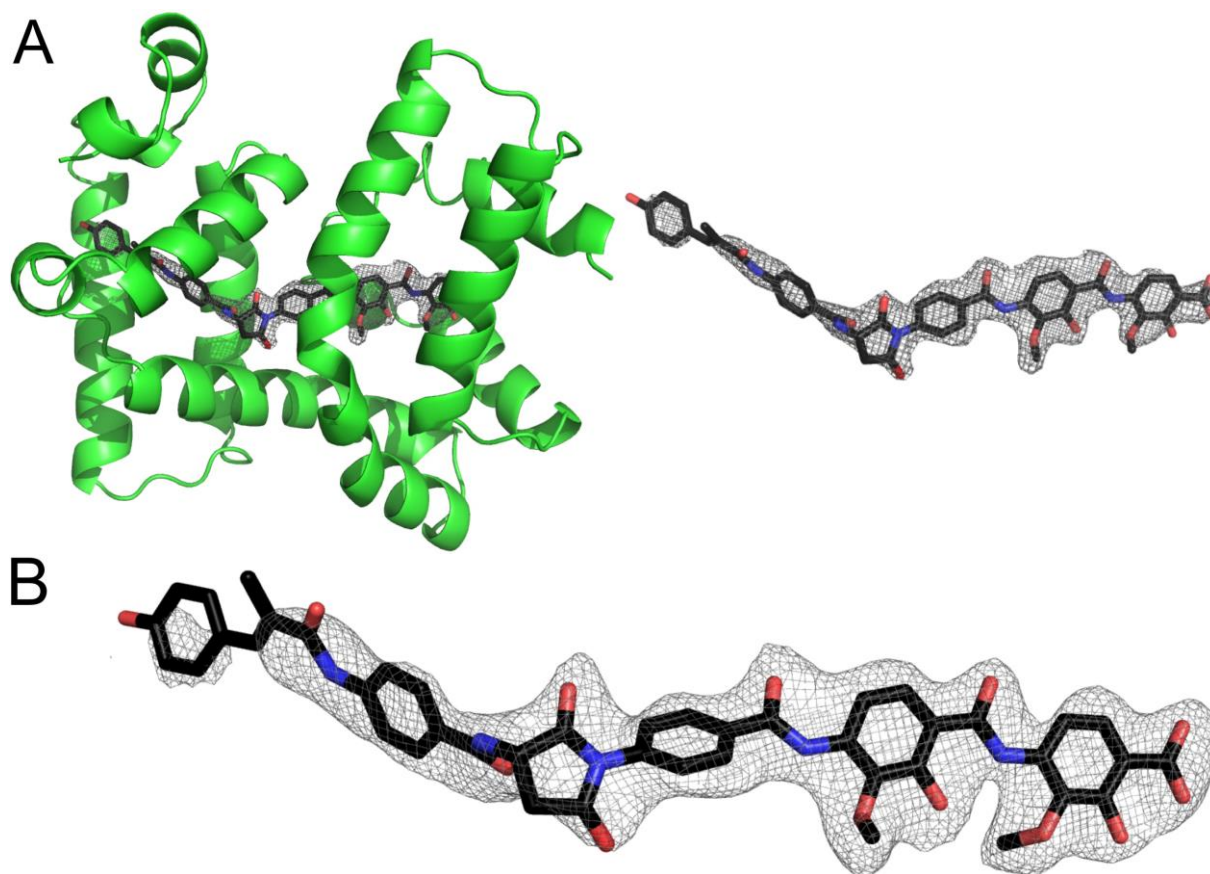


Figure S19: (A) Cartoon representation of the overall structure of the AlbA<sup>M131A</sup> - 3 complex (left, green). 3 is shown as sticks (Atom colors: Carbon black, nitrogen blue, oxygen red). The difference electron density ( $F_o - F_c$ ) was contoured at  $3\sigma$  with phases calculated from a model which was refined with no 3 present and is shown as a grey isomesh. (B) Polder map of the ligand contoured at  $3\sigma$ .

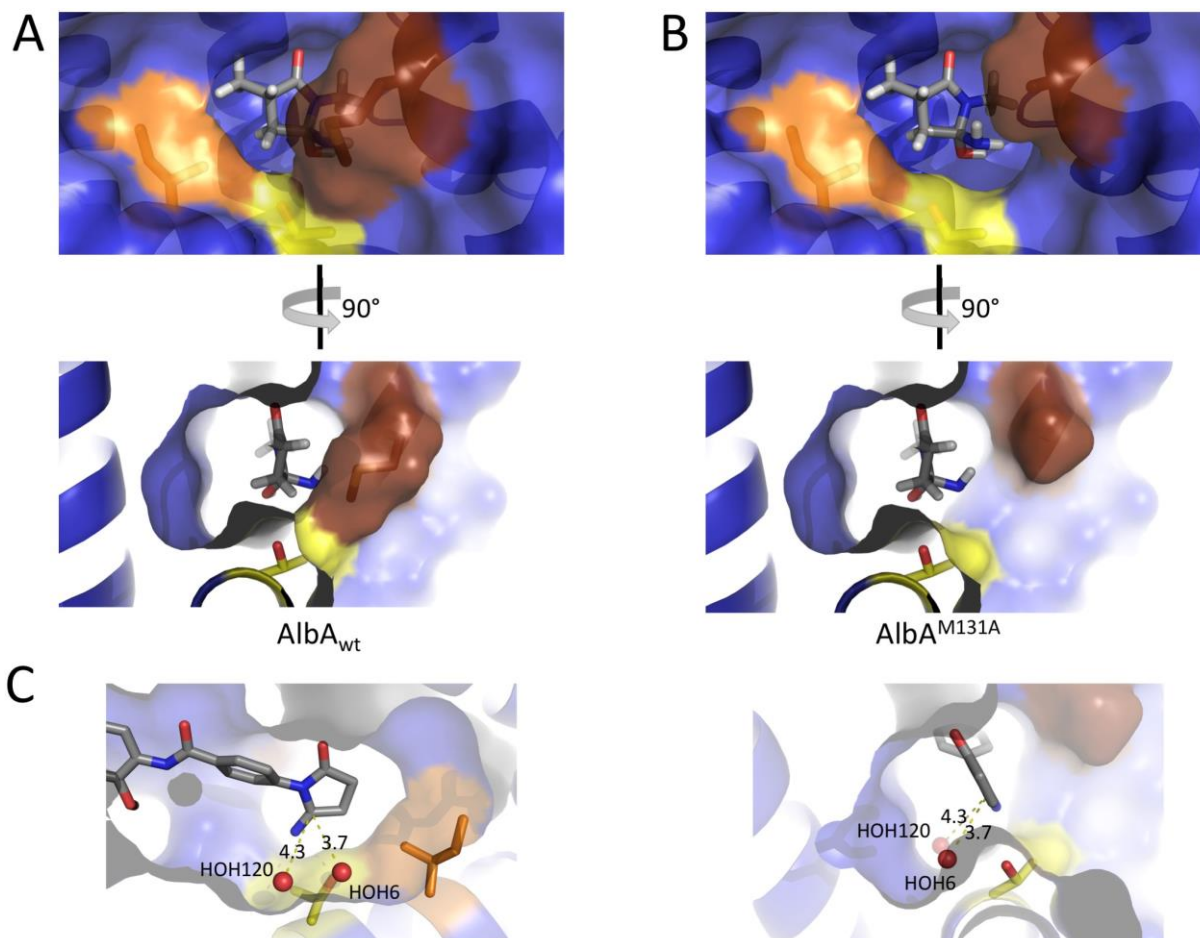


Figure S20: (A) Surface representation of the AlbA<sub>wt</sub> – albicidin complex structure (blue) with residues in contact with the newly formed ketimine moiety highlighted (I95 orange, T99 yellow, M131 brown). The transition state (4a/b) is modeled in place of the ketimine moiety (grey sticks). In the wt protein 4a/b cannot form since it would clash with M131. (B) Same as left, but based on the complex structure of AlbA<sup>M131A</sup>. Here the transition state between 2 and 3 does not clash with the protein and could thus be formed while 2 is still bound to the protein. (C) Two ordered water molecules (red spheres) are placed in ideal positions for nucleophilic attack on the ketimine carbon.

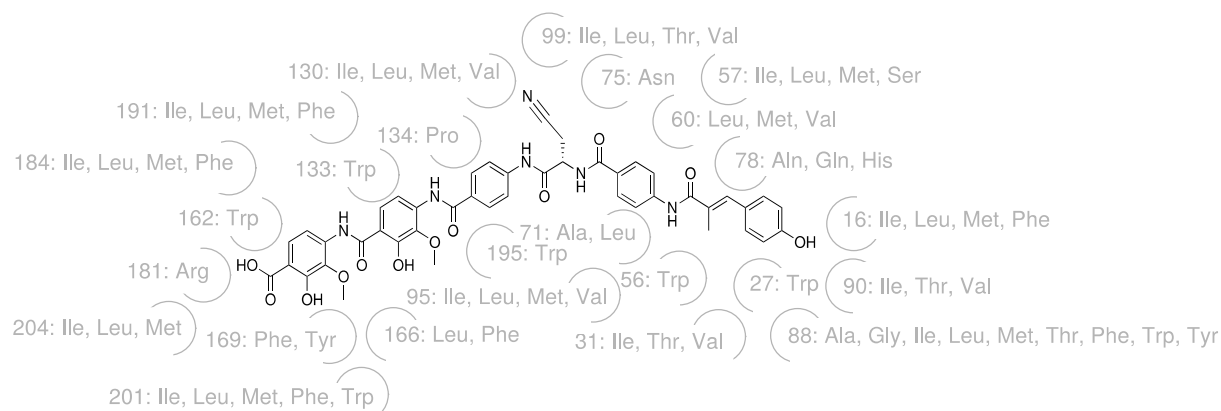


Figure S21: Analysis of the AlbA complex structure allowed the definition of a contact pattern. This includes the conservation requirements for all residues in direct contact with albicidin.

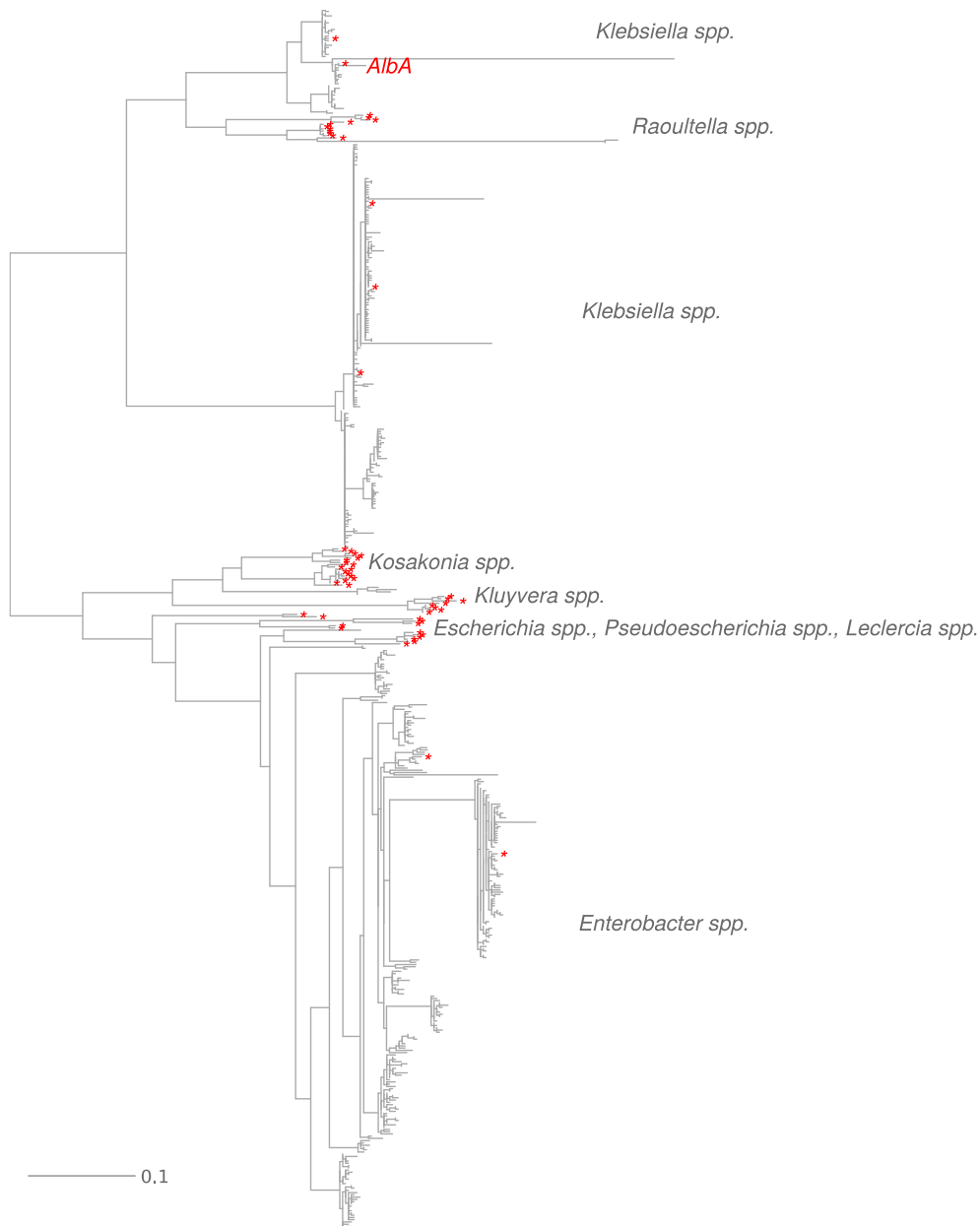


Figure S22: Phylogenetic analysis of Alba-like sequences in publicly available genomes. A search of the non-redundant database for Alba homologs using default BLAST parameters returned 1,906 sequences from beta-, gamma-, and delta-proteobacteria, all of which were annotated as mercuric ion resistance (*MerR*) family transcription regulators. Next, we the contact pattern (Figure S23) to filter the obtained homologs. Phylogenetic analysis reveals that this pattern can occasionally be found in *Klebsiella* and *Enterobacter* species, as well as in several closely related sequences from *Raoultella*, *Kosakonia*, *Kluyvera*, *Escherichia*, *Pseudoescherichia*, and *Leclercia* spp.. The distribution of the relevant sequences in the phylogenetic tree suggests that in *Klebsiella* and *Enterobacter* the pattern evolved on several independent occasions, whereas in *Raoultella*, *Kosakonia*, *Kluyvera*, *Escherichia*, *Pseudoescherichia*, and *Leclercia*, close homology coupled with the virtual omnipresence of sequences containing the contact pattern suggests introduction of the protein into these species via horizontal gene transfer as a means of resistance against albicidin. Additionally, the *Raoultella* branch is embedded into the *Klebsiella* branch, which suggests *Klebsiella* as the source of the horizontal gene transfer into *Raoultella*. For other species, the phylogeny is not well resolved at the branching point, so the source of the transfer is more difficult to pinpoint. Red stars indicate proteins with a conserved contact pattern as defined in Figure S21. A vector graphic of this tree is also available separately online.

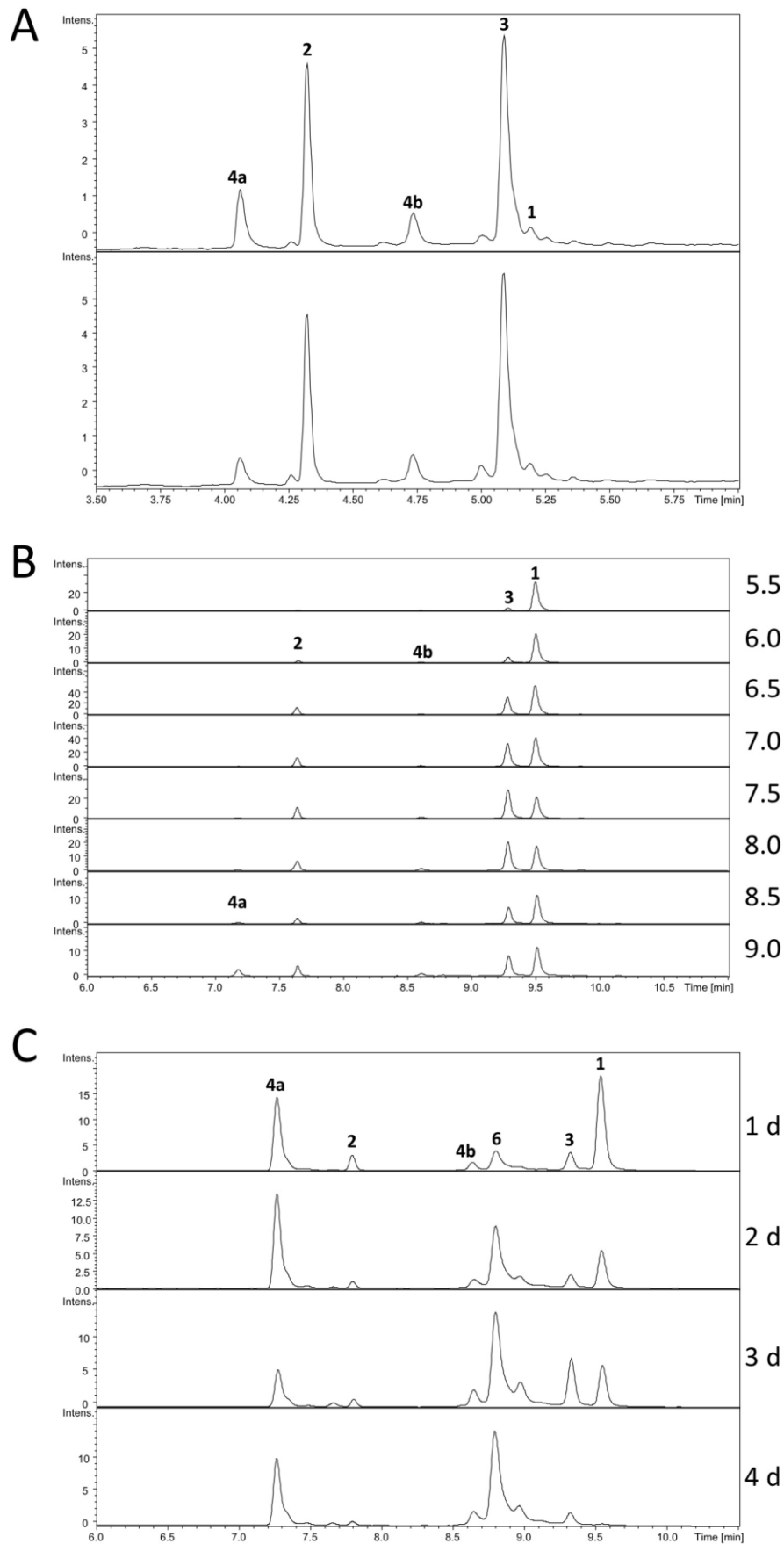


Figure S23: (A) Comparison of the effects of AlBA<sup>M131Q</sup> (top) and AlBA<sup>M131K</sup> (bottom) on albicidin. Protein and albicidin were incubated for 16 h at 37 °C prior to LC-MS analysis. (B) pH-dependency of albicidin cyclization when incubated with AlBA<sup>M131Q</sup>. (C) AlBA<sup>M131Q</sup> was incubated with albicidin at a 1:10 ratio and samples were taken every 24 h.



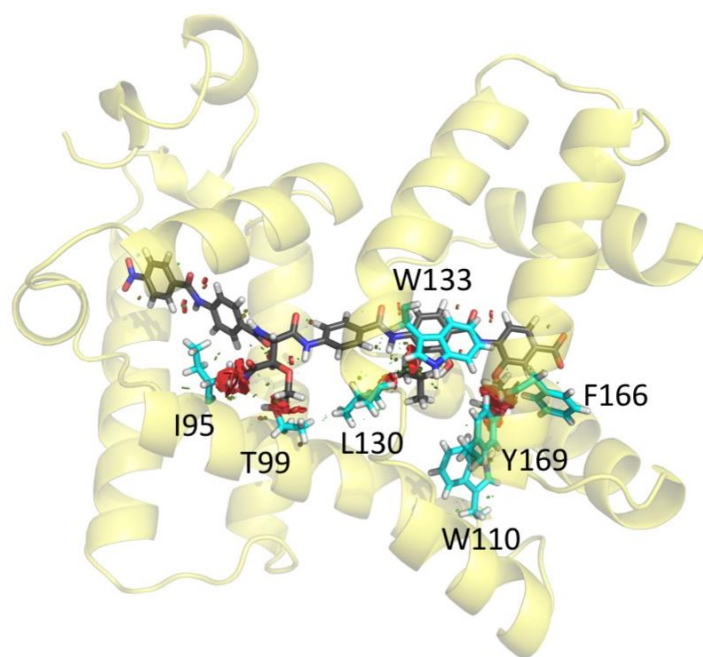
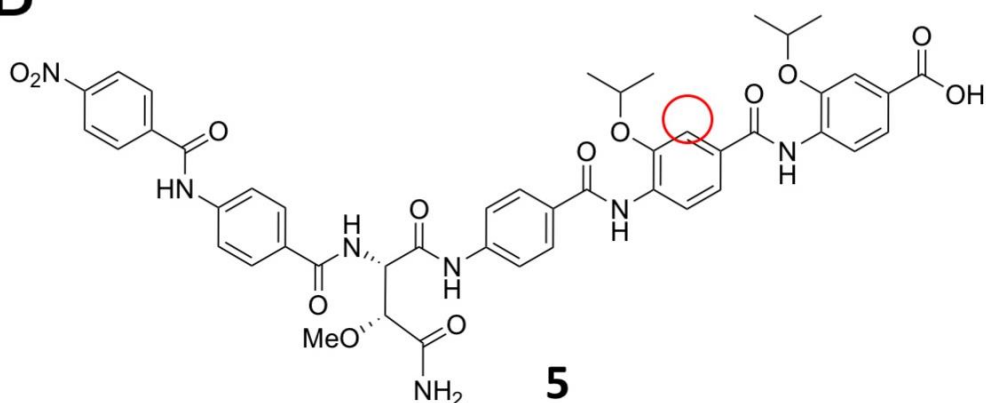
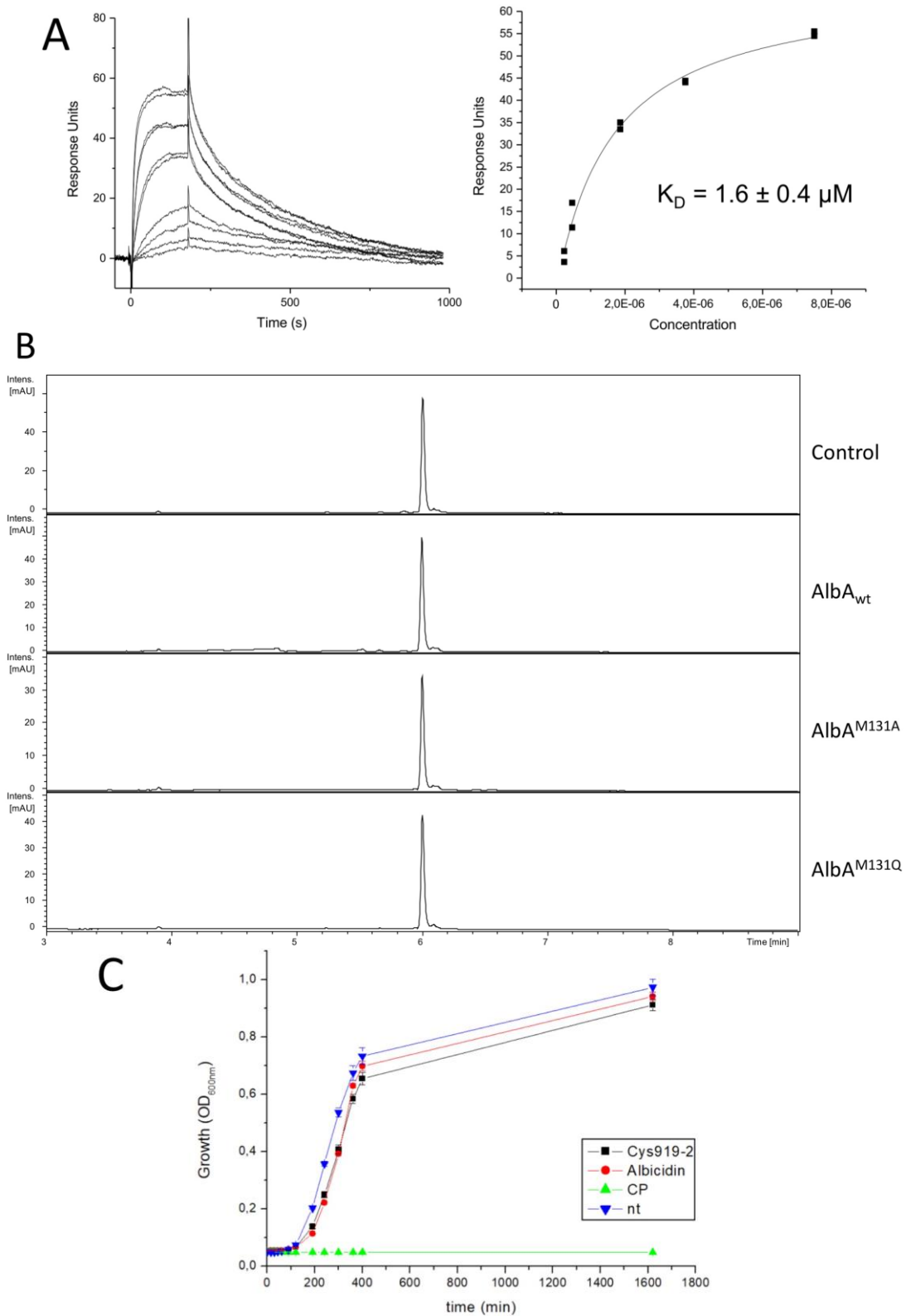
**A****B**

Figure S24: (A) Model of  $AlbA_{wt}$  in complex with cystobactamid 919-2. Cystobactamid is shown as sticks (carbon atoms grey, oxygen atoms red, nitrogen atoms blue, hydrogen atoms white) and clashes with the protein as red discs. Protein residues responsible for the clashes are shown as cyan sticks and labeled, the protein is shown as a pale yellow cartoon. We used the  $AlbA_{wt}$ -albicidin complex structure as a template to model an  $AlbA$ -cystobactamid 919-2 complex (Figure 4A). Overall, cystobactamid fits quite well into the substrate binding tunnel. However, three cystobactamid-specific modifications lead to severe clashes with the protein: The isopropoxy group of terminal PABA unit clashes severely with the side-chains of F166 and Y169 and lightly with the side-chain of W110. The second isopropoxy group does not clash as severely, but the side-chains of L130 and W133 would have to move approximately  $1 \text{ \AA}$  for a reasonable fit. The central  $\beta$ -methoxy-asparagine linker clashes severely with I95 and lightly with T99 owing to its  $\beta$ -methoxy group. Without  $\beta$ -methoxylation an orientation of the side-chain with minimal clashes can be found. This implies that for compound optimization via medicinal chemistry bulky modifications in these three positions could be used to evade  $AlbA$ -mediated resistance. Since the substrate binding tunnel runs through the hydrophobic core of the protein, mutations in  $AlbA$  to counteract these modifications are possible but likely rare - spontaneous complementary mutations would be required to maintain a functional protein. (B) Structure of the cystobactamid derivative used by Rostock et al, lacking a hydroxy group in the position indicated by a red circle when compared to cystobactamid 919-2.



Unpaired t-test

	Two-tailed P value	Statistically significant
919-2 vs. non-treated	0.0354	Yes
Albicidin vs. non-treated	0.156	No
919-2 vs. albicidin	0.1054	No

Figure S25: (A) SPR of  $AlbA_{wt}$  with cystobactamid 919-2. (B) Cystobactamid 919-2 is not converted by  $AlbA_{wt}$  or selected mutants after a four day incubation at 37 °C. (C) Cystobactamid 919-2 but not albicidin show a statistically significant effect on *K.pneumoniae* growth.

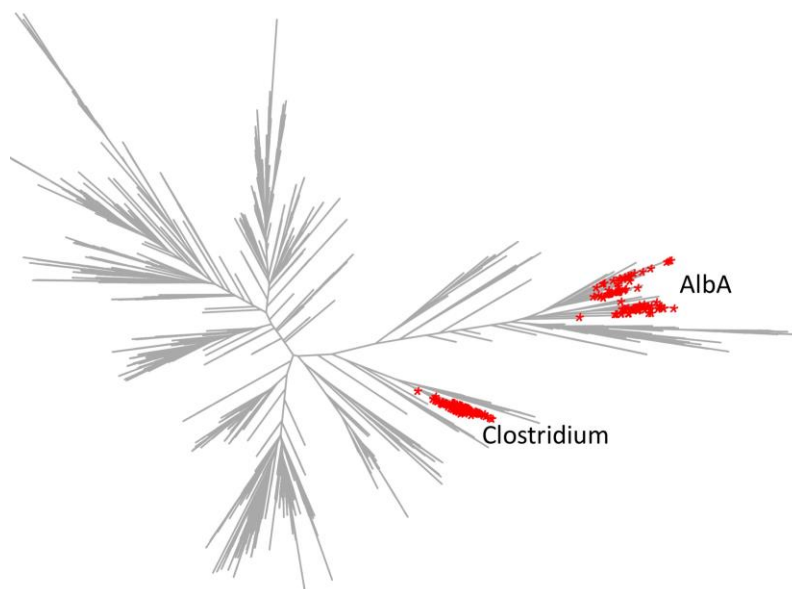


Figure S26: Distribution of the TipA-S duplication amongst all TipA sequences currently publicly available. In addition to the AlbA clade one can find a closely related “TipA-TipA” protein family in *Clostridia*.

Table S1: Data collection and refinement statistics (molecular replacement).

	AlbA <sub>wt</sub>	AlbA <sub>wt</sub> albicidin complex	AlbA T99V	AlbA M131A albicidin complex
<b>Data collection</b>				
Space group	P 43 21 2	C 2 2 21	P 43 21 2	C 2 2 21
Cell dimensions				
<i>a, b, c</i> (Å)	70.58 70.58 95.10	54.14 123.18 161.06	69.92 69.92 95.22	53.26 123.50 161.17
$\alpha, \beta, \gamma$ (°)	90.00 90.00 90.00	90.00 90.00 90.00	90.00 90.00 90.00	90.00 90.00 90.00
Resolution (Å)	47.55 - 1.90 (1.94 - 1.90)	48.92 - 1.55 (1.58 - 1.55)	49.44 - 2.60 (2.71 - 2.60)	49.01 - 2.20 (2.27 - 2.20)
<i>R</i> <sub>merge</sub>	0.050 (0.901)	0.041 (0.589)	0.113 (0.896)	0.117 (0.811)
<i>I</i> / $\sigma I$	18.3 (2.0)	16.3 (2.2)	12.7 (2.0)	7.5 (2.7)
Completeness (%)	99.8 (100.00)	99.6 (99.9)	99.7 (98.2)	99.7 (99.8)
Redundancy	7.0 (7.1)	4.6 (4.7)	7.0 (6.8)	5.1 (4.8)
<b>Refinement</b>				
Resolution (Å)	44.20 - 1.9	47.37 - 1.55	39.35 - 2.598	49.01 - 2.2
No. reflections	19504 (1906)	77989 (7724)	7698 (740)	27360 (2670)
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.1943 / 0.2065	0.1886 / 0.2073	0.2181 / 0.2576	0.2105 / 0.2399
No. atoms	1931	4146	1853	3766
Protein	1848	3571	1848	3540
Ligand/ion	-	133	-	134
Water	83	442	5	92
<i>B</i> -factors	42.22	34.60	58.99	50.77
Protein	42.20	34.14	59.00	50.86
Ligand/ion	-	32.08	-	51.41
Water	42.65	39.02	54.21	46.64
R.m.s. deviations				

Bond lengths (Å)	0.008	0.009	0.005	0.013
Bond angles (°)	1.16	1.03	0.93	1.25

Values in parentheses are for highest-resolution shell

### Phasing of AlbA<sub>wt</sub> using Phenix Autosol:

#### Substructure search

No. of sites	15
Figure of merit	0.392
Overall score	52.52 +/- 8.22

#### Density modification

R-factor	0.3006
Map skew	0.21
Corr. Of local RMS density	0.84

## Chapter 5: Discussion

This thesis deals with two aspects of NP research, namely the biosynthesis of NPs and bacterial defence strategies against them. The first half of the discussion is dedicated to the role of two hydrolases, BotH and BotAH, found in the bottromycins BGCs; whereas the second half deals with the autoregulatory resistance strategy used by *Klebsiella oxytoca* against a potent NP antibacterial compound, albicidin.

### 1 Functional and Structural Characterization of BotAH and BotH

Bottromycins have been identified as highly modified RiPPs with potent antibacterial activity<sup>1-8</sup>. Fortuitously, the identified mode of action of bottromycins differs from other known antibiotics - bottromycins inhibit bacterial protein synthesis by binding to a novel target, the A site of the prokaryotic 50S ribosome<sup>9-12</sup>. This novel target-site in conjunction with potent antibacterial activity, which extends to MRSA and VRE, makes them an extremely attractive target for antibiotic research<sup>3,13</sup>. Nevertheless, progress has been hampered due to difficulties associated with the chemical synthesis and derivatization of naturally occurring bottromycins<sup>3,4,13,14</sup>. These problems can be overcome by large-scale, bottom-up engineering approaches, which require an understanding of the underlying principles of bottromycin biosynthesis<sup>15</sup>. The identification of the gene cluster responsible for the biosynthesis of bottromycins, followed by *in vivo* and/or *in vitro* functional characterization of these genes has been a significant step forward in that direction<sup>5-8,16-19</sup>. However, the function of two genes, *botH* and *botAH*, remained unclear. Therefore, in-depth structural and functional characterization of these two proteins was carried out to facilitate a better understanding of their role in the bottromycin pathway.

BotAH is a putative metallo-dependent amidohydrolase that based on gene deletion studies was suggested to cooperate with BotCD to form the macrocycle<sup>16</sup>. However, *in vitro* studies demonstrated that the absence of BotAH did not impede macrocyclization - BotCD alone was sufficient to carry out this modification with no observable enhancement of enzymatic activity upon addition of BotAH<sup>18,19</sup>. To resolve the apparent disparity between *in vivo* and *in vitro* studies, BotAH was functionally and structurally characterized. The reported *in vitro* characterization of BotAH clearly demonstrated that it is responsible for the removal of the

follower peptide and this activity required the presence of divalent metal ions. BotAH was also shown to be highly selective for heterocyclized, macrocyclized intermediate, and therefore is likely to act after BotC and BotCD in the bottromycin pathway. Moreover, once the follower peptide has been removed by BotAH the macrocycle cannot be reopened by BotCD, placing BotAH as the gatekeeper between the early and late-stage modification steps. This observation helps to explain the previously suggested cooperative role of BotAH/CD in *in vivo* studies. The knockout of BotAH most likely leads to an accumulation of heterocyclized, macrocyclized intermediate that is acted upon by BotCD to reopen the macrocycle, thereby resulting in highly similar mass spectral network reported for both BotAH and BotCD knockouts<sup>16</sup>. Taken together the data presented here not only establishes the precise role of BotAH in bottromycin biosynthesis, but also help rationalize the disparity between the *in vitro* and *in vivo* studies.

The reported crystal structure of BotAH sheds light on its substrate specificity and mechanism of action. The structure revealed that BotAH belongs to the subtype I amidohydrolase family - the binuclear metal centre embedded in the active site is ligated to the protein through electrostatic interactions with the side chains of six amino acids (4 His, 1 Asp and 1 Lys). Analogous to an amidohydrolase from the *Escherichia coli*, dihydroorotase (DHO), binding of the peptide substrate to BotAH may be facilitated by the interaction of the carbonyl group of the scissile peptide bond with the divalent ions. This interaction has been noted to polarize the amide bond, thereby making the carbonyl carbon more electrophilic. Asp348, located in the active site of BotAH, may then activate the nucleophilic water for the hydrolysis step. This role of Asp348 as a general base is supported by the loss of activity observed for Asp348Asn mutation. The nucleophilic attack results in the formation of an oxyanion that is stabilized by as of yet unknown residues in the BotAH ligand binding site. It is worth mentioning that somewhat different mechanisms have also been reported for other amidohydrolases, and therefore in-depth mechanistic studies are warranted to fully understand the mechanism employed by BotAH for the cleavage of follower peptide. In the BotAH structure an extended, wide substrate-binding site is observed that is predicted to be suitable for binding both, the heterocycle and macrocycle, explaining the observed *in vitro* preference for the heterocyclized and macrocyclized substrate over only heterocyclized substrate. Moreover, this preference also hints towards a mode of binding that may proceed via the interaction with the macrocycle acting as an anchor to help position the peptide bond present between the heterocycle and the follower peptide for hydrolysis. Future studies focusing on the relative stabilization effect of the hydrolysis products on BotAH may better explain the mode of substrate binding. Interestingly,

the substrate binding site comprises of highly conserved residues, indicating that BotAH may have evolved to specifically accept large cyclic peptides. In general, RiPP enzymes exhibit relaxed substrate specificity, but in all the bottromycin BGCs identified to date (except for bottromycin D) the core peptide sequence is fully conserved. To understand if this is a consequence of stringent BotAH gatekeeping, the effect of single point mutations in the core peptide of BotA on BotAH activity were tested. Data suggests that BotAH is promiscuous in the context of single amino acid changes, but has high fidelity for macrocyclized and heterocyclized substrates. This is not surprising given the highly conserved nature of the BotAH substrate binding site. In the absence of a complex structure of BotAH with substrate it is difficult to rationalize the role of each residue in substrate binding and recognition - this work will certainly provide much needed insight into BotAH ligand promiscuity, and possibly extend it further by protein reengineering.

With the function of BotAH established, focus was shifted towards the last remaining uncharacterized hydrolase present in the bottromycins BGCs. The high-resolution structure of BotH revealed it to be a member of the  $\alpha/\beta$ -hydrolase (ABH) family. However, comparison of BotH with its closest structural homolog (PDB: 2XUA) led to an interesting observation: the active site residues are either mutated or missing in BotH, and this non-functional active site is found in a large hydrophobic cavity that appeared to be suitable for binding to the BotAH product. Concurrently, work on the enzymatic production of the substrate for BotCYP also revealed that heterocyclized and macrocyclized-BotAH product undergoes slow spontaneous epimerization (**3a** and **3b**; **3a** << **3b**; page no. 72 & 94). The incubation of this substrate with BotH resulted in a significant change in the relative abundance of the epimers (**3a** > **3b**). Marfey's reagent and analysis using high-resolution LC-MS confirmed that **3a** is D-Asp; this epimer was subsequently shown *in vivo* to be the preferred substrate of BotCYP to carry out oxidative decarboxylation of the heterocycle (Adam, S. and Franz, L.; unpublished work). Therefore, these observations confirm the role of BotH in bottromycins biosynthesis as the enzyme responsible for the epimerization of L- to D-Asp.

Among RiPPs, epimerization is best studied for lanthipeptides and proteusins<sup>20,21</sup>. In lanthipeptides it proceeds via a two-step process. The first step involves the formation of dehydroalanine (Dha) or dehydrobutyrine (Dhb) - from either serine or threonine - as an intermediate by a dehydratase<sup>22-26</sup>. Usually the dehydrated intermediate undergoes intramolecular Michael-type addition of Cys thiol by a cyclase that results in the inversion of

the stereocenter<sup>20,21,25,27</sup>. However, a subclass of lantipeptides contains D-amino acids that are formed by hydrogenation of Dha and/or Dhb<sup>28-30</sup>. In contrast to lanthipeptides, proteusins make use of a radical-SAM enzymes for the conversion of L-Val, L-Ile and L-Ala into their D-configured counterparts<sup>31,32</sup>. Since BotH shares no homology to any of these proteins, we also solved the structure of BotH in complex with **3a**, which sheds light on the mechanism behind **3** epimerization. In the absence of catalytic residues around C $\alpha$  proton of Asp7 that may be involved in proton abstraction or addition, a ligand-assisted deprotonation-protonation mechanism has been proposed. The binding of **3b** traps a water molecule within hydrogen bond distance of the thiazoline nitrogen. In the complex structure of BotH and **3a**, the carboxyl of Asp7 is positioned such that it may serve as a base to abstract the C $\alpha$  proton from itself, which triggers imine-enamine tautomerization with proton transfer from the trapped water molecule to the thiazoline nitrogen. The resulting hydroxide ion, which is held in position via hydrogen bond interaction with the thiazoline's carboxy group and protonated nitrogen, can then extract hydrogen back from the thiazoline nitrogen. This leads to reprotonation of the enamine by the side chain of Asp7 and  $\alpha$ -carbon stereoinversion. The possible involvement of the Asp7-carboxyl group in epimerization is based on the observed inability of BotH to epimerize substrates containing Ala or Asn at the Asp position. This may also explain why *O*-methylation of this Asp is the last step in bottromycin pathway. The proposed mechanism of epimerization highlights the importance of the correct orientation of the thiazoline carboxy group as a proton shuttle during epimerization. This may also explain the slow rate of epimerization observed for **3b** in the absence of BotH. In light of this, four BotH residues (Met43, Phe109, Met174, Ile224) seem to be critical as they are involved in the formation of a small hydrophobic groove in which the thiazoline resides. The mutation of these with less bulky residues may lead to a more flexible thiazoline, potentially unable to trap the water molecule. This should in turn lead to a loss of BotH epimerization activity, further supporting the proposed epimerization mechanism.

The BotH structure in complex with **3a** represents the first reported structure of bottromycin biosynthetic enzymes with ligands, and has proved invaluable in providing a better understanding of enzyme-substrate flexibility (page no. 97 & 105). BotH can handle mutations at most of the positions in and around the macrocycle. The Phe residue at position 6 of the core peptide is bound in a rather hydrophobic pocket, held in place by hydrophobic interactions with the surrounding residues. The pocket is big enough to accommodate larger hydrophobic residues at this position, which explains the successful conversion of Phe6Tyr and Phe6Trp substrates. The inability of BotH to process Phe6Ala hints towards the importance of bulky hydrophobic



residues at this position of the core peptide in substrate binding. In light of this, it would be interesting to see if this mutation has any impact on the overall binding affinity of the substrate to BotH. The valines at position 3, 4 and 5 of the core peptide are pointing away from the binding pocket towards the bulk solvent, and therefore, BotH exhibits considerable flexibility at these positions. However, care must be taken when introducing multiple mutations simultaneously, so not to disturb the intramolecular hydrophobic interactions formed by Val residues. The flexibility at core peptide position 2 is limited due to the positioning of the Pro residue in a small pocket. Hence, Pro can only be substituted with small amino acids e.g. Ala and Gly. Two remaining positions (1 and 8) were not tested, as mutations at these positions are determinantal to the activity of the enzymes preceding BotH in the pathway.

Given **3a/b** are close homologs of bottromycin A2, the binding affinity of four different bottromycins to BotH was also tested. BotH was able to bind bottromycins with  $K_D$  in the high nM to low  $\mu$ M range. The reported high-resolution crystal structures of bottromycin-BotH complexes help explain the mode of binding. Bottromycins bind in a manner similar to **3**, with minimal change in the overall structure of the protein. The bound ligand is stabilized by an extensive network of hydrogen and hydrophobic interactions, explaining the observed high-affinity of bottromycins for BotH. Unsurprisingly, bottromycin A2 was also observed to act as an orthosteric inhibitor of epimerization, suggesting that BotH has secondary non-enzymatic roles in bottromycin pathway i.e. self-immunity of the producer strain and feedback inhibition. The overexpression of BotT (putative transporter) in a bottromycin sensitive strain resulted in a slight increase in resistance. This resistance was further amplified when BotT was overexpressed along with BotH, implying that BotH exert its secondary role in conferring self-resistance via interaction with the transporter protein (Santos-Aberturas, J. and Truman, AW.; unpublished). However, at the moment the nature of this interaction is still not clear, but *in vivo* studies are currently ongoing to shed light on it. Although our understanding of the mechanisms involved in the biosynthesis of RiPPs is continuously increasing, the way they are exported out of the producing cells using export machineries is poorly documented. Currently, McjD from microcin J25 (MccJ25) is the only functionally and structurally characterized RiPP transporter. This transporter exclusively recognizes and transports MccJ25, and this stringency is presumably dictated by the specific interactions between key residues present on MccJ25 with McjD. Based on these findings, one could speculate that bottromycins, upon binding to BotH, adapt a conformation that presents MccJ25-like key residues for optimal interaction with BotT - compared to the published NMR structure, bottromycin A2 undergoes significant

conformational change as a result of binding to BotH (page no. 105). The epimerization of Asp7 has been shown to be important for the downstream enzyme, BotCYP, as it preferentially consumes D-Asp substrate (Adam, S. and Franz, L.; unpublished work). Thus, feedback inhibition of BotH by mature bottromycins may offer a direct and faster means to regulate NP production than altered gene expression. These findings hint towards a possible subcellular colocalization of these two enzymes, and therefore, it would be interesting to study if BotH and BotCYP interact with each other or not. Taken together, the secondary functions of BotH may have a profound impact on the production yield of bottromycins in heterologous hosts, which is one of the bottlenecks in the development of these compound for pharmaceutical use. Overexpression of BotT has already been shown to improve production titer 20 times in a *S. coelicolor* host, and this increase was linked to the role of BotT in host self-resistance<sup>5</sup>. Thus, it is likely that the observed enhancement in resistance upon co-overexpression of BotT and BotH could potentially lead to a further increase of bottromycins production yields.

In conclusion, work presented here has helped to establish the primary function of two proteins, BotAH and BotH, in the bottromycin pathway, and shed light on their mechanism of action. In addition, the work detailing the substrate flexibility of these enzymes will be important in future efforts to derivatize bottromycins. However, several questions pertaining to the secondary role(s) of BotH need to be addressed, namely the nature of the BotH and BotT interaction, the role of bottromycins residues in export, and the effect of co-overexpression of BotH and BotT on heterologous yields.

## **2 Functional and Structural Characterization of AlbA**

Albicidin is a phytotoxic molecule produced by the Gram-negative plant pathogen *Xanthomonas albilineans* that is responsible for leaf scald disease in sugarcane plants<sup>33,34</sup>. Additionally, albicidin is bactericidal in the nanomolar range against Gram-positive and Gram-negative bacteria<sup>35-37</sup>. The mechanism of action of albicidin has been identified: it inhibits DNA replication by targeting bacterial DNA gyrase (topoisomerase II), an enzyme which is essential for supercoiling of DNA during replication<sup>35,38,39</sup>. Despite potent antibacterial activity and no cytotoxicity to mammalian cells, the development of albicidin as a new family of antibacterial drugs suffers from one major drawback - the presence of resistance mechanisms against it, including MerR-like protein, AlbA, in *Klebsiella oxytoca*<sup>40-44</sup>. Previously, this protein had been suggested to confer resistance by rapidly removing free albicidin from cells by binding the antibiotic with strong affinity ( $K_D$  64 nM)<sup>45</sup>. Additionally, an alanine scan identified several

key residues involved in albicidin binding, and it was found that ligand-binding induced major conformational changes<sup>46</sup>. However, the understanding of how AlbA mediated neutralization of albicidin suffered from the lack of high-resolution structures to fully understand how AlbA exert its function.

The crystal structure of AlbA revealed that it adopts an all- $\alpha$ -helical fold. This does not agree with previously reported circular dichroism (CD) spectrometry studies, as Zhang and co-workers estimated AlbA to be composed of at least 21%  $\beta$ -sheets that are not observed in the reported crystal structure<sup>46</sup>. This is most likely due to the presence of misfolded AlbA and/or instability of AlbA in the buffer used for CD spectrometry leading to the formation of  $\beta$ -sheet-rich fibrillar aggregates that can interfere with protein secondary structure estimation<sup>47-49</sup>. The structure also revealed that AlbA is composed of a tandem architecture, AlbA<sub>1-114</sub> and AlbA<sub>115-221</sub>, with each domain of the protein being structurally similar to TipA<sub>S</sub>. This fusion of two TipA<sub>S</sub>-like proteins results in the formation of a long putative ligand binding tunnel, running across the entire length of AlbA. The structure of AlbA in complex with albicidin confirmed that albicidin does indeed occupy the ligand binding tunnel formed by the two TipA<sub>S</sub> domains, stabilized via several residues previously identified to be important for albicidin binding. Unlike previously reported significant conformational changes upon ligand binding by CD spectroscopy, the overall structure of AlbA is virtually unchanged after binding albicidin<sup>46</sup>. Despite the limitations of the CD spectroscopy studies in elucidating protein structure it can be argued that the AlbA structures reported here do not shed light on the protein dynamics, as crystal structures only report the ensemble average<sup>50</sup>. It is well known that proteins typically populate a number of structural states, displaying a range of conformations that are important for its biological function<sup>51</sup>. Therefore, it is likely that AlbA, just like TipA<sub>S</sub>, displays inherent and ligand induced protein dynamics<sup>52</sup>. Work by Süssmuth and co-workers, have extend the knowledge of AlbA-albicidin binding; it was reported that the NMR spectra of AlbA shows extensive line-broadening in the absence of albicidin, presumably reflecting the internal dynamics of the protein<sup>53</sup>. In addition, they also evaluated the binding contributions of each TipA<sub>S</sub> domain of AlbA by using albicidin and its truncated derivatives. AlbA<sub>115-221</sub> was found to be the major contributor in the stabilization of albicidin in the binding tunnel. Taken together, these studies led the authors to propose a possible binding mechanism that involved the fixation of albicidin to the AlbA<sub>115-221</sub> domain via a salt-bridge interaction which then helps to orient the ligand and allows subsequent association to the other domain, AlbA<sub>1-114</sub><sup>53</sup>. Such cooperative binding has also been observed for another MerR-like protein, BmrB<sup>54</sup>. The characterization of

protein dynamics continues to be a major challenge in the field of structural biology, but two non-traditional methods which can help understand the ligand-induced protein dynamics at high resolution, and are gaining popularity, include time-resolved crystallography at X-ray free-electron lasers (XFEL) and cryo-electron microscopy (Cryo-EM)<sup>55,56</sup>. The combination of these two techniques, along with molecular dynamics (MD) simulations, may provide a more complete picture of events associated with AlbA interaction with albicidin<sup>57</sup>.

Unexpectedly, a close inspection of the AlbA-albicidin complex structure also revealed that the bound albicidin had been modified. Subsequent, extensive *in vitro* characterization confirmed that the modification is not a crystallization artifact or a result of radiation damage incurred during X-ray data collection, but a result of AlbA interaction with albicidin. This was overlooked by Süßmuth and co-workers, who went on to report an unmodified albicidin in their AlbA-albicidin complex structure<sup>53</sup>. To help better understand the underlying mechanism of the AlbA-induced modification of albicidin, in-depth mutational and structural studies are also reported. These studies resulted in the identification of three key residues: residues Thr99 and Ile95 form a stable bottleneck that is responsible for forcing the nitrile group of albicidin into a position that promotes the formation of a five-membered ring, whereas Met131 protects the newly formed ring from bulk solvent. This modification contributes to drug resistance, as modified albicidin was found to be less active against a panel of bacteria, especially *Escherichia coli* for which a decrease of activity by 300 x was observed. Currently, it is not clear if the decrease in activity is a result of poor interaction with the bacterial target, DNA gyrase and/or a consequence of reduced uptake by the outer-membrane channel, Tsx<sup>36,37,44</sup>. These findings helped redefine our understanding of the neutralization of albicidin by AlbA - the protein does not only sequester the antibiotic but also modifies it into a lot less potent compound. Given proteins have a half-life in cells, this mechanism of albicidin neutralization may also help reduce the gene expression burden in the host cells, as a single protein can act on numerous albicidin molecules. This mechanism is unlike the TipA system, where thiopeptide-like antibiotics bind irreversibly to the protein via a covalent interaction<sup>58</sup>.

The structural homology of AlbA with TipA<sub>S</sub> also led to the re-evaluation of the *alba* gene. Akin to the TipA system, AlbA was also found to exist in two forms: a truncated version, AlbA<sub>S</sub>, that contains the ligand binding domain formed by two TipA<sub>S</sub>-like domains; and a full-length protein, AlbA<sub>L</sub>, composed of AlbA<sub>S</sub> connected to a putative DNA binding domain via a coiled-coiled region - an observation also noted by Süßmuth and co-workers<sup>53</sup>. Subsequently,

*in vivo* studies confirmed that the exposure of an *alba*-positive bacterium to albicidin leads to an upregulation of both AlbA<sub>L</sub> and AlbA<sub>S</sub>; thus, AlbA can be considered an autoregulated antibiotic resistance system. However, at the moment it is still not clear how binding of albicidin to AlbA<sub>L</sub> leads to upregulation of the *alba* gene. Given AlbA's similarity to the TipA system, it can be assumed to act via a similar mechanism<sup>52,59</sup>. The binding of ligand to AlbA<sub>L</sub> may promote its dimerization, which stabilizes the AlbA<sub>L</sub> - RNA polymerase - promoter complex, thereby inducing transcription. In the TipA system, ligand binding has been reported to induce the expression of additional genes, possibly other MDR genes<sup>60,61</sup>. Whether this is the case for the AlbA system as well is currently not known, although the susceptibility of *Escherichia coli* BL21 cells overexpressing AlbA<sub>S</sub> hint towards the involvement of additional factors in albicidin resistant strains containing AlbA system.

Given the ligand promiscuity of the TipA system, AlbA binding to an albicidin related antibiotic was also tested. Cystobactamid, unlike albicidin, was not modified by AlbA, probably due to the presence of a central iso-asparagine moiety that prevents cyclization<sup>62-65</sup>. Moreover, in *in vivo* studies using cystobactamid resistant *Klebsiella pneumoniae*, which harbors *alba* gene, the exposure of cystobactamid did not elicit an upregulation of the AlbA system. This may be a consequence of its significantly reduced affinity to AlbA, compared to albicidin, or a mode of interaction that prevents AlbA<sub>L</sub> mediated upregulation of the *alba* gene. Since *Klebsiella pneumoniae* is fully resistant to cystobactamid it is also likely that other mechanism(s) of resistance are employed by the bacteria, which require further investigation. The comparison of cystobactamid resistant strains before and after exposure to cystobactamid by RNA-seq may help identify these resistance mechanism(s)<sup>66</sup>. In addition, it would also be interesting to check if cystobactamid resistant strains that contain *alba* gene, harbor mutations in the putative dimerization domain of TipA<sub>L</sub>. As explained earlier, dimerization of TipA<sub>L</sub>-like transcriptional regulators is a prerequisite for gene upregulation, and mutations in the dimerization region of AlbA-like proteins have been reported to have either an enhanced activation or an enhanced repression phenotype<sup>67</sup>. Therefore, mutations in the AlbA<sub>L</sub> dimerization domain may compensate for the reduced affinity of cystobactamid for AlbA. The different effect of compounds on transcription of the *alba* gene point to one thing: unique structural features of cystobactamid represent natural chemical modifications of a NP against AlbA-mediated neutralization of antibiotics. Since AlbA-like drug resistance systems have also been identified in a number of pathogenic bacteria, the derivatization of albicidin must take these structural features into account to overcome such resistance systems.

In summary, work presented here contributed to a general understanding of resistance strategies of bacteria against antibiotics. The in-depth structural and biochemical characterization of the AlbA resistance system will be useful for future antibiotic-development efforts. A step in that direction may be the optimization of albicidin to incorporate cystobactamid-like structural features. Besides antibiotic development, a better understanding AlbA system may also have significant potential in biotechnology: (1) as biosensors to help identify new potent albicidin-like compounds, and (2) as biocontrol against economically important leaf scald disease in sugarcane via production of transgenic plants<sup>68,69</sup>.

### **3 Concluding Remarks**

Antibiotic resistance has become a global health threat, especially due to the emergence, spread, and persistence of multidrug-resistance bacteria. To tackle this threat, continued and improved discovery of new NP antibiotics, along with the detailed understanding of their biosynthesis and mechanism of action are of great urgency. Microbes have evolved to defend their ecological niche from other microorganisms, often using NPs with antibacterial properties as chemical weapons. This poses an inherent problem in the use of NP as antibiotics i.e. the prevalence of pre-existing resistance strategies in microbes, which can be disseminated if the compound finds a broad use. Therefore, the discovery of a new NP with antibacterial property must be followed by the identification and understanding of pre-existing resistance strategies against it, if any, to allow for the rational development of the compound to circumvent drug resistance.

## 4 References

- 1 Nakamura, S., Omura, S., Nishimura, T., Tanaka, N. & Umezawa, H. Derivatives of bottromycin A2 and their biological activity. *J Antibiot (Tokyo)* **20**, 162-166 (1967).
- 2 Nakamura, S., Yajima, T., Lin, Y. & Umezawa, H. Isolation and characterization of bottromycins A2, B2, C2. *J Antibiot (Tokyo)* **20**, 1-5 (1967).
- 3 Kobayashi, Y. *et al.* Bottromycin derivatives: efficient chemical modifications of the ester moiety and evaluation of anti-MRSA and anti-VRE activities. *Bioorg Med Chem Lett* **20**, 6116-6120, doi:10.1016/j.bmcl.2010.08.037 (2010).
- 4 Yamada, T. *et al.* Synthesis and Evaluation of Antibacterial Activity of Bottromycins. *J Org Chem* **83**, 7135-7149, doi:10.1021/acs.joc.8b00045 (2018).
- 5 Huo, L., Rachid, S., Stadler, M., Wenzel, S. C. & Muller, R. Synthetic biotechnology to study and engineer ribosomal bottromycin biosynthesis. *Chem Biol* **19**, 1278-1287, doi:10.1016/j.chembiol.2012.08.013 (2012).
- 6 Crone, W. J. K., Leeper, F. J. & Truman, A. W. Identification and characterisation of the gene cluster for the anti-MRSA antibiotic bottromycin: expanding the biosynthetic diversity of ribosomal peptides. *Chemical Science* **3**, 3516-3521, doi:10.1039/C2SC21190D (2012).
- 7 Hou, Y. *et al.* Structure and biosynthesis of the antibiotic bottromycin D. *Org Lett* **14**, 5050-5053, doi:10.1021/ol3022758 (2012).
- 8 Gomez-Escribano, J. P., Song, L., Bibb, M. J. & Challis, G. L. Posttranslational  $\beta$ -methylation and macrolactamidation in the biosynthesis of the bottromycin complex of ribosomal peptide antibiotics. *Chemical Science* **3**, 3522-3525, doi:10.1039/C2SC21183A (2012).
- 9 Kinoshita, T. & Tanaka, N. On the site of action of bottromycin A2. *J Antibiot (Tokyo)* **23**, 311-312, doi:10.7164/antibiotics.23.311 (1970).
- 10 Otaka, T. & Kaji, A. Mode of action of bottromycin A2. Release of aminoacyl- or peptidyl-tRNA from ribosomes. *J Biol Chem* **251**, 2299-2306 (1976).
- 11 Otaka, T. & Kaji, A. Mode of action of bottromycin A2: effect on peptide bond formation. *FEBS Lett* **123**, 173-176, doi:10.1016/0014-5793(81)80280-3 (1981).
- 12 Otaka, T. & Kaji, A. Mode of action of bottromycin A2: effect of bottromycin A2 on polysomes. *FEBS Lett* **153**, 53-59, doi:10.1016/0014-5793(83)80118-5 (1983).
- 13 Shimamura, H. *et al.* Structure determination and total synthesis of bottromycin A2: a potent antibiotic against MRSA and VRE. *Angew Chem Int Ed Engl* **48**, 914-917, doi:10.1002/anie.200804138 (2009).
- 14 Ackermann, S., Lerchen, H. G., Habich, D., Ullrich, A. & Kazmaier, U. Synthetic studies towards bottromycin. *Beilstein J Org Chem* **8**, 1652-1656, doi:10.3762/bjoc.8.189 (2012).
- 15 Fischbach, M. & Voigt, C. A. Prokaryotic gene clusters: a rich toolbox for synthetic biology. *Biotechnol J* **5**, 1277-1296, doi:10.1002/biot.201000181 (2010).
- 16 Crone, W. J. *et al.* Dissecting Bottromycin Biosynthesis Using Comparative Untargeted Metabolomics. *Angew Chem Int Ed Engl* **55**, 9639-9643, doi:10.1002/anie.201604304 (2016).
- 17 Mann, G. *et al.* Structure and Substrate Recognition of the Bottromycin Maturation Enzyme BotP. *Chembiochem* **17**, 2286-2292, doi:10.1002/cbic.201600406 (2016).
- 18 Schwalen, C. J. *et al.* In Vitro Biosynthetic Studies of Bottromycin Expand the Enzymatic Capabilities of the YcaO Superfamily. *J Am Chem Soc* **139**, 18154-18157, doi:10.1021/jacs.7b09899 (2017).

- 19 Franz, L., Adam, S., Santos-Aberturas, J., Truman, A. W. & Koehnke, J. Macroamidine Formation in Botromycins Is Catalyzed by a Divergent YcaO Enzyme. *J Am Chem Soc* **139**, 18158-18161, doi:10.1021/jacs.7b09898 (2017).
- 20 Ogasawara, Y. & Dairi, T. Peptide Epimerization Machineries Found in Microorganisms. *Front Microbiol* **9**, 156, doi:10.3389/fmicb.2018.00156 (2018).
- 21 Repka, L. M., Chekan, J. R., Nair, S. K. & van der Donk, W. A. Mechanistic Understanding of Lanthipeptide Biosynthetic Enzymes. *Chem Rev* **117**, 5457-5520, doi:10.1021/acs.chemrev.6b00591 (2017).
- 22 Xie, L. *et al.* Lacticin 481: in vitro reconstitution of lantibiotic synthetase activity. *Science* **303**, 679-681, doi:10.1126/science.1092600 (2004).
- 23 Chatterjee, C. *et al.* Lacticin 481 synthetase phosphorylates its substrate during lantibiotic production. *J Am Chem Soc* **127**, 15332-15333, doi:10.1021/ja0543043 (2005).
- 24 Ortega, M. A. *et al.* Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB. *Nature* **517**, 509-512, doi:10.1038/nature13888 (2015).
- 25 Tang, W., Jimenez-Oses, G., Houk, K. N. & van der Donk, W. A. Substrate control in stereoselective lanthionine biosynthesis. *Nat Chem* **7**, 57-64, doi:10.1038/nchem.2113 (2015).
- 26 Garg, N., Salazar-Ocampo, L. M. & van der Donk, W. A. In vitro activity of the nisin dehydratase NisB. *Proc Natl Acad Sci U S A* **110**, 7258-7263, doi:10.1073/pnas.1222488110 (2013).
- 27 Cheng, F., Takala, T. M. & Saris, P. E. Nisin biosynthesis in vitro. *J Mol Microbiol Biotechnol* **13**, 248-254, doi:10.1159/000104754 (2007).
- 28 Yang, X. & van der Donk, W. A. Post-translational Introduction of D-Alanine into Ribosomally Synthesized Peptides by the Dehydroalanine Reductase NpnJ. *J Am Chem Soc* **137**, 12426-12429, doi:10.1021/jacs.5b05207 (2015).
- 29 Mu, D., Montalban-Lopez, M., Deng, J. & Kuipers, O. P. Lantibiotic Reductase LtnJ Substrate Selectivity Assessed with a Collection of Nisin Derivatives as Substrates. *Appl Environ Microbiol* **81**, 3679-3687, doi:10.1128/AEM.00475-15 (2015).
- 30 Cotter, P. D. *et al.* Posttranslational conversion of L-serines to D-alanines is vital for optimal production and activity of the lantibiotic lacticin 3147. *Proc Natl Acad Sci U S A* **102**, 18584-18589, doi:10.1073/pnas.0509371102 (2005).
- 31 Parent, A. *et al.* Mechanistic Investigations of PoyD, a Radical S-Adenosyl-l-methionine Enzyme Catalyzing Iterative and Directional Epimerizations in Polytheonamide A Biosynthesis. *J Am Chem Soc* **140**, 2469-2477, doi:10.1021/jacs.7b08402 (2018).
- 32 Fuchs, S. W. *et al.* A Lanthipeptide-like N-Terminal Leader Region Guides Peptide Epimerization by Radical SAM Epimerases: Implications for RiPP Evolution. *Angew Chem Int Ed Engl* **55**, 12330-12333, doi:10.1002/anie.201602863 (2016).
- 33 Robert G. Birch, S. S. P. Correlation between albicidin production and chlorosis induction by *Xanthomonas albilineans*, the sugarcane leaf scald pathogen. *Physiological and Molecular Plant Pathology*, **30**, 199-206 (1987).
- 34 Birch RG, P. S. The relation of blocked chloroplast differentiation to sugarcane leaf scald disease. *Phytopathology* **73**, 1368-1374 (1983).
- 35 Gratz, S. *et al.* Synthesis and Antimicrobial Activity of Albicidin Derivatives with Variations of the Central Cyanoalanine Building Block. *ChemMedChem* **11**, 1499-1502, doi:10.1002/cmde.201600163 (2016).
- 36 Birch, R. G. & Patil, S. S. Preliminary characterization of an antibiotic produced by *Xanthomonas albilineans* which inhibits DNA synthesis in *Escherichia coli*. *J Gen Microbiol* **131**, 1069-1075, doi:10.1099/00221287-131-5-1069 (1985).



- 37 Hashimi, S. M. Albicidin, a potent DNA gyrase inhibitor with clinical potential. *J Antibiot (Tokyo)* **72**, 785-792, doi:10.1038/s41429-019-0228-2 (2019).
- 38 Robert G. Birch, S. S. P. Evidence that an albicidin-like phytotoxin induces chlorosis in sugarcane leaf scald disease by blocking plastid DNA replication. *Physiological and Molecular Plant Pathology*, **2**, 207-214 (1987).
- 39 Hashimi, S. M., Wall, M. K., Smith, A. B., Maxwell, A. & Birch, R. G. The phytotoxin albicidin is a novel inhibitor of DNA gyrase. *Antimicrob Agents Chemother* **51**, 181-187, doi:10.1128/AAC.00918-06 (2007).
- 40 Basnayake, W. V. & Birch, R. G. A gene from *Alcaligenes denitrificans* that confers albicidin resistance by reversible antibiotic binding. *Microbiology* **141** ( Pt 3), 551-560, doi:10.1099/13500872-141-3-551 (1995).
- 41 Zhang, L. & Birch, R. G. The gene for albicidin detoxification from *Pantoea dispersa* encodes an esterase and attenuates pathogenicity of *Xanthomonas albilineans* to sugarcane. *Proc Natl Acad Sci U S A* **94**, 9984-9989, doi:10.1073/pnas.94.18.9984 (1997).
- 42 Walker, M. J., Birch, R. G. & Pemberton, J. M. Cloning and characterization of an albicidin resistance gene from *Klebsiella oxytoca*. *Mol Microbiol* **2**, 443-454, doi:10.1111/j.1365-2958.1988.tb00050.x (1988).
- 43 Vieweg, L. *et al.* The Albicidin Resistance Factor AlbD Is a Serine Endopeptidase That Hydrolyzes Unusual Oligoaromatic-Type Peptides. *J Am Chem Soc* **137**, 7608-7611, doi:10.1021/jacs.5b04099 (2015).
- 44 Birch, R. G., Pemberton, J. M. & Basnayake, W. V. Stable albicidin resistance in *Escherichia coli* involves an altered outer-membrane nucleoside uptake system. *J Gen Microbiol* **136**, 51-58, doi:10.1099/00221287-136-1-51 (1990).
- 45 Zhang, L., Xu, J. & Birch, R. G. High affinity binding of albicidin phytotoxins by the AlbA protein from *Klebsiella oxytoca*. *Microbiology* **144** ( Pt 2), 555-559, doi:10.1099/00221287-144-2-555 (1998).
- 46 Weng, L. X. *et al.* Molecular and conformational basis of a specific and high-affinity interaction between AlbA and albicidin phytotoxin. *Appl Environ Microbiol* **71**, 1445-1452, doi:10.1128/AEM.71.3.1445-1452.2005 (2005).
- 47 Kusumoto, Y., Lomakin, A., Teplow, D. B. & Benedek, G. B. Temperature dependence of amyloid beta-protein fibrillization. *Proc Natl Acad Sci U S A* **95**, 12277-12282, doi:10.1073/pnas.95.21.12277 (1998).
- 48 Gursky, O. & Aleshkov, S. Temperature-dependent beta-sheet formation in beta-amyloid A $\beta$ (1-40) peptide in water: uncoupling beta-structure folding from aggregation. *Biochim Biophys Acta* **1476**, 93-102, doi:10.1016/s0167-4838(99)00228-9 (2000).
- 49 Arutyunyan, A. M., Rafikova, E. R., Drachev, V. A. & Dobrov, E. N. Appearance of “ $\beta$ -Like” Circular Dichroism Spectra on Protein Aggregation That Is not Accompanied by Transition to  $\beta$ -Structure. *Biochemistry (Moscow)* **66**, 1378-1380, doi:10.1023/a:1013337930104 (2001).
- 50 Chruszcz, M., Wlodawer, A. & Minor, W. Determination of protein structures--a series of fortunate events. *Biophys J* **95**, 1-9, doi:10.1529/biophysj.108.131789 (2008).
- 51 Erik G Marklund, J. L. B. Weighing-up protein dynamics: the combination of native mass spectrometry and molecular dynamics simulations. *Current Opinion in Structural Biology* **54**, 50-58 (2019).
- 52 Habazettl, J. *et al.* Structural basis and dynamics of multidrug recognition in a minimal bacterial multidrug resistance system. *Proc Natl Acad Sci U S A* **111**, E5498-5507, doi:10.1073/pnas.1412070111 (2014).
- 53 Rostock, L. *et al.* Molecular insights into antibiotic resistance - how a binding protein traps albicidin. *Nat Commun* **9**, 3095, doi:10.1038/s41467-018-05551-4 (2018).

- 54 Bachas, S., Kohrs, B. & Wade, H. Unconventional Coupling between Ligand Recognition and Allosteric Control in the Multidrug Resistance Gene Regulator, BmrR. *ChemMedChem* **12**, 426-430, doi:10.1002/cmdc.201700017 (2017).
- 55 Pandey, S. *et al.* Time-resolved serial femtosecond crystallography at the European XFEL. *Nat Methods* **17**, 73-78, doi:10.1038/s41592-019-0628-z (2020).
- 56 Bonomi, M. & Vendruscolo, M. Determination of protein structural ensembles using cryo-electron microscopy. *Curr Opin Struct Biol* **56**, 37-45, doi:10.1016/j.sbi.2018.10.006 (2019).
- 57 Salmaso, V. & Moro, S. Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview. *Front Pharmacol* **9**, 923, doi:10.3389/fphar.2018.00923 (2018).
- 58 Chiu, M. L. *et al.* Characterization of the covalent binding of thiostrepton to a thiostrepton-induced protein from *Streptomyces lividans*. *Biochemistry* **35**, 2332-2341, doi:10.1021/bi952073e (1996).
- 59 Kahmann, J. D. *et al.* Structural basis for antibiotic recognition by the TipA class of multidrug-resistance transcriptional regulators. *EMBO J* **22**, 1824-1834, doi:10.1093/emboj/cdg181 (2003).
- 60 Chiu, M. L. *et al.* Broad spectrum thiopeptide recognition specificity of the *Streptomyces lividans* TipAL protein and its role in regulating gene expression. *J Biol Chem* **274**, 20578-20586, doi:10.1074/jbc.274.29.20578 (1999).
- 61 Murakami, T., Holt, T. G. & Thompson, C. J. Thiostrepton-induced gene expression in *Streptomyces lividans*. *J Bacteriol* **171**, 1459-1466, doi:10.1128/jb.171.3.1459-1466.1989 (1989).
- 62 Baumann, S. *et al.* Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity. *Angew Chem Int Ed Engl* **53**, 14605-14609, doi:10.1002/anie.201409964 (2014).
- 63 Huttel, S. *et al.* Discovery and Total Synthesis of Natural Cystobactamid Derivatives with Superior Activity against Gram-Negative Pathogens. *Angew Chem Int Ed Engl* **56**, 12760-12764, doi:10.1002/anie.201705913 (2017).
- 64 Cheng, B., Muller, R. & Trauner, D. Total Syntheses of Cystobactamids and Structural Confirmation of Cystobactamid 919-2. *Angew Chem Int Ed Engl* **56**, 12755-12759, doi:10.1002/anie.201705387 (2017).
- 65 Planke, T. *et al.* Cystobactamids 920-1 and 920-2: Assignment of the Constitution and Relative Configuration by Total Synthesis. *Org Lett* **21**, 1359-1363, doi:10.1021/acs.orglett.9b00058 (2019).
- 66 Liang, H. & Zeng, E. RNA-Seq Experiment and Data Analysis. *Methods Mol Biol* **1366**, 99-114, doi:10.1007/978-1-4939-3127-9\_9 (2016).
- 67 Brown, N. L., Stoyanov, J. V., Kidd, S. P. & Hobman, J. L. The MerR family of transcriptional regulators. *FEMS Microbiol Rev* **27**, 145-163, doi:10.1016/S0168-6445(03)00051-2 (2003).
- 68 Keusgen, M. Biosensors: new approaches in drug discovery. *Naturwissenschaften* **89**, 433-444, doi:10.1007/s00114-002-0358-3 (2002).
- 69 Zhang, L., Xu, J. & Birch, R. G. Engineered detoxification confers resistance against a pathogenic bacterium. *Nat Biotechnol* **17**, 1021-1024, doi:10.1038/13721 (1999).

## Photorhabdus luminescens lectin A (PIIA)

JBC ARTICLE



### Photorhabdus luminescens lectin A (PIIA): A new probe for detecting $\alpha$ -galactoside-terminating glycoconjugates

Received for publication, August 17, 2017, and in revised form, September 25, 2017. Published, Papers in Press, September 28, 2017, DOI 10.1074/jbc.M117.812792

Ghamdan Beshr<sup>†‡§¶</sup>, Asfandyar Sikandar<sup>¶||</sup>, Eva-Maria Jemiller<sup>\*\*</sup>, Nikolai Klymiuk<sup>\*\*</sup>, Dirk Hauck<sup>†‡§</sup>, Stefanie Wagner<sup>†‡§</sup>, Eckhard Wolf<sup>\*\*</sup>, Jesko Koehnke<sup>¶||</sup>, and Alexander Titz<sup>†‡§¶</sup>

From the <sup>†</sup>Divisions of Chemical Biology of Carbohydrates and <sup>||</sup>Structural Biology of Biosynthetic Enzymes, Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), D-66123 Saarbrücken, the <sup>§</sup>Deutsches Zentrum für Infektionsforschung (DZIF), Standort Hannover-Braunschweig, the <sup>¶</sup>Department of Pharmacy, Saarland University, 66123 Saarbrücken, and the <sup>\*\*</sup>Chair for Molecular Animal Breeding and Biotechnology, Gene Center and Department of Veterinary Sciences, Ludwig Maximilian University of Munich, 81377 Munich, Germany

Edited by Gerald W. Hart

Lectins play important roles in infections by pathogenic bacteria, for example, in host colonization, persistence, and biofilm formation. The Gram-negative entomopathogenic bacterium *Photorhabdus luminescens* symbiotically lives in insect-infecting *Heterorhabditis* nematodes and kills the insect host upon invasion by the nematode. The *P. luminescens* genome harbors the gene *plu2096*, coding for a novel lectin that we named PIIA. We analyzed the binding properties of purified PIIA with a glycan array and a binding assay in solution. Both assays revealed a strict specificity of PIIA for  $\alpha$ -galactoside-terminating glycoconjugates. The crystal structures of apo PIIA and complexes with three different ligands revealed the molecular basis for the strict specificity of this lectin. Furthermore, we found that a 90° twist in subunit orientation leads to a peculiar quaternary structure compared with that of its ortholog LecA from *Pseudomonas aeruginosa*. We also investigated the utility of PIIA as a probe for detecting  $\alpha$ -galactosides. The  $\alpha$ -Gal epitope is present on wild-type pig cells and is the main reason for hyperacute organ rejection in pig to primate xenotransplantation. We noted that PIIA specifically recognizes this epitope on the glycan array and demonstrated that PIIA can be used as a fluorescent probe to detect this epitope on primary porcine cells *in vitro*. In summary, our biochemical and structural analyses of the *P. luminescens* lectin PIIA have disclosed the structural basis for PIIA's high specificity for  $\alpha$ -galactoside-containing ligands, and we show that PIIA can be used to visualize the  $\alpha$ -Gal epitope on porcine tissues.

*Photorhabdus luminescens* is a Gram-negative  $\gamma$ -proteobacterium belonging to the Enterobacteriaceae family. In its complex life cycle it lives symbiotically in the intestine of Heterorh-

abditidae entomopathogenic nematodes and pathogenically in insect larvae upon nematode invasion. *P. luminescens* was first isolated in 1977 as a symbiont bacterium of Heterorhabditidae nematodes and classified initially as *Xenorhabdus luminescens* (1, 2) and later renamed *P. luminescens* (3). The genus *Photorhabdus* consists of the four species *P. luminescens*, *P. temperata*, *P. heterorhabditis*, and *P. asymbiotica* (4–6), with the latter species being pathogenic to humans (7). Both *Xenorhabdus* and *Photorhabdus* species enter a wide range of insect larvae via Steinernematidae and Heterorhabditidae nematodes, respectively (4, 8). Once the nematodes enter into the insect at their infective juvenile developmental stage, their bacterial symbionts are released into the insect's blood, and both bacteria and nematode are able to kill the larvae within 48 h (9).

The complete genome of *P. luminescens* subsp. *laumondii* TTO1 was published in 2003 by Duchaud *et al.* (10). Compared with other bacteria, it shows a high number of genes predicted as toxins, and some of them, such as *Tca* and *Tcd*, have been extensively studied (10–12). Often, bacterial carbohydrate-binding proteins, *i.e.* lectins, also act as toxins in addition to their function as adhesins to enable host colonization. A number of predicted lectins are present in the genome of *P. luminescens*, but only one example, the fucose-binding protein PLL (13), has been experimentally characterized. In *Pseudomonas aeruginosa*, a ubiquitous opportunistic pathogen with a high current medical need for new therapeutics (14), the two lectins LecA and LecB are both toxins and adhesins with roles in biofilm formation and persistent infection (15–17). LecB (18, 19)-type proteins are relatively common, and several LecB homologs have also been previously characterized in other bacterial species such as *Ralstonia solanaceum* (20), *Burkholderia cenocepacia* (21–23), and *Chromobacterium violaceum* (24). The *P. luminescens* genome harbors several genes of predicted lectins that are homologs of LecB (25). In contrast, homologs of LecA are comparably rare and are not present in the genomes of the former species. In *P. luminescens*, a single gene encoding for a LecA homolog (*plu2096*) was predicted by Duchaud *et al.* (10).

Here, we report the biochemical and structural characterization of the *P. luminescens* lectin PIIA. We demonstrate that PIIA is highly specific for  $\alpha$ -galactoside-containing ligands. By determining several crystal structures of PIIA in

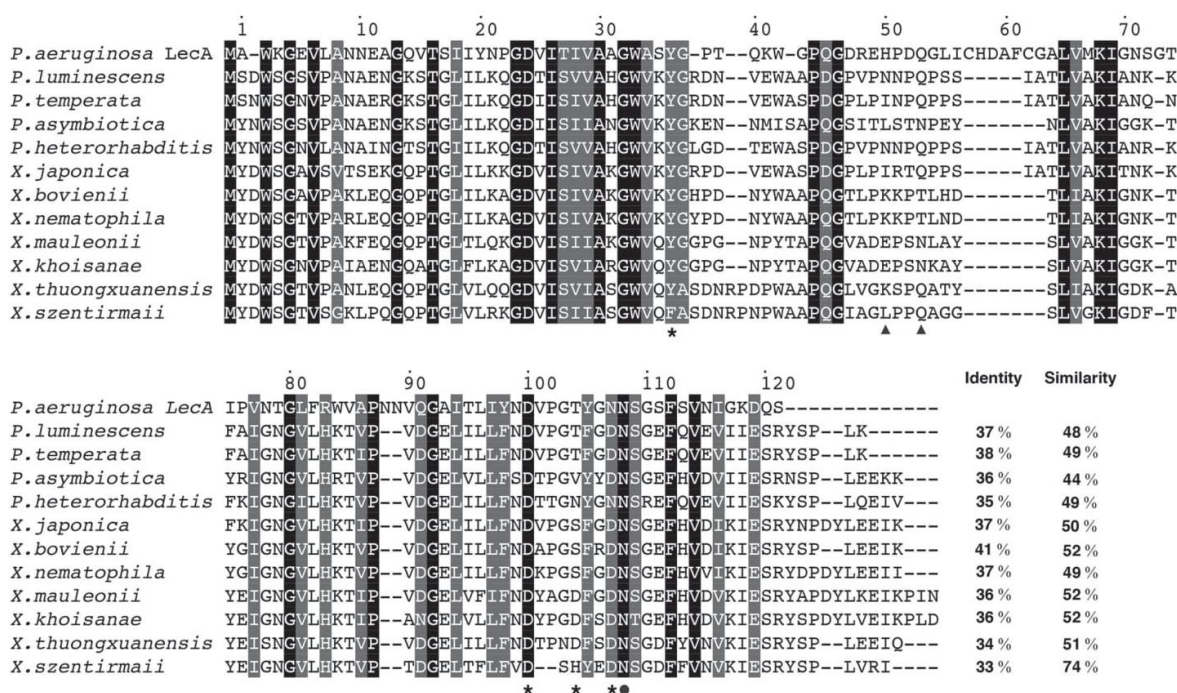
The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

This article contains supplemental Figs. S1–S11, Tables S1–S2 and Scheme S1. The atomic coordinates and structure factors (codes 5OFZ, 5ODU, 5OFX, and 5OFI) have been deposited in the Protein Data Bank (<http://www.pdb.org/>).

<sup>1</sup> Recipient of Emmy Noether Fellowship KO4116/3-1 from Deutsche Forschungsgemeinschaft. To whom correspondence may be addressed. E-mail: jesko.koehnke@helmholtz-hzi.de.

<sup>2</sup> Recipient of Helmholtz Association Grant VH-NG-934 and Deutsche Forschungsgemeinschaft Grant Ti 756/2-1. To whom correspondence may be addressed. E-mail: alexander.titz@helmholtz-hzi.de.

## Photorhabdus lectin A — PIIA



**Figure 1. Sequence alignment of LecA from *P. aeruginosa* with hypothetical LecA-like proteins from *Photorhabdus* and *Xenorhabdus* species (one single ortholog per organism selected based on highest identity to LecA).** Strictly conserved amino acids are shaded black, and similarly conserved amino acids are shaded gray. Black dot, amino acid of LecA involved in Ca<sup>2+</sup> binding; black triangles, amino acids of LecA involved in sugar binding; asterisks, amino acids of LecA involved in both Ca<sup>2+</sup> and sugar binding. Amino acid numbering follows the LecA crystal structure where the N-terminal methionine is lacking. The depicted protein sequence of *P. luminescens* (PIIA) is encoded by the *plu2096* gene.

complex with ligands, we were able to rationalize PIIA's strong preference for  $\alpha$ -galactosides. We further demonstrate that PIIA can serve as a detection tool for the specific visualization of the  $\alpha$ -Gal epitope present on porcine tissue. This epitope is responsible for hyperacute rejection of pig to primate organ xenotransplants.

## Results and discussion

### Identification and production of PIIA

Although *P. aeruginosa* LecB orthologs have been widely studied, orthologs of LecA have not been characterized in detail. We searched the publicly accessible NCBI database using protein blast and the protein sequence of LecA from strain PAO1 as a template. A moderate number of orthologs was retrieved from only a few different genera of Gram-negative bacteria. These were mainly entomopathogenic *Photorhabdus* and *Xenorhabdus* species, as well as human opportunistic pathogens from the *Enterobacter* spp. and a few other pathogenic bacterial species (see supplemental Fig. S1).

In *P. luminescens*, the gene *plu2096* was previously proposed as the coding gene for a LecA-like protein (10) and later confirmed (26) to be a galactose-binding lectin. An alignment of the retrieved orthologs of LecA in each *Photorhabdus* and *Xenorhabdus* species shows a high degree of similarity within these entomopathogenic species (Fig. 1). Although the residues involved in metal and ligand binding in LecA are relatively conserved, distinct differences to *P. aeruginosa* LecA were

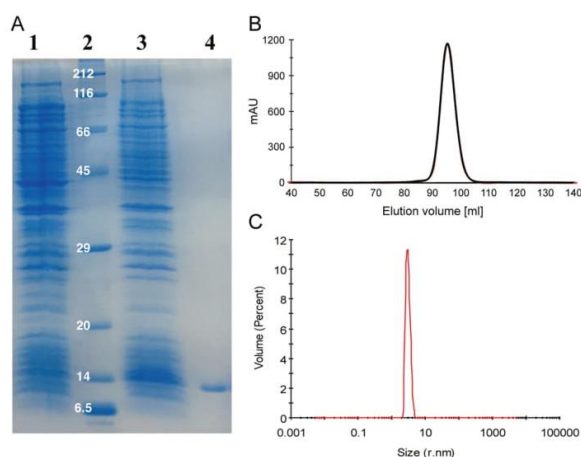
observed; LecA contains an insert flanked by two cysteines that are absent in PIIA, and the *Photorhabdus* and *Xenorhabdus* homologs contain an additional C-terminal tail of 5–13 amino acids, which is partially conserved (Fig. 1).

PIIA has 37% sequence identity to LecA, and all amino acids whose side chains are involved in calcium ion binding are conserved except for one Asn (LecA) to Asp (PIIA) variation. Amino acids involved in carbohydrate recognition are only partially conserved.

### Recombinant production and purification of PIIA

The *plu2096* gene was amplified from genomic DNA of *P. luminescens* subsp. *laumondii* TTO1 and cloned into the pET22b(+) vector. Recombinant expression in *Escherichia coli* BL21(DE3) yielded a protein product at a 13-kDa apparent molecular mass by SDS-PAGE (Fig. 2A), which corresponds to its predicted size of 12.95 kDa. Because of the sequence homology to galactose-binding LecA, we subjected the cell lysate to galactosylated Sepharose (27), and PIIA was retained on this affinity resin. Subsequent elution was achieved using galactose in the elution buffer. Purification yielded ~6 mg of PIIA/liter of bacterial culture, which was later improved to 19 mg/liter by purification on melibiose-coupled Sepharose (see below).

Because lectins often oligomerize and LecA forms a tetramer (28), we analyzed PIIA's multimeric state using



**Figure 2.** A, recombinant expression and affinity purification of PIIA analyzed by SDS-PAGE (15%). *E. coli* whole-cell extracts of uninduced (lane 1) and IPTG-induced cultures (lane 3), and purified PIIA (lane 4), molecular mass marker in kDa (lane 2) are shown. B, Sepharose size-exclusion chromatogram of PIIA; C, DLS analysis of PIIA.

size-exclusion chromatography (SEC).<sup>3</sup> The protein's observed apparent molecular mass was 27 kDa, which suggested a PIIA dimer (Figs. 2B and supplemental Fig. S2). When we used the more robust technique of dynamic light scattering (DLS), we observed a homogeneous sample with an apparent molecular mass of 52.1 kDa (Fig. 2C). The tetramer LecA was also studied by DLS under identical conditions, and its apparent molecular mass of 52.4 kDa (supplemental Fig. S3) matched the value for PIIA and thus suggests the tetramerization of PIIA in solution. The observed differences for PIIA in SEC and DLS could result from weak interactions between PIIA and the glycan-based Sepharose resin, altering the observed molecular mass to a smaller size.

**Carbohydrate-binding specificity of PIIA using the CFG glycan array**

To assess the carbohydrate-binding specificity of PIIA, the protein was fluorescently labeled using fluorescein isothiocyanate (FITC), and binding to a glycan array containing over 600 carbohydrate epitopes was performed at the Consortium for Functional Genomics (Fig. 3A). The LecA homolog PIIA showed a strict specificity toward glycans with terminal  $\alpha$ -galactosides, whereas  $\beta$ -galactosides and other carbohydrates only showed very weak or no binding. The highest apparent binding affinity was detected for the bivalent  $\alpha$ -Gal-terminating *N*-glycan (Gal- $\alpha$ -1,3-Gal- $\beta$ -1,4-GlcNAc)<sub>2</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> (glycan nos. 360 and 550) or its difucosylated derivative glycan no. 368 bearing blood group B antigens at the non-reducing end and the two monovalent disaccharides Gal- $\alpha$ -1,3-GalNAc (glycan no. 112) and Gal- $\alpha$ -1,4-GlcNAc (glycan no. 123).

The highest apparent binding was observed for bivalent Gal- $\alpha$ -1,3-Gal- $\beta$ -1,4-GlcNAc terminating glycans (glycans no. 360, 368, and 550). This glycan structure, called the  $\alpha$ -Gal epitope (29), is a ubiquitous constituent of glycans in non-primate mammals and new world monkeys. The nematode *Parelaphostrongylus tenuis* also contains *N*-glycans decorated with this epitope (30). This epitope is mainly responsible for hyperacute rejection of porcine organ transplants in humans during xenotransplantation (29, 31). Interestingly, the corresponding monovalent glycans (no. 105 and 115) showed a 4–5-fold lower binding signal (Table 1), indicating that PIIA binds carbohydrates multivalently as known for its ortholog LecA. Interestingly, when the same monovalent epitopes were presented on a shorter spacer (Sp0) with one mannose between spacer and epitope (no. 516 and 517), binding was reduced further.

Apart from the divalent ligands, only the two disaccharides, Gal- $\alpha$ -1,3-GalNAc and Gal- $\alpha$ -1,4-GlcNAc, showed a high binding signal among the monovalent series (Fig. 3A). These two ligands are monovalently displayed and may reveal the intrinsic specificity of PIIA, because other monovalent ligands showed only weak or no binding to PIIA on this glycan array. Interestingly, these two ligands displayed much stronger binding than analogous Gal- $\alpha$ -1,3-Gal and Gal- $\alpha$ -1,2-Gal, suggesting an important role of the acetamide moiety in the penultimate residue for binding to PIIA (Table 1). Among the monovalent ligands on the array, the observed linkage specificity of PIIA was broad for glycans containing terminal Gal- $\alpha$ -1,3 and Gal- $\alpha$ -1,4 linkages. The single present Gal- $\alpha$ -1,6-linked ligand (Gal- $\alpha$ -1,6-Glc) was moderately bound, whereas the single Gal- $\alpha$ -1,2-linked ligand (Gal- $\alpha$ -1,2-Gal) was not recognized by PIIA (Table 1).

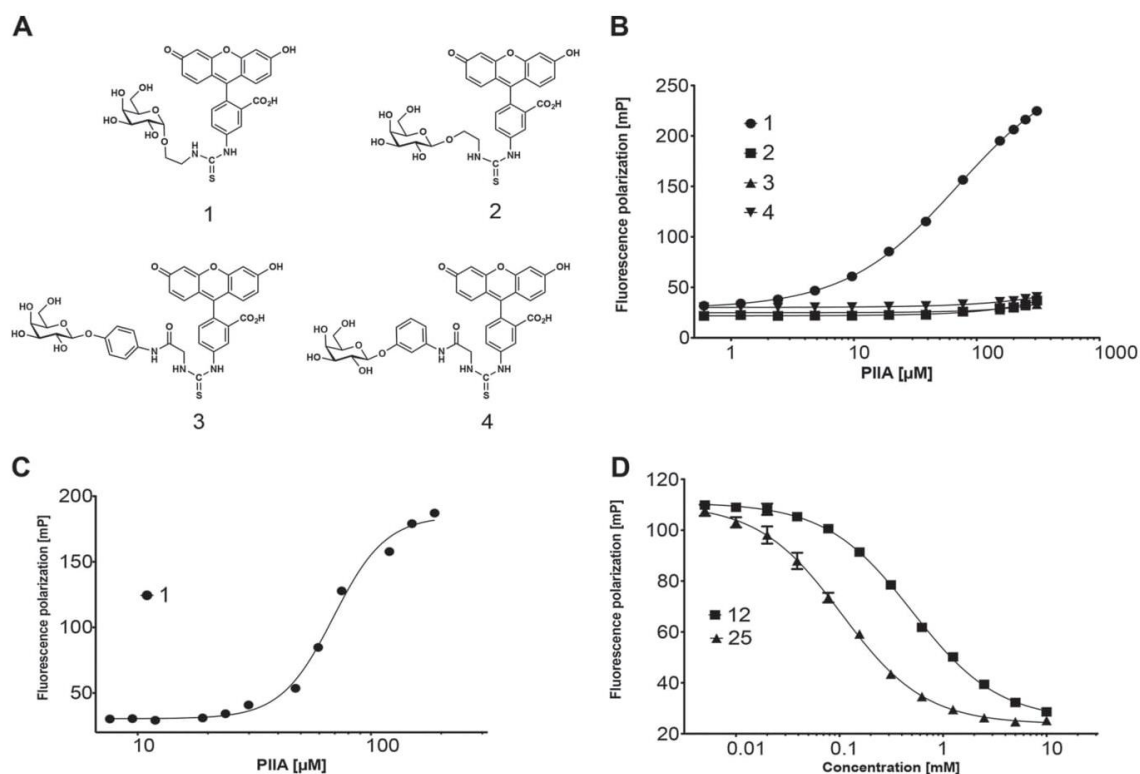
To compare the carbohydrate specificity of PIIA with its previously characterized ortholog LecA (32), glycan array binding data of monovalent glycan ligands for both lectins was normalized and plotted (Fig. 3B). Notably, LecA showed the best binding to Gal- $\alpha$ -1,4-Gal- $\beta$ -1,4-GlcNAc, which is only weakly recognized by PIIA. This glycan is part of the glycosphingolipid Gb3, which when bound by LecA triggers membrane bending, a process that was proposed as an entry pathway for *P. aeruginosa* invasion into the host cell (33). In contrast, PIIA showed high apparent binding to the epitopes Gal- $\alpha$ -1,3-GalNAc and Gal- $\alpha$ -1,4-GlcNAc, whereas LecA only shows moderate apparent binding as observed for a number of other  $\alpha$ -galactosides on the glycan array (Fig. 3B). In summary, LecA is a rather promiscuous receptor for a variety of monovalent galactosides. In addition to PIIA's binding to the bivalent *N*-glycan structures described above, PIIA was rather specific for Gal- $\alpha$ -1,3-GalNAc and Gal- $\alpha$ -1,4-GlcNAc, one or both of which may be the natural ligand of PIIA.

**Development of a competitive binding assay for PIIA**

To rapidly assess and quantify the binding specificity of PIIA, we developed a competitive binding assay for PIIA by utilizing fluorescence polarization, which is based on our previous work (19, 21, 27, 34) for four different lectins. Four different FITC-labeled *D*-galactosides (27) were titrated with increasing amounts of PIIA (Fig. 4, A–C). All three  $\beta$ -linked galactosides

<sup>3</sup> The abbreviations used are: SEC, size-exclusion chromatography; PDB, Protein Data Bank; r.m.s.d., root mean square deviation; BisTris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol; DLS, dynamic light scattering; GTKO,  $\alpha$ -1,3-galactosyltransferase knock-out; SCNT, somatic cell nuclear transfer; IPTG, isopropyl 1-thio- $\beta$ -D-galactopyranoside; BAC, bacterial artificial chromosome.





**Figure 4.** Establishing a carbohydrate-binding assay for PIIA in solution. *A*, structure of fluorescent ligands 1–4 based on D-galactose. *B*, titration of fluorescent ligands 1–4 with PIIA. *C*, dissociation constant for 1 was obtained from a four-parameter fitting procedure to the dose-dependent increase in fluorescence polarization ( $K_d$ ,  $62.7 \pm 3.8 \mu\text{M}$ ). *D*, competitive inhibition of the binding of 1 to PIIA with methyl  $\alpha$ -D-galactoside (**12**,  $\text{IC}_{50} = 0.52 \pm 0.07$  mM) and raffinose (**25**,  $\text{IC}_{50} = 0.11 \pm 0.01$  mM). One representative titration experiment of triplicates on one plate is shown. Dissociation constant and standard deviations given were obtained from at least three independent replicates of triplicates on three plates each.

We also tested a set of oligosaccharides containing  $\alpha$ -galactosyl residues for competitive binding to PIIA. Gal- $\alpha$ -1,3-Gal (**20**,  $\text{IC}_{50} = 0.90$  mM) and Gal- $\alpha$ -1,4-Gal (**21**,  $\text{IC}_{50} = 1.08$  mM) showed a 2-fold lower binding affinity to PIIA than melibiose (Gal- $\alpha$ -1,6-Glc, **22**,  $\text{IC}_{50} = 0.39$  mM), whereas Gal- $\alpha$ -1,2-Gal (**19**) was only weakly active and resulted in  $\sim 50\%$  inhibition at 10 mM. A comparable binding specificity for Gal- $\alpha$ -1,6-Glc has been reported for LecA from *P. aeruginosa* (37). The plant trisaccharide raffinose (**25**) contains a terminal melibiose motif and showed the highest binding to PIIA among all tested compounds with an  $\text{IC}_{50}$  of 0.11 mM. This ubiquitous plant galactoside has also shown an inhibitory effect on *P. aeruginosa* biofilms and inhibits LecA in a similar affinity range ( $K_d = 32 \mu\text{M}$ ) (37, 38). The tetrasaccharide stachyose (**26**) is another plant derivative of raffinose with an additional 1,6-linked  $\alpha$ -galactoside moiety. For PIIA an  $\text{IC}_{50}$  of 0.34 mM was observed, indicating that longer oligosaccharides do not improve the binding affinity to PIIA, which is different for LecA. Bivalent oligosaccharides containing the  $\alpha$ -Gal antigen Gal- $\alpha$ -1,3-Gal- $\beta$ -1,4-GlcNAc were identified as the apparent best ligands of the glycan array (Fig. 3A). The corresponding monovalent trisaccharide **27** was tested in our competitive binding assay and showed only  $\sim 70\%$  inhibition of PIIA at 10 mM. This weak binding is in good agreement with the glycan array data where the

monovalent  $\alpha$ -Gal epitope had reduced binding to PIIA compared with its bivalent structure (see Table 1).



























Because Gal- $\alpha$ -1,3-GalNAc and Gal- $\alpha$ -1,4-GlcNAc were identified as the monovalent ligands with the highest apparent affinity on the glycan array, we tested the corresponding biotinylated disaccharides **23** and **24**, respectively. Both soluble glycans differ only in spacer identity and for **23** also in the anomeric configuration from those glycans used for the production of the glycan array. Surprisingly, they were as active as the comparably biotinylated  $\alpha$ -galactosyl monosaccharide **13**, displaying  $\text{IC}_{50}$  values from 0.59 to 0.66 mM.

The human blood group antigen P<sub>1</sub> (**28**) (39) was also moderately recognized by PIIA on the glycan array (ligand 121, supplemental Table S1) and in the competitive binding assay soluble monovalent **28** showed a moderate binding affinity ( $\text{IC}_{50} = 1.80$  mM) to this lectin. Because PIIA was shown to bind to  $\alpha$ -Gal residues but not to GalNAc residues (see above) and it recognized the blood group B-terminating ligand 368 on the glycan array (Fig. 3A), we also tested soluble blood group B antigens **29** and **30**. These oligosaccharides were inhibitors of PIIA with moderate potency ( $\text{IC}_{50}$  of 1.18 and 1.35 mM, respectively). This specificity of PIIA is in contrast to LecA, which binds to B and A antigens (27). Thus, PIIA could be used as a new reagent for rapid identification of blood group B sero-

**Table 2**

**Evaluation of natural and synthetic inhibitors of PIIA using the competitive binding assay**

Averages and standard deviations were obtained from three independent experiments. n.i.: no inhibition observed up to 10 mM. SP: spacer  $-(CH_2)_3NH-CO(CH_2)_2NH-$ .

Compound	IC <sub>50</sub> [mM]	Compound	IC <sub>50</sub> [mM]	Compound	IC <sub>50</sub> [mM]	Compound	IC <sub>50</sub> [mM]
	n.i.		0.59 ± 0.05		47.8 ± 0.8% inhibition @ 10 mM		0.11 ± 0.01
	n.i.		0.93 ± 0.129		0.90 ± 0.05		0.34 ± 0.05
	n.i.		n.i.		1.08 ± 0.004		68.5 ± 0.04% inhibition @ 10 mM
	n.i.		1.32 ± 0.262		0.39 ± 0.02		1.80 ± 0.06
	1.57 ± 0.08		1.15 ± 0.434		0.65 ± 0.004		1.18 ± 0.05
	0.86 ± 0.22		0.47 ± 0.138		0.66 ± 0.02		1.35 ± 0.26
	n.i.						
	0.52 ± 0.07						

types such as the currently used GS-IB4 isolectin from *Griffonia simplicifolia* (40).

**PIIA, a lectin with a unique tetrameric structure**

Apo-PIIA crystallizes in space group P<sub>3</sub>2<sub>1</sub> and crystals diffracted to 1.7 Å. Data collection and refinement statistics for all presented PIIA structures can be found in Table 3. The core of PIIA consists of two four-stranded anti-parallel β-sheets (Fig. 5). We did not observe the canonical Ca<sup>2+</sup> ion found in other C-type lectins at the sugar-binding site, which may be a result of the crystallization buffer that contained a high concentration of citrate known to chelate calcium ions.

The asymmetric unit contained two PIIA dimers, which form tetramers with symmetry mates in accordance with DLS data. The C-terminal five-residue extension (Fig. 1) of the four protomers are engaged in well-defined interactions leading to a 90° twist in the tetramer (Fig. 6A). Of the tail residues (YSPLK), Tyr-118 packs hydrophobically against Pro-120, and Ser-119 forms two hydrogen bonds with the tetramer partner (Ser-119 side-chain hydroxyl with side-chain amino group of Lys-82, Ser-119 carbonyl with the main chain of Thr-83). Residue Leu-121 is inserted into a tailored hydrophobic pocket of the tetramer partner, which is composed of residues Leu-22, Ile-28, Ala-60, Ile-68, Phe-73, Ile-75, Val-79, Val-84, and Leu-90 (Fig. 6B). When we compared the structure of PIIA with the struc-

ture of the well-studied protein LecA, we found the structures to be very similar (Cα r.m.s.d. of 0.67 over 78 atoms, supplemental Fig. S11). The main differences are found in the region between β3 and β7 (Fig. 5C), which has a profound impact on carbohydrate binding (see below).

The sequence alignment of PIIA with LecA from *P. aeruginosa* and the LecA homologs from *Photorhabdus* and *Xenorhabdus* species showed that PIIA and LecA homologs from the latter two species possess an extension at their C termini. This extension led to the surprising result that the dimerization of dimers is twisted by 90° in PIIA. In contrast, LecA does not possess the five C-terminal residues. As a result, the LecA tetramer is planar and formed by the tail-to-tail arrangement of two dimers (Fig. 6, C and D). We have no indication as to the biological significance of this arrangement, but we believe that the interactions between the tails of the tetramer partners will lead to a significant stabilization of the tetrameric assembly.

**Structural basis of α-galactoside specificity of PIIA**

To understand the α-galactoside specificity of PIIA, we determined the crystal structures of several PIIA-carbohydrate complexes in the presence of 3 mM calcium chloride in the crystallization buffer (Fig. 7): PIIA in complex with the monosaccharide methyl α-D-galactoside (12), the trisaccharide raffinose (25), and an α-D-galactoside linked to fluorescein (1).



**Table 3**  
Data collection and refinement statistics

Statistics for the highest-resolution shell are shown in parentheses. r.m.s. is root mean square.

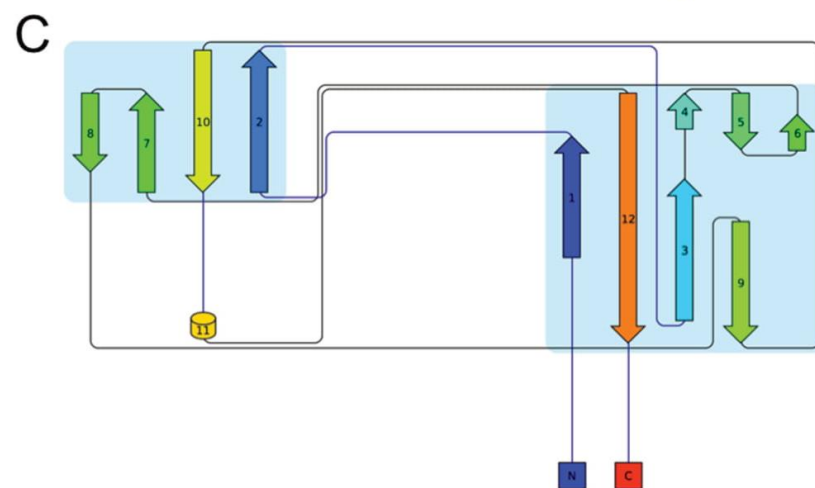
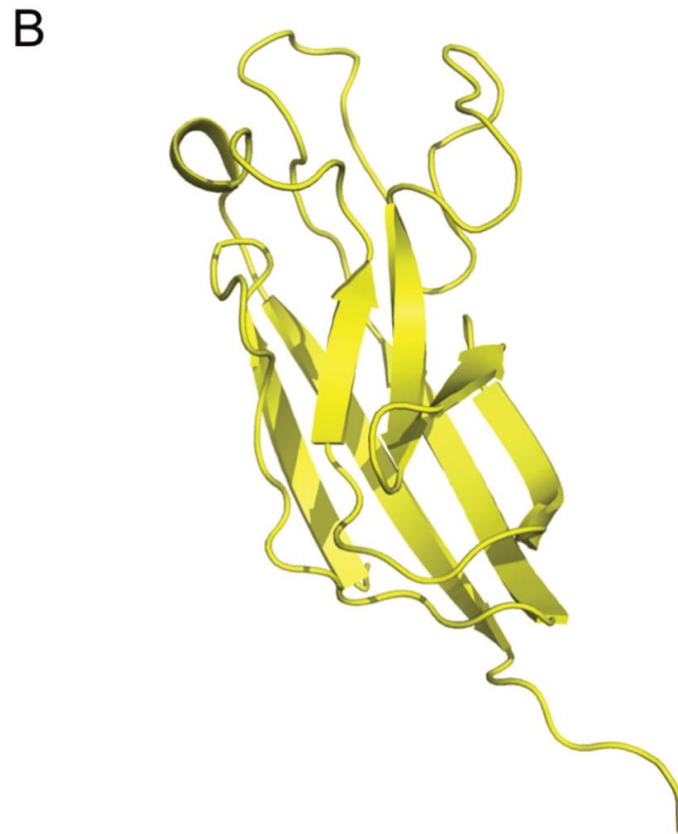
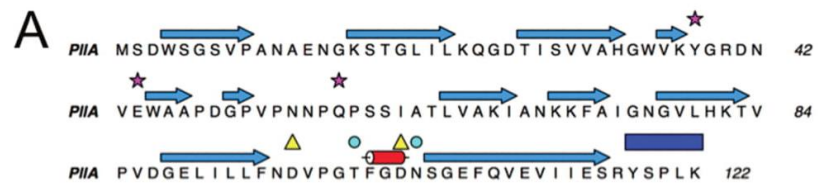
	PIIA, apo, PDB 5OFZ	PIIA, Me- $\alpha$ -Gal (12), PDB 5ODU	PIIA, raffinose (25), PDB 5OFX	PIIA, fluorescent ligand 1, PDB 5OFI
Resolution range	46.05–1.75 (1.81–1.75)	47.2–1.56 (1.62–1.56)	44.09–1.75 (1.81–1.75)	43.76–2.0 (2.07–2.0)
Space group	P 32 2 1	P 1 21 1	P 1	P 21 21 21
Unit cell	92.1 92.1 164.2 90 90 120	62.9 103.3 76.1 90 93.0 90	59.0 63.0 75.9 101.1 112.8 94.4	48.6 134.3 153.0 90 90 90
Total reflections	725,065 (34,414)	597,157 (60,933)	539,432 (54,563)	126,530 (12,830)
Unique reflections	81,583 (7832)	133,175 (13,311)	89,118 (9042)	65,944 (6644)
Multiplicity	8.9 (4.4)	4.5 (4.6)	6.1 (6.0)	1.9 (1.9)
Completeness (%)	99.59 (96.57)	96.18 (95.63)	90.73 (92.05)	95.91 (97.62)
Mean $I/\sigma(I)$	26.62 (3.32)	11.19 (1.22)	9.08 (2.02)	8.23 (2.54)
Wilson $B$ -factor	23.67	14.93	13.78	23.65
$R$ -merge	0.04649 (0.3966)	0.09883 (1.261)	0.1569 (0.7841)	0.05735 (0.2516)
$R$ -meas	0.04921 (0.451)	0.1121 (1.427)	0.1719 (0.8578)	0.08111 (0.3558)
$R$ -pim	0.01592 (0.2106)	0.05232 (0.6619)	0.06949 (0.345)	0.05735 (0.2516)
CC1/2	0.999 (0.886)	0.998 (0.358)	0.994 (0.78)	0.996 (0.907)
CC*	1 (0.969)	1 (0.726)	0.999 (0.936)	0.999 (0.975)
Reflections used in refinement	81,572 (7830)	132,517 (13,152)	89,095 (9038)	65,903 (6639)
Reflections used for $R$ -free	4183 (452)	6491 (608)	4458 (428)	3335 (326)
$R$ -work	0.2006 (0.2858)	0.2050 (0.3530)	0.1849 (0.2669)	0.1864 (0.2446)
$R$ -free	0.2202 (0.3105)	0.2334 (0.3668)	0.2133 (0.2986)	0.2216 (0.3016)
CC(work)	0.943 (0.876)	0.960 (0.626)	0.954 (0.867)	0.957 (0.905)
CC(free)	0.937 (0.823)	0.958 (0.620)	0.940 (0.834)	0.939 (0.820)
No. of non-hydrogen atoms	4138	8639	8749	8306
Macromolecules	3608	7264	7264	7264
Ligands		112	246	180
Solvent	530	1263	1239	862
Protein residues	480	968	968	968
r.m.s. (bonds)	0.009	0.003	0.004	0.004
r.m.s. (angles)	1.00	0.58	0.69	0.61
Ramachandran favored (%)	97.46	98.00	97.69	97.37
Ramachandran allowed (%)	2.54	2.00	2.31	2.63
Ramachandran outliers (%)	0.00	0.00	0.00	0
Rotamer outliers (%)	0.26	0.76	0.51	0.25
Clashscore	3.35	3.92	2.18	1.85
Average $B$ -factor	30.15	22.83	19.00	29.55
Macromolecules	28.81	20.95	17.24	28.84
Ligands		27.65	21.05	29.75
Solvent	39.27	33.22	28.88	35.49
No. of TLS groups	28	50	76	46

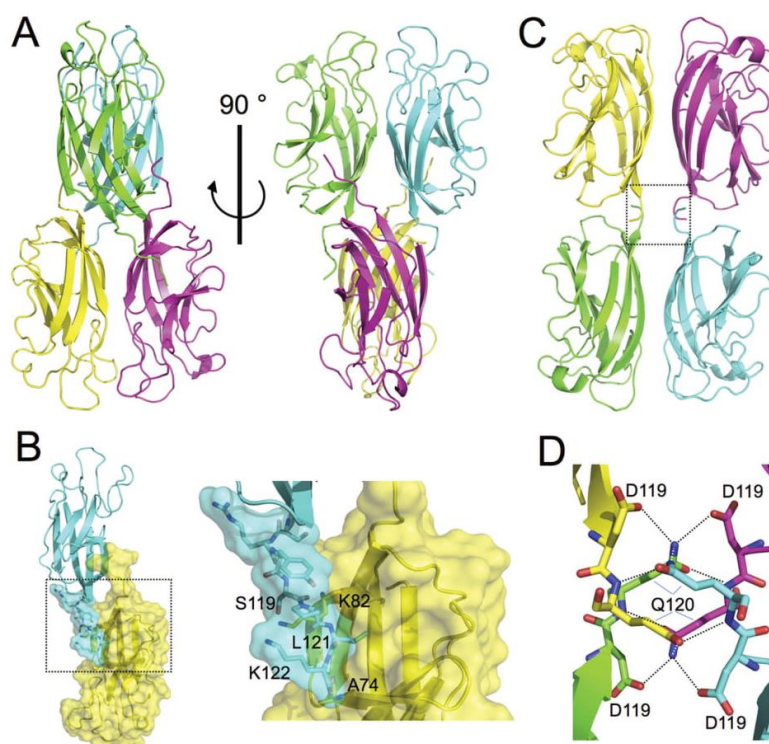
Complex crystals of PIIA with **12** were obtained by co-crystallization. The resulting crystals belonged to space group P2<sub>1</sub> and were diffracted to 1.56 Å. The overall structure of PIIA does not change upon complex formation ( $C\alpha$  r.m.s.d. of 0.12 Å), and we observed unambiguous electron density for Ca<sup>2+</sup> and the ligand in each of the eight protomers in the asymmetric unit. The canonical Ca<sup>2+</sup> ion at the sugar-binding site is coordinated by the side chains of Asp-96, Thr-100, Asp-103, and Asn-104 as well as the main-chain carbonyl oxygen atoms of Tyr-38 and Thr-100. The ligand is not involved in crystal contacts and oriented in the same way in each protomer. It sits in a shallow binding pocket and is engaged in a total of 10 hydrogen bonds (protein and the Ca<sup>2+</sup> ion): the anomeric oxygen (O1) and galactoside C2 hydroxyl with the side chain of Glu-44; sugar C3 hydroxyl with the Ca<sup>2+</sup> ion, the side chain of Asp-103, and the main chain of Tyr-38; sugar C4 hydroxyl with the Ca<sup>2+</sup> ion, the side chain of Asp-96, and the main chain of Tyr-38; sugar ring O5 with the side chain of Gln-57; and sugar C6 hydroxyl with the side chain of Gln-57. From this complex structure, it is clear that PIIA is only able to bind to  $\alpha$ -galactosides; the side chains of the two amino acids Val-43 and Asn-55 that are absent in LecA form a barrier and any  $\beta$ -linkage would result in a clash with the PIIA surface (Fig. 8). The structure also allowed us to rationalize why D-galactosamine (**10**) is a better binder than D-galactose (**9**), and why N-acetylgalactosamine (**11**) shows no detectable binding. By swapping the C2 hydroxyl group for an amino

group, the sugar can now engage in an additional hydrogen bond with the side chain of Asp-103 (see supplemental Fig. S8). Acetylation of the amino group leads to a clash with the protein and thus abolishes binding.

Complex crystals of PIIA with raffinose were also obtained by co-crystallization. The resulting crystals belonged to space group P1 and diffracted to 1.75 Å. As expected, the overall structure of PIIA does not change upon complex formation ( $C\alpha$  r.m.s.d. of 0.12 Å). There are eight protomers in the asymmetric unit, and we observed unambiguous electron density for raffinose in all of them. Although some of the raffinose molecules are involved in crystal contacts, the orientation and shape of the trisaccharide is virtually identical in each protomer. The orientation of the galactose moiety does not differ between the monosaccharide and raffinose structures, and the same hydrogen bonds are formed. The  $\alpha$ -1,6-linkage leads glucose away from the PIIA surface, but the C4 hydroxyl forms one hydrogen bond with the side chain of Gln-57. Interestingly, the final fructose moiety is pointing back toward the PIIA surface, giving the raffinose an overall horseshoe shape. The fructose C3 and C4 hydroxyls are engaged in a hydrogen bond with the side chain of Glu-44. Glu-44 is also involved in hydrogen bond formation with the galactose C2 hydroxyl, thus linking the two ends of the raffinose horseshoe resulting in an additional intraligand hydrogen bond between fructose C6 hydroxyl with galactose C2 hydroxyl.

*Photorhabdus lectin A* — PIIA





**Figure 6. Overall structure of PIIA and comparison to LecA.** *A*, schematic representation of the PIIA tetramer. Two parallel dimers (yellow/magenta and green/cyan) form tail-to-tail dimers with a 90° twist. *B*, detailed view of the PIIA tail-to-tail interface. We observe two hydrogen bonds between the side chains of tail Ser-119 (cyan) and Lys-82 (yellow) and the C terminus of the tail and the backbone nitrogen of Ala-74 (yellow). In addition, tail residue Leu-121 is inserted into a hydrophobic pocket of its binding partner. *C*, LecA tetramer is planar, formed by tail-to-tail dimerization of two parallel dimers (yellow/magenta and green/cyan). *D*, much shorter tail of LecA provides several stabilizing hydrogen bonds (dashed lines), but the interactions are not sufficient to cause a twist of the two dimers relative to each other.

Because we used fluorescent probe **1** in our competitive binding assays, we wanted to understand how the probe binds to PIIA. Complex crystals of PIIA with **1** were also obtained by co-crystallization. The resulting crystals belonged to space group  $P2_12_12_1$  and diffracted to 2.0 Å. The overall structure of the eight protomers in the asymmetric unit did not differ significantly from the apo structure ( $C\alpha$  r.m.s.d. of 0.15 Å), and we observed unambiguous density for **1** in four of the protomers (Fig. 7, *E* and *F*, and supplemental Fig. S10). The interactions of the galactose include all of those observed in the other two structures. Through fortuitous crystal packing, we were able to obtain good electron density for the fluorophore and were able to fit it. The ordered nature of the fluorophore is the result of  $\pi$ -stacking between the tricyclic ring systems of two molecules of **1** bound to symmetry mates.

The strict specificity toward  $\alpha$ -galactosides is unique for PIIA when compared with LecA. From the crystal structure of PIIA with **12**, it becomes clear that  $\beta$ -galactosides cannot be recognized without a steric clash with the protein surface of PIIA (Fig. 8*A*). In contrast, LecA opens a shallow cleft close to its anomeric center that allows the accommodation of

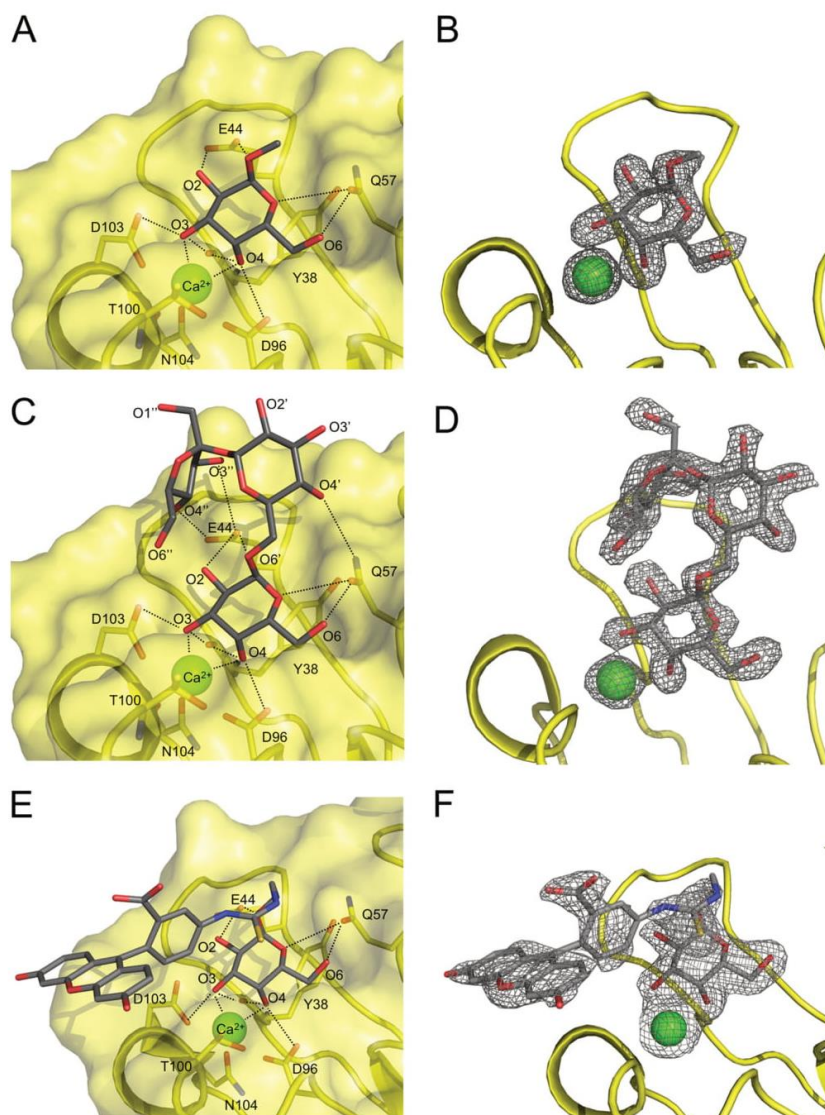
large  $\beta$ -linked aglycons, such as in 4-nitrophenyl  $\beta$ -D-galactoside. From a superposition of the binding site amino acid residues of PIIA with LecA, it can be deduced that the additional amino acids Val-43 and Asn-55 present in PIIA are responsible for preventing PIIA from binding  $\beta$ -galactosides (Fig. 8*B*).

#### Application of PIIA for the detection of the $\alpha$ -Gal epitope

PIIA showed the highest apparent binding on the glycan array to a biantennary *N*-glycan structure carrying the  $\alpha$ -Gal epitope on its antenna. This antigen (Gal- $\alpha$ -1,3-Gal- $\beta$ -1,4-GlcNAc) is a ubiquitous epitope in non-primate mammals and new world monkeys. This carbohydrate structure is the major factor of hyperacute rejection of xenotransplanted organs in humans (29, 41, 42). In pigs, genetic engineering resulted in animals lacking the corresponding galactosyltransferase thus reducing the risk of severe immune responses (31, 43, 44). By using modern techniques such as CRISPR/Cas to engineer animals or animal tissue lacking the  $\alpha$ -Gal epitope, quality controls for the complete suppression of the biosynthetic machineries are of crucial importance. Currently, the isolectin GS-IB4

**Figure 5.** *A*, sequence of PIIA. Secondary structure elements are shown above the sequence (blue arrows,  $\beta$ -strands; red barrel,  $\alpha$ -helix). Residues responsible for sugar binding are highlighted with magenta stars,  $Ca^{2+}$ -binding residues with cyan circles, and amino acids coordinating both as yellow triangles. Tail residues unique to PIIA and its close homologs are highlighted with a blue box. *B*, schematic representation of a PIIA apo monomer. *C*, fold diagram for the structure shown in *B*.

*Photorhabdus lectin A — PIIA*

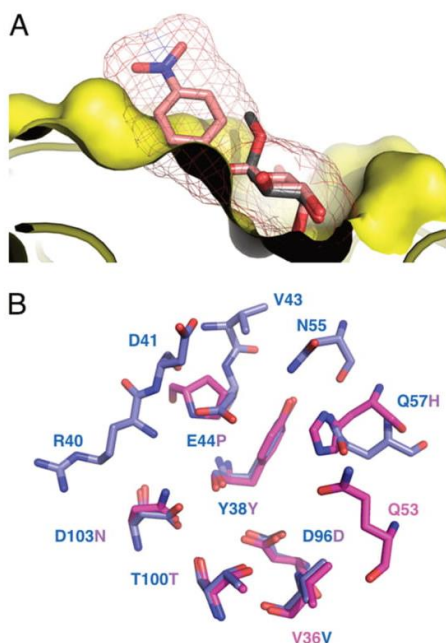


**Figure 7. PIIA-carbohydrate complex structures.** *A* and *B*, PIIA bound to methyl  $\alpha$ -D-galactoside (**12**). This interaction is stabilized through 10 hydrogen bonds (*dashed lines*). Eight of them are between the ligand and the protein, and two are provided by the  $\text{Ca}^{2+}$  ion. *C* and *D*, PIIA bound to raffinose (**25**). In addition to the hydrogen bonds observed in *A*, the glucose moiety forms a hydrogen bond with the side chain of Gln-57, whereas terminal fructofuranoside forms two hydrogen bonds with the side chain of Glu-44, which results in the ligand adopting a horseshoe shape. *E* and *F*, PIIA bound to fluorescein tracer **1**. No interactions with the protein are observed beyond the carbohydrate moiety. The fluorescein can only be observed as the result of fortuitous crystal contacts in half of the monomers in the asymmetric unit. PIIA is shown as a *yellow schematic/surface representation*, ligand as *gray sticks*, oxygen atoms in *red*, nitrogen atoms in *blue*, sulfur atoms in *yellow*, and  $\text{Ca}^{2+}$  ions as *green spheres*. Difference electron density ( $F_o - F_c$ ) contoured to  $3\sigma$  with phases calculated from a model that was refined in the absence of metal ions is shown as *gray isomesh* (*B*, *D*, and *F*).

purified from the plant *G. simplicifolia* is used as a tool to identify a wide range of  $\alpha$ -galactoside epitopes, among which is the  $\alpha$ -Gal epitope (45). Because of the high selectivity of PIIA, this bacterial lectin could be an alternative to the currently used GS-IB4.

The crystal structure of GS-IB4 in complex with the terminal disaccharide Gal- $\alpha$ -1,3-Gal (**20**) as a methyl glycoside shows extensive interactions between the terminal galactose residue and the protein but no contacts with the reducing-end galac-

tose moiety (46). To compare the recognition features of both proteins to this epitope, we have docked the methyl glycoside of **20** into the carbohydrate-binding site of PIIA (supplemental Fig. S9). In this computed structure, the terminal saccharide moiety forms extensive contacts with the lectin receptor. In contrast to GS-IB4, the reducing-end galactose established two more hydrogen bonds with Glu-44 and Asn-55 of PIIA, which may serve as an explanation for the high specificity of PIIA for the Gal- $\alpha$ -1,3-Gal epitope.



**Figure 8. Rationalizing PIIA  $\alpha$ -galactoside specificity.** *A*, representation of the PIIA (yellow)-binding pocket with methyl  $\alpha$ -D-galactoside (**12**, gray sticks). 4-Nitrophenyl  $\beta$ -D-galactoside (salmon sticks and isomesh, taken from PDB 3ZYF) was superposed onto  $\alpha$ -D-galactoside. Because of the restricted ligand-binding site of PIIA only  $\alpha$ -substituted ligands, leading away from the surface, can be accommodated, whereas  $\beta$ -substituted ligands clash. *B*, superposition of the binding site amino acid residues of PIIA (blue) with LecA (magenta), oxygen atoms, red; nitrogen atoms, blue. Residue numbers correspond to PIIA.

We thus tested the suitability of PIIA to detect the  $\alpha$ -Gal epitope in wild-type primary pig kidney cells and cells derived from the corresponding  $\alpha$ -1,3-galactosyltransferase knock-out (GTKO) animals (Fig. 9). Both PIIA and the current standard lectin GS-IB4 visualized the  $\alpha$ -Gal antigen in wild-type porcine cells similarly. Because the  $\alpha$ -Gal antigen is also present on glycolipids in red blood cells (41), we performed hemagglutination experiments with red blood cells (RBC) from wild-type and GTKO pigs. PIIA agglutinated wild-type porcine red blood cells but was unable to agglutinate RBCs from the GTKO pig (Fig. 10, *A* and *B*). In addition, we could further show that this agglutination was galactose-dependent and could be inhibited by the presence of raffinose (Fig. 10C).

**Conclusion**

The opportunistic pathogen *P. aeruginosa* utilizes the two soluble lectins LecA and LecB for infection of the host and biofilm formation. Although numerous LecB-like proteins have been characterized, LecA orthologs are scarce. Here, we show that various orthologs of LecA are present in the insect pathogenic bacteria from the *Photorhabdus* and *Xenorhabdus* species as well as in the human gut bacterium and pathogen *Enterobacter* spp. A high degree of similarity was observed among those orthologs with LecA having a sequence insert and lacking an otherwise conserved C-terminal tail.

The gene *plu2096* from the entomopathogenic bacterium *P. luminescens* was cloned and recombinantly produced in high

production yields. It encodes the galactose-binding lectin PIIA with 37% identity to LecA. The carbohydrate-binding specificity of PIIA was assessed on a glycan array containing over 600 different carbohydrate epitopes. Interestingly, PIIA showed very strict specificity toward  $\alpha$ -galactosides with high apparent binding to the  $\alpha$ -Gal epitope as well as to Gal- $\alpha$ -1,4-GlcNAc and Gal- $\alpha$ -1,3-GalNAc.

To date, the biological role of the Gal- $\alpha$ -1,4-GlcNAc epitope remains unclear, and natural sources have not been identified despite the fact that anti-Gal- $\alpha$ -1,4-GlcNAc antibodies are present in human serum (47). In contrast, Gal- $\alpha$ -1,3-GalNAc epitopes are present in nematodes and have, for example, been described in glycolipids from the worms *Ascaris suum* and *Caenorhabditis elegans* (48). Furthermore, Gal- $\alpha$ -1,3-GalNAc is a ubiquitous epitope present on glycoproteins of the nematode *Hemonchus contortus*, and vaccination of lambs with glycoproteins of the sheep parasite *H. contortus* specifically elicited anti-Gal- $\alpha$ -1,3-GalNAc IgG antibodies (49). In addition,  $\alpha$ -linked galactosyl residues have been identified in *C. elegans* N-glycans attached to mannose residues (50, 51) or attached to core fucose residues (52). Importantly, the nematodes *C. elegans* and *H. contortus* are both phylogenetically closely related to *Heterorhabditis*, whereas *A. suum* is more distantly related (53). It is thus reasonable to speculate that the Gal- $\alpha$ -1,3-GalNAc epitope recognized by the bacterial lectin PIIA is also present in the nematode *Heterorhabditis* and plays a role in bacterial attachment or symbiosis of *Photorhabdus* species with their native nematode hosts. Moreover, this epitope has been described as one terminal constituent of glycosphingolipids of the insect *Calliphora vicina* pupae (54) and members of the order of diptera, *i.e.* flies, are generally susceptible to infection with *Heterorhabditis* and *P. luminescens*. Therefore, it is possible that Gal- $\alpha$ -1,3-GalNAc is one natural ligand bound by PIIA both in the nematode symbiont and in infected insects.

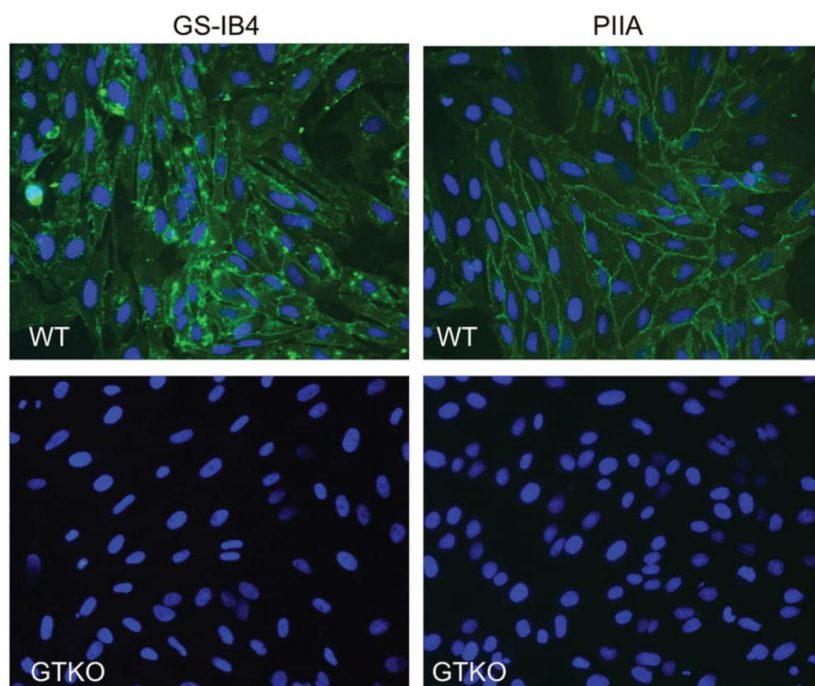
Based on the carbohydrate specificity of PIIA as determined by the glycan array, PIIA was tested in a porcine cell culture staining experiment for the detection of the  $\alpha$ -Gal epitope, the prime reason for hyperacute organ rejection in xenobiotic transplants. PIIA proved to be a suitable detection tool and specifically detected the  $\alpha$ -Gal epitope in porcine tissue and on red blood cells. This fact qualifies recombinantly produced PIIA for the efficacy assessment of methods to genetically manipulate cells, such as CRISPR/Cas, for the production of alternative animal cells, tissue, or organisms lacking the  $\alpha$ -Gal epitope as donors for xenotransplantation.

We are currently analyzing the biological role of PIIA in *P. luminescens* for its life cycle in nematodes and insects. Furthermore, it will be of interest to analyze the role of PIIA orthologs in the human pathogens *P. asymbiotica* and the *Enterobacter* spp., a group of bacteria that are part of the normal human gut flora with pathogenic potential.

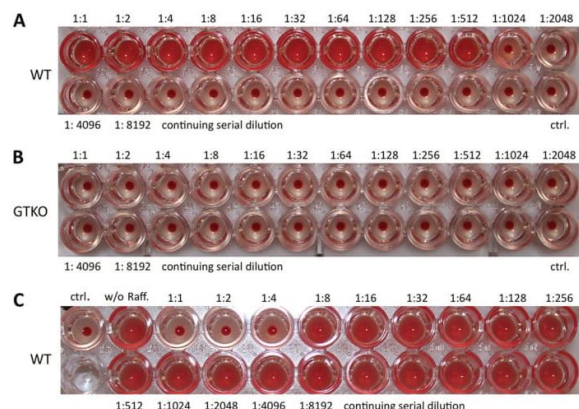
**Materials and methods**

**Chemicals**

Methyl  $\alpha$ -L-fucoside (**6**), methyl  $\alpha$ -D-mannoside (**7**), and D-galactose (**9**) were purchased from Dextra Laboratories (Reading, UK); D-galactosamine (**10**), methyl  $\alpha$ -D-galactoside



**Figure 9. Staining of primary porcine kidney cells from wild-type pigs (WT) and GGTA1 KO (GTKO) animals unable to produce the  $\alpha$ -Gal antigen.** Fluorescein-tagged PIIA or GS-IB4 were used as probes and detected the  $\alpha$ -Gal antigen in WT cells. Lectin concentration: PIIA, 50  $\mu$ g/ml; GS-IB4, 500  $\mu$ g/ml, 400 $\times$  magnification.



**Figure 10. Hemagglutination of porcine red blood cells by PIIA.** A, wild-type pig RBCs. B, GTKO pig RBCs. C, inhibition of PIIA-mediated agglutination of wild-type pig RBCs with raffinose.

(12), *p*-nitrophenyl  $\alpha$ -D-galactoside (16), 4-methylumbelliferyl  $\alpha$ -D-galactoside (17), 5-bromo-4-chloro-3-indolyl  $\alpha$ -D-galactoside (18), isopropyl  $\beta$ -D-1-thiogalactoside (IPTG) were from Carbosynth Ltd. (UK); methyl  $\beta$ -D-arabinoside (8) was from Tokyo Chemical Industry (Japan); methyl  $\alpha$ -D-glucoside (5), *N*-acetyl-D-galactosamine (11), and stachyose (26) were from Sigma (Germany); melibiose (22) was from MP Biomedicals Llc. (France); raffinose (25) was from Th. Geyer Laboratories (Germany); Gal- $\alpha$ -1,3-Gal (20), Gal- $\alpha$ -1,4-Gal (21), Xeno antigen (27), P1 antigen (28), and blood group B antigens (29, 30)

were from Elicityl OligoTech (France); Gal- $\alpha$ -SP-biotin (13), Gal- $\alpha$ -1,3-GalNAc- $\alpha$ -SP-biotin (23), and Gal- $\alpha$ -1,4-GlcNAc- $\beta$ -SP-biotin (24) were from Lectinity (Russia); and Gal- $\alpha$ -1,2-Gal- $\beta$ -1-OMe (19) was from Carbohydrate Synthesis (Oxford, UK). Fluorescent ligands 1–4 were synthesized as described (27).

Allyl  $\alpha$ -D-galactopyranoside (14) was synthesized from galactose in a Fischer-type glycosylation with allyl alcohol in presence of Amberlite IR120/H<sup>+</sup> (supplemental Scheme 1). The title compound was obtained by recrystallization. <sup>1</sup>H NMR (400 MHz, MeOH-*d*<sub>4</sub>)  $\delta$  5.98 (dddd, *J* = 17.2, 10.4, 6.1, 5.2 Hz, 1H, CH<sub>2</sub>CHCH<sub>2</sub>O-), 5.33 (dq, *J* = 17.2, 1.7 Hz, 1H, CH<sub>2</sub>CHCH<sub>2</sub>O-), 5.17 (dq, *J* = 10.4, 1.4 Hz, 1H, CH<sub>2</sub>CHCH<sub>2</sub>O-), 4.87 (d, *J* = 3.0 Hz, 1H, H1), 4.23 (ddt, *J* = 13.0, 5.3, 1.5 Hz, 1H, CH<sub>2</sub>CHCH<sub>2</sub>O-), 4.04 (ddt, *J* = 13.0, 6.1, 1.4 Hz, 1H, CH<sub>2</sub>CHCH<sub>2</sub>O-), 3.93–3.89 (m, 1H, H4), 3.85–3.78 (m, 1H, H5), 3.78–3.76 (m, 2H, H2, H3), 3.73–3.69 (m, 2H, H6). <sup>13</sup>C NMR (101 MHz, MeOH-*d*<sub>4</sub>)  $\delta$  135.65 (CH<sub>2</sub>CHCH<sub>2</sub>O-), 117.48 (CH<sub>2</sub>CHCH<sub>2</sub>O-), 99.46 (C1), 72.37 (C5), 71.51 (C2/3), 71.16 (C4), 70.21 (C2/3), 69.39 (CH<sub>2</sub>CHCH<sub>2</sub>O-), 62.78 (C6). Transcripts of NMR spectra are shown in supplemental Figs. 4 and 5).

For allyl 3-*O*-methyl- $\alpha$ -D-galactopyranoside (15), a microwave vial was filled with allyl galactoside 14 (57 mg, 0.26 mmol) and dibutyltin oxide (71 mg, 0.29 mmol), and the reagents were dried *in vacuo*. Dry PhMe/MeCN (5:1, 660  $\mu$ l) was added, and the suspension in the sealed tube was exposed to microwave irradiation for 20 min at 150  $^{\circ}$ C. The clear solution was allowed to cool to 50  $^{\circ}$ C, and to the resulting suspension was added Mel (405  $\mu$ l, 6.5 mmol) dropwise. The reaction was stirred for 48 h at

50 °C, after removal of the volatiles *in vacuo* and purification of the crude product by MPLC, the title compound was obtained as a pure product (40 mg, 65%). <sup>1</sup>H NMR (400 MHz, MeOH-*d*<sub>4</sub>) δ 6.08–5.89 (m, 1H, CH<sub>2</sub>CHCH<sub>2</sub>O-), 5.34 (dd, *J* = 17.2, 1.8 Hz, 1H, CH<sub>2</sub>CHCH<sub>2</sub>O-), 5.17 (dd, *J* = 10.4, 1.6 Hz, 1H, CH<sub>2</sub>CHCH<sub>2</sub>O-), 4.85 (d, *J* = 4.0 Hz, 1H, H1), 4.23 (dd, *J* = 13.0, 5.3, 1.6 Hz, 1H, CH<sub>2</sub>CHCH<sub>2</sub>O-), 4.13 (dd, *J* = 3.3, 1.2 Hz, 1H, H4), 4.04 (dd, *J* = 13.0, 6.1, 1.4 Hz, 1H, CH<sub>2</sub>CHCH<sub>2</sub>O-), 3.85 (dd, *J* = 10.1, 3.9 Hz, 1H, H2), 3.82–3.77 (m, 1H, H5), 3.76–3.63 (m, 2H, H6), 3.46 (s, 3H, CH<sub>3</sub>), 3.42 (dd, *J* = 10.1, 3.2 Hz, 1H, H3). <sup>13</sup>C NMR (101 MHz, MeOH-*d*<sub>4</sub>) δ 135.65 (CH<sub>2</sub>CHCH<sub>2</sub>O-), 117.50 (CH<sub>2</sub>CHCH<sub>2</sub>O-), 99.37 (C1), 81.15 (C3), 72.35 (C5), 69.38 (CH<sub>2</sub>CHCH<sub>2</sub>O-), 69.22 (C2), 66.91 (C4), 62.74 (C6), 57.24 (CH<sub>3</sub>). Transcripts of NMR spectra are shown in supplemental Figs. 6 and 7).

### Bioinformatics

A BLAST search (blastp) was done using the amino acid sequences of LecA from *P. aeruginosa* as query (accession number Q05097). The search was carried out choosing non-redundant protein sequence database with exclusion of *P. aeruginosa* (taxid: 287). The best 100 matches were chosen for an alignment using the COBAL tool (55). The aligned sequences were clustered using the SECATOR algorithm (56), which relies on BIONJ (57) to build the phylogenetic tree. The best LecA-like sequences (with lowest E-value) from each *Xenorhabdus* and *Photorhabdus* species were aligned with LecA using COBAL, and the conserved sequence was colored using Color Align Conservation (58). The amino acid sequence of hypothetical LecA homologs (with lowest E-value in *Xenorhabdus* and *Photorhabdus* species) were WP\_011146351.1 (*P. luminescens*), WP\_046975865.1 (*P. temperata*), WP\_065824676.1 (*P. asymbiotica*), WP\_054480913.1 (*P. heterorhabditis*), WP\_038256436.1 (*Xenorhabdus bovienii*), WP\_013184196.1 (*Xenorhabdus nematophila*), WP\_047963870.1 (*Xenorhabdus khoisanensis*), WP\_074019816.1 (*Xenorhabdus thuongxuanensis*), WP\_038237499.1 (*Xenorhabdus szentirmaii*), and GenBank<sup>TM</sup> number SFO04414.1 (*Xenorhabdus japonica*), and SFJ01328.1 (*Xenorhabdus mauleonii*).

### Cloning, expression, and purification of recombinant PIIA

Genomic DNA was isolated from *P. luminescens* subsp. *laumondii* TTO1 using GenElute Bacterial Genomic DNA Kit (Sigma). The *plu2096* gene sequence was amplified by PCR with Phusion polymerase (New England Biolabs, UK) and primers introducing NdeI (5'-GGAATTCCATATGTCT-GATTGGTCAGGAAG-3') and BamHI (5'-CGGGATCCT-TATTTAAAGGGGAGTATCGAG-3') restriction sites. After digestion of the expression vector pET22b(+) (Novagen, Germany) and the PCR product with NdeI and BamHI (New England Biolabs, UK), ligation of the insert was performed with T4 DNA ligase (New England Biolabs, UK) resulting in plasmid pET22b-*plIA*. The sequence was confirmed by sequencing (GATC Biotech, Germany) with primers T7 promoter (5'-TAATACGACTCACTATATAGG-3') and T7 terminator (5'-GCTAGTTATTGCTCAGCGG-3').

For expression, pET22b-*plIA* was transformed into chemically competent *E. coli* BL21(DE3), and the expression strain was selected on LB agar supplemented with ampicillin (100 μg

ml<sup>-1</sup>). 2 liters of LB supplemented with ampicillin (100 μg ml<sup>-1</sup>) were inoculated with a preculture and grown at 37 °C and 180 rpm to an A<sub>600</sub> of 0.5–0.6. Expression was induced with addition of IPTG (0.5 mM final concentration), and bacteria were then further cultured for 6 h at 30 °C and 180 rpm. The cells were harvested by centrifugation (3000 × *g*, 10 min), and the pellet was washed with TBS/Ca (20 mM Tris, 137 mM NaCl, 2.6 mM KCl, pH 7.4, supplemented with 100 μM CaCl<sub>2</sub>). The cells were resuspended in 25 ml of TBS/Ca with PMSF (1 mM) and lysozyme (0.4 mg ml<sup>-1</sup>) and subsequently disrupted by five cycles in a microfluidics homogenizer (Microfluidics Corp.). Cell debris was removed by centrifugation (10,000 × *g*, 60 min), and the supernatant was loaded onto a column containing galactosylated (59) or later melibiose-coupled Sepharose CL-6B. The column was washed with TBS/Ca, and PIIA was eluted by addition of 100 mM galactose or 100 mM raffinose to the buffer. The eluted fractions were extensively dialyzed against distilled water and then TBS/Ca buffer. The concentration was determined by UV absorbance at 280 nm using a calculated molar extinction coefficient of 19,480 M<sup>-1</sup> cm<sup>-1</sup>. The yield of purified PIIA was 6 mg (galactose-column) or 19 mg (melibiose-column) per liter of culture volume.

### Gel filtration

A HiLoad 16/600 Superdex 200 pg (GE Healthcare) was equilibrated with TBS/Ca buffer (20 mM Tris, 137 mM NaCl, 2.6 mM KCl, pH 7.4, supplemented with 1 mM CaCl<sub>2</sub>) with a flow rate of 1 ml/min. A calibration curve for molecular size estimation was generated by loading 10 μM of mixture of standard proteins (lysozyme, DNase I, ovalbumin, and BSA). Thereafter, 10 μM PIIA was loaded on the column and analyzed with the same flow rate.

### Dynamic light scattering (DLS) measurements

DLS measurements were performed on a Zetasizer Nano-ZS (Malvern Instruments, UK). Stock solutions were filtered with a syringe filter before measurements. 50 μl of PIIA or LecA (100 mM) in TBS/Ca (20 mM Tris, 137 mM NaCl, 2.6 mM KCl, pH 7.4, supplemented with 1 mM CaCl<sub>2</sub>) was measured at 25 °C.

### Fluorescent labeling of PIIA and glycan array analysis

PIIA (700 μl, 58 μM in Na<sub>2</sub>CO<sub>3</sub> buffer, pH 9.3) was incubated at room temperature under shaking (500 rpm) with fluorescein isothiocyanate (FITC, 33 μl, 3 mg ml<sup>-1</sup>, in sodium carbonate buffer, pH 9.3) for 1 h. Purification of the labeled protein was performed as described above for unlabeled PIIA; the protein concentration was determined as described previously for LecB-PA14 (19) using an extinction coefficient of 19,480 M<sup>-1</sup> cm<sup>-1</sup> for PIIA.

FITC-labeled PIIA was tested on the Consortium for Functional Glycomics (CFG) mammalian glycan array (Core H) version 5.3 containing 600 printed glycans in replicates of 6. Standard procedures of Core H (details see <http://www.functionalglycomics.org/glycomics/publicdata/selectedScreens.jsp>)<sup>4</sup> were run at 5 and 50 μg ml<sup>-1</sup> protein based on the protocol

<sup>4</sup> Please note that the JBC is not responsible for the long-term archiving and maintenance of this site or any other third party hosted site.

## Photorhabdus lectin A — PIIA

by Blixt *et al.* (60). Raw data of the PIIA binding experiments are available as supplemental Tables S1 and S2 as an XLS spreadsheet.

### Direct binding of fluorescent ligands 1–4 to PIIA

10  $\mu\text{l}$  of a serial dilution of PIIA in TBS/Ca (618–0.30  $\mu\text{M}$ ) was added in triplicate to a 384-well plate (Greiner Bio-One, Germany, catalog no. 781900). Then, 10  $\mu\text{l}$  of fluorescent ligand 1–4 dissolved in TBS/Ca were added to PIIA to a final concentration of 10 nM. After incubation for 1 h at room temperature, blank corrected fluorescence intensity was recorded using a PheraStar FS microplate reader (BMG Labtech GmbH, Germany) with excitation filters at 485 nm and emission filters at 535 nm, and fluorescence polarization was calculated. The data were analyzed using a four-parameter fit of the MARS Data Analysis Software (BMG Labtech GmbH, Germany). A minimum of three independent experiments on three plates was performed for each fluorescent ligand.

### Competitive binding assay for PIIA

10  $\mu\text{l}$  of a serial dilution of each tested compounds in TBS/Ca (20 to 0.01 mM) were added in triplicate to a 384-well plate (Greiner Bio-One, Germany, catalog no. 781900). Afterward, 10  $\mu\text{l}$  of PIIA and **1** were added to each well at final concentrations of 55  $\mu\text{M}$  and 10 nM, respectively. After incubation for 1 h at room temperature, fluorescence polarization was determined using a microplate reader as described above. The data were analyzed using a four-parameter fit of the MARS Data Analysis Software (BMG Labtech GmbH, Germany). A minimum of three independent experiments on three plates was performed for each compound.

### X-ray crystallography

Crystals of apo-PIIA were obtained in 1.6 M sodium citrate tribasic dihydrate, pH 6.5. To solve the PIIA–ligand complex structures, PIIA was co-crystallized in the presence of 10 mM ligand and 3 mM calcium chloride. Optimized crystals of PIIA-12, PIIA-25, and PIIA-1 were grown under conditions of 0.2 M ammonium acetate, 0.1 M BisTris buffer, pH 5.5, and 25% PEG 3350; 0.2 M magnesium acetate and 20% PEG 3350; and 0.15 M DL-malic acid and 20% PEG 3350, respectively. Diffraction data for all proteins was collected from single crystals at 100 K. Data for apo-PIIA and PIIA-12 were obtained at beamline ID23-2 (ESRF) at a wavelength of 0.873 Å, whereas data for the PIIA-25 and PIIA-1 were collected at beamline ID30-B (ESRF) at a wavelength of 0.967 Å. Data were processed using Xia2 (61) or XDS (62), and the structures were solved using PHASER (63) molecular replacement with LecA (PDB code 1L7L) as a search model. The models were manually rebuilt with COOT (64) and refined using PHENIX (65) and Refmac5 (66). The structures were validated using MolProbity, and all images were created using PyMOL (67).

### Molecular modeling

Docking was performed using PLANTS version 1.1 (68). The calculation of charge and energy minimization of the protein and ligand was performed with Molecular Operating Environment (MOE) version 2014.09 (Chemical Computing Group

Inc., Montreal, Quebec, Canada). Then, the standard docking procedure was used to dock D-galactosamine (**10**) and the methyl glycoside of **20** into the binding pocket of the apo-PIIA crystal structure. The docking site was limited to a 13 Å radius sphere centered in the mass center (coordination:  $X = -8.624$ ,  $Y = 15.131$ , and  $Z = 45.115$ ) of the crystallized protein. Asp-103, Asp-96, Gln-57, and Glu-44 were set as flexible residues in the input file.

### Generation of primary GTKO cells

Pigs lacking the GTKO were generated by disrupting the causative galactosyltransferase gene *GGTA1*, according to the procedure described in Klymiuk *et al.* (69). First, a bacterial artificial chromosome (BAC) containing the target region of the porcine genome, CH242-21F3, was modified by bacterial recombineering in a way that it contained a STOP box right after the START codon of *GGTA1*, resulting in the termination of protein translation as well as RNA transcription of the gene, and a floxed resistance cassette for neomycin selection. Then, this modified BAC was nucleofected into pig primary cells according to Richter *et al.* (70), and single-cell clones were generated under antibiotic selection and propagated to yield cells for DNA isolation and somatic cell nuclear transfer (SCNT). Single-cell clones were screened for homologous recombination by a quantitative PCR-based loss-of-wild-type allele approach, and cell clones that indicated a heterozygous modification of the *GGTA1* allele were used for SCNT to generate heterozygous knock-out pigs. After birth, one of the animals was sacrificed, and primary cells were cultivated and nucleofected with a plasmid encoding Cre recombinase. Again, single-cell clones were generated and now screened for the removal of the neomycin selection cassette. Another round of SCNT was performed to generate heterozygous KO animals lacking the neomycin selection cassette. Pigs were then maintained and bred to achieve homozygous GTKO pigs after two generations. Primary cells from such GTKO animals were isolated according to the procedure described by Richter *et al.* (70), and these cells were used for evaluating the specificity and sensitivity of the PIIA lectin.

### Lectin staining of porcine cells

For lectin staining,  $1 \times 10^4$  cells were seeded in 6-channel slides (IBIDI, Martinsried, Germany), coated with collagen type 1 (Serva Electrophoresis, Heidelberg, Germany), and cultivated under conventional conditions (70). When reaching a confluence of 80–100%, cells were stained for 15 min with 5  $\mu\text{g}/\text{ml}$  Hoechst 33342 and subsequently with FITC-labeled isolectin B4 (GS-IB4, Sigma, 500  $\mu\text{g}/\text{ml}$ ) or FITC-labeled PIIA (50  $\mu\text{g}/\text{ml}$ ) for 1 h at room temperature. After washing with PBS, cells were visualized in a fluorescence microscope (Axiovert 200, Zeiss).

### Hemagglutination of porcine red blood cells

Hemagglutination was done in analogy to a previously published protocol (34). Lithium-heparinized pig blood was centrifuged at  $1000 \times g$  for 5 min. Plasma was removed, and the pRBCs were washed with 45 ml of PBS three times. A 10% pRBC solution was prepared by diluting 1 ml of pRBCs with 9 ml of



PBS ( $A_{600} = 7$ ). Then, 50  $\mu$ l of PBS was added to each well of a 96-well plate. Thereafter, 50  $\mu$ l of PIIA (2.5 mg/ml) were added to the first well and mixed, and 50  $\mu$ l of this mixture was transferred to the second well. Serial dilution of PIIA was continued until 23 dilutions were obtained. 50  $\mu$ l of 10% pRBCs from WT or GTKO pig were added to each well, and the plate was incubated for 2 h at room temperature. Inhibition of PIIA-mediated WT pRBC agglutination was then tested with raffinose. A serial dilution of raffinose (20 mM) was mixed with the lowest lectin concentration showing agglutination (2.4  $\mu$ g/ml). After incubation for 30 min at room temperature, pRBCs were added. The plate was incubated for 2 h at room temperature.

**Author contributions**—A. T. conceived and coordinated the study. G. B., J. K., and A. T. wrote the paper. G. B. and A. T. designed, performed, and analyzed the experiments shown in Figs. 1–4 and Tables 1 and 2. G. B. performed docking experiments. D. H. performed chemical synthesis of compounds. S. W. designed and performed cloning of PIIA expression vector with assistance of G. B., A. S., and J. K. designed, performed, and analyzed the experiments shown in Figs. 5–8 and Table 3 with assistance of G. B., E. M. J., N. K., and E. W. designed, performed, and analyzed the experiments shown in Figs. 9 and 10. All authors reviewed the results and approved the final version of the manuscript.

**Acknowledgments**—We are grateful to Prof. Rolf Müller (HIPS Saarbrücken) for providing the *P. luminescens* strain used in this study and to the Consortium for Functional Glycomics (Core H, Protein–Glycan Interaction Resource of the Consortium for Functional Glycomics and National Institutes of Health Supporting Grant R24 GM098791) for the glycan array analysis of PIIA. We acknowledge use of the ESRF synchrotron (beamlines ID23-2 and ID30-B). GTKO pigs were produced with funding from Deutsche Forschungsgemeinschaft Grant TRR 127 “Biology of xenogeneic cell, tissue, and organ transplantation—from bench to bedside.”

**References**

1. Poinar, G. O., Thomas, G. M., and Hess, R. (1977) Characteristics of the specific bacterium associated with *Heterorhabditis bacteriophora* (Heterorhabditidae: Rhabditida). *Nematologica* **23**, 97–102
2. Thomas, G. M., and Poinar, J. R. (1979) *Xenorhabdus* gen. nov., a genus of entomopathogenic, nematophilic bacteria of the family Enterobacteriaceae. *Int. J. Syst. Evol. Microbiol.* **29**, 352–360
3. Boemare, N., Akhurst, R., and Mourant, R. (1993) DNA relatedness between *Xenorhabdus* spp. (Enterobacteriaceae), symbiotic bacteria of entomopathogenic nematodes, and a proposal to transfer *Xenorhabdus luminescens* to a new genus, *Photorhabdus* gen. nov. *Int. J. Syst. Evol. Microbiol.* **43**, 249–255
4. Forst, S., and Nealson, K. (1996) Molecular biology of the symbiotic-pathogenic bacteria *Xenorhabdus* spp., and *Photorhabdus* spp. *Microbiol. Rev.* **60**, 21–43
5. Fischer-Le Saux, M., Viallard, V., Brunel, B., Normand, P., and Boemare, N. E. (1999) Polyphasic classification of the genus *Photorhabdus* and proposal of new taxa: *P. luminescens* subsp. *luminescens* subsp. nov., *P. luminescens* subsp. *akhurstii* subsp. nov., *P. luminescens* subsp. *laumondii* subsp. nov., *P. temperata* sp. nov., *P. temperata* subsp. *temperata* subsp. nov., and *P. asymbiotica* sp. nov. *Int. J. Syst. Bacteriol.* **49**, 1645–1656
6. Ferreira, T., van Reenen, C. A., Endo, A., Tailliez, P., Pagès, S., Spröer, C., Malan, A. P., and Dicks, L. M. (2014) *Photorhabdus heterorhabditis* sp. nov., a symbiont of the entomopathogenic nematode *Heterorhabditis zealandica*. *Int. J. Syst. Evol. Microbiol.* **64**, 1540–1545

7. Hapeshi, A., and Waterfield, N. R. (2017) *Photorhabdus asymbiotica* as an insect and human pathogen. *Curr. Top. Microbiol. Immunol.* **402**, 159–177
8. Akhurst, R., and Boemare, N. (1990) in *Entomopathogenic Nematodes in Biological Control* (Gaugler, R., and Kaya, H. K., eds) pp. 75–90, CRC Press Inc., Boca Raton, FL
9. Akhurst, R., and Dunphy, G. (1993) in *Parasites and Pathogens of Insects* (Beckage, N. E., Thompson, S. A., and Federici, B. A., eds) Vol. 2, pp. 1–23, Academic Press, New York
10. Duchaud, E., Rusniok, C., Frangeul, L., Buchrieser, C., Givaudan, A., Taourit, S., Bocs, S., Boursaux-Eude, C., Chandler, M., Charles, J.-F., Dassa, E., Derose, R., Derzelle, S., Freyssinet, G., Gaudriault, S., et al. (2003) The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. *Nat. Biotechnol.* **21**, 1307–1313
11. Bowen, D., Rocheleau, T. A., Blackburn, M., Andreev, O., Golubeva, E., Bhartia, R., and French-Constant, R. H. (1998) Insecticidal toxins from the bacterium *Photorhabdus luminescens*. *Science* **280**, 2129–2132
12. Guo, L., Fatig, R. O., 3rd., Orr, G. L., Schafer, B. W., Strickland, J. A., Sukhupinda, K., Woodsworth, A. T., and Petell, J. K. (1999) *Photorhabdus luminescens* W-14 insecticidal activity consists of at least two similar but distinct proteins. Purification and characterization of toxin A and toxin B. *J. Biol. Chem.* **274**, 9836–9842
13. Kumar, A., Sýkorová, P., Demo, G., Dobeš, P., Hyršl, P., and Wimmerová, M. (2016) A novel fucose-binding lectin from *Photorhabdus luminescens* (PLL) with an unusual heptabladed  $\beta$ -propeller tetrameric structure. *J. Biol. Chem.* **291**, 25032–25049
14. Wagner, S., Sommer, R., Hinsberger, S., Lu, C., Hartmann, R. W., Empting, M., and Titz, A. (2016) Novel strategies for the treatment of *Pseudomonas aeruginosa* infections. *J. Med. Chem.* **59**, 5929–5969
15. Diggle, S. P., Stacey, R. E., Dodd, C., Cámara, M., Williams, P., and Winzer, K. (2006) The galactophilic lectin, LecA, contributes to biofilm development in *Pseudomonas aeruginosa*. *Environ. Microbiol.* **8**, 1095–1104
16. Tielker, D., Hacker, S., Loris, R., Strathmann, M., Wingender, J., Wilhelm, S., Rosenau, F., and Jaeger, K.-E. (2005) *Pseudomonas aeruginosa* lectin LecB is located in the outer membrane and is involved in biofilm formation. *Microbiology* **151**, 1313–1323
17. Garber, N. (1997) Specific Adherence Mechanisms in Microbiology and Immunology. *Nova Acta Leopoldina* 1997, 75
18. Mitchell, E., Houles, C., Sudakevitz, D., Wimmerova, M., Gautier, C., Pérez, S., Wu, A. M., Gilboa-Garber, N., and Imberty, A. (2002) Structural basis for oligosaccharide-mediated adhesion of *Pseudomonas aeruginosa* in the lungs of cystic fibrosis patients. *Nat. Struct. Biol.* **9**, 918–921
19. Sommer, R., Wagner, S., Varrot, A., Nycholat, C. M., Khaledi, A., Häussler, S., Paulson, J. C., Imberty, A., and Titz, A. (2016) The virulence factor LecB varies in clinical isolates: consequences for ligand binding and drug discovery. *Chem. Sci.* **7**, 4990–5001
20. Sudakevitz, D., Kostlánová, N., Blatman-Jan, G., Mitchell, E. P., Lerrer, B., Wimmerová, M., Katcoff, D. J., Imberty, A., and Gilboa-Garber, N. (2004) A new *Ralstonia solanacearum* high-affinity mannose-binding lectin RS-III structurally resembling the *Pseudomonas aeruginosa* fucose-specific lectin PA-III. *Mol. Microbiol.* **52**, 691–700
21. Beshr, G., Sommer, R., Hauck, D., Siebert, D. C. B., Hofmann, A., Imberty, A., and Titz, A. (2016) Development of a competitive binding assay for the *Burkholderia cenocepacia* lectin BC2L-A and structure activity relationship of natural and synthetic inhibitors. *Med. Chem. Commun.* **7**, 519–530
22. Lameignere, E., Shiao, T. C., Roy, R., Wimmerova, M., Dubreuil, F., Varrot, A., and Imberty, A. (2010) Structural basis of the affinity for oligomannosides and analogs displayed by BC2L-A, a *Burkholderia cenocepacia* soluble lectin. *Glycobiology* **20**, 87–98
23. Marchetti, R., Malinowska, L., Lameignère, E., Adamova, L., de Castro, C., Cioci, G., Stanetty, C., Kosma, P., Molinaro, A., Wimmerova, M., Imberty, A., and Silipo, A. (2012) *Burkholderia cenocepacia* lectin A binding to heptoses from the bacterial lipopolysaccharide. *Glycobiology* **22**, 1387–1398
24. Zinger-Yosovich, K., Sudakevitz, D., Imberty, A., Garber, N. C., and Gilboa-Garber, N. (2006) Production and properties of the native *Chromobacterium violaceum* fucose-binding lectin (CV-III) compared to homo-

- logous lectins of *Pseudomonas aeruginosa* (PA-III) and *Ralstonia solanacearum* (RS-III). *Microbiology* **152**, 457–463
25. Imberty, A., Wimmerová, M., Mitchell, E. P., and Gilboa-Garber, N. (2004) Structures of the lectins from *Pseudomonas aeruginosa*: insight into the molecular basis for host glycan recognition. *Microbes Infect.* **6**, 221–228
  26. McMahon, K. (2009) Structural and Functional Characterisation of Lectins from the PA-IL Superfamily. Ph.D. thesis, Dublin City University, Dublin, Ireland
  27. Joachim, I., Rikker, S., Hauck, D., Ponader, D., Boden, S., Sommer, R., Hartmann, L., and Titz, A. (2016) Development and optimization of a competitive binding assay for the galactophilic low affinity lectin LecA from *Pseudomonas aeruginosa*. *Org. Biomol. Chem.* **14**, 7933–7948
  28. Cioci, G., Mitchell, E. P., Gautier, C., Wimmerová, M., Sudakevitz, D., Pérez, S., Gilboa-Garber, N., and Imberty, A. (2003) Structural basis of calcium and galactose recognition by the lectin PA-IL of *Pseudomonas aeruginosa*. *FEBS Lett.* **555**, 297–301
  29. Galili, U. (2005) The  $\alpha$ -Gal epitope and the anti-Gal antibody in xenotransplantation and in cancer immunotherapy. *Immunol. Cell Biol.* **83**, 674–686
  30. Duffy, M. S., Morris, H. R., Dell, A., Appleton, J. A., and Haslam, S. M. (2006) Protein glycosylation in *Parelaphostrongylus tenuis*—first description of the Gal $\alpha$ 1–3Gal sequence in a nematode. *Glycobiology* **16**, 854–862
  31. Eksler, B., and Cooper, D. K. (2010) Overcoming the barriers to xenotransplantation: prospects for the future. *Expert Rev. Clin. Immunol.* **6**, 219–230
  32. Blanchard, B., Nurisso, A., Hollville, E., Tétaud, C., Wiels, J., Pokorná, M., Wimmerová, M., Varrot, A., and Imberty, A. (2008) Structural basis of the preferential binding for globo-series glycosphingolipids displayed by *Pseudomonas aeruginosa* lectin I. *J. Mol. Biol.* **383**, 837–853
  33. Eierhoff, T., Bastian, B., Thuenauer, R., Madl, J., Audfray, A., Aigal, S., Juillot, S., Rydell, G. E., Müller, S., de Bentzmann, S., Imberty, A., Fleck, C., and Römer, W. (2014) A lipid zipper triggers bacterial invasion. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 12895–12900
  34. Hauck, D., Joachim, I., Frommeyer, B., Varrot, A., Philipp, B., Möller, H. M., Imberty, A., Exner, T. E., and Titz, A. (2013) Discovery of two classes of potent glycomimetic inhibitors of *Pseudomonas aeruginosa* LecB with distinct binding modes. *ACS Chem. Biol.* **8**, 1775–1784
  35. Wohlschläger, T., Buttschi, A., Grassi, P., Sutov, G., Gauss, R., Hauck, D., Schmieder, S. S., Knobel, M., Titz, A., Dell, A., Haslam, S. M., Hengartner, M. O., Aebi, M., and Künzler, M. (2014) Methylated glycans as conserved targets of animal and fungal innate defense. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E2787–E2796
  36. Sommer, R., Hauck, D., Varrot, A., Imberty, A., Künzler, M., and Titz, A. (2016) O-Alkylated heavy atom carbohydrate probes for protein X-ray crystallography: Studies towards the synthesis of methyl 2-O-methyl-L-selenofucopyranoside. *Beilstein J. Org. Chem.* **12**, 2828–2833
  37. Chen, C. P., Song, S. C., Gilboa-Garber, N., Chang, K. S., and Wu, A. M. (1998) Studies on the binding site of the galactose-specific agglutinin PA-IL from *Pseudomonas aeruginosa*. *Glycobiology* **8**, 7–16
  38. Kim, H.-S., Cha, E., Kim, Y., Jeon, Y. H., Olson, B. H., Byun, Y., and Park, H.-D. (2016) Raffinose, a plant galactoside, inhibits *Pseudomonas aeruginosa* biofilm formation via binding to LecA and decreasing cellular cyclic diguanylate levels. *Sci. Rep.* **6**, 25318
  39. Landsteiner, K., and Levine, P. (1927) Further observations on individual differences of human blood. *Exp. Biol. Med.* **24**, 941–942
  40. Khan, F., Khan, R. H., Sherwani, A., Mohmood, S., and Azfer, M. A. (2002) Lectins as markers for blood grouping. *Med. Sci. Monit.* **8**, RA293–RA300
  41. Macher, B. A., and Galili, U. (2008) The Gal $\alpha$ 1,3Gal $\beta$ 1,4GlcNAc-R ( $\alpha$ -Gal) epitope: a carbohydrate of unique evolution and clinical relevance. *Biochim. Biophys. Acta* **1780**, 75–88
  42. Li, S., Waer, M., and Billiau, A. D. (2009) Xenotransplantation: role of natural immunity. *Transpl. Immunol.* **21**, 70–74
  43. Yamada, K., Yazawa, K., Shimizu, A., Iwanaga, T., Hisashi, Y., Nuhn, M., O'Malley, P., Nobori, S., Vagefi, P. A., Patience, C., Fishman, J., Cooper, D. K., Hawley, R. J., Greenstein, J., Schuurman, H.-J., et al. (2005) Marked prolongation of porcine renal xenograft survival in baboons through the use of  $\alpha$ 1,3-galactosyltransferase gene-knockout donors and the cotransplantation of vascularized thymic tissue. *Nat. Med.* **11**, 32–34
  44. Klymiuk, N., Aigner, B., Brem, G., and Wolf, E. (2010) Genetic modification of pigs as organ donors for xenotransplantation. *Mol. Reprod. Dev.* **77**, 209–221
  45. Puga Yung, G. L., Li, Y., Borsig, L., Millard, A.-L., Karpova, M. B., Zhou, D., and Seebach, J. D. (2012) Complete absence of the  $\alpha$ Gal xenoantigen and isoglobotrihexosylceramide in  $\alpha$ 1,3galactosyltransferase knock-out pigs. *Xenotransplantation* **19**, 196–206
  46. Tempel, W., Tschampel, S., and Woods, R. J. (2002) The xenograft antigen bound to *Griffonia simplicifolia* lectin 1-B(4). X-ray crystal structure of the complex and molecular dynamics characterization of the binding site. *J. Biol. Chem.* **277**, 6615–6621
  47. Obukhova, P., Rieben, R., and Bovin, N. (2007) Normal human serum contains high levels of anti-Gal  $\alpha$ 1–4GlcNAc antibodies. *Xenotransplantation* **14**, 627–635
  48. Gerdt, S., Dennis, R. D., Borgonie, G., Schnabel, R., and Geyer, R. (1999) Isolation, characterization and immunolocalization of phosphorylcholine-substituted glycolipids in developmental stages of *Caenorhabditis elegans*. *Eur. J. Biochem.* **266**, 952–963
  49. van Stijn, C. M., van den Broek, M., Vervelde, L., Alvarez, R. A., Cummings, R. D., Tefsen, B., and van Die, I. (2010) Vaccination-induced IgG response to Gal $\alpha$ 1–3GalNAc glycan epitopes in lambs protected against *Haemonchus contortus* challenge infection. *Int. J. Parasitol.* **40**, 215–222
  50. Yan, S., Jin, C., Wilson, I. B., and Paschinger, K. (2015) Comparisons of *Caenorhabditis* fucosyltransferase mutants reveal a multiplicity of isomeric N-glycan structures. *J. Proteome Res.* **14**, 5291–5305
  51. Yan, S., Brecker, L., Jin, C., Titz, A., Dragosits, M., Karlsson, N. G., Jantsch, V., Wilson, I. B., and Paschinger, K. (2015) Bisecting galactose as a feature of N-glycans of wild-type and mutant *Caenorhabditis elegans*. *Mol. Cell. Proteomics* **14**, 2111–2125
  52. Yan, S., Bleuler-Martinez, S., Plaza, D. F., Künzler, M., Aebi, M., Joachim, A., Razzazi-Fazeli, E., Jantsch, V., Geyer, R., Wilson, I. B., and Paschinger, K. (2012) Galactosylated fucose epitopes in nematodes: increased expression in a *Caenorhabditis* mutant associated with altered lectin sensitivity and occurrence in parasitic species. *J. Biol. Chem.* **287**, 28276–28290
  53. Meldal, B. H., Debenham, N. J., De Ley, P., De Ley, I. T., Vanfleteren, J. R., Vierstraete, A. R., Bert, W., Borgonie, G., Moens, T., Tyler, P. A., Austen, M. C., Blaxter, M. L., Rogers, A. D., and Lamshead, P. J. (2007) An improved molecular phylogeny of the Nematoda with special emphasis on marine taxa. *Mol. Phylogenet. Evol.* **42**, 622–636
  54. Dennis, R. D., Geyer, R., Egge, H., Menges, H., Stirn, S., and Wiegandt, H. (1985) Glycosphingolipids in insects. Chemical structures of ceramide monosaccharide, disaccharide, and trisaccharide from pupae of *Calliphora vicina* (Insecta: Diptera). *Eur. J. Biochem.* **146**, 51–58
  55. Papadopoulos, J. S., and Agarwala, R. (2007) COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* **23**, 1073–1079
  56. Wicker, N., Perrin, G. R., Thierry, J. C., and Poch, O. (2001) Secator: a program for inferring protein subfamilies from phylogenetic trees. *Mol. Biol. Evol.* **18**, 1435–1441
  57. Gascuel, O. (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* **14**, 685–695
  58. Stothard, P. (2000) The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *BioTechniques* **28**, 1102
  59. Fornstedt, N., and Porath, J. (1975) Characterization studies on a new lectin found in seeds of *Vicia ervilia*. *FEBS Lett.* **57**, 187–191
  60. Blixt, O., Head, S., Mondala, T., Scanlan, C., Huflejt, M. E., Alvarez, R., Bryan, M. C., Fazio, F., Calarese, D., Stevens, J., Razi, N., Stevens, D. J., Skehel, J. J., van Die, I., Burton, D. R., Wilson, I. A., Cummings, R., Bovin, N., Wong, C.-H., and Paulson, J. C. (2004) Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 17033–17038
  61. Winter, G. (2010) xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190
  62. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132

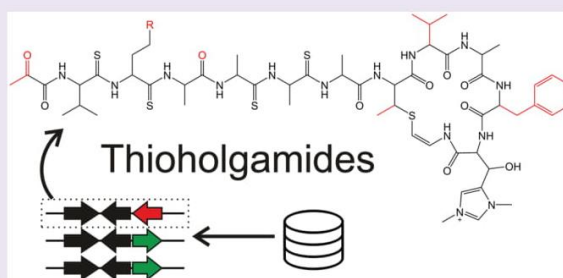
63. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr* **40**, 658–674
64. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol. Crystallogr.* **66**, 486–501
65. Adams, P. D., Afonine, P. V., Bunkóczy, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol. Crystallogr.* **66**, 213–221
66. Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol. Crystallogr.* **67**, 355–366
67. Schrödinger L. (2015) *The PyMOL Molecular Graphics System*, Version 1.8, New York
68. Korb, O., Stütze, T., and Exner, T. E. (2006) PLANTS: application of ant colony optimization to structure-based drug design. *Lecture Notes in Computer Science* **4150**, 247–258
69. Klymiuk, N., Mundhenk, L., Kraehe, K., Wuensch, A., Plog, S., Emrich, D., Langenmayer, M. C., Stehr, M., Holzinger, A., Kröner, C., Richter, A., Kessler, B., Kurome, M., Eddicks, M., Nagashima, H., *et al.* (2012) Sequential targeting of CFTR by BAC vectors generates a novel pig model of cystic fibrosis. *J. Mol. Med.* **90**, 597–608
70. Richter, A., Kurome, M., Kessler, B., Zakhartchenko, V., Klymiuk, N., Nagashima, H., Wolf, E., and Wuensch, A. (2012) Potential of primary kidney cells for somatic cell nuclear transfer mediated transgenesis in pig. *BMC Biotechnol.* **12**, 84

## Thioholgamides: Thioamide-Containing Cytotoxic RiPP Natural Products

Louise Kjaerulff,<sup>†</sup> Asfandiyar Sikandar,<sup>‡</sup> Nestor Zaburanyi,<sup>†,§</sup> Sebastian Adam,<sup>‡</sup> Jennifer Herrmann,<sup>†,§</sup> Jesko Koehnke,<sup>‡,Ⓢ</sup> and Rolf Müller<sup>\*,†,§,Ⓢ</sup><sup>†</sup>Department of Microbial Natural Products, <sup>‡</sup>Structural Biology of Biosynthetic Enzymes, Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research and Pharmaceutical Biotechnology at Saarland University, Saarland University Campus, Building E8.1, 66123 Saarbrücken, Germany<sup>§</sup>German Centre for Infection Research (DZIF), Partner Site Hannover-Braunschweig, 38124, Braunschweig, Germany

## Supporting Information

**ABSTRACT:** Thioviridamide is a structurally unique ribosomally synthesized and post-translationally modified peptide that contains several thioamide bonds and is active against a number of cancer cell lines. In the search for naturally occurring thioviridamide analogs, we employed genome mining that led to the identification of several related gene clusters. Chemical screening followed by cultivation and isolation yielded thioholgamides A and B, two new additions to the thioviridamide family with several amino acid substitutions, a different N-capping moiety, and with one less thioamide bond. Thioholgamides display improved cytotoxicity in the submicromolar range against a range of cell lines and an  $IC_{50}$  of 30 nM for thioholgamide A against HCT-116 cells. Herein, we report the isolation and structural elucidation of thioholgamides A and B, a proposed biosynthetic cluster for their production, and their bioactivities against a larger panel of microorganisms and cancer cell lines.



Cancer is among the leading causes of death worldwide. The number of new cases is steadily increasing, and it is expected to continue to do so in the coming years.<sup>1</sup> Therefore, there is an omnipresent need to develop new chemotherapeutics, preferably with fewer side effects and high target specificity.

As much as 75% of anticancer drugs and 69% of anti-infectives are derived from or inspired by natural products, which remain one of the most promising sources of drug lead molecules due to their inherently privileged structures.<sup>2,3</sup> Microorganisms are the most practical producers of secondary metabolites for drug development because they are relatively easy to grow and upscale. In particular, the Gram-positive streptomycetes are a major source of antibiotics used in the clinic (e.g., chloramphenicol, fosfomicin, daptomycin, and streptomycin),<sup>4,5</sup> but they also produce secondary metabolites with other activities,<sup>6</sup> including the immunosuppressant rapamycin<sup>7–9</sup> and the chemotherapeutics doxorubicin<sup>10</sup> and dactinomycin.<sup>11</sup> Despite many years of mining for biologically active molecules from *Streptomyces* species, there is still much potential hidden under this genus.<sup>12</sup>

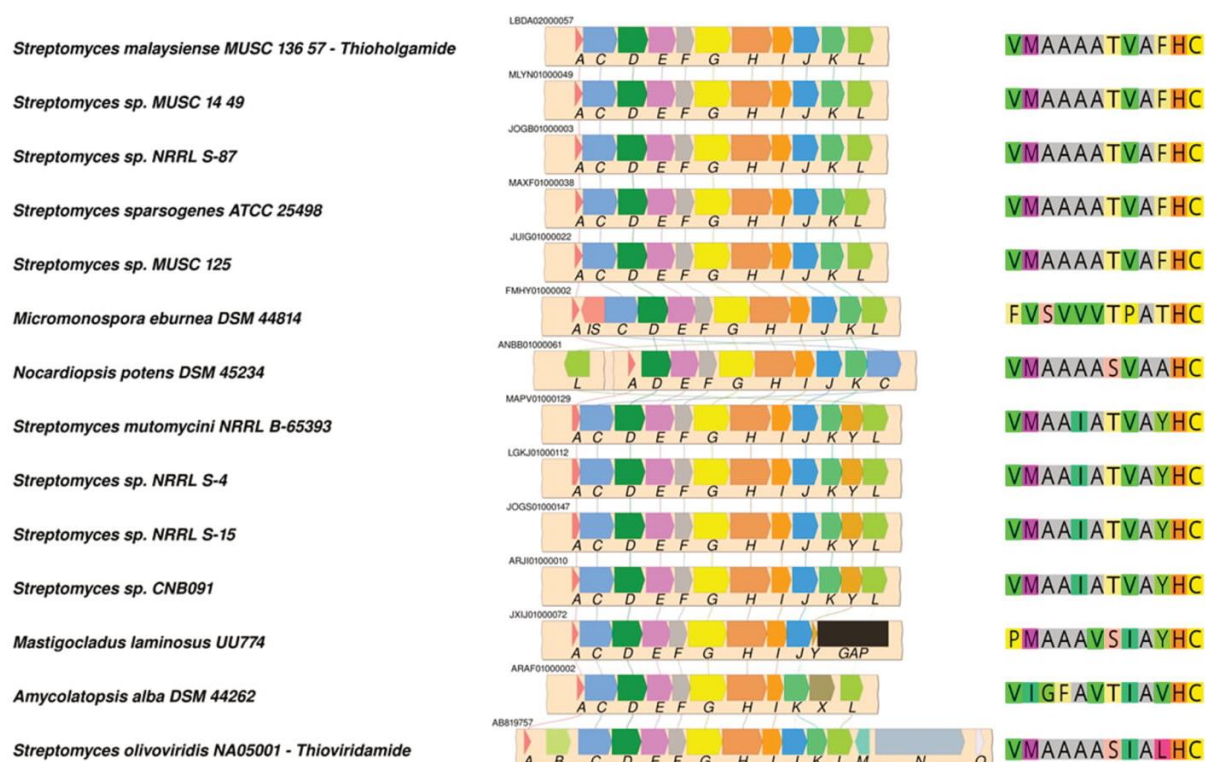
Ribosomally synthesized and post-translationally modified peptides (RiPPs) are a relatively recent addition to the major classes of natural products, including the well-known alkaloids, polyketides, nonribosomal peptides, and terpenoids.<sup>13,14</sup> By modifications of a precursor peptide, the RiPP biosynthetic

machinery can produce various intriguing compound families, from the lantibiotics to the thiopeptides (or thiazolyl peptides) containing thiazoles and other heterocycles. Two articles from 2006 by Hayakawa et al. described the isolation and characterization of thioviridamide from *Streptomyces olivoviridis*,<sup>15,16</sup> a RiPP with five thioamide bonds and a permanent positive charge. In 2013, the biosynthetic gene cluster was identified,<sup>17</sup> and in 2015, a derivative was identified (resulting from heterologous expression) with lactic acid as an N-terminal modification.<sup>18</sup> By definition, the thiopeptide RiPP family does not include any peptides with thioamide bonds (containing N–C=S, like thioviridamide), which are extremely rare in natural products. To date, only three compound families of bacterial origin contain this type of thioamide (N–C=S) bond: thioviridamide and its analog JBIR-140, the methanobactins (also RiPPs),<sup>19</sup> and closthioamide.<sup>20</sup> In addition, the small molecule cycasthioamide from the plant *Cycas revolute* contains a single thioamide bond.<sup>21</sup> Thioamides generally have higher stability towards hydrolysis than standard amides, and isosteric replacement has been a tool for medicinal chemists to change degradation and ADME properties of biologically active small molecules.<sup>22</sup> In some cases, this can lead to new or enhanced

Received: August 7, 2017

Accepted: October 5, 2017

Published: October 5, 2017



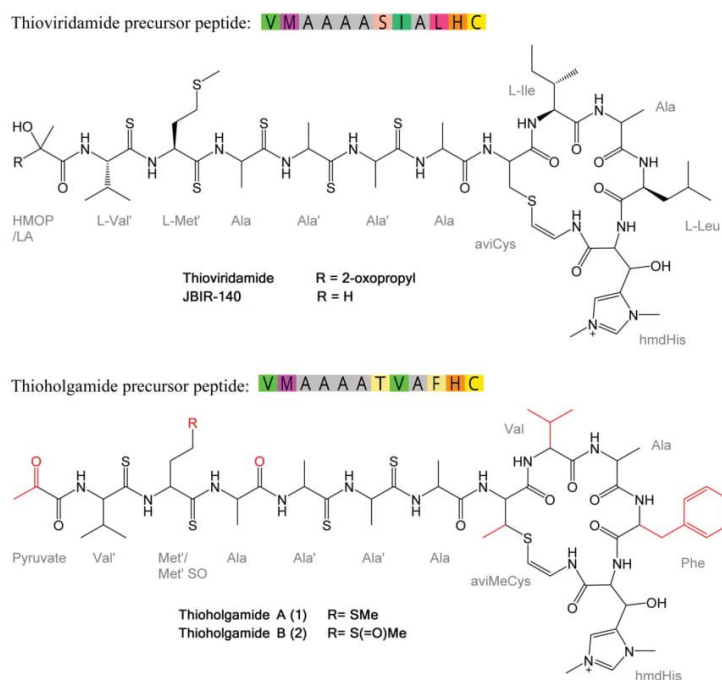
**Figure 1.** Proposed thioholgamide biosynthetic gene cluster (top) from *Streptomyces malaysiense* MUSC 136 57 compared to homologous gene clusters from other organisms by the SimpleSynteny program. A, precursor peptide; B, putative SARP family regulator; C, hypothetical protein; D, hypothetical protein; E, hypothetical protein; F, putative flavoprotein decarboxylase; G, putative methyltransferase; H, hypothetical protein; I, TfuA-like core domain-containing protein; J, putative phytyl-CoA dioxygenase family protein; K, putative proteinase; L, hypothetical protein; M, putative regulatory protein; N, putative LuxR family transcriptional regulator; O, hypothetical protein; X, putative SDR family oxidoreductase; Y, putative methyltransferase; IS, putative ISS/IS1182 family transposase; GAP, sequencing gap. The peptide sequence encoded in the precursor peptide is shown on the right.

bioactivity,<sup>19</sup> but it can also completely abolish the activity, as was seen for closthioamide when its hexaoox analog closamide was synthesized and found inactive.<sup>20</sup>

## RESULTS AND DISCUSSION

Fascinated by the chemistry, biosynthesis, and bioactivity of thioviridamide-like molecules, we set out to identify novel members of this compound class by genome mining. Searching publicly available genome sequences of actinomycetes and other microbes for homologous biosynthetic gene clusters (see the Supporting Information), we identified 13 genomes (Figure 1) with loci that closely resemble the biosynthetic gene cluster published for thioviridamide.<sup>17</sup> All genome sequences but one belonged to Actinobacteria (*Streptomyces* spp., *Amycolatopsis alba*, *Micromonospora eburnea*, and *Nocardioopsis potens*), and one strain was a cyanobacterium (*Mastigocladus laminosus*). The newly found biosynthetic gene clusters showed varying gene synteny. Of all biosynthetic gene clusters identified, five had nearly identical gene compositions. The others showed varying degrees of gene rearrangements, insertions, and deletions (Figure 1). Homologs of the core thioviridamide biosynthetic proteins TvaA, TvaC, TvaD, TvaE, TvaF, TvaG, TvaH, and TvaI were found in all 13 genomic loci, while homologs of TvaJ, TvaK, and TvaL were located in 12 of 13 loci. Genes *tvaB*, *tvaM*, *tvaN*, and *tvaO* had no homologs in examined

sequences, suggesting that they are either not involved, not crucial for biosynthesis, or are located elsewhere in the respective genomes. Of the examined biosynthetic loci, several showed the existence of additional proteins that were not present in the query sequences but nevertheless might be involved in the biosynthesis of the respective compounds: A putative SDR family oxidoreductase (designated as X-homolog) and a putative methyltransferase (designated as Y-homolog) (Figure 1). The variability of biosynthetic loci prompted us to obtain and cultivate four strains (DSM 100712, DSM 45234, DSM 44262, and NRRL S-87). Subsequent LCMS analyses of the secondary metabolite profiles led us to identify *Streptomyces malaysiense* MUSC 136 57 (DSM 100712) as a prolific producer of a series of thioviridamide-like compounds, even though pairwise sequence identity to thioviridamide biosynthetic genes was merely between 23 and 53%. These compounds with masses in the same range as thioviridamide (~1,300 Da) also contained a similarly high number of sulfur atoms, as seen from the isotopic patterns, and fractions containing these compounds showed strong cytotoxicity against the human cervix carcinoma cell line KB-3.1. Upscaling, isolation, and analysis by LCMS and NMR spectroscopy revealed new peptides in the thioviridamide family with structural changes in N-terminal modification, amino acid sequence, and the number of thioamide bonds. Extraction and fractionation of combined 150 mL cultures amounting to a total



**Figure 2.** Structures of thioviridamide, JBIR-140, and thioholgamides A (1) and B (2). The methionine in 1 is oxidized to methionine sulfoxide in 2. Differences in the thioholgamides compared to thioviridamide are highlighted in red.

cultivation volume of 6 L allowed the purification and structural characterization of thioholgamides A (1, 11.3 mg) and B (2, 4.3 mg) (see details in the Supporting Information), whereas the putative thioholgamides C and D were merely observed by LCMS and partially characterized based on MS<sup>2</sup> fragmentation patterns.

HRESIMS analysis of 1 displayed the molecular ion [M]<sup>+</sup> at *m/z* 1305.4901, consistent with the molecular formula C<sub>56</sub>H<sub>85</sub>N<sub>14</sub>O<sub>10</sub>S<sub>6</sub><sup>+</sup> ( $\Delta = 0.7$  ppm), which underlined the connection to thioviridamide (exact mass 1345.5244 and formula C<sub>56</sub>H<sub>93</sub>N<sub>14</sub>O<sub>10</sub>S<sub>7</sub><sup>+</sup>)<sup>16</sup> but also suggested that there must be some difference between the structures based on the number of hydrogen and sulfur atoms. This was confirmed by analysis of the 2D NMR data (DQF-COSY, HSQC, HMBC, and ROESY) of 1 (see NMR tables in the Supporting Information). Whereas NMR data for thioviridamide and JBIR-140 were acquired in methanol-*d*<sub>4</sub>,<sup>16,18</sup> 1 displayed limited solubility and broad resonances in and close to the backbone of the macrocyclic ring system. Data in DMSO-*d*<sub>6</sub> showed less structural flexibility, but the best results were obtained in D<sub>2</sub>O (both data sets are reported in the Supporting Information). Proton spin systems of the individual amino acids of 1 were connected via DQF-COSY correlations, and <sup>13</sup>C connectivities were determined from the HSQC spectrum. Spin systems were then connected via HMBC and ROESY NMR data, and it was soon clear that the alanine residue next to methionine had a standard peptide bond (173 ppm) and not a thioamide bond as for thioviridamide. Key correlations supporting this claim are given in detail in the Supporting Information. The remaining sulfur atoms were placed as in thioviridamide, as seen from the four <sup>13</sup>C=S resonances around 205 ppm, the methionine residue (e.g., singlet from the -S-Me at 2.07 and 14.9 ppm), and the cysteine-based bridge to Thr giving a [(*Z*)-2-aminovinyl]-3-

methylcysteine (aviMeCys) moiety to form the macrocyclic ring. From the gene sequences coding for the precursor peptide (VMAAATVAFHC), two further amino acids in the macrocyclic ring were expected to be exchanged (from isoleucine to leucine and from leucine to phenylalanine), and this was corroborated by the NMR data. Furthermore, the N-terminal modification of the peptide was identified as pyruvate, as seen from an acetate-like methyl group (2.42 and 25.0 ppm) with correlations to two carbonyls at 198.3 and 162.6 ppm (the latter also correlating to H<sub>α</sub> of Met). Altogether, this resulted in the structure of 1 as seen below (Figure 2).

HRESIMS analysis of 2 displayed a molecular ion at *m/z* 1321.4862 ([M]<sup>+</sup>), consistent with the molecular formula C<sub>56</sub>H<sub>85</sub>N<sub>14</sub>O<sub>11</sub>S<sub>6</sub><sup>+</sup> ( $\Delta = 1.2$  ppm), and thus possessed just one additional oxygen compared to 1. This was found in the methionine residue, where oxidation to methionine sulfoxide resulted in higher chemical shifts of the adjacent C<sub>γ</sub> and C<sub>ε</sub> resonances (Me here at 2.68 and 37.2 ppm). This sulfoxide analog was observed in the raw extract of *Streptomyces malaysiense* MUSC 136 57 (DSM 100712) and therefore we expect it to be a genuine secondary metabolite and not a product of the workup procedure.

Two further analogs at 1289.4963 (C<sub>56</sub>H<sub>85</sub>N<sub>14</sub>O<sub>9</sub>S<sub>6</sub><sup>+</sup>,  $\Delta = 0.9$  ppm) and *m/z* 1307.5077 (C<sub>56</sub>H<sub>87</sub>N<sub>14</sub>O<sub>10</sub>S<sub>6</sub><sup>+</sup>,  $\Delta = 1.3$  ppm) contain one less oxygen and one less double bond equivalent, respectively. On the basis of MS<sup>2</sup> fragmentation, the missing oxygen stems from loss of the hydroxy group in hdmHis, and the compound with *m/z* 1307 is believed to be an analog with a different N-capping group, as MS<sup>2</sup> fragmentation for the molecule is identical from the macrocycle to (excl.) the methionine moiety (see MS<sup>2</sup> spectra and interpretation in the Supporting Information).

Amino acid analysis of **1** was attempted by Marfey's method and showed the presence of L-Val, L-Phe, and both L- and D-Ala. The methionine appeared as two very small peaks of both configurations, which, along with the relative amounts observed for Ala, Val, and Phe, suggests that the thioamide bonds are not readily hydrolyzed, even when heated for prolonged periods in hydrochloric acid.

Crude extracts of *Streptomyces malaysiense* MUSC 136 57 displayed strong cytotoxic activity against KB-3.1 cells and additionally inhibited the growth of some bacteria such as *Staphylococcus aureus*. We therefore assessed the biological activity of thioholgamides A and B with a panel of cell lines and microorganisms. Thioholgamide A showed markedly increased activity compared to thioviridamide and JBIR-140 (both being active against cancer cell lines in the two-digit micromolar range<sup>18</sup>), with IC<sub>50</sub>s against human SW480 colon carcinoma, Jurkat acute T cell leukemia, and KB-3.1 cervix carcinoma cell lines of 0.11, 0.53, and 0.54 μM, respectively. Human SKOV-3 ovarian adenocarcinoma and U937 histiocytic lymphoma cells were somewhat less sensitive. The most promising results were obtained with the human colon carcinoma cell line HCT-116 with IC<sub>50</sub>s for thioholgamides A and B of 30 and 510 nM, respectively (Table 1). In most assays, thioholgamide B displayed approximately 10-fold lower activity compared to thioholgamide A.

**Table 1. Cytotoxic Activity of Thioholgamides A (1) and B (2)**

cell line	IC <sub>50</sub> [μM]	
	1	2
HCT-116	0.03	0.51
Jurkat	0.53	5.28
KB-3.1	0.54	4.49
SKOV-3	2.49	20.89
SW480	0.11	12.17
U937	2.19	16.94

Besides their submicromolar activity against different cancer cell lines, thioholgamides A and B showed only moderate inhibitory activity against Gram-positive bacteria at 4–32 μg mL<sup>-1</sup> and were particularly active against *Mycobacterium smegmatis* (minimum inhibitory concentration, MIC 1–2 μg mL<sup>-1</sup>). The compounds displayed no activity against Gram-negative bacteria and eukaryotic microorganisms (Table 2). In contrast to their activity against human cancer cell lines, both compounds were almost equipotent in antibacterial testing, which might hint towards a different mechanism of action in bacteria and cancer cells.

In summary, the thioholgamides were described as new members of a RiPP family of thiopeptides also containing thioviridamide and JBIR-140,<sup>15,18</sup> with multiple structural differences and enhanced cytotoxic activities. In addition to the difference in their amino acid makeup, which is the direct result of precursor peptide sequence variations, the additional structural differences are likely the result of the divergent enzymes involved in biosynthesis. On the basis of the constructs used for the production of thioviridamides in a heterologous host, it has been suggested that genes *tvaA*–*tvaO* are necessary for thioviridamide production. Since we observe the same types of chemical modifications in thioholgamides, we propose that genes “A” to “L” are sufficient. A list of their pairwise sequence identities on a protein level and putative

**Table 2. Minimum Inhibitory Concentrations (MIC) of Thioholgamides A (1) and B (2) against a Panel of Microorganisms**

indicator strain	MIC [μg/mL]	
	1	2
<i>Bacillus subtilis</i> DSM 10	4	4
<i>Micrococcus luteus</i> DSM 1790	4–8	4–8
<i>Mycobacterium smegmatis</i> mc <sup>2</sup> 155	1–2	1–2
<i>Staphylococcus aureus</i> Newman	8	32
<i>Chromobacterium violaceum</i> DSM 30191	>64	>64
<i>Escherichia coli</i> DSM 1116	>64	>64
<i>Escherichia coli</i> (ToIC-deficient)	>64	>64
<i>Pseudomonas aeruginosa</i> PA14	>64	>64
<i>Candida albicans</i> DSM 1665	>64	>64
<i>Mucor hiemalis</i> DSM 2656	>64	>64
<i>Wickerhamomyces anomalous</i> DSM 6766	>64	>64

functions can be found in Table S7.2. We did not find homologs of *tvaM*, *tvaN*, or *tvaO* in the genome of the thioholgamide producer. Their absence in all other gene clusters, which could be identified as potential producers of thioviridamide-like compounds, is a further indicator that they are not required for biosynthesis (Figure 1). Thioholgamide A was approximately 10-fold more active than thioholgamide B, thioviridamide, and JBIR-140.<sup>18</sup> Further characterization of their biological potential will include more comprehensive profiling with a larger number of cell lines of different origin. Studies toward understanding the molecular basis of thioholgamide are underway. The sequence of the proposed thioholgamide biosynthetic gene cluster has been deposited in the GenBank database under accession ID MF593843.

## METHODS

General experimental procedures as well as cultivation and isolation protocols are available in the Supporting Information. Biological assay methods as previously described.<sup>23</sup>

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscchembio.7b00676.

Experimental procedures; strains and cultivation; isolation of thioholgamides A (1) and B (2); NMR tables and MS<sup>2</sup>; Marfey's amino acid analysis; NMR spectra; genome mining (PDF)

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: Rolf.Mueller@helmholtz-hzi.de.

### ORCID

Jesko Koehnke: 0000-0002-7153-1365

Rolf Müller: 0000-0002-1042-5665

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

R. Müller would like to acknowledge funding from DFG Forschergruppe FOR 1406. J. Koehnke would like to thank the Deutsche Forschungsgemeinschaft for an Emmy Noether fellowship (KO4116/3-1).

## ■ REFERENCES

- (1) Stewart, B. W., and Wild, C., Eds. (2014) *World Cancer Report 2014*, International Agency for Research on Cancer, Lyon, France.
- (2) Newman, D. J. (2008) Natural products as leads to potential drugs: an old process or the new hope for drug discovery? *J. Med. Chem.* *51*, 2589–2599.
- (3) Cragg, G. M., and Newman, D. J. (2013) Natural products: a continuing source of novel drug leads. *Biochim. Biophys. Acta, Gen. Subj.* *1830*, 3670–3695.
- (4) Cassir, N., Rolain, J.-M., and Brouqui, P. (2014) A new strategy to fight antimicrobial resistance: the revival of old antibiotics. *Front. Microbiol.* *5*, 551.
- (5) Lucas, X., Senger, C., Erxleben, A., Gruning, B. A., Doring, K., Mosch, J., Flemming, S., and Gunther, S. (2013) StreptomeDB: a resource for natural compounds isolated from *Streptomyces* species. *Nucleic Acids Res.* *41*, D1130–D1136.
- (6) Weber, I., Welzel, K., Pelzer, S., Vente, A., and Wohlleben, W. (2003) Exploiting the genetic potential of polyketide producing streptomycetes. *J. Biotechnol.* *106*, 221–232.
- (7) Vézina, C., Kudelski, A., and Sehgal, S. N. (1975) Rapamycin (AY-22,989), a new antifungal antibiotic. I. Taxonomy of the producing streptomycete and isolation of the active principle. *J. Antibiot.* *28*, 721–726.
- (8) Sehgal, S. N., Baker, H., and Vézina, C. (1975) Rapamycin (AY-22,989), a new antifungal antibiotic. II. Fermentation, isolation and characterization. *J. Antibiot.* *28*, 727–732.
- (9) Swindells, D. N., White, P. S., and Findlay, J. A. (1978) The X-ray crystal structure of rapamycin, C<sub>51</sub>H<sub>79</sub>NO<sub>13</sub>. *Can. J. Chem.* *56*, 2491–2492.
- (10) Tacar, O., Sriamornsak, P., and Dass, C. R. (2013) Doxorubicin: an update on anticancer molecular action, toxicity and novel drug delivery systems. *J. Pharm. Pharmacol.* *65*, 157–170.
- (11) Waksman, S. A., and Woodruff, H. B. (1940) Bacteriostatic and Bactericidal Substances Produced by a Soil Actinomycetes. *Exp. Biol. Med.* *45*, 609–614.
- (12) Ziemert, N., Alanjary, M., and Weber, T. (2016) The evolution of genome mining in microbes - a review. *Nat. Prod. Rep.* *33*, 988–1005.
- (13) Amison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., Camarero, J. A., Campopiano, D. J., Challis, G. L., Clardy, J., Cotter, P. D., Craik, D. J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P. C., Entian, K.-D. D., Fischbach, M. A., Garavelli, J. S., Göransson, U., Gruber, C. W., Haft, D. H., Hemscheidt, T. K., Hertweck, C., Hill, C., Horswill, A. R., Jaspars, M., Kelly, W. L., Klinman, J. P., Kuipers, O. P., Link, A. J., Liu, W., Marahiel, M. A., Mitchell, D. A., Moll, G. N., Moore, B. S., Müller, R., Nair, S. K., Nes, I. F., Norris, G. E., Olivera, B. M., Onaka, H., Patchett, M. L., Piel, J., Reaney, M. J. T., Rebuffat, S., Ross, R. P., Sahl, H.-G. G., Schmidt, E. W., Selsted, M. E., Severinov, K., Shen, B., Sivonen, K., Smith, L., Stein, T., Süßmuth, R. E., Tagg, J. R., Tang, G. L., Truman, A. W., Vederas, J. C., Walsh, C. T., Walton, J. D., Wenzel, S. C., Willey, J. M., and van der Donk, W. (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* *30*, 108–160.
- (14) Ortega, M. A., and van der Donk, W. A. (2016) New Insights into the Biosynthetic Logic of Ribosomally Synthesized and Post-translationally Modified Peptide Natural Products. *Cell Chem. Biol.* *23*, 31–44.
- (15) Hayakawa, Y., Sasaki, K., Adachi, H., Furihata, K., Nagai, K., and Shin-Ya, K. (2006) Thioviridamide, a Novel Apoptosis Inducer in Transformed Cells from *Streptomyces olivoviridis*. *J. Antibiot.* *59*, 1–5.
- (16) Hayakawa, Y., Sasaki, K., Nagai, K., Shin-Ya, K., and Furihata, K. (2006) Structure of Thioviridamide, a Novel Apoptosis Inducer from *Streptomyces olivoviridis*. *J. Antibiot.* *59*, 6–10.
- (17) Izawa, M., Kawasaki, T., and Hayakawa, Y. (2013) Cloning and Heterologous Expression of the Thioviridamide Biosynthesis Gene Cluster from *Streptomyces olivoviridis*. *Appl. Environ. Microbiol.* *79*, 7110–7113.
- (18) Izumikawa, M., Kozono, I., Hashimoto, J., Kagaya, N., Takagi, M., Koiwai, H., Komatsu, M., Fujie, M., Satoh, N., Ikeda, H., and Shin-Ya, K. (2015) Novel thioviridamide derivative? JBIR-140: heterologous expression of the gene cluster for thioviridamide biosynthesis. *J. Antibiot.* *68*, 533–536.
- (19) Banala, S., and Süßmuth, R. D. (2010) Thioamides in nature: in search of secondary metabolites in anaerobic microorganisms. *ChemBioChem* *11*, 1335–1337.
- (20) Lincke, T., Behnken, S., Ishida, K., Roth, M., and Hertweck, C. (2010) Closthioamide: an unprecedented polythioamide antibiotic from the strictly anaerobic bacterium *Clostridium cellulolyticum*. *Angew. Chem., Int. Ed.* *49*, 2011–2013.
- (21) Pan, M., Mabry, T. J., Beale, J. M., and Mamiya, B. M. (1997) Nonprotein amino acids from *Cycas revoluta*. *Phytochemistry* *45*, 517–519.
- (22) Reiner, A., Wildemann, D., Fischer, G., and Kiefhaber, T. (2008) Effect of thiopeptide bonds on alpha-helix structure and stability. *J. Am. Chem. Soc.* *130*, 8079–8084.
- (23) Kjaerulff, L., Raju, R., Panter, F., Scheid, U., Garcia, R., Herrmann, J., and Müller, R. (2017) Pyxipyrrolones: Structure elucidation and biosynthesis of cytotoxic myxobacterial metabolites. Novel cytotoxic myxobacterial metabolites. Structure elucidation and biosynthesis proposal. *Angew. Chem., Int. Ed.* *56*, 9614–9618.



# Carolacton inhibits folate-dependent C1 metabolism



## ARTICLE

DOI: 10.1038/s41467-017-01671-5

OPEN

## The natural product carolacton inhibits folate-dependent C1 metabolism by targeting FoLD/MTHFD

Chengzhang Fu<sup>1</sup>, Asfandyar Sikandar<sup>2</sup>, Jannik Donner <sup>3</sup>, Nestor Zaburannyi<sup>1</sup>, Jennifer Herrmann<sup>1,4</sup>, Michael Reck<sup>3</sup>, Irene Wagner-Döbler<sup>3</sup>, Jesko Koehnke <sup>2</sup> & Rolf Müller<sup>1,4</sup>

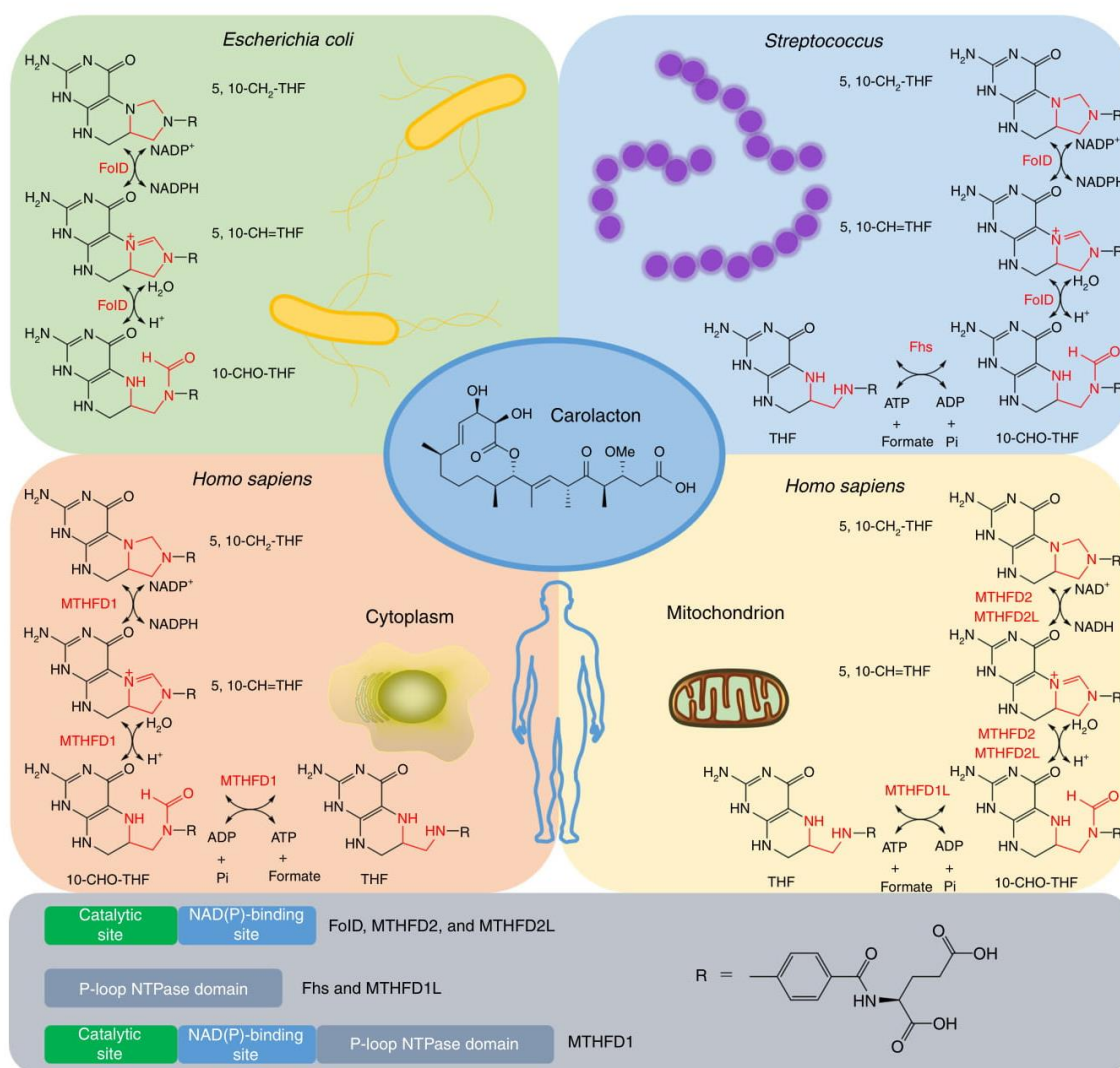
The natural product carolacton is a macrolide keto-carboxylic acid produced by the myxobacterium *Sorangium cellulosum*, and was originally described as an antibacterial compound. Here we show that carolacton targets FoLD, a key enzyme from the folate-dependent C1 metabolism. We characterize the interaction between bacterial FoLD and carolacton biophysically, structurally and biochemically. Carolacton binds FoLD with nanomolar affinity, and the crystal structure of the FoLD–carolacton complex reveals the mode of binding. We show that the human FoLD orthologs, MTHFD1 and MTHFD2, are also inhibited in the low nM range, and that micromolar concentrations of carolacton inhibit the growth of cancer cell lines. As mitochondrial MTHFD2 is known to be upregulated in cancer cells, it may be possible to use carolacton as an inhibitor tool compound to assess MTHFD2 as an anti-cancer target.

<sup>1</sup>Department of Microbial Natural Products, Helmholtz Centre for Infection Research and Department of Pharmaceutical Biotechnology, Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Saarland University, Campus E8.1, 66123 Saarbrücken, Germany. <sup>2</sup>Helmholtz Institute for Pharmaceutical Research Saarland, Workgroup Structural Biology of Biosynthetic Enzymes, Helmholtz Centre for Infection Research, Saarland University, Campus E8.1, 66123 Saarbrücken, Germany. <sup>3</sup>Helmholtz Center for Infection Research (HZI), Group Microbial Communication, 38124 Braunschweig, Germany. <sup>4</sup>German Centre for Infection Research (DZIF), Partner Site Hannover, 38124 Braunschweig, Germany. Chengzhang Fu, Asfandyar Sikandar and Jannik Donner contributed equally to this work. Correspondence and requests for materials should be addressed to I.W-Döb. (email: [Irene.wagnerdoebler@helmholtz-hzi.de](mailto:Irene.wagnerdoebler@helmholtz-hzi.de)) or to J.K. (email: [jesko.koehnke@helmholtz-hzi.de](mailto:jesko.koehnke@helmholtz-hzi.de)) or to R.Mül. (email: [rolf.mueller@helmholtz-hzi.de](mailto:rolf.mueller@helmholtz-hzi.de))

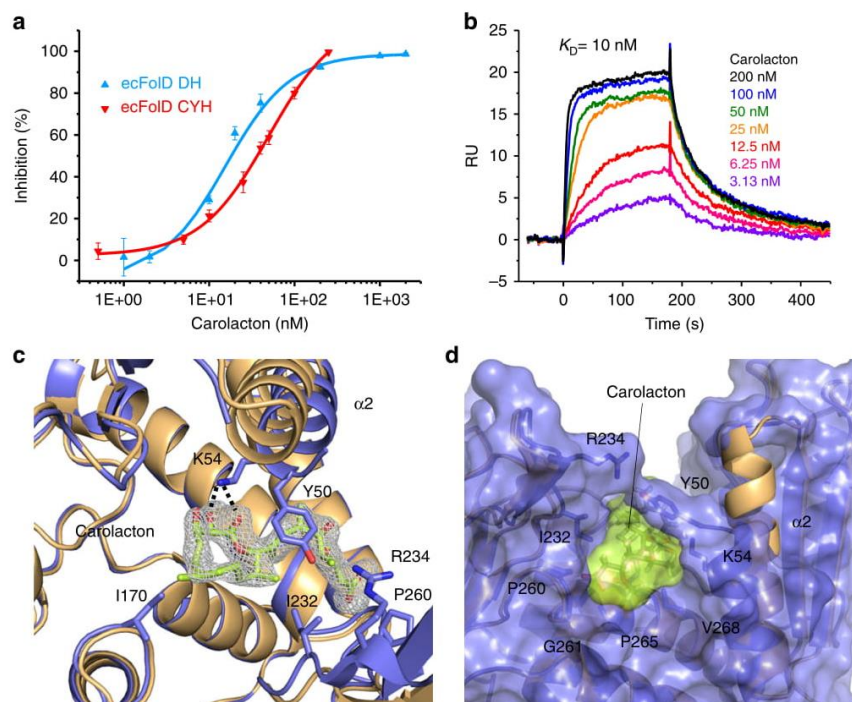
Natural products have served as a source of novel drug leads for thousands of years, but for many of them their mechanism of action (MoA) remains unknown<sup>1</sup>. The secondary metabolite carolacton is a macrolide keto-carboxylic acid produced by the myxobacterium *Sorangium cellulosum* with an elusive molecular target<sup>2</sup>. Carolacton is a potent inhibitor of biofilm formation in the human pathogen *Streptococcus mutans* and a growth inhibitor of *S. pneumoniae*<sup>3</sup>. It also inhibits the efflux pump mutant *Escherichia coli*Δ*tolC*<sup>2, 4</sup>. Carolacton has moderate activity on various fungi but showed no acute toxicity to L929 mouse cells<sup>2, 4</sup>. Although carolacton seemed to be a very narrow spectrum antibiotic, its interesting anti-biofilm and anti-fungal activity aroused our interest, and we aimed to uncover its MoA.

Here we report the identification and validation of FoID as the carolacton target in *E. coli*Δ*tolC* and streptococci. FoID occupies a

central position in the folate-dependent C1 metabolism. Folate (vitamin B9) is an essential co-factor in all cells, but it is synthesized only by bacteria and plants<sup>5</sup>. The folate-dependent C1 metabolism is highly conserved in all domains of life and it provides the key building blocks for growth, most importantly nucleic acids, amino acids, provitamins and formylated methionine tRNA for translation initiation<sup>6</sup>. FoID is a dual function enzyme: it catalyses a reversible NADP<sup>+</sup>-dependent dehydrogenation step (5,10-methylenetetrahydrofolate (5,10-CH<sub>2</sub>-THF) dehydrogenase (DH)) and a subsequent cyclohydrolysis step (5,10-methenyltetrahydrofolate (5,10-CH=THF) cyclohydrolyase (CYH)) (Fig. 1)<sup>7</sup>. We demonstrate that carolacton inhibits both reactions catalyzed by FoID at nM concentrations. Determination of the FoID/carolacton complex crystal structure allows us to rationalize how mutations in FoID can confer resistance to carolacton.



**Fig. 1** The enzymatic routes to synthesize 10-CHO-THF in different organisms. In *E. coli*, 10-CHO-THF is synthesized by FoID, while in several other bacteria such as Streptococci, 10-CHO-THF can be produced via an alternative enzyme named Fhs. In humans, there are different enzymes in the cytoplasm and the mitochondrion. In the cytoplasm, MTHFD1, which is a tri-functional enzyme, exerts FoID-DH, FoID-CYH and Fhs functions. In the mitochondrion, while MTHFD2 and MTHFD2L have FoID functions, MTHFD1L plays the same role as Fhs. The chemical structure of FoID inhibitor carolacton is shown in the centre



**Fig. 2** Carolacton is a potent and tightly binding inhibitor of FolD. **a**  $IC_{50}$  determination for carolacton against DH and CYH activities of ecFolD. Data are presented as means  $\pm$  s.e.m of three independent replicates.  $IC_{50}$ s were obtained via logistic dose-response fitting. The one-way ANOVA test was used for statistical analysis,  $P < 0.01$ . **b** SPR analysis of carolacton binding to ecFolD, RU resonance units. Carolacton shows strong binding to ecFolD. The estimated  $K_D$  was obtained by fitting the association and dissociation signals with a 1:1 interaction model using the Biacore X100 Evaluation Software. **c** Top view of the FolD-carolacton complex structure. Cartoon representations of apo-FolD (gold) and FolD in complex with carolacton (blue, lime) are superposed to show the movement and partial dissolution of helix  $\alpha 2$ . Residues involved in carolacton binding are shown as sticks. The difference electron density ( $F_o - F_c$ ) contoured to  $3\sigma$  with phases calculated from a model that was refined in the absence of carolacton is shown as a grey isomesh. **d** Surface representation of FolD bound to carolacton, side-view. Colours correspond to **c**. Partial dissolution of helix  $\alpha 2$  as a result of the interaction of Y50, which forms a lid on carolacton that is further stabilized by R234, can be seen

Higher organisms such as humans also possess FolD orthologs: the cytosolic trifunctional protein MTHFD1, which has DH and CYH, but also a formate tetrahydrofolate synthase (Fhs) domain, maintains the metabolism of methylene-, methenyl- and formyl-THF (10-CHO-THF)<sup>8–10</sup>. In mitochondria, the DH and CYH activities are provided by the FolD analogues MTHFD2 and MTHFD2L (Fig. 1)<sup>9, 10</sup>. The enzymes involved in folate-dependent C1 metabolism, therefore, represent powerful targets for the inhibition of fast growing cells and have been targeted by anticancer drugs such as clinically used methotrexate, which inhibits dihydrofolate reductase<sup>11</sup>. Because of its central role in the C1 metabolism, FolD was the subject of numerous studies: Although synthetic folate-analogue inhibitors were found that strongly inhibited FolD in vitro, none were active on whole bacterial cells<sup>12–15</sup>. Several studies have reported inhibitors of the FolD analogues MTHFD1 or MTHFD2, some of which display in vivo activity against human cells<sup>15–17</sup>. However, either poor activity was observed (e.g. LY345899,  $IC_{50}$  128  $\mu$ M) or the compounds were unspecific (e.g. LY231514, which principally inhibits thymidylate synthase) (Supplementary Fig. 1)<sup>16, 18, 19</sup>.

To gauge carolacton's potential for inhibiting the human FolD variants, we test it against the two human FolD orthologs, MTHFD1 and MTHFD2, which are also inhibited in the low nM range. When we tested carolacton's effects on a panel of human cell lines, we observed activity in the  $\mu$ M range. Since MTHFD2 is differentially upregulated in cancer tissue, it presents an attractive target for anti-cancer compounds. We believe that carolacton, a

non-substrate inhibitor, may be used as a tool compound to assess MTHFD2 as an anti-cancer target<sup>20, 21</sup>.

## Results

**FolD is the carolacton target.** To facilitate the identification of the molecular target of carolacton, *E. coli* $\Delta$ tolC (*E. coli* without outer membrane efflux protein TolC) was chosen to develop carolacton-resistant mutants. After 1 week, carolacton-resistant mutants arose spontaneously on agar plates supplemented with four times the carolacton minimum inhibitory concentration (MIC = 0.125  $\mu$ g/mL). Whole-genome sequencing of five independent mutants unveiled that only the gene encoding FolD was mutated (Supplementary Tables 1 and 2). Four different mutations conferring resistance were identified: G8S (observed in two mutants), K54N, Q98H and K54\_K56delinsK (hereafter referred to as  $\Delta$ K54R55) (Supplementary Figs. 2 and 3). To validate FolD as the carolacton target and to understand the potential MoA, *E. coli* $\Delta$ tolC FolD (ecFolD) was overexpressed and purified. A SDS-PAGE of all proteins used in this study, as well as their LC-MS analysis, can be found in Supplementary Fig. 4. The purified ecFolD showed DH and CYH activity comparable to previously reported bacterial FolDs<sup>13, 22, 23</sup>, including the DH activity of *E. coli* FolD<sup>24</sup>. DH and CYH activities were found to be much stronger than reported for ecFolD in two other studies<sup>25, 26</sup> (Supplementary Table 3). This discrepancy has been noted by other authors and possible reasons include impurities<sup>24, 26</sup> and

excessive enzyme concentrations used for enzyme kinetics<sup>25</sup>. When carolacton was added to the reaction, we observed strong inhibition of both steps catalysed by ecFolD (DH and CYH) in a carolacton concentration-dependent manner (Fig. 2a). Carolacton also showed competitive inhibition with both substrates and the cofactor involved in DH and CYH catalytic steps, which were 5,10-CH<sub>2</sub>-THF ( $K_i = 21$  nM), NADP<sup>+</sup> ( $K_i = 11$  nM) and 5,10-CH=THF ( $K_i = 32$  nM) (Supplementary Table 4), as the apparent Michaelis constant ( $K_M$ ) increased when carolacton was added (Supplementary Fig. 5). These findings supported that ecFolD is indeed the target of carolacton. To confirm that both molecules interact, we then tested ecFolD for its ability to bind carolacton using surface plasmon resonance (SPR). The observed interaction between ecFolD and carolacton was very strong ( $K_D = 10$  nM), further confirming ecFolD as the carolacton target (Fig. 2b and Supplementary Table 5).

**The structure of the ecFolD–carolacton complex.** To understand how carolacton binds to ecFolD and thus rationalize the mutations giving rise to carolacton resistance, we determined the crystal structure of ecFolD in complex with carolacton. The originally reported crystallization conditions for ecFolD<sup>27</sup> yielded only poorly diffracting crystals in our hands. Reductive lysine methylation ecFolD<sup>Meth</sup> gave a new crystal form that diffracted well. Since a key active-site residue of ecFolD is a lysine (K54)<sup>15</sup>, and we found that mutations at this position conferred carolacton resistance (see resistant mutants above), we thought that K54 could be involved in carolacton binding, and thus sought to protect this residue from methylation. To this end, we incubated ecFolD with an excess of carolacton before lysine methylation. After the reaction, carolacton was removed via size exclusion chromatography and the resultant ecFolD<sup>Meth</sup> tested for activity. We observed only a slight decrease in enzyme activity (Supplementary Fig. 6a) and no significant change of carolacton's ability to inhibit ecFolD DH activity (Supplementary Fig. 6b). To further analyse whether we had methylated active-site K54, we first determined the apo-structure of ecFolD<sup>Meth</sup>. The protein crystallized in space group P2<sub>1</sub> and crystals diffracted to 1.9 Å resolution (PDB ID 5o28). Full data collection and refinement statistics for all structures can be found in Supplementary Table 6. ecFolD<sup>Meth</sup> showed the canonical dimeric arrangement observed in the structures of ecFolD (PDB ID 1B0A)<sup>27</sup> and its orthologs<sup>17, 28</sup> and the overall structure was virtually unchanged when compared to unmethylated protein (Supplementary Fig. 7). We observed unambiguous electron density for methylated lysines at positions K4, 22, 194, 212 and 222. With the exception of K22, all of them can be found at crystal contacts (Supplementary Fig. 8). We observed density for unmethylated active-site K54, supporting our biochemical data that indicated this lysine was not methylated (Supplementary Figs. 6 and 8). We soaked the apo crystals of ecFolD<sup>Meth</sup> with 2 mM carolacton and harvested them after 16 h. The soaked crystals diffracted to 2.1 Å (PDB ID 5o22) and showed strong additional density at the protein's active site in three of the four monomers of the asymmetric unit (protomers A, B and D, the picture in protomer C is skewed due to a crystal contact). Carolacton can be fitted into the additional density (Fig. 2c) and refines well (Supplementary Fig. 8). Binding of carolacton to FolD has a minimal impact on FolD's overall structure (C $\alpha$  rmsd 0.27 Å) and is facilitated by three hydrogen bonds (K54  $\epsilon$ -amino group with carolacton O8 and O33 and G261 main chain N with carolacton O29) and several hydrophobic interactions (Y50, I170, I232, P260, P265 and V268) (Fig. 2c, d and Supplementary Fig. 9). The hydrophobic interaction between carolacton and Y50 causes a partial disruption of helix  $\alpha$ 2 (residues 45–56) and allows the Y50 hydroxyl group to

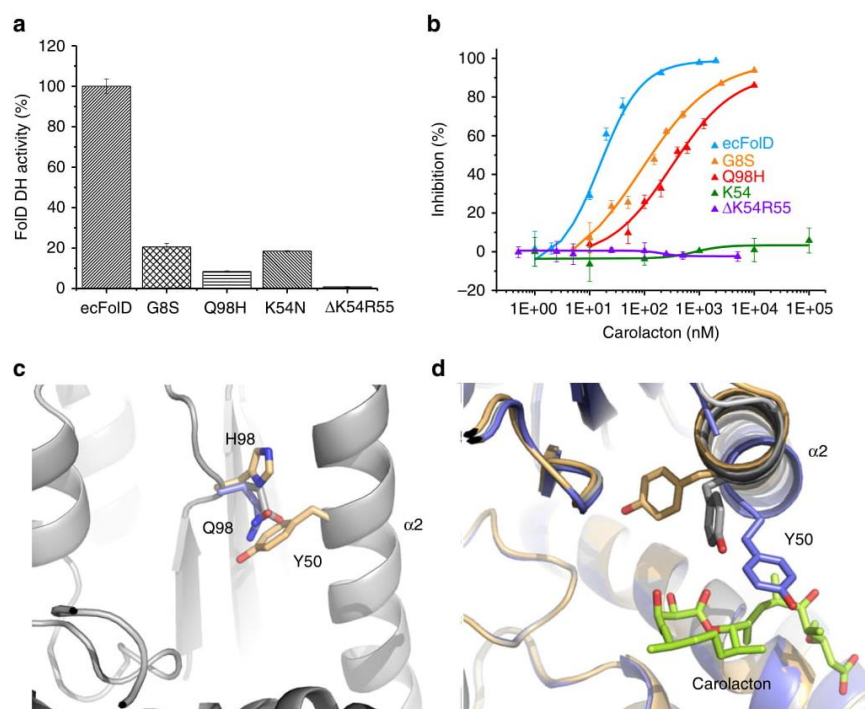
form a hydrogen bond with the side-chain of R234 (Fig. 2d). When carolacton is bound, the N-terminal residues of  $\alpha$ 2 and a part of the loop connecting  $\beta$ 1 with  $\alpha$ 2 (residues 41–44) become disordered (residues 44–49). It has been noted by others that this part of FolD is unstable<sup>15, 17</sup>, and we observe only poor density for residues 44–56 in the ecFolD<sup>Meth</sup> apo structure (protomers A, B and D; protomer C behaves differently due to a crystal contact). We believe that this reflects a catalytic mechanism, in which a meta-stable helix ( $\alpha$ 2) harbours a residue essential for catalysis. Upon substrate (or carolacton) binding, Y50 engages in hydrophobic contact that completely destabilizes residues 44–49, while the other half of the helix is stabilized (residues 50–56) (Fig. 2d).

When the FolD–carolacton complex structure is superposed onto the human ortholog of FolD (PDB ID 1DIA), which was co-crystallized with NADP<sup>+</sup>, and substrate analogue LY249543, it can be seen that carolacton is likely to prevent binding of both cofactor and substrates (Supplementary Fig. 10a, b). In a recent study, a number of carolacton analogues were synthesized and tested for their effects on *S. mutans* biofilm formation. None of the reported compounds had improved bioactivity<sup>29</sup>. We modelled the best-performing compound ('carylacton', see structure in Supplementary Fig. 1) based on our complex structure and discovered that carylacton would be able to form better hydrophobic interactions with Y50 than carolacton. The length of the carylacton tail, however, is one methylene group too long to allow hydrogen bonding of the carboxy terminus with G261 (Supplementary Fig. 10c). Our complex structure should therefore help future synthetic efforts.

**Effects of mutations on ecFolD and carolacton binding.** The effects of mutations K54N and G8S on carolacton binding can easily be explained by the complex crystal structure. K54 provides two hydrogen bonds to anchor carolacton in the active site of ecFolD (Fig. 2c) and its mutation to N will severely affect the binding of carolacton to ecFolD. G8 is not in direct contact with carolacton, but its position and orientation in FolD implies that the addition of any side-chain will cause a clash with C-terminal helix  $\alpha$ 11 (Supplementary Fig. 11). G261, which forms a hydrogen bond with carolacton, is part of this helix and we assume that the resulting movement of helix  $\alpha$ 11 leads to a clash with the carboxyl group of carolacton. Due to the very tight fit of carolacton in this area, the clash cannot be remedied, but instead leads to severely reduced binding of carolacton to ecFolD. Helix  $\alpha$ 2 is vital for binding of carolacton to FolD since it contains both K54 and Y50. It was therefore not unexpected that the deletion of two residues from this helix (mutant ecFolD $\Delta$ K54R55) severely affected binding of carolacton to FolD. The final mutation, Q98H, could not be rationalized using the complex structure and we wondered if this mutation does in fact affect binding of carolacton to ecFolD or is an experimental artefact.

To test experimentally the effects of the mutations that we discovered in carolacton resistant *E. coli*  $\Delta$ tolC, we expressed and purified all four mutant proteins: ecFolDG8S, ecFolDK54N, ecFolD $\Delta$ K54R55 and ecFolDQ98H. When we tested the mutants for DH activity, we found that all of them were attenuated, but with varying severity. The mutant K54N still retains ~19% of the wild-type (wt) DH activity, but  $\Delta$ K54R55 has only ~1% DH activity left. The two-residue deletion might severely change the NADP<sup>+</sup> binding, which can be deduced from the dramatically increased  $K_M$  of NADP<sup>+</sup>. The DH activity of G8S and Q98H is about 21 and 8% of the wt, respectively (Fig. 3a and Supplementary Table 7).

Mutants K54N and  $\Delta$ K54R55 showed no detectable CYH activity, suggesting that K54 is essential for this reaction. The kinetic properties of the CYH activity of mutants G8S and Q98H



**Fig. 3** Effects of FOLD mutations on DH activity and carolacton binding. **a** Residual DH activity of ecFolD mutants compared to ecFolD. Data are presented as means  $\pm$  s.e.m of three independent replicates. DH specific activities were calculated based on the  $V_{max}$  obtained via Michaelis–Menten fitting. The one-way ANOVA test was used for statistical analysis,  $P < 0.01$ . **b**  $IC_{50}$  determination for carolacton against the DH activity of ecFolD and mutants. Data are presented as means  $\pm$  s.e.m of three independent replicates.  $IC_{50}$ s were obtained via logistic dose-response fitting. The one-way ANOVA test was used for statistical analysis,  $P < 0.01$ . **c** Position of Y50 in ecFolDQ98H (gold). The position of Q98 in the apo (grey) and complex (blue) structures prevents the side-chain of Y50 from rotating out of position. **d** Position of Y50 in the apo (grey), complex (blue) and Q98H (gold) structures. The 90° rotation of Y50 in the Q98H mutant prevents Y50 from forming the hydrophobic lid on carolacton (lime sticks)

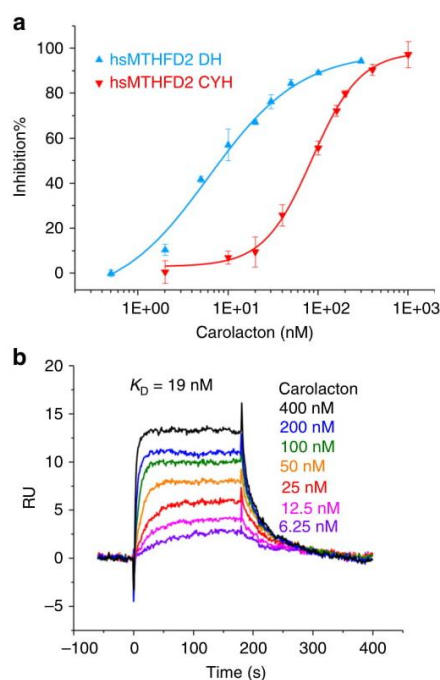
were not investigated since only very low CYH activity was detected. We estimate residual activity of  $\sim 1\%$  based on the requirement to increase the enzyme concentration  $\sim 100$ -fold to achieve wt turnover (Supplementary Fig. 12). The mutant *E. coli*  $\Delta tolC$  strains likely survive without CYH activity because the turnover of 5,10-CH=THF to 10-CHO-THF can occur spontaneously<sup>30</sup>. These biochemical results are consistent with the growth curves measured for these carolacton-resistant *E. coli*  $\Delta tolC$  strains: the one bearing  $\Delta K54R55$  mutation is the poorest growing mutant. The K54N mutant strain also grows very slowly even though its DH activity is not impaired as much as in the  $\Delta K54R55$  mutant, which might be due to the total loss of CYH activity (Supplementary Fig. 13).

The mutants G8S and Q98H could still be inhibited by carolacton, but with much higher  $IC_{50}$  compared to the wt (Fig. 3b and Supplementary Table 8). The central role of K54 is highlighted by the fact that the residual DH activity of K54N and  $\Delta K54R55$  is completely insensitive to carolacton (Fig. 3b), in accordance with the SPR data that indicated that carolacton does not bind to K54N (Supplementary Fig. 14a). It appears that, under selection pressure, viable mutants had sacrificed CYH activity (reaction can be spontaneous), but required retention of FOLD's DH activity since no alternative pathway exists in *E. coli*<sup>31</sup>.

The binding of carolacton to ecFolD Q98H was seriously attenuated as reflected by the SPR data (Supplementary Fig. 14b, c). To understand how the mutation Q98H affects the binding of carolacton, we determined the crystal structure of this mutant (PDB ID 5o2a). After lysine methylation, this protein crystallized

in the same space group as the wt protein and crystals diffracted to 1.9 Å. The overall structure of ecFolD<sup>Meth</sup>Q98H is not altered when compared to wt as a result of the mutation ( $C\alpha$  rmsd 0.11 Å) and the effect of the Q98H mutation was unexpected. The presence of H98 allows the side-chain of Y50 to rotate 90° compared to the apo wt ecFolD structure. It appears that in the wt protein, Q98 helps to position the side-chain of Y50 such that it can easily rotate to engage in hydrophobic interactions with carolacton (or substrate) upon binding. In Q98H, the side-chain of Y50 is able to rotate farther from the active site and packs against residues V38, L40, L96 and H98. The result is a fully ordered helix  $\alpha 2$  with good density for all residues, which we believe reflects a stabilization of the protein. The absence of the hydrophobic interaction between carolacton and Y50 leads to significantly weakened binding, which we believe to be reflected by the much faster association and dissociation of carolacton in SPR experiments with ecFolDQ98H (Fig. 3c, d, Supplementary Fig. 14b and Supplementary Table 5).

**Effects of carolacton on *S. pneumoniae* FOLD.** Carolacton was originally discovered as an anti-streptococcal compound<sup>2, 4</sup>, and the FOLD ortholog from *S. pneumoniae* (spFOLD) shares 49% sequence identity with ecFolD. The key residues involved in carolacton binding are conserved, and carolacton binds tightly to spFOLD ( $K_D = 27$  nM) (Supplementary Fig. 14d and Supplementary Table 5). Carolacton also showed competitive inhibition of spFOLD with slightly higher inhibition constants than ecFolD,



**Fig. 4** Carolacton effects on the human enzyme hsMTHFD2. **a**  $IC_{50}$  determination for carolacton against DH and CYH activity of hsMTHFD2. Data are presented as means  $\pm$  s.e.m. of three independent replicates.  $IC_{50}$ s were obtained via logistic dose-response fitting. The one-way ANOVA test was used for statistical analysis,  $P < 0.01$ . **b** SPR analysis of carolacton binding to hsMTHFD2. Carolacton shows strong binding to hsMTHFD2. The estimated  $K_D$  was obtained by fitting the association and dissociation signals with a 1:1 interaction model using the Biacore X100 Evaluation Software

which were still in the low nM range (Supplementary Figs. 15 and 16 and Supplementary Tables 4 and 8). Analysis of the inhibition of FolD in streptococci is complicated by the fact that they, unlike *E. coli*, possess an alternative route for one-carbon metabolism via Fhs (Fig. 1)<sup>31, 32</sup>.

**Effects of carolacton on the human FolD orthologs.** As mentioned above, FolD orthologs are also present in humans. The human enzymes hsMTHFD1\_DC and hsMTHFD2 both share 42% sequence identity to ecFolD. The key residues involved in carolacton binding are fully conserved (Supplementary Fig. 3), and the structures of hsMTHFD1\_DC and hsMTHFD2 are conserved well when compared to ecFolD (pairwise  $\alpha$  rmsds of 1.09 and 0.84 Å, respectively, Supplementary Fig. 17). We were therefore confident that carolacton could also serve as an inhibitor of the human enzymes. This would be of particular interest since the mitochondrial hsMTHFD2 is overexpressed in many tumor cells and thus its inhibition might cause a selective effect on tumor growth<sup>20</sup>. We therefore expressed and purified hsMTHFD1\_DC and hsMTHFD2 to investigate them biochemically. Not surprisingly, their inhibition by carolacton was similar to that observed for ecFolD (Fig. 4a and Supplementary Fig. 15). The  $K_i$  of carolacton against all three substrates and the cofactor were also determined and are in the same range as those for ecFolD (Supplementary Fig. 18 and Supplementary Tables 4 and 8). When the binding of carolacton to hsMTHFD2 was analysed by SPR, we found strong binding on par with the

Cell line	$EC_{50}$ ( $\mu$ M)	Maximum inhibition
HCT-116	24.6 $\pm$ 0.8	>80%
KB-3.1	11.3 $\pm$ 3.6	>80%
KB-V.1	41.8 $\pm$ 14.5	>80%
U-937	>100	n.a.
U-937 (folate-depleted)	41.7 $\pm$ 15.2	65–75%

	MIC ( $\mu$ g/mL)
<i>E. coli</i> (TolC-deficient)	0.13
<i>E. coli</i> (TolC-deficient) + PMBN <sup>a</sup>	$\leq$ 0.03
<i>E. coli</i> DSM-1116 (WT)	>64
<i>E. coli</i> DSM-1116 (WT) + PMBN <sup>a</sup>	64 to >64
<i>E. coli</i> DSM-1116 (WT) + PA $\beta$ N <sup>b</sup>	2 <sup>c</sup>
<i>E. coli</i> DSM-1116 (WT) + PMBN <sup>a</sup> + PA $\beta$ N <sup>b</sup>	2 <sup>c</sup>

<sup>a</sup> + 3  $\mu$ g/mL polymyxin B nonapaptide (PMBN) for permeabilization  
<sup>b</sup> + 20  $\mu$ g/mL phenylalanine-arginine  $\beta$ -naphthylamide (PA $\beta$ N) efflux inhibitor  
<sup>c</sup> No full inhibition

bacterial proteins ( $K_D = 19$  nM) (Fig. 4b and Supplementary Table 5).

**Carolacton is active against human cancer cell lines.** Since the human FolD orthologs are strongly inhibited by carolacton, we were curious to see if carolacton would show activity on human cancer cell lines. When tested against HCT-116, KB-3.1 and KB-V.1 cells, carolacton displayed an  $EC_{50}$  of 25, 11 and 42  $\mu$ M, respectively. In each case, the maximum inhibition of cell growth was >80%. The inhibition of human epidermoid carcinoma KB-3.1 cells by carolacton is four times stronger than that of the corresponding multidrug-resistant (MDR) cell line KB-V.1 (Table 1). In the latter cell line, the *mdr1* gene is amplified, which encodes the drug transporter P-glycoprotein (Pgp); this suggests that Pgp may export carolacton from the cells. In the case of U-937 cells, carolacton displayed no activity at concentrations up to 100  $\mu$ M. However, when this cell line is grown in folate-depleted medium, an  $EC_{50}$  of 42  $\mu$ M and a maximum inhibition of 70% are observed. This apparent folate-dependent activity on U-937 cells supports that human FolD orthologs are the cellular targets of carolacton in human cancer cells.

**Efflux hampers carolacton activity.** Carolacton shows no activity against wt *E. coli*, but is highly active against efflux protein-deficient *E. coli*  $\Delta$ tolC. It also showed a four-fold decrease in activity against Pgp-overexpressing KB-V.1 cells, in comparison with KB-3.1 cells. To probe the role of efflux on carolacton's cellular activity, we treated *E. coli* DSM-1116 (wt) and *E. coli*  $\Delta$ tolC with a permeability enhancer and/or an efflux pump inhibitor (Table 2). When *E. coli* DSM-1116 was treated with a combination of carolacton and sub-inhibitory concentrations of polymyxin B nonapaptide (PMBN) to increase outer membrane permeability<sup>33</sup>, the MIC was still at the upmost range of the assay (at least 64  $\mu$ g/mL). The same assay repeated with carolacton-susceptible *E. coli*  $\Delta$ tolC showed a four-fold increase of carolacton activity. In contrast, when *E. coli* DSM-1116 was treated with carolacton and the efflux inhibitor phenylalanine-arginine  $\beta$ -naphthylamide (PA $\beta$ N)<sup>34</sup>, the carolacton MIC decreased to 2  $\mu$ g/mL. In addition, combining PMBN, carolacton and PA $\beta$ N did not enhance the effect of carolacton further. These data support that

carolacton can enter cells, but it is excreted by efflux proteins, which seems to be the main challenge to be addressed when attempting to increase the cellular activity of carolacton.

## Discussion

We have identified the bifunctional enzyme FoLD from the folate-dependent C1 metabolism as the target of carolacton. Through the analyses of carolacton-resistant *E. coli*  $\Delta tolC$  mutants, biochemical experiments and the crystal structure of the FoLD-carolacton complex, we were able to identify the key residues involved in carolacton binding (Y50, K54, and G261). All the residues mutated in carolacton-resistant *E. coli*  $\Delta tolC$  isolates affected FoLD's DH activity only partially, but almost completely abolished its CYH activity. This reflects the fact that the DH activity is vital (no alternative pathway exists in *E. coli*), while the CYH reaction can also occur spontaneously. However, carolacton is effluxed in wt *E. coli* as evidenced by the activity on *E. coli*  $\Delta tolC$  mutants, and the data show a strong enhancing effect of the efflux pump inhibitor PA $\beta$ N.

Due to the high degree of sequence conservation between FoLD and human mitochondrial hMTHFD2, especially of the key carolacton-binding residues, we tested carolacton as a possible inhibitor of hMTHFD2. This protein is overexpressed in many tumor cells<sup>20</sup>, and we observed strong inhibition and binding by carolacton in vitro. To see if carolacton could serve as a starting point for FoLD/hMTHFD2 inhibitors that display activity on whole cells, we tested the effect of carolacton on several cancer cell lines. Carolacton displayed moderate activity in these assays. Since carolacton is able to inhibit both reactions carried out by this enzyme and perturbs cofactor binding, we believe that carolacton, a non-substrate inhibitor, may be used as a tool compound to assess MTHFD2 as an anti-cancer target in the future.

## Methods

**Strains and plasmids.** The *E. coli*  $\Delta tolC$  strain used in carolacton-resistant mutant development is an antibiotic-sensitive *E. coli* strain with deficient outer membrane channel protein TolC. It was the same strain used in the previous carolacton studies<sup>2,4</sup>. *E. coli* DH10B<sup>35</sup> or DH5 $\alpha$  (Invitrogen) strains were used for gene cloning and *E. coli* BL21 (DE3) (Novagen) was used for protein expression. *E. coli* DSM-1116 was ordered from German Collection of Microorganisms and Cell Cultures (DSMZ). The protein expression plasmid pHis-TEV<sup>36</sup> was used for cloning and expression of all *folD* gene orthologues studied in this paper (Supplementary Table 9).

**Carolacton-resistant mutant development.** First, the minimum inhibitory concentration (MIC) of carolacton on the *E. coli*  $\Delta tolC$  strain was determined. We started to develop the resistant mutants on CASO agar (peptone from casein 15 g/L, peptone from soybean 5 g/L, NaCl 5 g/L and agar 15 g/L; pH 7.3) plates supplemented with 4 $\times$  the MIC concentration of carolacton (0.5  $\mu$ g/mL) as selection pressure. Different numbers of *E. coli*  $\Delta tolC$  cells from  $1 \times 10^7$  to  $10^9$  were spread on caso agar plates with carolacton. After 7 days, *E. coli*  $\Delta tolC$  colonies, which were resistant to carolacton, started to form on the agar plates. The mutants picked were inoculated in liquid MHB medium (beef extract 2 g/L, peptone from casein 17.5 g/L and corn starch 1.5 g/L, pH 7.4) supplemented with carolacton. The genomic DNA of wt *E. coli*  $\Delta tolC$  and five carolacton resistant mutants were extracted and the genomes sequenced.

**Whole-genome sequencing and variant calling.** Genomic DNA of five independent mutant strains and one control *E. coli*  $\Delta tolC$  strain was sequenced using Illumina Paired-End technology on the MiSeq instrument at the Helmholtz Centre for Infection Research (Braunschweig, Germany). Characteristics of the obtained raw-data sequencing reads are shown in Supplementary Table 1. Raw data were imported into the Geneious 9.1.3<sup>37</sup> software package and trimmed of low-quality parts with an error probability threshold of 0.05. It was aligned against the *E. coli*  $\Delta tolC$  reference sequence using the 'Low sensitivity' option of the Geneious 'Map to reference...' sequence aligner. This produced assembly files in which whole genome mean sequencing coverages varied in the range of 108–116. Assembly files were then converted to consensus sequences by the 'Generate consensus sequence...' option in the Geneious software and 'Highest quality' consensus calling option. Six resulting consensus sequences were aligned to each other and to the reference genome sequence by the 'progressiveMauve' algorithm of the MAUVE whole-genome sequence alignment tool<sup>38</sup>. The variant calling was

done by comparing the consensus sequences of each mutant sample against the control *E. coli*  $\Delta tolC$  WT consensus sequence. The complete list of mutations is shown in Supplementary Table 2.

**Gene cloning and protein purification.** All *folD* gene orthologues studied in this paper were cloned into the pHis-TEV protein expression vector. The *folD* genes of *E. coli*  $\Delta tolC$  and its carolacton-resistant mutants were amplified from the genomic DNA of the corresponding strains. The *folD* gene of *S. pneumoniae* was amplified from the genomic DNA of the TIGR4 strain<sup>39</sup>. The cDNAs encoding MTHFD1 and MTHFD2 were synthesized based on the available human genomic DNA sequences (gene ID 4522 and 10197, respectively; DNA synthesized by ATG:bio-synthesis). All primers and restriction sites are listed in Supplementary Table 10. The resulting protein expression plasmids were verified by enzyme restriction digestion and DNA sequencing before being transformed into *E. coli* BL21 (DE3) for protein overexpression and purification. For optimal protein purification, a systematic buffer test including ten different buffers was conducted for each protein by using the KingFisher mL (ThermoFisher scientific) magnetic beads purification system. For large-scale protein purification, a single colony was picked into LB liquid medium containing 50  $\mu$ g/mL kanamycin to make an overnight culture. The overnight culture was inoculated 1 to 100 into fresh LB medium supplemented with 50  $\mu$ g/mL kanamycin and was grown at 37 °C until the optical density at 600 nm (OD<sub>600</sub>) reached 0.6. Then, the culture was transferred to 16 °C for half-an-hour to cool the culture before 0.1 mM IPTG was added to induce protein expression. The cells were harvested and lysed by sonication after 16 h shaking at 16 °C. The proteins were purified by immobilized metal ion affinity chromatography (IMAC) on a 5 mL HisTrap HP column (GE healthcare). A HiPrep 26/10 desalting column (GE healthcare) was used to remove imidazole. When removal of the N-terminal 6x His-tag was required, TEV protease was added to the imidazole free protein solution and incubated at 4 °C overnight, followed by another hour at room temperature. The digestion mixture was loaded on a HisTrap HP column for a second time to bind the undigested protein, His-tag and TEV protease. The flow-through, which contained the protein, was collected, concentrated and loaded onto a gel filtration column HiLoad 16/600 Superdex 200 pg (GE healthcare) to further remove impurities and FoLD aggregates. Finally, all purified proteins used in this study were checked by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and subjected to liquid chromatography–mass spectrometry (LC-MS) to check purity and assess protein molecular weights. All buffers for protein purification are listed in Supplementary Table 11.

**Enzyme assay conditions.** (6*R,S*)-5,10-Methylene-5,6,7,8-tetrahydrofolic acid ((6*R,S*)-5,10-CH<sub>2</sub>-THF) calcium salt and (6*R,S*)-5,10-Methenyl-5,6,7,8-tetrahydrofolic acid ((6*R,S*)-5,10-CH=THF) chloride were purchased from Schircks Laboratories (Bauma, Switzerland) and used as substrates for FoLD dehydrogenase and cyclohydrolyase enzyme assays, respectively. Only the *R*-isomer, which accounts for 50% of (6*R,S*)-5,10-CH<sub>2</sub>-THF and (6*R,S*)-5,10-CH=THF, respectively, is used by the enzyme. (6*R,S*)-5,10-CH<sub>2</sub>-THF was dissolved in N<sub>2</sub>-sparged basic buffer (50 mM Tris-HCl (pH 8.0), 100 mM  $\beta$ -mercaptoethanol) as described by Varshney et al.<sup>25</sup>. (6*R,S*)-5,10-CH=THF was dissolved in DMSO as a 100 mM stock solution.

Because FoLD is a bifunctional enzyme, dehydrogenase and cyclohydrolyase activities were determined for FoLD enzyme kinetics, respectively. The dehydrogenase activity of FoLD was assayed for its substrate 5,10-CH<sub>2</sub>-THF and cofactor NADP<sup>+</sup> (in the case of MTHFD2, NAD<sup>+</sup> was used as the cofactor) based on monitoring the formation of 5,10-CH=THF, while the cyclohydrolyase activity of FoLD was determined for its substrate 5,10-CH=THF by monitoring the hydrolysis of 5,10-CH<sub>2</sub>-THF. The enzyme kinetics for the dehydrogenase activity of all FoLDs except hMTHFD1\_DC were determined at 30 °C in 50 mM Tris-HCl (pH 7.5), 30 mM  $\beta$ -mercaptoethanol. The hMTHFD1\_DC dehydrogenase activity was measured in 25 mM MOPS (pH 7.3), 30 mM  $\beta$ -mercaptoethanol at 30 °C. To measure the enzyme kinetic constants for 5,10-CH<sub>2</sub>-THF, for example in the case of ecFoLD, the NADP concentration was fixed at 1 mM and the concentration of 5,10-CH<sub>2</sub>-THF varied from 5 to 1500  $\mu$ M (*R*-isomer). Similarly, to measure the kinetic constants of ecFoLD for NADP<sup>+</sup>, the concentration of 5,10-CH<sub>2</sub>-THF (*R*-isomer) was fixed at 1 mM and the concentration of NADP<sup>+</sup> varied from 10 to 1500  $\mu$ M. For different FoLDs, the varied concentration range of substrates could be different. The 50  $\mu$ L reactions of dehydrogenase assays were initiated by adding appropriate amounts enzyme (4 nM ecFoLD, 4 nM spFoLD, 15 nM hMTHFD1\_DC, 20 nM hMTHFD2, 20 nM G8S, 20 nM Q98H, 10 nM K54N and 200 nM  $\Delta$ K54R55 in the corresponding assays) and terminated with 50  $\mu$ L 1 M HCl after 2 min incubation at 30 °C. To monitor the formation of 5,10-CH=THF, the absorbance of the reaction mixture was monitored at 350 nm in micro UV cuvettes (BRAND, Essex, Connecticut, USA). The concentration of 5,10-CH=THF produced in the reaction was determined using an extinction co-efficient of 0.0249  $\mu$ M<sup>-1</sup> cm<sup>-1</sup>. Cyclohydrolyase activity for each FoLD was assayed in the same buffer as dehydrogenase activity at 30 °C but for 30 s. The concentration of FoLD used in each assay was 5 nM of ecFoLD, spFoLD and hMTHFD1\_DC, respectively, and 2 nM of hMTHFD2. The consumption of 5,10-CH=THF was measured by recording the absorbance of the reactions at 355 nm using the TECAN Infinite 200 PRO equipped with a monochromator. To measure the enzyme kinetic constants, different concentrations of 5,10-CH=THF (varied from 0.5  $\mu$ M to 100  $\mu$ M) were used in 100  $\mu$ L reaction mixtures. The inhibition constants (*K<sub>i</sub>*) of carolacton on

Fold were determined by measuring the apparent  $K_m$  when adding certain amounts of carolacton to the enzyme reactions. For ecFolD, 10 nM carolacton was added in the dehydrogenase assay and 50 nM carolacton was used in the cyclohydrolase assay. For spFolD, 25 nM carolacton was added in both dehydrogenase assay and cyclohydrolase assay. For hsMTHFD1\_DC, 30 nM carolacton was added in the dehydrogenase assay and 10 nM in the cyclohydrolase assay. For hsMTHFD2, 20 nM carolacton was used to determine the apparent  $K_m$  for 5,10-CH<sub>2</sub>-THF and 5,10-CH<sub>2</sub>-THF, while 10 nM carolacton was used to determine the apparent  $K_m$  for NADP. To compare the inhibition effects of carolacton between the different FolDs, the half maximal inhibitory concentration ( $IC_{50}$ ) of carolacton on the same amount of enzyme was determined for FolDs involved in this study. The enzyme concentration used to determine dehydrogenase activity  $IC_{50}$  was 10 nM, and the enzyme used for cyclohydrolase  $IC_{50}$  measurements was 5 nM. All enzyme kinetics measurements were performed as independent triplicates. Data processing and fitting of curves was done using OriginPro 2016 (OriginLab, Northampton, Massachusetts, US). Analysis of variance (ANOVA) in OriginPro 2016 software gave statistical test reports after fitting curves.

**Surface plasmon resonance assay.** All surface plasmon resonance (SPR) experiments described in this study were performed on a Biacore X100 system. Different FolDs with or without His-tag were coupled on CM5 sensor chips (GE Healthcare) by the amine coupling method using a kit from GE Healthcare Life Sciences (Freiburg, Germany). A pH scouting process was performed before protein immobilization to test the most appropriate buffer and the protein concentration for protein coupling. Finally, 20 µg/mL of ecFolD in 10 mM maleate buffer (pH 6.8), 30 µg/mL of ecFolD Q98H in 10 mM maleate buffer (pH 6.3), 40 µg/mL of ecFolD K54N in 10 mM maleate buffer (pH 5.8), 20 µg/mL of spFolD in 10 mM maleate buffer (pH 6.0) and 15 µg/mL of MTHFD2 (His-tag cleaved) in 10 mM maleate buffer (pH 6.8) were used for coupling. In the protein immobilization procedure, the contact time was calculated based on the pH scouting results to achieve ~3000–5000 relative resonance units (RU). The buffer used in the SPR assays was 1× HBS-P buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 0.05% Tween 20). The 1 mM stock solution of carolacton was prepared in 1× HBS-P buffer and diluted to different concentration ranges by two-fold serial dilution. In the SPR assay, the association time was 180 s and the dissociation time was 600 s. Data analysis was performed using the Biacore evaluation software. The SPR curves were exported to be re-plotted in OriginPro 2016 (OriginLab, Northampton, Massachusetts, US).

**Antimicrobial screening.** The determination of MICs was performed as described elsewhere<sup>40</sup>. In brief, *E. coli* strains were treated with carolacton in serial dilution in either the presence or absence of sub-inhibitory concentrations of polymyxin B nonapeptide (PMBN; 3 µg/mL) and/or phenylalanine-arginine β-naphthylamide (PAβN; 20 µg/mL) for 16 h at 37 °C in Mueller–Hinton broth. The MICs were assessed by visual inspection of the plates.

**Cytotoxicity screening.** Half maximal effective concentrations ( $EC_{50}$ ) on human cancer cell lines were determined as described elsewhere<sup>41</sup>. Human cancer cell lines were maintained as recommended by the depositor (German Collection of Microorganisms and Cell Cultures, DSMZ) and were treated for 5 d with carolacton in serial dilution. The viability of adherent cell types, HCT-116 (ACC 581), KB-3.1 (ACC-158) and KB-V.1 (ACC-149) was assessed by the addition of MTT tetrazolium salt and absorbance measurement at 570 nm. Histiocytic lymphoma U-937 cells (ACC-5) were maintained in RPMI 1640 medium and transferred into folate-free medium prior to the assay. Cell viability was determined by measuring alamar blue fluorescence emission at 590 nm. The values were referenced and  $EC_{50}$  values were determined by sigmoidal curve fitting. The cell lines KB-3.1 and KB-V.1 are originally derived from HeLa, whereas the latter is an MDR subclone of KB-3.1. They were used to directly compare the activity of carolacton on an *mdr1* overexpressing and a sensitive cell line. The effect of folate depletion was assessed using the U-937 cell line as these lymphoma cells are described to express high levels of MTHFD2, however, carolacton was virtually inactive under normal culture conditions.

**Reductive methylation of surface lysine residues.** To protect active site lysine K54, a six-fold molar excess of carolacton was added to the protein and the complex incubated on ice for 16 h before the reaction. Surface lysine residues were then methylated using the JBS Methylation Kit (Jena Bioscience) according to the manufacturer's instructions and Supplementary Information. Methylated ecFolD (ecFolD<sup>Meth</sup>) and ecFolD<sup>Meth</sup>Q98H were applied to a Superdex 200 gel filtration column (GE Healthcare) pre-equilibrated with gel filtration buffer (150 mM NaCl, 10 mM HEPES (pH 7.4) and 1 mM TCEP), and then concentrated to 5 mg/mL.

**X-ray crystallography.** Crystals of ecFolD<sup>Meth</sup> were obtained at 18 °C by using the sitting drop vapour diffusion method. Protein (300 nL) at 5 mg/mL was added to 150 nL reservoir solution: 0.2 M sodium acetate, 0.1 M sodium cacodylate pH 6.5 and 30% (w/v) PEG 8000. To solve the ecFolD-carolacton complex structure, ecFolD<sup>Meth</sup> crystals were soaked overnight in the presence 2 mM carolacton.

Optimized crystals of apo-ecFolD<sup>Meth</sup>Q98H were grown under conditions of 0.2 ammonium sulphate, Bis-Tris pH 6.0 and 25% PEG 3350. Diffraction data for all proteins were collected from single crystals at 100 K. Data for ecFolD<sup>Meth</sup> were obtained at Beamline X06DA at a wavelength of 1 Å (Swiss Light Source), while data for the ecFolD<sup>Meth</sup>-carolacton complex and ecFolD<sup>Meth</sup>Q98H crystals were collected at beamline ID30-A3 (ESRF) at a wavelength of 0.967 Å. Data were processed using Xia2<sup>42</sup> or XDS<sup>43</sup> and the structures solved using PHASER<sup>44</sup>. Molecular replacement with ecFolD (PDB ID 1DIA) as a search model. The models were manually rebuilt in COOT<sup>45</sup> and refined using PHENIX<sup>46</sup> and Refmac5<sup>47</sup> (Supplementary Table 12). The structures were validated using MolProbity, and all images presented were created using PyMOL<sup>48</sup>. Interaction diagrams were created using Ligplot<sup>49</sup>.

**Data availability.** Atomic coordinates and structure factors are deposited in the RCSB Protein Data Bank with accession codes 5O28 (ecFolD apo), 5O22 (ecFolD carolacton complex) and 5O2A (ecFolDQ98H). Other relevant data supporting the findings of this study are available in this published article and its Supplementary Information files or from the corresponding authors upon request.

Received: 23 August 2017 Accepted: 5 October 2017

Published online: 16 November 2017

## References

- Cragg, G. M. & Newman, D. J. Natural products: a continuing source of novel drug leads. *Biochim. Biophys. Acta* **1830**, 3670–3695 (2013).
- Jansen, R. et al. Carolacton—a macrolide ketocarboxylic acid that reduces biofilm formation by the caries- and endocarditis-associated bacterium *Streptococcus mutans*. *Eur. J. Org. Chem.* **2010**, 1284–1289 (2010).
- Donner, J. et al. The biofilm inhibitor Carolacton inhibits planktonic growth of virulent pneumococci via a conserved target. *Sci. Rep.* **6**, 29677 (2016).
- Kunze, B. et al. Damage of *Streptococcus mutans* biofilms by carolacton, a secondary metabolite from the myxobacterium *Sorangium cellulosum*. *BMC Microbiol.* **10**, 199 (2010).
- Green, J. M. & Matthews, R. G. Folate biosynthesis, reduction, and polyglutamylation and the interconversion of folate derivatives. *EcoSal Plus* **2**, doi:10.1128/ecosalplus.3.6.3.6 (2007).
- Ducker, G. S. & Rabinowitz, J. D. One-carbon metabolism in health and disease. *Cell. Metab.* **25**, 27–42 (2017).
- Blakley, R. L., Benkovic, S. J. & Whitehead, V. M. *Folates and Pterins* (Wiley, New York, Chichester, 1984).
- Appling, D. R. Compartmentation of folate-mediated one-carbon metabolism in eukaryotes. *FASEB J.* **5**, 2645–2651 (1991).
- Christensen, K. E. & MacKenzie, R. E. Mitochondrial one-carbon metabolism is adapted to the specific needs of yeast, plants and mammals. *Bioessays* **28**, 595–605 (2006).
- Christensen, K. E. & MacKenzie, R. E. in *Folic Acid and Folates* (Gerald L., editor), pp. 393–410 (Academic Press, 2008).
- Rajagopalan, P. T. R. et al. Interaction of dihydrofolate reductase with methotrexate: ensemble and single-molecule kinetics. *Proc. Natl Acad. Sci. USA* **99**, 13481–13486 (2002).
- Eadsforth, T. C. et al. Assessment of *Pseudomonas aeruginosa* N5,N10-methylenetetrahydrofolate dehydrogenase-cyclohydrolase as a potential antibacterial drug target. *PLoS ONE* **7**, e35973 (2012).
- Eadsforth, T. C., Maluf, F. V. & Hunter, W. N. Acinetobacter baumannii FolD ligand complexes – potent inhibitors of folate metabolism and a re-evaluation of the structure of LY374571. *FEBS J.* **279**, 4350–4360 (2012).
- Eadsforth, T. C. et al. Characterization of 2,4-diamino-6-oxo-1,6-dihydropyrimidin-5-yl Ureido based inhibitors of trypanosoma brucei FolD and testing for antiparasitic activity. *J. Med. Chem.* **58**, 7938–7948 (2015).
- Schmidt, A. et al. Structures of three inhibitor complexes provide insight into the reaction mechanism of the human methylenetetrahydrofolate dehydrogenase/cyclohydrolase. *Biochemistry* **39**, 6325–6335 (2000).
- Tonkinson, J. L. et al. The antiproliferative and cell cycle effects of 5,6,7, 8-tetrahydro-N<sub>5</sub>,N<sub>10</sub>-carbonylfolic acid, an inhibitor of methylenetetrahydrofolate dehydrogenase, are potentiated by hypoxanthine. *J. Pharmacol. Exp. Ther.* **287**, 315–321 (1998).
- Gustafsson, R. et al. Crystal structure of the emerging cancer target MTHFD2 in complex with a substrate-based inhibitor. *Cancer Res.* **77**, 937–948 (2017).
- McDonald, A. C. et al. A phase I and pharmacokinetic study of LY231514, the multitargeted antifolate. *Clin. Cancer Res.* **4**, 605–610 (1998).
- Shih, C. et al. LY231514, apyrrolo2,3-dipyrimidine-based antifolate that inhibits multiple folate-requiring enzymes. *Cancer Res.* **57**, 1116–1123 (1997).
- Nilsson, R. et al. Metabolic enzyme expression highlights a key role for MTHFD2 and the mitochondrial folate pathway in cancer. *Nat. Commun.* **5**, 3128 (2014).



21. Tedeschi, P. M., Vazquez, A., Kerrigan, J. E. & Bertino, J. R. Mitochondrial methylenetetrahydrofolate dehydrogenase (MTHFD2) overexpression is associated with tumor cell proliferation and is a novel target for drug development. *Mol. Cancer Res.* **13**, 1361–1366 (2015).
22. Clark, J. E. & Ljungdahl, L. G. Purification and properties of 5,10-methylenetetrahydrofolate cyclohydrolase from *Clostridium formicoaceticum*. *J. Biol. Chem.* **257**, 3833–3836 (1982).
23. Wohlfarth, G., Geerligs, G. & Diekert, G. Purification and characterization of NADP(+)-dependent 5,10-methylenetetrahydrofolate dehydrogenase from *Peptostreptococcus productus marburg*. *J. Bacteriol.* **173**, 1414–1419 (1991).
24. D'Ari, L. & Rabinowitz, J. C. Purification, characterization, cloning, and amino acid sequence of the bifunctional enzyme 5,10-methylenetetrahydrofolate dehydrogenase/5,10-methylenetetrahydrofolate cyclohydrolase from *Escherichia coli*. *J. Biol. Chem.* **266**, 23953–23958 (1991).
25. Sah, S. & Varshney, U. Impact of mutating the key residues of a bifunctional 5,10-methylenetetrahydrofolate dehydrogenase-cyclohydrolase from *Escherichia coli* on its activities. *Biochemistry* **54**, 3504–3513 (2015).
26. Dev, I. K. & Harvey, R. J. A complex of N5,N10-methylenetetrahydrofolate dehydrogenase and N5,N10-methylenetetrahydrofolate cyclohydrolase in *Escherichia coli*. Purification, subunit structure, and allosteric inhibition by N10-formyltetrahydrofolate. *J. Biol. Chem.* **253**, 4245–4253 (1978).
27. Shen, B. W. et al. The crystal structure of a bacterial, bifunctional 5,10 methylene-tetrahydrofolate dehydrogenase/cyclohydrolase. *Protein Sci.* **8**, 1342–1349 (1999).
28. Allaire, M., Li, Y., MacKenzie, R. E. & Cygler, M. The 3-D structure of a folate-dependent dehydrogenase/cyclohydrolase bifunctional enzyme at 1.5 Å resolution. *Structure* **6**, 173–182 (1998).
29. Solinski, A. E. et al. Diverted total synthesis of carolacton-inspired analogs yields three distinct phenotypes in streptococcus mutans biofilms. *J. Am. Chem. Soc.* **139**, 7188–7191 (2017).
30. Maden, B. E. H. Tetrahydrofolate and tetrahydromethanopterin compared. Functionally distinct carriers in C1 metabolism. *Biochem. J.* **350**, 609–629 (2000).
31. Sah, S., Aluri, S., Rex, K. & Varshney, U. One-carbon metabolic pathway rewiring in *Escherichia coli* reveals an evolutionary advantage of 10-formyltetrahydrofolate synthetase (Fhs) in survival under hypoxia. *J. Bacteriol.* **197**, 717–726 (2015).
32. Crowley, P. J., Gutierrez, J. A., Hillman, J. D. & Bleiweis, A. S. Genetic and physiologic analysis of a formyl-tetrahydrofolate synthetase mutant of *Streptococcus mutans*. *J. Bacteriol.* **179**, 1563–1572 (1997).
33. Vaara, M., Viljanen, P., Vaara, T. & Mäkelä, P. H. An outer membrane-disorganizing peptide PMBN sensitizes *E. coli* strains to serum bactericidal action. *J. Immunol.* **132**, 2582–2589 (1984).
34. Renau, T. E. et al. Inhibitors of efflux pumps in *Pseudomonas aeruginosa* potentiate the activity of the fluoroquinolone antibacterial levofloxacin. *J. Med. Chem.* **42**, 4928–4931 (1999).
35. Grant, S. G., Jessee, J., Bloom, F. R. & Hanahan, D. Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc. Natl Acad. Sci. USA* **87**, 4645–4649 (1990).
36. Liu, H. & Naismith, J. H. A simple and efficient expression and purification system using two newly constructed vectors. *Protein Expr. Purif.* **63**, 102–111 (2009).
37. Kearse, M. et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
38. Darling, A. C. E., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **14**, 1394–1403 (2004).
39. Tettelin, H. et al. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* **293**, 498–506 (2001).
40. Müller, R. et al. Discovery and total synthesis of natural cystobactamid derivatives with superior activity against Gram-negative pathogens. *Angew. Chem. Int. Ed.* **56**, 12760–12764 (2017).
41. Steinmetz, H. et al. Isolation, structure elucidation, and (Bio)Synthesis of Haprolid, a cell-type-specific myxobacterial cytotoxin. *Angew. Chem. Int. Ed.* **55**, 10113–10117 (2016).
42. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190 (2010).
43. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
44. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
45. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
46. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
47. Skubak, P., Murshudov, G. N. & Pannu, N. S. Direct incorporation of experimental phase information in model refinement. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2196–2201 (2004).
48. The PyMOL Molecular Graphics System, Version 1.8.2.1 Schrödinger, LLC.
49. Laskowski, R. A. & Swindells, M. B. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* **51**, 2778–2786 (2011).

### Acknowledgements

This work was funded by the German Federal Ministry of Education and Research (BMBF) in the program e:bio (grant number 031 A299). We acknowledge use of the ESRF (beamline ID30A) and SLS (beamline X06DA) synchrotrons. J.K. thanks the Deutsche Forschungsgemeinschaft for an Emmy Noether fellowship (KO4116/3–1).

### Author contributions

C.F. developed the carolacton-resistant mutants, identified the carolacton target, carried out protein expression and purification, biochemical assays and SPR experiments, interpreted data and wrote the paper. A.S. carried out protein expression and purification, crystallization and x-ray crystallography experiments and interpreted data. N.Z. analysed the genome sequences, identified the mutations and interpreted data. J.H. determined the biological activity, helped with SPR studies and interpreted data. J.D. and M.R. helped with data interpretation. I. W.-D. interpreted data and helped to write the paper. J.K. carried out x-ray crystallography experiments, interpreted data and wrote the paper. R.M. conceived and designed the project, interpreted data and wrote the paper.


### Additional information

Supplementary Information accompanies this paper at doi:10.1038/s41467-017-01671-5.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

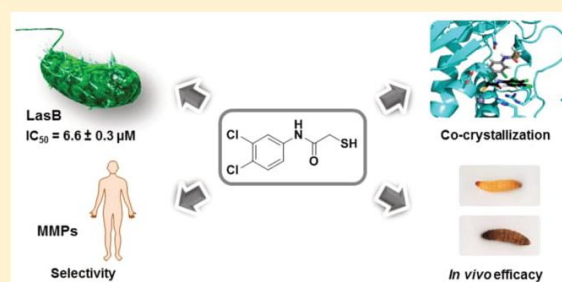
© The Author(s) 2017

Binding Mode Characterization and Early *in Vivo* Evaluation of Fragment-Like Thiols as Inhibitors of the Virulence Factor LasB from *Pseudomonas aeruginosa*Andreas M. Kany,<sup>†,‡</sup> Asfandyar Sikandar,<sup>‡,‡</sup> Jörg Haupenthal,<sup>†</sup> Samir Yahiaoui,<sup>†,§</sup> Christine K. Maurer,<sup>†</sup> Ewgenij Proschak,<sup>||</sup> Jesko Köhnke,<sup>‡,§</sup> and Rolf W. Hartmann<sup>\*,†,§,||</sup><sup>†</sup>Department of Drug Design and Optimization, Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Campus E8.1, 66123, Saarbrücken, Germany<sup>‡</sup>Workgroup Structural Biology of Biosynthetic Enzymes, Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Campus E8.1, 66123, Saarbrücken, Germany<sup>§</sup>Department of Pharmacy, Pharmaceutical and Medicinal Chemistry, Saarland University, Campus E8.1, 66123, Saarbrücken, Germany<sup>||</sup>Institute of Pharmaceutical Chemistry, Goethe University Frankfurt, Max-von-Laue-Straße 9, 60438, Frankfurt, Germany

## Supporting Information

**ABSTRACT:** The increasing emergence of antibiotic resistance necessitates the development of anti-infectives with novel modes of action. Targeting bacterial virulence is considered a promising approach to develop novel antibiotics with reduced selection pressure. The extracellular collagenase elastase (LasB) plays a pivotal role in the infection process of *Pseudomonas aeruginosa* and therefore represents an attractive antivirulence target. Mercaptoacetamide-based thiols have been reported to inhibit LasB as well as collagenases from clostridia and bacillus species. The present work provides an insight into the structure–activity relationship (SAR) of these fragment-like LasB inhibitors, demonstrating an inverse activity profile compared to similar inhibitors of clostridial collagenase H (ColH). An X-ray cocrystal structure is presented, revealing distinct binding of two compounds to the active site of LasB, which unexpectedly maintains an open conformation. We further demonstrate *in vivo* efficacy in a *Galleria mellonella* infection model and high selectivity of the LasB inhibitors toward human matrix metalloproteinases (MMPs).

**KEYWORDS:** antibiotic resistance, antivirulence agent, elastase, LasB, binding mode, selectivity, *Galleria mellonella*



The increasing emergence of resistant bacteria poses a threat to public health, especially in the case of Gram-negative species.<sup>1,2</sup> *Pseudomonas aeruginosa* is one of the three most problematic pathogens on the WHO priority list.<sup>1</sup> It is the reason for many hospital-acquired infections as well as fatal lung infections in cystic fibrosis and bronchiectasis patients.<sup>3,4</sup> To combat the rise of antibiotic-resistant *Pseudomonas aeruginosa* infections, novel treatment options are urgently needed.<sup>5,6</sup>

A promising new approach to reduce selection pressure is to target bacterial virulence in order to disarm pathogens rather than to kill them.<sup>7–10</sup> *Pseudomonas aeruginosa* produces numerous virulence factors contributing to disease progression, which provide attractive anti-infective targets.<sup>11–14</sup> The extracellular collagenase elastase (LasB) is a major virulence factor, playing a crucial role for the pathogenicity of *P. aeruginosa*.<sup>15</sup> The enzyme is a zinc–metalloprotease with high structural similarity to thermolysin.<sup>16,17</sup> One of its main functions is the cleavage of components of the connective tissue

like elastin<sup>15</sup> or collagen<sup>18</sup> to allow the bacteria to colonize a niche in the host. Tissue damage is further caused by disruption of cell-to-cell junctions.<sup>19,20</sup> Additionally, LasB enables *P. aeruginosa* to evade the human immune response by cleaving, i.e., IgG,<sup>21</sup> cytokines,<sup>22</sup> surfactant proteins A and D,<sup>23</sup> complement factor C3,<sup>24</sup> or pulmonary defense receptor PAR2.<sup>25</sup> Consequently, this protease represents an attractive anti-infective target and its extracellular localization facilitates drug discovery as permeation of the Gram-negative cell wall is not needed.<sup>9,26</sup>

LasB belongs to the thermolysin (M4) family of enzymes.<sup>27</sup> For this class, it has been reported that the rather open active site cleft of the protease adopts a more closed conformation upon inhibitor binding.<sup>28</sup> To date, several zinc-chelating inhibitors of LasB have been described,<sup>29–32</sup> including

Received: January 8, 2018

Published: February 27, 2018

compounds bearing a mercaptoacetamide motif attached either to small peptides<sup>33</sup> or to aniline.<sup>34</sup> Crystallographic data shedding light on the binding mode of these thiols to the protease has not been available yet. We have recently described *N*-aryl mercaptoacetamide-based compounds as very potent clostridial collagenase inhibitors with high selectivity toward human matrix metalloproteinases (MMPs).<sup>35</sup> The compounds contain a prodrug-like thiocarbamate-motif, which liberates free thiols as the active form after hydrolysis in buffer. The best collagenase H (ColH) inhibitor **1** (Figure 1) was also found to inhibit LasB.

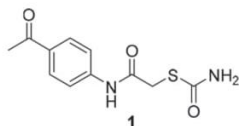


Figure 1. Structure of most potent ColH inhibitor **1**.

In this work, we report a functional screening followed by LC-MS validation for LasB inhibition of the focused TimTec ActiTarg-P library previously used to discover the ColH inhibitors. The only inhibitor resulting from the screening was an *N*-aryl mercaptoacetamide. Encouraged by the promising inhibitory activity and selectivity, we wanted to gain further insight into LasB inhibition by this compound class. We describe the activity and selectivity profiles of a broad range of *N*-aryl mercaptoacetamides. An X-ray crystal structure of a mercaptoacetamide-based inhibitor in complex with LasB is presented revealing an unprecedented open conformation of the active site, which harbors two inhibitor molecules. To the best of our knowledge, this is the first description of inhibitor binding to LasB which does not lead to a closure of the active site cleft.

## RESULTS AND DISCUSSION

**Screening for Novel LasB Inhibitors.** In order to expand the chemical space of LasB inhibitors, we performed a functional screening based on the FRET-based *in vitro* assay developed by Nishino and Powers.<sup>36</sup> We used a protease inhibitor-enriched library which, after removing structures known as PAINS,<sup>37</sup> comprised 1192 low molecular weight compounds. In addition, we included 330 fragments (Maybridge Fragment Library) into the screening. The only compound showing more than 50% inhibition when tested at 100  $\mu$ M was mercaptoacetamide **2** (Figure 2). Several false-positive hits had to be excluded because of quenching of substrate fluorescence.

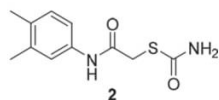


Figure 2. Structure of screening hit **2**.

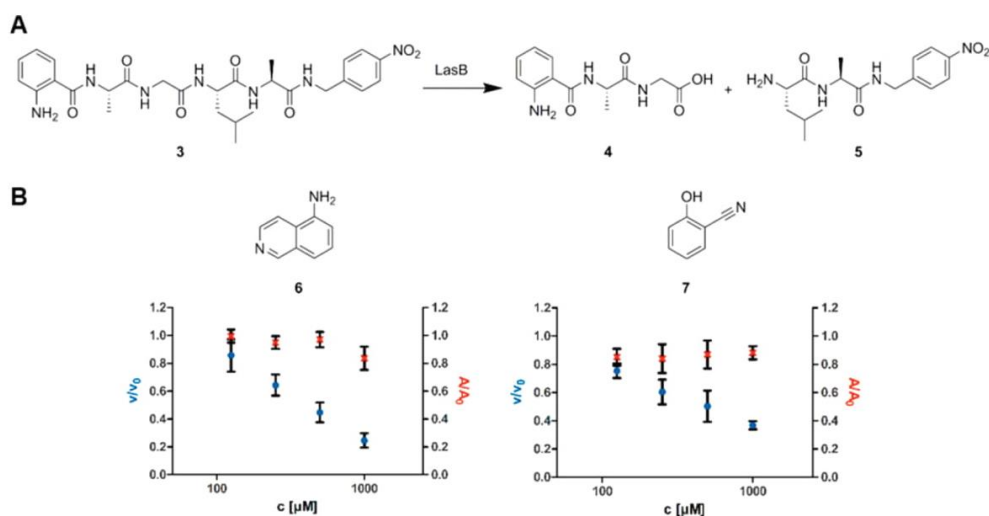
**Development of an LC-MS-Based Readout for the FRET Assay.** In the FRET-based proteolytic assay, active LasB cleaves a quenched substrate, which results in an increase of fluorescence.<sup>36</sup> Enzyme inhibition leads to reduced substrate cleavage and consequently to reduced fluorescence. However, compounds interfering with the fluorophore by quenching effects<sup>38</sup> can pretend enzyme inhibition, resulting in false-

positive hits.<sup>39,40</sup> Several quenching compounds were found, especially among fragments. To clarify if reduced fluorescence was caused by protease inhibition or to quenching effects, we developed an LC-MS-based readout for the FRET Assay. Elastase cleaves the synthetic substrate Abz-Ala-Gly-Leu-Ala-Nba (**3**) at the Gly-Leu bond,<sup>36,41</sup> forming cleavage products **4** and **5** (Figure 3A) which could be separated chromatographically. Using the published LasB inhibitor phosphoramidon<sup>42</sup> as a positive control, the FRET assay results were excellently reproduced using an LC-MS-based readout which was based on the mass peak of cleavage product **5** (Figure S1). Applying this technology, we could detect several false positives whose apparent inhibition in the FRET assay was only due to quenching. Two examples are shown in Figure 3B.

**Structure-Activity Relationship of *N*-Aryl Mercaptoacetamides.** To elucidate the structure-activity relationship (SAR) of *N*-aryl mercaptoacetamides, 35 derivatives were purchased or synthesized (for further information, see the Supporting Information) and tested for LasB inhibition applying the FRET-based *in vitro* assay (Tables 1 and S1). Nonpolar aromatic substituents, especially halogens, turned out to be favorable for activity while polar hydrogen-bond accepting moieties led to significantly reduced inhibition. For the *o*-, *m*-, and *p*-Cl derivatives **9**, **12**, and **17** as well as *m*- and *p*-CH<sub>3</sub> derivatives **13** and **21**, no substantial effect of the substitution position on activity was found. The same holds true for the methoxy (**11**, **15**, **23**) and *o*- and *p*-phenyl (**10**, **18**) analogues. Introduction of a second substituent improved the IC<sub>50</sub> at least 2-fold. Addition of methyl or chlorine substituents to **17** resulted in the best compounds **26**, **27**, and **28**, displaying IC<sub>50</sub> values in the one-digit micromolar range. An exception is dimethoxy derivative **32**, which showed an activity similar to the monomethoxy analogues **11**, **15**, and **23**. In comparison to the initial screening hit **2**, the activity could be improved by more than 2.5-fold. Two examples showed that the introduction of a third substituent was not beneficial for inhibition: Introduction of a methyl group to compound **26** led to a 2-fold drop in activity (**34**), while no difference was observed between di- and trimethyl derivatives **29** and **33**.

**Comparison to the SAR for ColH Inhibition.** Interestingly, the here reported SAR shows an inverse activity profile of the *N*-aryl mercaptoacetamides compared to ColH. We previously reported polar, hydrogen bond-accepting substituents, especially in *para*-position to the aniline function, to be most potent for ColH inhibition.<sup>35</sup> Contrary to that, oxygen-containing compounds like, e.g., the methoxy-derivatives **11**, **15**, **23**, and **32** showed considerably weaker LasB inhibition compared to their halogen-substituted analogs. Strikingly, our best ColH inhibitor **1** is one of the weakest inhibitors described in this study. The position of the substituent substantially affects inhibition of ColH but not of LasB.

**Confirmation of Thiols as Active Compounds.** As previously described,<sup>35</sup> thiocarbamates are rapidly cleaved to the corresponding free thiols in aqueous buffers (Figure 4A). Consequently, to confirm that the observed weaker activity on LasB compared to clostridial collagenases was not due to slower hydrolysis in LasB assay buffer, we performed the recently reported LC-MS-based stability test. Fast hydrolysis of **17** in 50 mM Tris, pH 7.2, was observed with a thiocarbamate half-life of  $3.7 \pm 0.1$  min (Figure 4B). A third component was formed which is most likely the disulfide oxidation product. Compound hydrolysis proceeded faster than described for the ColH assay

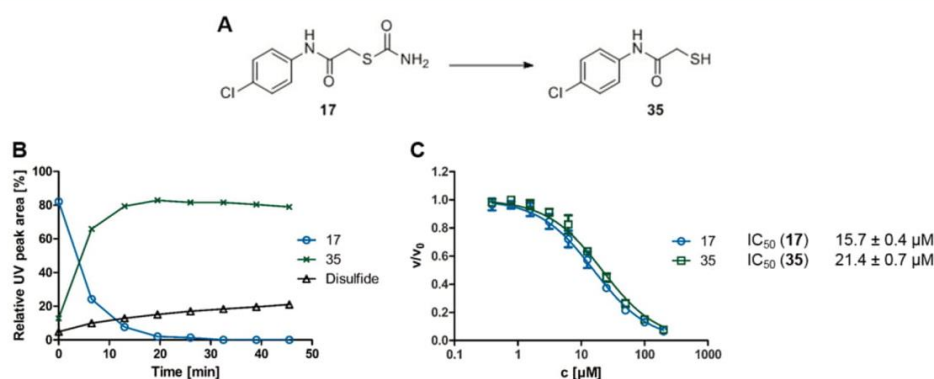


**Figure 3.** (A) Cleavage of FRET substrate Abz–Ala–Gly–Leu–Ala–Nba (3) into products 4 and 5 by LasB. Abz = 2-aminobenzoyl; Nba = 4-nitrobenzylamide. (B) Structure and comparison of FRET (blue,  $v/v_0$ ) vs LC-MS results (red,  $A/A_0$ ) for fragments 6 and 7.

**Table 1. Chemical Structures and LasB Inhibition of a Series of *N*-aryl Mercaptoacetamides**

Cp.	R	IC <sub>50</sub> ( $\mu\text{M}$ )	Cp.	R	IC <sub>50</sub> ( $\mu\text{M}$ )	Cp.	R	IC <sub>50</sub> ( $\mu\text{M}$ )
8	2-Br	11.4 $\pm$ 0.2	17	4-Cl	15.7 $\pm$ 0.4	26	2-CH <sub>3</sub> -3-Cl	5.9 $\pm$ 0.3
9	2-Cl	14.1 $\pm$ 0.5	18	4-Ph	19.8 $\pm$ 1.4	27	3,4-di-Cl	6.2 $\pm$ 0.3
10	2-Ph	25.0 $\pm$ 0.8	19	4-Br	22.8 $\pm$ 1.1	28	2-CH <sub>3</sub> -5-Cl	7.2 $\pm$ 0.2
11	2-OCH <sub>3</sub>	51.7 $\pm$ 4.0	20	4-I	32.9 $\pm$ 1.3	29	2,4-di-CH <sub>3</sub>	12.1 $\pm$ 0.4
12	3-Cl	19.4 $\pm$ 0.5	21	4-CH <sub>3</sub>	36.1 $\pm$ 0.7	30	2,3-di-CH <sub>3</sub>	12.3 $\pm$ 0.4
13	3-CH <sub>3</sub>	47.5 $\pm$ 1.4	22	4-OC <sub>2</sub> H <sub>5</sub>	47.6 $\pm$ 1.1	2	3,4-di-CH <sub>3</sub>	16.0 $\pm$ 1.9
14	3-F	59.2 $\pm$ 0.9	23	4-OCH <sub>3</sub>	47.7 $\pm$ 1.0	31	3-Cl-4-CH <sub>3</sub>	16.3 $\pm$ 1.0
15	3-OCH <sub>3</sub>	61.7 $\pm$ 2.1	1	4-COCH <sub>3</sub>	73.1 $\pm$ 2.5	32	2,4-di-OCH <sub>3</sub>	54.0 $\pm$ 2.7
16	3-NHSO <sub>2</sub> CH <sub>3</sub>	127.2 $\pm$ 3.5	24	4-COOCH <sub>3</sub>	81.5 $\pm$ 2.5	33 <sup>af</sup>	2,4,6-tri-CH <sub>3</sub>	11.9 $\pm$ 0.4
			25	4-(CH <sub>2</sub> -1,2,4-triazole)	85.6 $\pm$ 2.3	34 <sup>af</sup>	3-Cl-2,6-di-CH <sub>3</sub>	11.9 $\pm$ 0.4

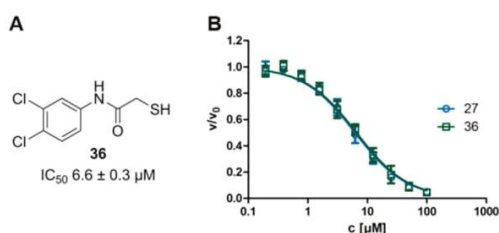
<sup>a</sup>Compounds 33 and 34 were synthesized as free thiols.



**Figure 4.** (A) Conversion of thiocarbamate 17 into corresponding free thiol 35. (B) LC-MS stability assay showing fast hydrolysis of 17 into 35 in 50 mM Tris, pH 7.2 (10% methanol), at 37 °C. (C) *In vitro* results for 17 and 35 are in accordance with the LC-MS results.

at 22.5 °C,<sup>35</sup> indicating temperature dependence of hydrolysis. Additionally, we found the same *in vitro* activity of selected thiols compared to their prodrugs, confirming the results from

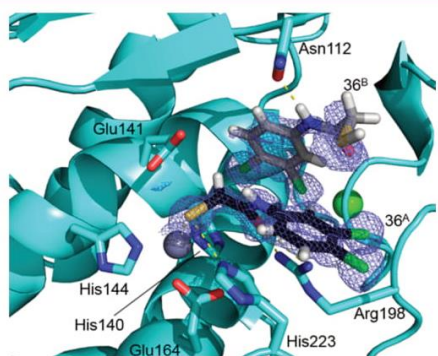
the LC-MS assay that compound activation occurs within the preincubation time of our assay (Figures 4C and 5). Unlike thiols, which can be oxidized to disulfides,<sup>43</sup> these thiocarb



**Figure 5.** (A) Structure of thiol **36**. (B) FRET assay results for **36** and prodrug **27**. Nonlinear regression was performed with the Hill Slope constrained to 1.

mate prodrugs have the advantage of being stable toward oxidation. As expected, the inactivity of dithiocarbamates **75–82** (Table S1) could be explained by their stability toward hydrolysis (Figure S2).

**Binding Mode of Compound 36 to LasB.** To elucidate the binding mode of the *N*-aryl mercaptoacetamides to LasB, we cocrystallized compound **27** with LasB purified from *Pseudomonas aeruginosa* PA14 culture supernatant. The putative LasB–**27** complex crystallized in space group  $P2_1$ , and crystals diffracted to 1.3 Å resolution (Figure 6). The structure was



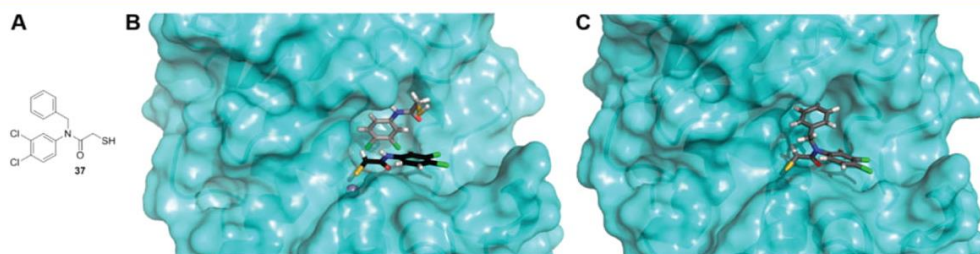
**Figure 6.** Structure of LasB in complex with **36**. Cartoon representations of LasB (cyan) in complex with **36** (black, gray). The difference electron density ( $F_o - F_c$ ) contoured to  $3\sigma$  with phases calculated from a model that was refined in the absence of **36** is shown as a yellow isomesh. The active-site zinc ion is shown as a gray sphere and calcium ion, as a green sphere. Residues involved in binding of  $36^A$  and  $36^B$  are shown as sticks.

solved by molecular replacement using the published LasB apo structure (PDB ID 1EZM) as a search model. Full details of the

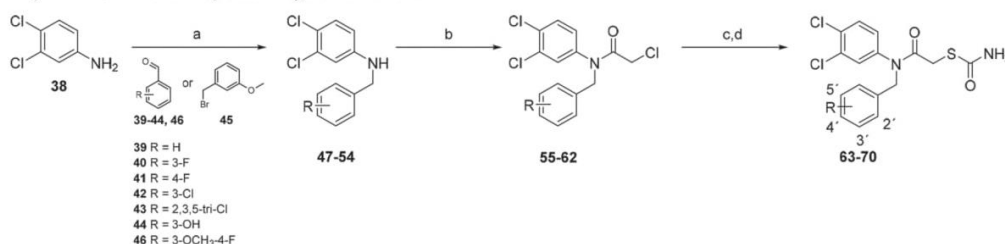
data collection and refinement statistics can be found in Table S2.

Hydrolysis of **27** to **36** was expected from the stability tests and from the *in vitro* assay using the corresponding thiol **36**, and indeed, there was no electron density for the thiocarbamate group observed. These observations confirm that the prodrug moiety does not contribute to the inhibitory activity. To our surprise, instead of one molecule, the active site of LasB contained unambiguous electron density for two molecules of **36**, arranged in an antiparallel fashion (hereafter, referred to as  $36^A$  and  $36^B$ ). In the LasB apo structure, the active-site zinc ion is coordinated by the side chains of His140, His144, and Glu164 as well as one water molecule. The free thiol of  $36^A$  displaces the water molecule to complete the tetrahedral coordination sphere (sulfur–zinc distance of 2.3 Å). The carbonyl oxygen of  $36^A$  forms a bidentate hydrogen bond with Arg198 in the S1' binding site (2.2 Å/2.3 Å), while the side chain of His223 forms a hydrogen bond with the thiol and amide nitrogen (2.4 Å/3.5 Å). The amide nitrogen of the second molecule hydrogen bonds with Asn112 in the edge strand (2.0 Å), while its aromatic core lies in the lipophilic S2' binding pocket. The tolerance of substitution at different positions without crucial changes in activity can be explained by the relatively large size of the binding site in comparison to the rather narrow binding pocket of ColH. Calculations performed with Molecular Operating Environment (MOE) software<sup>44</sup> enabled us to determine differences in the acidity of the thiol groups: Zinc coordination significantly increased thiol acidity, reducing the  $pK_a$  value from 9.0 in solution to 4.0. These results indicate full ionization of this sulfur atom when bound to the active site. In contrast, no significant change in the acidity of the nonzinc bound thiol could be determined, indicating a protonated state of this sulfur atom. It appears probable that binding of  $36^A$  is required for  $36^B$  to bind to the protein by providing a hydrophobic surface anchored to the active-site zinc ion  $36^B$  with which it can interact. Regarding the Hill Slope of 1 in the *in vitro* assay for **36** (Figure 5B), it seems that only one binding event is necessary for full inhibition. This indicates that a second molecule might not be required for the activity of the compound and its absence in solution cannot be excluded.

Elastase belongs to the M4 family of peptidases, showing high similarity to thermolysin from *Bacillus thermoproteolyticus*.<sup>27</sup> Its C-terminal domain is composed mostly of  $\alpha$ -helices, while the N-terminal domain is formed predominantly by antiparallel  $\beta$ -sheets. The active site cleft is positioned within the hinge region in between the two domains.<sup>17</sup> Interestingly, binding of  $36^A$  and  $36^B$  does not lead to a closure of the active site cleft as it has been reported after binding of phosphoramidon<sup>45</sup> and another peptidic inhibitor.<sup>46</sup> In fact,



**Figure 7.** (A) Structure of **37**, the *N*-benzyl derivative of active thiol **36**. (B) Surface representation of LasB bound to two molecules of **27**. (C) Modeling of **37**, the *N*-benzyl derivative of the active thiol **36** based on the X-ray crystal structure.

Scheme 1. Synthesis of *N*-Benzyl Mercaptoacetamides<sup>a</sup>

<sup>a</sup>Reagents and conditions: (a) sodium triacetoxyborohydride, DCM, RT, 20 h (47–52, 54), or 3-methoxybenzylbromide, K<sub>2</sub>CO<sub>3</sub>, DMF, 120 °C, 20 h (53); (b) chloroacetyl chloride, acetone, 0 °C to r.t., 1.5 h; (c) ammonium thiocyanate, ethanol, 80 °C, 2 h; (d) sulfuric acid, acetic acid, 0 °C, 30 min.

the open conformation observed in the apo structure is virtually unperturbed ( $C_{\alpha}$  rmsd of just 0.24 over 290 residues, Figure S3). It seems likely that the antiparallel binding of two molecules of **36** prevents the active site from closing and thus allows the enzyme to be addressed in an open conformation. This offers a completely new avenue for the design of LasB inhibitors.

In the work of Zhu et al., a primed binding mode of *N*-aryl mercaptoacetamides to the LasB active site was proposed on the basis of docking studies.<sup>34</sup> We experimentally confirmed the inhibitor to be placed in the primed binding pocket. However, there are substantial differences in the orientation of the inhibitor, mainly owing to the presence of two molecules. The reported docking studies suggested additional chelation of the zinc atom by the carbonyl group of the inhibitor, which we demonstrated to hydrogen bond with Arg198. The proposed hydrogen bond between Asn112 and the amide nitrogen of the inhibitor is indeed present, yet it is formed by **36<sup>B</sup>** and not by the zinc chelating **36<sup>A</sup>**.

**Structure-Based Optimization.** Binding of a single molecule to the binding site instead of two might have an entropic benefit resulting in improved inhibition. Hence, we tried to replace the second, nonzinc chelating molecule and to grow the zinc-binding molecule deeper into the binding pocket, using the crystal structure of the LasB–**36** complex. Modeling approaches showed a benzyl group to be appropriate to fill the part of the binding pocket occupied by the aromatic core of the nonzinc binding molecule (Figure 7).

A straightforward approach was used to synthesize *N*-benzyl derivatives of compound **27** (Scheme 1): *N*-benzyl substituted anilines **47–54** were obtained via reductive amination, using aniline **38** and a variety of benzaldehydes (**39–44,46**). Amide coupling with chloroacetyl chloride led to intermediates **55–62**. Replacement of the  $\alpha$ -chlorine by a thiocyanate group was followed by hydrolysis to the respective thiocarbamates **63–70** using a mixture of concentrated sulfuric acid and acetic acid.<sup>47</sup>

Introduction of a benzylic group at the position of the amide nitrogen of **27** led to an active compound with an IC<sub>50</sub> of 20.4 ± 0.9 μM (**63**, Table 2). The activity was approximately 3-fold lower compared to the nonbenzylated compound (Table 1). We introduced halogens at several positions of the benzyl group (**64–67**) to improve hydrophobic interactions with the lipophilic part of the binding pocket. The activity was slightly increased compared to **63**; however, with IC<sub>50</sub>'s in the two-digit micromolar range, our compounds were still less active than **27**. The polar phenol and methoxyphenol derivatives **68** and **69** inhibited LasB in the same range as the halogenated compounds **64–67**, indicating that the lipophilic part of the

Table 2. Structure and LasB Inhibition of *N*-Substituted Derivatives **63–74**

Cp.	R	IC <sub>50</sub> (μM)
63	Ph	20.4 ± 0.9
64	3'-F-Ph	12.6 ± 0.4
65	4'-F-Ph	15.3 ± 0.6
66	3'-Cl-Ph	17.3 ± 0.8
67	2',3',5'-tri-Cl-Ph	15.9 ± 0.9
68	3'-OH-Ph	17.6 ± 0.6
69	3'-OCH <sub>3</sub> -Ph	14.3 ± 0.6
70	3'-OCH <sub>3</sub> -4'-F-Ph	18.3 ± 0.9
71	H	15.9 ± 0.7
72	<i>n</i> -C <sub>4</sub> H <sub>9</sub>	15.0 ± 0.5
73	<i>t</i> -C <sub>4</sub> H <sub>9</sub>	54.5 ± 2.3
74	C <sub>6</sub> H <sub>11</sub>	27.6 ± 1.4

binding pocket might not be reached. Combination of the 4'-fluorine of **65** and the 3'-methoxy group of **69** did not lead to an additive effect (**70**).

Given that the activity could not be improved by introducing rigid aromatic functions, we investigated conformationally more flexible alkyl substituents. Synthesis was achieved following the same procedure (Scheme 1) using the respective alkyl aldehydes, giving compounds **71–74** (Table 2). Similar to the benzyl compounds **63–70**, introduction of alkyl substituents at the amide nitrogen led to a loss of activity, especially in case of the rather bulky neopentyl and cyclohexylmethyl substituents (**73** and **74**). Considering that replacement of the amide hydrogen by the methyl group already resulted in a more than 2-fold loss of activity (**71**), we concluded that the hydrogen bond formed between Asn112 and the second molecule in the binding pocket plays an important role for compound binding.

To assess the impact of *N*-substitution on thiocarbamate half-life, the stability of **68** and **74** was analyzed. Again, we proved that thiocarbamates were cleaved rapidly; thus, the reduced activity was not caused by slower compound activation in assay buffer (Figure S4).

**Selectivity toward Further Proteases.** MMPs are ubiquitously present in the human body, playing pivotal roles in the progress of various diseases but also exerting beneficial effects on human health. The unselective inhibition of various MMPs was the reason for the failure of several protease

inhibitors, while selectively inhibiting specific MMPs remains a challenging task in protease drug discovery.<sup>48–53</sup> Several bacterial protease inhibitors lacking MMP selectivity have been reported.<sup>54–56</sup> Thiol-based LasB inhibitors have been tested toward selected MMPs.<sup>33,34</sup> In this context, the inhibition of range of six MMPs with structural variation in their S1' binding pocket<sup>57</sup> (deep: MMP-3 and -14; intermediate: MMP-2 and -8; shallow: MMP-1 and -7) was analyzed for 27. As we previously demonstrated for other *N*-aryl mercaptoacetamides,<sup>35</sup> 27 did not inhibit this broad range of MMPs at concentrations up to 100  $\mu\text{M}$  either (Figure S5).

In contrast to these antitargets, bacterial metallo- $\beta$ -lactamases represent attractive additional targets, since cleavage of  $\beta$ -lactam antibiotics by these proteases is a crucial mechanism of antibiotic resistance.<sup>58</sup> Inhibition of metallo- $\beta$ -lactamases by thiol-based inhibitors has been reported to restore the activity of  $\beta$ -lactam antibiotics.<sup>59–62</sup> To this end, we tested IMP-7, a metallo- $\beta$ -lactamase present in clinical isolates of *P. aeruginosa* for *in vitro* inhibition by prodrug 27 and thiol 36. Interestingly, the enzyme was found to be inhibited by 27 ( $1.16 \pm 0.07 \mu\text{M}$ ) as well as 36 ( $0.86 \pm 0.06 \mu\text{M}$ ).

**Cytotoxicity Assays.** For a potential therapeutic application in humans, we analyzed the cytotoxicity of thiocarbamate 17 as well as thiols 35 and 36 toward two different human cell lines (Table 3). As described previously,<sup>35</sup> the mercaptoaceta-

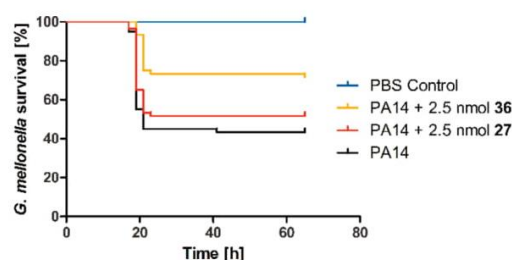
**Table 3. Cytotoxicity of 17, 35, and 36 in HEP G2 and HEK293 Cells**

Cp.	concn [ $\mu\text{M}$ ]	reduction of viability [%]	
		HEP G2	HEK293
17	100	26 $\pm$ 16	51 $\pm$ 12
35	100	28 $\pm$ 12 <sup>a</sup>	51 $\pm$ 4
36	100	25 $\pm$ 3	22 $\pm$ 5
doxorubicin	1	50 $\pm$ 5 <sup>a</sup>	49 $\pm$ 7
rifampicin	100	29 $\pm$ 5 <sup>a</sup>	19 $\pm$ 3
batimastat	100	13 $\pm$ 7 <sup>a</sup>	2 $\pm$ 2

<sup>a</sup>Values taken from Schönauer et al.<sup>35</sup>

mides had only a low cytotoxic effect on HEP G2 cells at a concentration as high as 100  $\mu\text{M}$ . Likewise, the cytotoxic effect on HEK293 cells was also low at this high concentration.

**Galleria mellonella Infection Model.** Stimulated by the promising *in vitro* results, the *in vivo* efficacy of our compounds was investigated. Infection models with *Galleria mellonella* larvae have previously been employed by us to assess novel treatment options for *Pseudomonas aeruginosa*-induced infections<sup>63,64</sup> and were reported to highly correlate with mouse models.<sup>65</sup> We analyzed the effect of prodrug 27 and active thiol 36 by coinjecting them with PA14. The larvae were challenged with PA14 concentrations corresponding to two times the LD<sub>50</sub>.<sup>65</sup> Compared to PA14-infected larvae receiving no treatment, 2.5 nmol of 36 significantly increased the survival of *G. mellonella* from 43% to 72% after 65 h (Figure 8). Notably, not only the survival time was increased, as reported for LasB inhibitors in a *C. elegans* model,<sup>34</sup> but also the survival rate. However, administering the same concentration of thiocarbamate prodrug did not lead to a significant effect, indicating an insufficient release of the thiol in *G. mellonella* hemolymph. These findings prove the potential of LasB inhibition to effectively reduce *P. aeruginosa* pathogenicity.



**Figure 8.** Survival curves for PA14-challenged *Galleria mellonella* larvae receiving treatment with 2.5 nmol 27 or 36 in comparison to larvae receiving treatment with PBS only. Curves represent results from at least three independent measurements. The survival rate was significantly higher for larvae treated with 36 ( $p = 0.0003$ , log-rank test) but not with 27. The survival rate for larvae treated with the compounds in PBS was 100%.

## CONCLUSION

In this study, a functional screening of a protease-inhibitor enriched library and a fragment library was performed using a well-established FRET assay, with the aim to expand the chemical space of LasB inhibitors. An LC-MS-based counter-screen was developed, allowing identification of false-positives. Only one real screening hit was identified, an *N*-aryl mercaptoacetamide. Compounds of this class are known to be inhibitors of LasB as well as of ColH from *C. histolyticum*. We analyzed the inhibition of a further 35 derivatives of this class of fragment-like molecules. Interestingly, the SAR was inverse to the activity profile we discovered for ColH, allowing rational optimization of the compound class to selectively inhibit either LasB or ColH. The X-ray crystal structure of a LasB–inhibitor complex was solved, revealing a primed binding mode of two thiol molecules to the active site of the protease. Importantly, our results show that inhibitor binding does not necessarily lead to a closure of the binding pocket, as it has been described for thermolysin-like proteases. These findings pave the way for the development of novel LasB inhibitors targeting the open conformation of the enzyme, providing an important starting point for lead optimization. The structural information obtained by X-ray crystallography was used to grow the fragments by introducing benzyl or alkyl groups. These compounds indeed inhibited LasB, yet the activity was not improved. Supposedly, the conformation of the *N*-substituted compounds is not ideal to replace the second inhibitor molecule in the binding pocket. With the  $\beta$ -lactamase IMP-7, we discovered an additional target of the *N*-aryl mercaptoacetamides, which is attractive for the development of novel anti-infectives. Furthermore, *in vivo* efficacy was demonstrated using a *Galleria mellonella* infection model. The survival rate of PA14-infected larvae was increased significantly when treated with our best thiol. Considering the fragment-like character of our inhibitors and the fact that they represent antivirulence agents instead of traditional antibiotics, this is a remarkable effect. These results underline the potential of reducing bacterial pathogenicity to develop novel antibacterial drugs, which are urgently needed to combat antibiotic resistance. *N*-aryl mercaptoacetamides prove particularly interesting for the development of such drugs, since they display high selectivity toward human MMPs.

## ■ EXPERIMENTAL SECTION

**Expression and Purification of LasB.** *Pseudomonas aeruginosa* PA14 were grown in lysogeny broth medium for 72 h at 37 °C. Cells were removed by centrifugation (5000 rpm, 4 °C, 60 min), and the supernatant was filtered through a bottle-top-filter (0.20  $\mu\text{m}$ ). Purification was performed according to the method described by Morihara et al.<sup>66</sup> with few modifications. The precipitate formed by acetone treatment was dissolved in water and dialyzed against buffer A (20 mM Tris, pH 8.0). The dialyzed sample was then loaded onto a Hitrap Q HP column (GE Healthcare, Little Chalfont, UK) and washed back to baseline in buffer A, before being eluted with a gradient from buffer A into buffer A2 (20 mM Tris, pH 8.0, 1 M NaCl). The fractions with the strongest activity were pooled together (*in vitro* assay described below) and loaded onto a Superdex 200 gel filtration column (GE Healthcare) that was pre-equilibrated with buffer B (20 mM Tris, pH 8.0, 2 mM  $\text{CaCl}_2$ ). Protein purity and activity was assessed by SDS-PAGE and the *in vitro* inhibition assay.

**In Vitro Inhibition Assay.** For the initial screening, Elastase was purchased from Elastin Products Company (Owensville, MO, USA) or Merck (Darmstadt, Germany). Later, purified elastase prepared according to the procedure described above was used. The fluorogenic substrate 2-Aminobenzoyl-Ala-Gly-Leu-Ala-4-Nitrobenzylamide<sup>36</sup> was purchased from Peptides International (Louisville, KY, USA). Fluorescence intensity was measured for 60 min at 37 °C in black 384-well microtiter plates (Greiner BioOne, Kremsmünster, Austria) using a CLARIOstar microplate reader (BMG Labtech, Ortenberg, Germany) with an excitation wavelength of  $340 \pm 15$  nm and an emission wavelength of  $415 \pm 20$  nm. The assay was performed in a final volume of 50  $\mu\text{L}$  of assay buffer (50 mM Tris, pH 7.2, 2.5 mM  $\text{CaCl}_2$ , 0.075% Pluronic F-127, 5% DMSO) containing LasB at a final concentration of 10 nM (commercial batch) or 0.3 nM (purified batch) and the substrate at 150  $\mu\text{M}$ . Before substrate addition, compounds were preincubated with the enzyme for 15 min at 37 °C. Experiments were performed in duplicates and repeated for at least two times. Blank controls without enzyme were performed. After blank subtraction, the slope of samples containing inhibitors ( $v$ ) was divided by the slope of a simultaneously started uninhibited enzymatic reaction ( $v_0$ ).  $\text{IC}_{50}$  values were determined with nonlinear regression using GraphPad Prism 5 (Graph Pad Software, San Diego, CA, USA) and are given as mean values  $\pm$  standard deviation (SD). The slope factor was constrained to 1.

**LC-MS-Based Readout for the FRET Assay.** The FRET-based fluorescence assay was performed according to the procedure described above using 50 mM Tris, pH 7.2. At the end of the measurement, the enzymatic reaction was stopped by adding formic acid at a final concentration of 2%. Simultaneously, amitryptiline was added as internal standard. The resulting mixture was diluted 1:10 in a mixture of 10% acetonitrile in Milli-Q water containing 2% formic acid, resulting in a 10  $\mu\text{M}$  amitryptiline concentration. The analyses were performed using a TF UltiMate 3000 binary RSLC UHPLC (Thermo Fisher, Dreieich, Germany) equipped with a degasser, a binary pump, an autosampler, and a thermostated column compartment and a MWD, coupled to a TF TSQ Quantum Access Max mass spectrometer with heated electrospray ionization source (HESI-II). For gradient elution, an Accucore RP-MS column (150  $\times$  2.1 mm, 2.6  $\mu\text{m}$ , Thermo

Fisher, Dreieich, Germany) was used with a mobile phase consisting of acetonitrile containing 1% formic acid (FA; v/v; eluent A) and water containing 1% FA (v/v; eluent B) at a flow rate of 400  $\mu\text{L}/\text{min}$  under the following conditions: 0–0.9 min 10% A, 0.9–2.5 min 10–50% A, 2.5–5.5 min 50–70% A, 5.5–6 min hold, and 6–6.5 min 10% A, giving a total run time of 6.5 min. The injection volume was 10  $\mu\text{L}$ . The divert valve was set to 1.6 min. The autosampler temperature was set to 6 °C. The following MS conditions were used: electrospray ionization (ESI), positive mode, sheath gas, nitrogen at a flow rate of 35 arbitrary units; auxiliary gas, nitrogen at flow rate of 10 arbitrary units; vaporizer temperature, 50 °C; ion transfer capillary temperature, 270 °C; capillary offset, 15 V; spray voltage, 3000 V. The mass spectrometer was operated in the SIM mode with the following masses: 4,  $m/z$  266.1 (tube lens offset 120 V); amitryptiline,  $m/z$  278.1 (tube lens offset 90 V); 5,  $m/z$  337.0 (tube lens offset 120 V); 3,  $m/z$  584.1 (tube lens offset 120 V) with a scan width of  $m/z$  2.0 and a scan time of 0.1 s, respectively. Measurements were performed in duplicates and repeated for at least two times. Observed retention times were as follows: 4, 2.00 min; 5, 3.52 min; amitryptiline, 4.02 min; 3, 4.37 min. MS-peak areas were determined using TF Xcalibur Software. Peak areas were normalized by ISTD peak area (giving A) and divided by the peak area of the respective ISTD normalized sample without inhibitor ( $A_0$ ).  $\text{IC}_{50}$  values were determined using the method described above (GraphPad Prism 5 software).

**LC-MS-Based Stability Assay.** The assay was performed as described previously<sup>35</sup> using 50 mM Tris, pH 7.2, and a temperature of 37 °C.

**Screening Library.** The protease inhibitor enriched screening library (ActiTarget-P) was purchased from TimTec (Newark, DE, USA). Fragments were obtained from the Maybridge Fragment Library (Maybridge, Altrincham, UK). Compounds were stored as DMSO stock solutions.

**Chemistry.** All reagents were used from commercial suppliers without further purification. Procedures were not optimized regarding yield. NMR spectra were recorded on a Bruker Fourier 300 (300 MHz) spectrometer. Chemical shifts are given in parts per million (ppm) and referenced against the residual proton,  $^1\text{H}$ , or carbon,  $^{13}\text{C}$ , resonances of the >99% deuterated solvents as internal reference. Coupling constants ( $J$ ) are given in Hertz. Data are reported as follows: chemical shift, multiplicity (s = singlet, d = doublet, t = triplet, m = multiplet, br = broad and combinations of these) coupling constants, and integration. Mass spectrometry was performed on a SpectraSystems-MSQ LCMS system (Thermo Fisher, Dreieich, Germany). Flash chromatography was performed using the automated flash chromatography system CombiFlash Rf+ (Teledyne Isco, Lincoln, NE, USA) equipped with RediSepRf silica columns (Axel Semrau, Sprockhövel Germany) or Chromabond Flash C18 columns (Macherey-Nagel, Düren, Germany). Purity of compounds synthesized by us was determined by LCMS using the area percentage method on the UV trace recorded at a wavelength of 254 nm and found to be >95%. Thiols and thiocarbamates were synthesized according to the procedures described previously.<sup>35</sup> *N*-benzyl and *N*-aryl mercaptoacetamides were synthesized according to Schemes S1 and S2 using General Procedures 1–3 as described in more detail in the Supporting Information.

**Human MMP Inhibition Assay.** The catalytic domains of MMP-1, -2, -3, -7, -8, and -14 along with the SensoLyte 520 Generic MMP Activity Kit were purchased from AnaSpec



(Fremont, CA, USA). The assay was performed as described previously using Batimastat as a positive control,<sup>35</sup> according to the guidelines of the manufacturer.

**$\beta$ -Lactamase Inhibition Assay.** Activity assays for IMP-7 were carried out, as described by Klingler et al.<sup>60</sup> Final protein concentrations of 0.1 nM in a 50 mM HEPES buffer (pH 7.5, 0.01% Triton X-100). Substrate (Fluorocillin (Invitrogen, Darmstadt, Germany)) was dissolved in assay buffer to a final concentration of 888 nM. Test compounds were dissolved and prediluted in DMSO (final concentration: 1%). In a black polystyrol 96-well plate (Corning), an amount of 1  $\mu$ L of the respective inhibitor solution at different concentrations was incubated with 89  $\mu$ L of IMP-7 containing buffer for 30 min at room temperature. Ten  $\mu$ L of substrate solution was added. The readout of the emitted fluorescence was started immediately (45 s for 30 cycles) using a Tecan Infinite F200Pro (Tecan Group Ltd.; excitation at 495 nm and emission at 525 nm). Blank controls were performed without enzyme. Positive controls were performed with enzyme but without inhibitor. The inhibitory activity of each test compound was measured in three independent experiments. For calculation of IC<sub>50</sub> values, data obtained from measurements with eight different inhibitor concentrations were used. For the evaluation of the sigmoidal dose response equation (variable slope with four parameters), GraphPad Prism 5 (GraphPad Software, La Jolla, CA, USA) was used.

**Cytotoxicity Assays.** Hep G2 or HEK293 cells ( $2 \times 10^5$  cells per well) were seeded in 24-well, flat-bottomed plates. Culturing of cells, incubations, and OD measurements were performed as described previously<sup>67</sup> with small modifications. Twenty-four hours after seeding the cells, the incubation was started by the addition of compounds in a final DMSO concentration of 1%. The living cell mass was determined after 48 h. At least two independent measurements were performed for each compound.

**Galleria mellonella Virulence Assay.** The virulence assay was performed as described by Lu et al.<sup>63</sup> with some modifications: *Galleria mellonella* larvae (TruLarv) were purchased from BioSystems Technology (Exeter, United Kingdom). Injections were performed using a LA120 syringe pump (Landgraf Laborsysteme, Langenhagen, Germany) equipped with 1 mL Injekt-F tuberculin syringes (B. Braun, Melsungen, Germany) and Sterican 0.30  $\times$  12 mm, 30G  $\times$  1.5 needles (B. Braun). The following treatment conditions were applied: (a) sterile PBS solution, (b) PA14 suspension, (c) 2.5 nmol of **27** in "b", (d) 2.5 nmol of **36** in "b", (e) 2.5 nmol of **27** in "a", and (f) 1.25 nmol of **36** in "a". Samples a–e contained 1% DMSO and f, 0.5%. For each treatment, data from at least three independent measurements were combined.

**X-ray Crystallography.** LasB was concentrated to 10–12 mg/mL and mixed with inhibitor **27** at a final concentration of 1 mM. Complex crystals were grown by the sitting drop method using a reservoir solution containing 0.2 M magnesium chloride and 30% (w/v) PEG 3350. Crystals were cryoprotected in glycerol, and diffraction data was collected from single crystals at 100 K at beamline ID23-2 (ESRF) at a wavelength of 0.873 Å. Data was processed using Xia2,<sup>68</sup> and the structure was solved using PHASER<sup>69</sup> molecular replacement with *Pseudomonas aeruginosa* elastase (PAE, PDB ID 1EZM) as a search model. The solution was manually rebuilt with COOT<sup>70</sup> and refined using PHENIX<sup>71</sup> and Refmac5.<sup>72</sup> The final refined structure of LasB in complex with compound **27** was deposited in the Protein Data Bank (PDB) as entry 6F8B.

**Molecular Modeling.** Molecular modeling was performed with Molecular Operating Environment 2015.10 (MOE) software (Chemical Computing Group, Montreal, Canada) using standard parameters. In the cocrystal structure of LasB and **36**, inhibitor molecule **36**<sup>B</sup> was removed and **36**<sup>A</sup> was grown using the Builder function. The final structure was energy-minimized using the QuickPrep function. AMBER10-EHT was used as a force field.

## ■ ASSOCIATED CONTENT

### 📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsinfectdis.8b00010.

Supporting Tables S1 and S2, Schemes 1 and 2, and Figures S1–S5, giving structures and stability data of inactive compounds, X-ray data collection and refinement statistics, FRET and LC-MS results for phosphoramidon, stability data for **68** and **74**, MMP inhibition assay, experimental procedures, and spectral data for synthetic compounds (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: Rolf.Hartmann@helmholtz-hzi.de.

### ORCID

Andreas M. Kany: 0000-0001-7580-3658  
Jörg Hauptenthal: 0000-0003-3991-2800  
Samir Yahiaoui: 0000-0001-5134-5007  
Ewgenij Proschak: 0000-0003-1961-1859  
Jesko Köhnke: 0000-0002-7153-1365  
Rolf W. Hartmann: 0000-0002-5871-5231

### Author Contributions

<sup>1</sup>A.M.K. and A.S. contributed equally.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We thank Jeannine Jung, Simone Amann, Tabea Schramm, and Lilia Weizel for technical support, Dr. Jens Eberhard and Dr. Giuseppe Allegretta for help with the establishment of the LC-MS assay, Dr. Teresa Röhrig for support with the *G. mellonella* assay, and Dr. Martin Empting for help with molecular modeling. E.P. thanks German Research Foundation (DFG, Heisenberg-Professur PR1405/4-1) for financial support.

## ■ REFERENCES

- (1) World Health Organization. (2017) Antibacterial Agents in Clinical Development. In *An analysis of the antibacterial clinical development pipeline, including tuberculosis*, World Health Organization, Geneva.
- (2) Taubes, G. (2008) The bacteria fight back. *Science* 321, 356–361.
- (3) Hancock, R. E. W., and Speert, D. P. (2000) Antibiotic resistance in *Pseudomonas aeruginosa*. Mechanisms and impact on treatment. *Drug Resist. Updates* 3, 247–255.
- (4) Sordé, R., Pahissa, A., and Rello, J. (2011) Management of refractory *Pseudomonas aeruginosa* infection in cystic fibrosis. *Infect. Drug Resist.* 4, 31–41.
- (5) Mesaros, N., Nordmann, P., Plésiat, P., Roussel-Delvallez, M., van Eldere, J., Glupczynski, Y., van Laethem, Y., Jacobs, F., Lebecque, P., Malfroot, A., Tulkens, P. M., and van Bambeke, F. (2007)

*Pseudomonas aeruginosa*. Resistance and therapeutic options at the turn of the new millennium. *Clin. Microbiol. Infect.* 13, 560–578.

- (6) Cooper, M. A., and Shlaes, D. (2011) Fix the antibiotics pipeline. *Nature* 472, 32.
- (7) Dickey, S. W., Cheung, G. Y. C., and Otto, M. (2017) Different drugs for bad bugs. Antivirulence strategies in the age of antibiotic resistance. *Nat. Rev. Drug Discovery* 16, 457–471.
- (8) Heras, B., Scanlon, M. J., and Martin, J. L. (2015) Targeting virulence not viability in the search for future antibacterials. *Br. J. Clin. Pharmacol.* 79, 208–215.
- (9) Lewis, K. (2013) Platforms for antibiotic discovery. *Nat. Rev. Drug Discovery* 12, 371–387.
- (10) Rasko, D. A., and Sperandio, V. (2010) Anti-virulence strategies to combat bacteria-mediated disease. *Nat. Rev. Drug Discovery* 9, 117–128.
- (11) Strateva, T., and Mitov, I. (2011) Contribution of an arsenal of virulence factors to pathogenesis of *Pseudomonas aeruginosa* infections. *Ann. Microbiol.* 61, 717–732.
- (12) Wagner, S., Sommer, R., Hinsberger, S., Lu, C., Hartmann, R. W., Empting, M., and Titz, A. (2016) Novel Strategies for the Treatment of *Pseudomonas aeruginosa* Infections. *J. Med. Chem.* 59, 5929–5969.
- (13) Storz, M. P., Maurer, C. K., Zimmer, C., Wagner, N., Brengel, C., de Jong, J. C., Lucas, S., Müsken, M., Häussler, S., Steinbach, A., and Hartmann, R. W. (2012) Validation of PqsD as an anti-biofilm target in *Pseudomonas aeruginosa* by development of small-molecule inhibitors. *J. Am. Chem. Soc.* 134, 16143–16146.
- (14) Kamal, A. A. M., Maurer, C. K., Allegretta, G., Haupenthal, J., Empting, M., and Hartmann, R. W. (2017) Quorum Sensing Inhibitors as Pathoblockers for *Pseudomonas aeruginosa* Infections: A New Concept in Anti-Infective Drug Discovery. In *Topics in Medicinal Chemistry*, pp 1–26, Springer, Berlin, Heidelberg.
- (15) Wretling, B., and Pavlovskis, O. R. (1983) *Pseudomonas aeruginosa* elastase and its role in *Pseudomonas* infections. *Clin. Infect. Dis.* 5 (Suppl 5), S998–S1004.
- (16) Morihara, K. (1964) Production of elastase and proteinase by *Pseudomonas aeruginosa*. *J. Bacteriol.* 88, 745–757.
- (17) Thayer, M. M., Flaherty, K. M., and McKay, D. B. (1991) Three-dimensional structure of the elastase of *Pseudomonas aeruginosa* at 1.5-Å resolution. *J. Biol. Chem.* 266, 2864–2871.
- (18) Heck, L. W., Morihara, K., McRae, W. B., and Miller, E. J. (1986) Specific cleavage of human type III and IV collagens by *Pseudomonas aeruginosa* elastase. *Infect. Immun.* 51, 115–118.
- (19) Golovkine, G., Faudry, E., Bouillot, S., Voulhoux, R., Attrée, I., and Huber, P. (2014) VE-cadherin cleavage by LasB protease from *Pseudomonas aeruginosa* facilitates type III secretion system toxicity in endothelial cells. *PLoS Pathog.* 10, e1003939.
- (20) Azghani, A. O. (1996) *Pseudomonas aeruginosa* and epithelial permeability. Role of virulence factors elastase and exotoxin A. *Am. J. Respir. Cell Mol. Biol.* 15, 132–140.
- (21) Holder, I. A., and Wheeler, R. (1984) Experimental studies of the pathogenesis of infections owing to *Pseudomonas aeruginosa*. Elastase, an IgG protease. *Can. J. Microbiol.* 30, 1118–1124.
- (22) Parmely, M., Gale, A., Clabaugh, M., Horvat, R., and Zhou, W. W. (1990) Proteolytic inactivation of cytokines by *Pseudomonas aeruginosa*. *Infect. Immun.* 58, 3009–3014.
- (23) Mariencheck, W. I., Alcorn, J. F., Palmer, S. M., and Wright, J. R. (2003) *Pseudomonas aeruginosa* elastase degrades surfactant proteins A and D. *Am. J. Respir. Cell Mol. Biol.* 28, 528–537.
- (24) Schmidtchen, A., Holst, E., Tapper, H., and Björck, L. (2003) Elastase-producing *Pseudomonas aeruginosa* degrade plasma proteins and extracellular products of human skin and fibroblasts, and inhibit fibroblast growth. *Microb. Pathog.* 34, 47–55.
- (25) Dulon, S., Leduc, D., Cottrell, G. S., D'Alayer, J., Hansen, K. K., Bunnnett, N. W., Hollenberg, M. D., Pidard, D., and Chignard, M. (2005) *Pseudomonas aeruginosa* elastase disables proteinase-activated receptor 2 in respiratory epithelial cells. *Am. J. Respir. Cell Mol. Biol.* 32, 411–419.
- (26) Graef, F., Vukosavljevic, B., Michel, J.-P., Wirth, M., Ries, O., De Rossi, C., Windbergs, M., Rosilio, V., Ducho, C., Gordon, S., and Lehr, C.-M. (2016) The bacterial cell envelope as delimiter of anti-infective bioavailability - An in vitro permeation model of the Gram-negative bacterial inner membrane. *J. Controlled Release* 243, 214–224.
- (27) Adekoya, O. A., and Sylte, I. (2009) The thermolysin family (M4) of enzymes. Therapeutic and biotechnological potential. *Chem. Biol. Drug Des.* 73, 7–16.
- (28) Holland, D. R., Tronrud, D. E., Pley, H. W., Flaherty, K. M., Stark, W., Jansonius, J. N., McKay, D. B., and Matthews, B. W. (1992) Structural comparison suggests that thermolysin and related neutral proteases undergo hinge-bending motion during catalysis. *Biochemistry* 31, 11310–11316.
- (29) Adekoya, O. A., Sjøli, S., Wuxiuer, Y., Bilot, I., Marques, S. M., Santos, M. A., Nuti, E., Cercignani, G., Rossello, A., Winberg, J.-O., and Sylte, I. (2015) Inhibition of pseudolysin and thermolysin by hydroxamate-based MMP inhibitors. *Eur. J. Med. Chem.* 89, 340–348.
- (30) Fullagar, J. L., Garner, A. L., Struss, A. K., Day, J. A., Martin, D. P., Yu, J., Cai, X., Janda, K. D., and Cohen, S. M. (2013) Antagonism of a zinc metalloprotease using a unique metal-chelating scaffold. Tropolones as inhibitors of *P. aeruginosa* elastase. *Chem. Commun. (Cambridge, U. K.)* 49, 3197–3199.
- (31) Burns, F. R., Paterson, C. A., Gray, R. D., and Wells, J. T. (1990) Inhibition of *Pseudomonas aeruginosa* elastase and *Pseudomonas keratitis* using a thiol-based peptide. *Antimicrob. Agents Chemother.* 34, 2065–2069.
- (32) Kessler, E., Israel, M., Landshman, N., Chechick, A., and Blumberg, S. (1982) In vitro inhibition of *Pseudomonas aeruginosa* elastase by metal-chelating peptide derivatives. *Infect. Immun.* 38, 716–723.
- (33) Cathcart, G. R. A., Quinn, D., Greer, B., Harriott, P., Lynas, J. F., Gilmore, B. F., and Walker, B. (2011) Novel inhibitors of the *Pseudomonas aeruginosa* virulence factor LasB. A potential therapeutic approach for the attenuation of virulence mechanisms in pseudomonal infection. *Antimicrob. Agents Chemother.* 55, 2670–2678.
- (34) Zhu, J., Cai, X., Harris, T. L., Gooyit, M., Wood, M., Lardy, M., and Janda, K. D. (2015) Disarming *Pseudomonas aeruginosa* virulence factor LasB by leveraging a *Caenorhabditis elegans* infection model. *Chem. Biol.* 22, 483–491.
- (35) Schönauer, E., Kany, A. M., Haupenthal, J., Hüsecken, K., Hoppe, I. J., Voos, K., Yahiaoui, S., Elsässer, B., Ducho, C., Brandstetter, H., and Hartmann, R. W. (2017) Discovery of a Potent Inhibitor Class with High Selectivity toward Clostridial Collagenases. *J. Am. Chem. Soc.* 139, 12696–12703.
- (36) Nishino, N., and Powers, J. C. (1980) *Pseudomonas aeruginosa* elastase. Development of a new substrate, inhibitors, and an affinity ligand. *J. Biol. Chem.* 255, 3482–3486.
- (37) Baell, J. B., and Holloway, G. A. (2010) New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* 53, 2719–2740.
- (38) Lakowicz, J. R. (2010) *Principles of Fluorescence Spectroscopy*, 3rd ed., Springer, New York, NY.
- (39) Kongkamnerd, J., Milani, A., Cattoli, G., Terregino, C., Capua, I., Beneduce, L., Gallotta, A., Pengo, P., Fassina, G., Monthakantirat, O., Umehara, K., De-Eknamkul, W., and Miertus, S. (2011) The quenching effect of flavonoids on 4-methylumbelliferone, a potential pitfall in fluorimetric neuraminidase inhibition assays. *J. Biomol. Screening* 16, 755–764.
- (40) Shapiro, A. B., Walkup, G. K., and Keating, T. A. (2009) Correction for interference by test samples in high-throughput assays. *J. Biomol. Screening* 14, 1008–1016.
- (41) Morihara, K., and Hiroshige, T. (1971) Comparative study of various neutral proteinases from microorganisms. Specificity with oligopeptides. *Arch. Biochem. Biophys.* 146, 291–296.
- (42) Morihara, K., and Tsuzuki, H. (1978) Phosphoramidon as an inhibitor of elastase from *Pseudomonas aeruginosa*. *Jpn. J. Exp. Med.* 48, 81–84.

- (43) Pramar, Y., Das Gupta, V., and Bethea, C. (1992) Stability of captopril in some aqueous systems. *J. Clin. Pharm. Ther.* 17, 185–189.
- (44) Labute, P. (2009) Protonate3D. Assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins: Struct., Funct., Genet.* 75, 187–205.
- (45) McKay, D. B., and Overgaard, M. T. (2009) *Pseudomonas aeruginosa* elastase with phosphoramidon; <https://www.rcsb.org/structure/3DBK>.
- (46) Bitto, E., and McKay, D. B. (2004) Elastase of *Pseudomonas aeruginosa* with an inhibitor; <https://www.rcsb.org/structure/1u4g>.
- (47) Kim, Y.-G., Lim, H. N., and Lee, K.-J. (2009) A new route to allyl thiols and allyl thiocarbamates from Baylis-Hillman adducts. *J. Heterocycl. Chem.* 46, 23–27.
- (48) Cathcart, J., Pulkoski-Gross, A., and Cao, J. (2015) Targeting Matrix Metalloproteinases in Cancer. *Bringing New Life to Old Ideas. Genes Dis* 2, 26–34.
- (49) Dufour, A., and Overall, C. M. (2013) Missing the target. Matrix metalloproteinase antitargets in inflammation and cancer. *Trends Pharmacol. Sci.* 34, 233–242.
- (50) Overall, C. M., and López-Otín, C. (2002) Strategies for MMP inhibition in cancer. Innovations for the post-trial era. *Nat. Rev. Cancer* 2, 657–672.
- (51) Sbardella, D., Fasciglione, G. F., Gioia, M., Ciaccio, C., Tundo, G. R., Marini, S., and Coletta, M. (2012) Human matrix metalloproteinases. An ubiquitous class of enzymes involved in several pathological processes. *Mol. Aspects Med.* 33, 119–208.
- (52) Turk, B. (2006) Targeting proteases. Successes, failures and future prospects. *Nat. Rev. Drug Discovery* 5, 785–799.
- (53) Vandenbroucke, R. E., and Libert, C. (2014) Is there new hope for therapeutic matrix metalloproteinase inhibition? *Nat. Rev. Drug Discovery* 13, 904–927.
- (54) Ilies, M., Banciu, M. D., Scozzafava, A., Ilies, M. A., Caproiu, M. T., and Supuran, C. T. (2003) Protease inhibitors. Synthesis of bacterial collagenase and matrix metalloproteinase inhibitors incorporating arylsulfonyleureido and 5-dibenzo-suberenyl/suberyl moieties. *Bioorg. Med. Chem.* 11, 2227–2239.
- (55) Scozzafava, A., and Supuran, C. T. (2002) Protease inhibitors. Synthesis of matrix metalloproteinase and bacterial collagenase inhibitors incorporating 5-amino-2-mercapto-1,3,4-thiadiazole zinc binding functions. *Bioorg. Med. Chem. Lett.* 12, 2667–2672.
- (56) Sjöli, S., Nuti, E., Camodeca, C., Bilotto, I., Rossello, A., Winberg, J.-O., Sylte, I., and Adekoya, O. A. (2016) Synthesis, experimental evaluation and molecular modelling of hydroxamate derivatives as zinc metalloproteinase inhibitors. *Eur. J. Med. Chem.* 108, 141–153.
- (57) Park, H. I., Jin, Y., Hurst, D. R., Monroe, C. A., Lee, S., Schwartz, M. A., and Sang, Q.-X. A. (2003) The intermediate S1' pocket of the endometase/matrixin-2 active site revealed by enzyme inhibition kinetic studies, protein sequence analyses, and homology modeling. *J. Biol. Chem.* 278, 51646–51653.
- (58) Bush, K., and Fisher, J. F. (2011) Epidemiological expansion, structural studies, and clinical challenges of new  $\beta$ -lactamases from gram-negative bacteria. *Annu. Rev. Microbiol.* 65, 455–478.
- (59) Buynak, J. D., Chen, H., Vogeti, L., Gadhachanda, V. R., Buchanan, C. A., Palzkill, T., Shaw, R. W., Spencer, J., and Walsh, T. R. (2004) Penicillin-derived inhibitors that simultaneously target both metallo- and serine-beta-lactamases. *Bioorg. Med. Chem. Lett.* 14, 1299–1304.
- (60) Klingler, F.-M., Wichelhaus, T. A., Frank, D., Cuesta-Bernal, J., El-Delik, J., Müller, H. F., Sjtus, H., Göttig, S., Koenigs, A., Pos, K. M., Pogoryelov, D., and Proschak, E. (2015) Approved Drugs Containing Thiols as Inhibitors of Metallo- $\beta$ -lactamases. Strategy To Combat Multidrug-Resistant Bacteria. *J. Med. Chem.* 58, 3626–3630.
- (61) Tehrani, K. H. M. E., and Martin, N. I. (2017) Thiol-Containing Metallo- $\beta$ -Lactamase Inhibitors Resensitize Resistant Gram-Negative Bacteria to Meropenem. *ACS Infect. Dis.* 3, 711–717.
- (62) Büttner, D., Kramer, J. S., Klingler, F.-M., Wittmann, S. K., Hartmann, M. R., Kurz, C. G., Kohnhäuser, D., Weizel, L., Brüggerhoff, A., Frank, D., Steinhilber, D., Wichelhaus, T. A., Pogoryelov, D., and Proschak, E. (2017) Challenges in the Development of a Thiol-Based Broad-Spectrum Inhibitor for Metallo- $\beta$ -Lactamases. *ACS Infect. Dis.*, DOI: 10.1021/acinfecdis.7b00129.
- (63) Lu, C., Maurer, C. K., Kirsch, B., Steinbach, A., and Hartmann, R. W. (2014) Overcoming the unexpected functional inversion of a PqsR antagonist in *Pseudomonas aeruginosa*. An in vivo potent antivirulence agent targeting pqs quorum sensing. *Angew. Chem., Int. Ed.* 53, 1109–1112.
- (64) Thomann, A., de Mello Martins, A. G. G., Bregel, C., Empting, M., and Hartmann, R. W. (2016) Application of Dual Inhibition Concept within Looped Autoregulatory Systems toward Antivirulence Agents against *Pseudomonas aeruginosa* Infections. *ACS Chem. Biol.* 11, 1279–1286.
- (65) Jander, G., Rahme, L. G., and Ausubel, F. M. (2000) Positive correlation between virulence of *Pseudomonas aeruginosa* mutants in mice and insects. *J. Bacteriol.* 182, 3843–3845.
- (66) Morihara, K., Tsuzuki, H., Oka, T., Inoue, H., and Ebata, M. (1965) *Pseudomonas aeruginosa* elastase: isolation, crystallization, and preliminary characterization. *J. Biol. Chem.* 240, 3295–3304.
- (67) Hauptenthal, J., Baehr, C., Zeuzem, S., and Piiper, A. (2007) RNase A-like enzymes in serum inhibit the anti-neoplastic activity of siRNA targeting polo-like kinase 1. *Int. J. Cancer* 121, 206–210.
- (68) Winter, G. (2010) xia2. An expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* 43, 186–190.
- (69) McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674.
- (70) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 66, 486–501.
- (71) Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX. A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 66, 213–221.
- (72) Skubák, P., Murshudov, G. N., and Pannu, N. S. (2004) Direct incorporation of experimental phase information in model refinement. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 60, 2196–2201.

# Hydroxamic acid-based LasB inhibitors

## Tackling *Pseudomonas aeruginosa* Virulence by a Hydroxamic Acid-Based LasB Inhibitor

Andreas M. Kany,<sup>†</sup> Asfandyar Sikandar,<sup>‡</sup> Samir Yahiaoui,<sup>†</sup> Jörg Haupenthal,<sup>†</sup> Isabell Walter,<sup>†</sup> Martin Empting,<sup>†</sup> Jesko Köhnke,<sup>‡</sup> and Rolf W. Hartmann<sup>\*,†,§</sup>

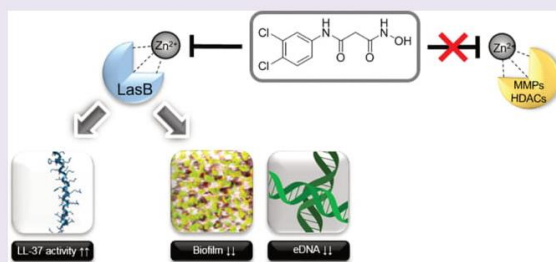
<sup>†</sup>Department of Drug Design and Optimization, Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Campus E8.1, 66123 Saarbrücken, Germany

<sup>‡</sup>Workgroup Structural Biology of Biosynthetic Enzymes, Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Campus E8.1, 66123 Saarbrücken, Germany

<sup>§</sup>Department of Pharmacy, Pharmaceutical and Medicinal Chemistry, Saarland University, Campus E8.1, 66123 Saarbrücken, Germany

### Supporting Information

**ABSTRACT:** In search of novel antibiotics to combat the challenging spread of resistant pathogens, bacterial proteases represent promising targets for pathoblocker development. A common motif for protease inhibitors is the hydroxamic acid function, yet this group has often been related to unspecific inhibition of various metalloproteases. In this work, the inhibition of LasB, a harmful zinc metalloprotease secreted by *Pseudomonas aeruginosa*, through a hydroxamate derivative is described. The present inhibitor was developed based on a recently reported, highly selective thiol scaffold. Using X-ray crystallography, the lack of inhibition of a range of human matrix metalloproteases could be attributed to a distinct binding mode sparing the S1' pocket. The inhibitor was shown to restore the effect of the antimicrobial peptide LL-37, decrease the formation of *P. aeruginosa* biofilm and, for the first time for a LasB inhibitor, reduce the release of extracellular DNA. Hence, it is capable of disrupting several important bacterial resistance mechanisms. These results highlight the potential of protease inhibitors to fight bacterial infections and point out the possibility to achieve selective inhibition even with a strong zinc anchor.



Proteases have proven to be attractive targets for the treatment of various diseases, including infections.<sup>1</sup> While antiviral protease inhibitors are in clinical use for the treatment of e.g. HIV or HCV, no bacterial protease inhibitors have been approved as anti-infective drugs yet.<sup>2,3</sup> However, to combat the spread of antibiotic resistance, new antibacterial agents with novel modes of action are urgently needed.<sup>4–6</sup> This applies especially for Gram-negative pathogens which are challenging to treat as their cell wall is difficult to permeate.<sup>7</sup> The development of pathoblockers which target bacterial virulence rather than killing bacteria is of growing interest in anti-infective drug discovery due to the reduced selection pressure such a strategy is supposed to have.<sup>8–12</sup> In this context, bacterial proteases represent attractive targets.<sup>2,3,13</sup> Notably, the only FDA-approved antivirulence drugs are immunoglobulins that target secreted virulence factors, highlighting the potential of extracellular targets to circumvent cell wall permeation problems.<sup>8</sup> The highly problematic Gram-negative pathogen *Pseudomonas aeruginosa* has been assigned critical priority by the WHO<sup>14</sup> and urgently requires novel treatment options because of increasing resistance.<sup>15,16</sup> *P. aeruginosa* is responsible for fatal lung infections in cystic fibrosis patients.<sup>17</sup> Among its numerous virulence factors representing potential

drug targets,<sup>12,16,18–20</sup> the zinc–metalloprotease elastase (LasB) is of specific interest, given its extracellular location.<sup>21</sup> LasB substantially contributes to disease progression in *P. aeruginosa* infected individuals by facilitating host invasion and immune evasion.<sup>22</sup> It was for example found to degrade and thereby inactivate the endogenous antimicrobial peptide (AMP) LL-37.<sup>23</sup> Furthermore, LasB was reported to be involved in the formation of *P. aeruginosa* biofilms either by periplasmic activation of nucleoside diphosphate kinase (NDK) required for alginate synthesis<sup>24</sup> or by upregulation of rhamnolipids.<sup>25</sup> The aggregation of bacteria in the biofilm matrix seriously impedes successful antibiotic treatment and blocks host defense mechanisms.<sup>26,27</sup>

Several thiol-based inhibitors with promising activity on LasB have been described.<sup>28–32</sup> A class of *N*-aryl mercaptoacetamides turned out to be particularly attractive because these thiols display high selectivity against a range of human matrix metalloproteinases (MMPs).<sup>32,33</sup> Thiol-containing compounds

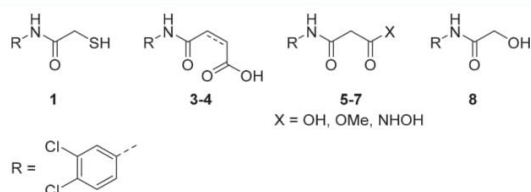
Received: March 19, 2018

Accepted: August 8, 2018

Published: August 8, 2018

are in clinical use for the treatment of various diseases.<sup>34–36</sup> However, a disadvantage of this class compared to other zinc-chelating inhibitors is the possible oxidation to the respective disulfides, resulting in inactivation of the compounds.<sup>37</sup>

In this study, we describe the synthetic replacement of the sulfhydryl function of LasB inhibitor **1**<sup>32</sup> ( $IC_{50}$ :  $6.6 \pm 0.3 \mu M$ ) by zinc-binding groups insensitive to oxidation (Figure 1).



**Figure 1.** Structures of lead LasB inhibitor **1** and compounds **3–8** synthesized in this article.

Among the compounds tested, a hydroxamic acid derivative was found to inhibit LasB in the low micromolar range as well. Hydroxamates have been reported as LasB inhibitors, but the described compounds lack selectivity against human MMPs.<sup>38,39</sup> In contrast, the inhibitor described in this work maintained the remarkable selectivity of thiol **1** against human MMPs, despite a binding mode equivalent to described hydroxamate–MMP complexes. The new compound was further able to reduce biofilm formation and eDNA release by *P. aeruginosa* and to restore the antimicrobial effect of LL-37.

## RESULTS AND DISCUSSION

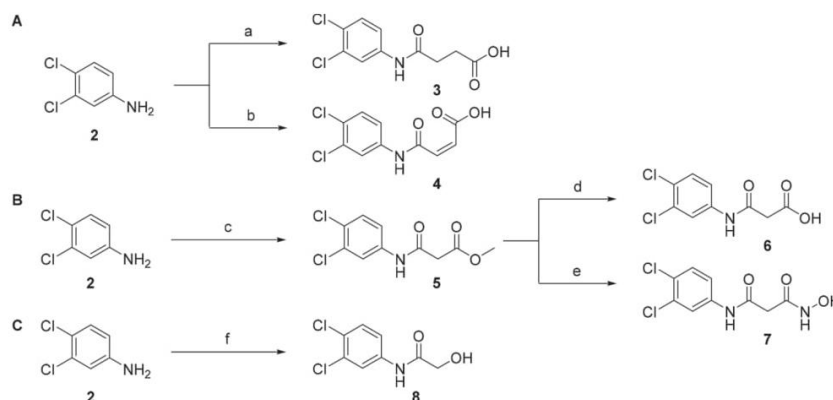
**Synthesis of Novel Compounds 3–8.** Among the variety of chemically diverse zinc binding groups in literature,<sup>40</sup> we focused on hydroxyl, carboxyl, and hydroxamate functions. By introduction of these relatively small zinc chelating groups, drastic changes in the size of the thiol function of **1** were avoided to allow the inhibitor backbone to preserve the previously observed binding mode.<sup>32</sup> Carboxylic acid derivatives **3**, **4**, and **6** were obtained either by reacting aniline **2** with succinic/maleic anhydride or via hydrolysis of methyl ester intermediate **5**. Similarly, **7** was synthesized by reacting **5** with

hydroxylamine (Schemes 1A and B). As an isosteric modification we further synthesized the alcohol derivative of **1**, compound **8**, using glycolic acid in a neat reaction (Scheme 1C).

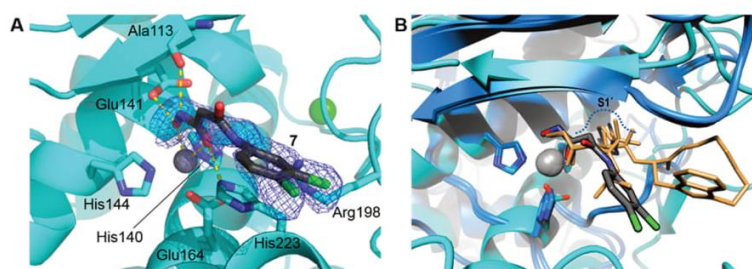
**Identification of Compound 7 as a Promising LasB Inhibitor.** Using a FRET-based inhibition assay,<sup>41</sup> it turned out that replacement of the thiol function of **1** by a hydroxy group led to a complete loss of activity when tested at  $600 \mu M$  (**8**). Compounds bearing a carboxylic acid in  $\gamma$ - (**3**, **4**) or in  $\beta$ -position (**6**) to the carbonyl group were also inactive. Contrary to that, the activity was maintained for the  $\beta$ -hydroxamic acid derivative **7**, displaying an  $IC_{50}$  of  $17.4 \pm 0.8 \mu M$  and a  $K_{i,app}$  of  $12.3 \pm 0.6 \mu M$ . This is slightly less active compared to the free thiol analogue **1**; however, the oxidation issue was resolved by replacing the thiol by a hydroxamic acid function.

**Binding Mode of 7 to LasB.** To rationalize whether the minor difference in activity compared to **1** was due to a different binding mode, the X-ray cocrystal structure of the LasB-7 complex was solved. The complex crystallized in space group  $P2_12_12_1$  and crystals diffracted to 2.1 Å resolution (Figure 2A). The structure was solved by molecular replacement using the published LasB structure (PDB ID 1EZM) as a search model. Full details of the data collection and refinement statistics can be found in Supplementary Table 1. Hydroxamate **7** was found to be orientated toward the primed binding site of the protease. As expected, the active site zinc atom is coordinated by both the carbonyl oxygen and the hydroxamide oxygen of **7**, leading to a distorted trigonal-bipyramidal geometry. The carbonyl oxygen further undergoes a weaker interaction with His223 (3.5 Å). Additionally, the hydroxamide oxygen forms a hydrogen bond with the adjacent Glu141 (2.5 Å), while the amide nitrogen interacts with the carbonyl group of Ala113 (3.0 Å). These observations are in excellent accordance with the reported binding of hydroxamate functions to MMP-7,<sup>42</sup> MMP-3<sup>43</sup> or to thermolysin.<sup>44</sup> Inhibitor binding to thermolysin-like proteases like LasB was described to lead to a closure of the binding pocket due to hinge-bending motion.<sup>45</sup> Intriguingly, thiol **1** has recently been discovered by us to keep the active site cleft in an open conformation due to the unexpected binding of two molecules to the primed binding site (Supplementary Figure S1).<sup>32</sup> In contrast, only one molecule of the hydroxamate binds to the

**Scheme 1.** Synthesis of 3–8<sup>a</sup>



<sup>a</sup>Reagents and conditions: (a) succinic anhydride, dioxane, 70 °C, 6 h; (b) maleic anhydride, dioxane, 70 °C, 6 h; (c) methyl malonyl chloride, Et<sub>3</sub>N, DCM, RT, 4 h; (d) NaOH, THF, RT, 24 h; (e) H<sub>2</sub>NOH, DIPEA, MeOH, 8 h reflux, 16 h RT; (f) glycolic acid, 130 °C, 24 h.



**Figure 2.** (A) Structure of LasB in complex with 7. Cartoon representations of LasB (cyan) in complex with 7 (black). The difference electron density ( $F_o - F_c$ ) contoured to  $3\sigma$  with phases calculated from a model that was refined in the absence of 7 is shown as a blue isomesh. The active-site zinc ion is shown as gray, and the calcium ion is shown as a green sphere. Residues involved in binding of 7 are shown as sticks, and their interactions depicted as yellow dashed lines. Zinc-liganding distances are 2.1 Å (OH) and 2.4 Å (NH). (B) Overlay of LasB (cyan) in complex with 7 (gray) and MMP-3/-7 (green) occupied by hydroxamate inhibitors 9 and 10 (brown, PDB codes 4G9L/1MMQ). Zinc ligands are highlighted.

protease, which undergoes the characteristic hinge-bending. Unlike the zinc-chelating thiol, the hydroxamate directly interacts with the edge strand via a hydrogen bond with the main chain oxygen of Ala113. This interaction presumably promotes closure of the active site cleft, which is hampered in case of thiol 1 by the second molecule interacting with Asn112. Due to the twisted orientation of the aromatic core of 7 compared to 1, a previously observed bidentate hydrogen bond with Arg198 in the S1' binding site is not possible. This observation could explain the slightly weaker activity of the hydroxamate compared to the thiol.

**Selectivity against Six MMPs and ADAM-17.** Strong zinc chelating groups such as hydroxamic acids can be the reason for poor selectivity against further metalloproteases, when compound binding is driven more by the chelating moiety than by the rest of the molecule.<sup>46</sup> In fact, the lack of selectivity against MMP antitargets has been one reason for the failure of various hydroxamate-based MMP inhibitors in clinical trials.<sup>47,48</sup> Considering the high similarity in zinc chelation by 7 and the hydroxamates in MMP-3 or MMP-7, it was investigated whether the previously demonstrated selectivity of *N*-aryl mercaptoacetamides toward six human MMPs<sup>32,33</sup> could be maintained. Fortunately, 7 did not inhibit these MMPs comprising members with differing depth of the S1' binding site,<sup>49</sup> including MMP-3 and -7 (Table 1). By contrast, the unselective inhibitor Batimastat<sup>50</sup> (Supplementary Figure S2) inhibited all tested enzymes in the low nanomolar range.

This of course prompted the question why, despite comparable interactions of the hydroxamate function, LasB was inhibited but MMP-7 was not. Logically, the selectivity

**Table 1. Residual Activity of Six MMPs and ADAM-17 in the Presence of 100  $\mu$ M 7 and IC<sub>50</sub> Values of Batimastat<sup>33a</sup>**

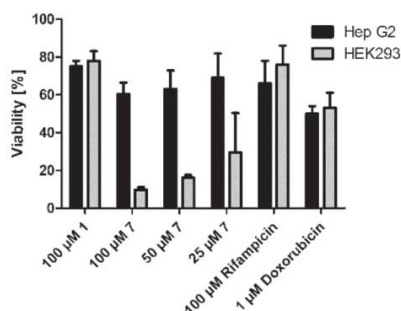
	7 residual activity at 100 $\mu$ M (%)	Batimastat IC <sub>50</sub> (nM)
MMP-1	91 ± 9	2.2 ± 0.1
MMP-2	87 ± 3	1.8 ± 0.1
MMP-3	84 ± 5	5.6 ± 0.9
MMP-7	98 ± 3	7.0 ± 0.2
MMP-8	73 ± 6	0.7 ± 0.2
MMP-14	98 ± 4	2.8 ± 0.2
ADAM-17	38 ± 12	ND
ADAM-17	61 ± 23*	ND

<sup>a</sup>Means and SD of at least two independent measurements are displayed. \*: test at 25  $\mu$ M. ND: not determined.

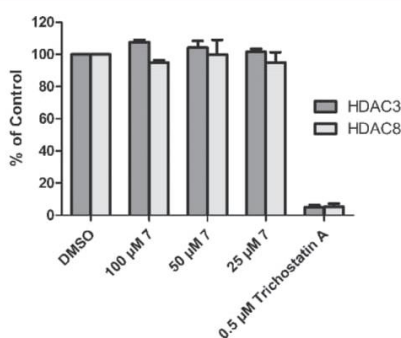
might be related to differences in the positioning of the inhibitor backbone in the pocket. To investigate this, published X-ray structures of MMPs in complex with hydroxamates were overlaid with the LasB-7 complex. The shallow S1' pocket of MMP-7 is occupied by an isobutyl moiety of inhibitor 9<sup>42</sup> and the deep S1' pocket of MMP-3 by an aromatic core of inhibitor 10<sup>43</sup> (for inhibitor structures, see Supplementary Figure S3). In contrast, the core of 7 does not bind to the respective pocket of LasB (Figure 2B). The ability to bind the S1' pocket of the respective MMP is a common feature of reported hydroxamate-based MMP inhibitors, which can also be the cause for a lack of selectivity.<sup>46,51</sup> Consequently, the high selectivity against various MMPs with differing depth of the S1' binding pocket might be explained by an inability of 7 to bind to this site of the protease. ADAMs (A disintegrin and metalloproteinase), a class of zinc-dependent proteases playing essential roles in muscle development, cell migration, or shedding, display additional antitargets for the presented LasB inhibitor. In this context, an *in vitro* inhibition assay toward ADAM-17 revealed that 7 exerted moderate inhibitory effects (Table 1). Consequently, further optimization regarding selectivity toward this antitarget is needed. Still, the promising selectivity toward a range of MMPs makes 7 an attractive starting point for LasB inhibitor development.

**Cytotoxicity Assays.** In addition to selectivity, it was also investigated whether the cytotoxicity of 7 toward human cell lines was as low as described for thiol 1.<sup>32</sup> This was of specific interest because cytotoxic properties are a known drawback of hydroxamates.<sup>52</sup> Notably, 7 had only low effects on the viability of HEP G2 cells, comparable to thiol 1 and rifampicin (Figure 3). The effect on HEK293 cells was more pronounced at the tested concentrations, which might be related to the inhibition of antitargets such as ADAM-17. For comparison, residual enzyme activities in the LasB *in vitro* assay are listed for each inhibitor concentration in Supplementary Table S2.

**Selectivity against HDACs.** Considering the cytotoxic effects of 7 against HEP G2 cells, it was investigated whether this observation could furthermore be related to inhibition of histone deacetylases (HDACs).<sup>53</sup> These zinc-dependent enzymes are involved in the epigenetic regulation of cell proliferation and differentiation.<sup>54</sup> Given that HDACs are known to be inhibited by hydroxamic acids such as vorinostat or trichostatin A,<sup>55</sup> assessing a potential inhibitory effect of hydroxamate 7 was of special interest. Figure 4 shows that 7 does not exert any inhibitory effect on HDAC3 and HDAC8, while these enzymes are efficiently inhibited by the positive



**Figure 3.** Cytotoxicity of 1, 7, and two reference compounds toward HEP G2 and HEK293 cells (values for 1 are taken from ref 32).

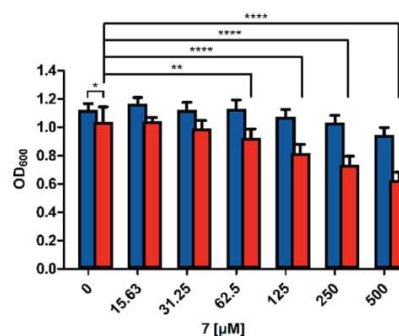


**Figure 4.** Residual activity of selected HDAC enzymes in the presence of 100 μM 7.

control trichostatin A (Supplementary Figure S2). The observed selectivity against these additional antitargets could be explained by the absence of a long linker separating the zinc chelating function from the aromatic core, as it is typical for hydroxamate-based HDAC inhibitors.<sup>52,56</sup>

**Restoration of LL-37 Antibacterial Activity against *P. aeruginosa* PA14.** LL-37 is an  $\alpha$ -helical human cathelicidin peptide which shows increased prevalence in cystic fibrosis patients.<sup>57</sup> Its effectiveness against *P. aeruginosa* is considerably reduced as it is susceptible toward cleavage by LasB.<sup>23</sup> To assess a potential restoration effect of 7 on the activity of LL-37, a bacterial growth assay with PA14 and LL-37 in presence/absence of 7 was performed (Figure 5). The antibacterial effect of LL-37 alone was only minor at 25 μg/mL ( $p = 0.0211$ ). However, when combined with the LasB inhibitor, it recovered its ability to reduce bacterial growth in a dose-dependent way. Significant reduction of the OD<sub>600</sub> was observed starting from a hydroxamate concentration of 62.5 μM ( $p = 0.0047$ ). At high concentrations, 7 itself slightly inhibited PA14 growth, yet to a much lower extent than in combination with LL-37 ( $p = 0.0255$  at 250 μM 7). These findings highlight the potential of LasB inhibitor 7 to restore a host defense mechanism which is otherwise hampered by LasB.

**Biofilm Volume and eDNA Reduction.** The biofilm matrix, a key element of *P. aeruginosa* resistance,<sup>26</sup> is composed of extracellular polysaccharides, lipids, proteins and, importantly, extracellular DNA (eDNA).<sup>58,59</sup> Inhibition of LasB was shown to inhibit biofilm formation.<sup>28</sup> Therefore, it was of great interest to investigate whether treatment of PA14 cultures with 7 would result in reduced biofilm formation as well. Indeed, the hydroxamate caused a concentration-dependent reduction

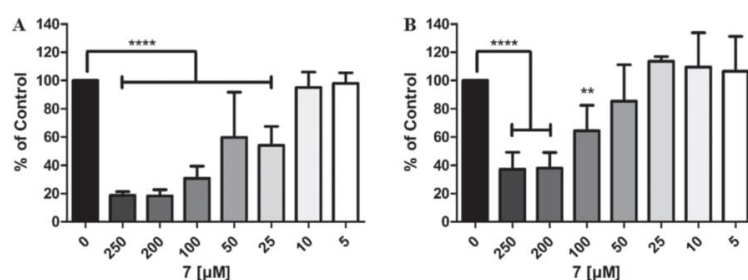


**Figure 5.** Growth of PA14 cultures incubated with 7 in absence (blue) and presence (red) of 25 μg/mL LL-37. Means and SD of three independent measurements are depicted. \* =  $p < 0.05$ , \*\* =  $p < 0.01$ , \*\*\*\* =  $p < 0.0001$ .

of eDNA release (Figure 6A) and of the overall biofilm volume (Figure 6B). Because significant inhibition was observed at concentrations lower than 250 μM, these effects can be attributed to the on-target activity of the compound and are not due to a reduction of bacterial growth or lysis of cells. Hence, LasB inhibitor 7 has demonstrated important pathoblocker activity. It is able to interfere with crucial factors leading to bacterial resistance of *P. aeruginosa* toward antibiotics and host defense molecules.<sup>27,60</sup> To the best of our knowledge, no direct correlation between LasB inhibition and a reduction of eDNA release has been reported to date. Regarding the two mechanisms discussed in literature, the reduced biofilm formation is either due to a reduced NDK-mediated alginate synthesis<sup>24</sup> or rhamolipid-mediated synthesis.<sup>25</sup> Which of these mechanisms inhibitor 7 interferes with cannot be concluded. Yet, these findings could hint at the capacity of 7 to permeate at least the outer membrane of the Gram-negative cell wall, as far as an inhibition of periplasmic NDK<sup>24</sup> is concerned.

## CONCLUSION

Based on our recent findings that *N*-aryl mercaptoacetamides are promising highly selective inhibitors of the virulence factor LasB from *P. aeruginosa*, the lead inhibitor was modified by changing the zinc-chelating moiety to a hydroxamate. Using X-ray crystallography, it was shown that similar to the thiol analogue 1, hydroxamate 7 occupied the primed binding site. Contrary to our recent observations for 1, only one inhibitor molecule was bound to the protease, which could undergo the characteristic hinge-bending motion resulting in a closed conformation of the enzyme. This was attributed to the ability of the inhibitor to interact with catalytic zinc and the edge strand alike. Despite the high similarity of zinc binding to hydroxamates inhibiting MMP-7, inhibitor 7 was unexpectedly able to maintain the remarkable selectivity of 1 toward a range of MMPs. This could be rationalized by the fact that the S1' pocket was not occupied by 7, unlike that observed for MMP inhibitors. These findings show that despite the selectivity issues related to hydroxamic acids, selectivity can indeed be achieved by sparing the S1' pocket. Because cytotoxic effects toward mammalian cell lines were observed, the activity of 7 was investigated regarding other antitargets, revealing a moderate inhibition of ADAM-17 and high selectivity against HDAC3 and 8. Consequently, there will be a special focus on selectivity and cytotoxicity during future compound optimiza-



**Figure 6.** Reduction of eDNA release by treatment of PA14 cultures with 7. (A) Reduction of overall PA14 biofilm volume by 7. (B) Columns represent mean and SD of at least four independent measurements. \*\* =  $p < 0.01$ , \*\*\*\* =  $p < 0.0001$ .

tion. Below an antibacterial concentration of 250  $\mu\text{M}$ , LasB was efficiently inhibited *in vitro*. This inhibition could be translated into more complex, cellular assays, highlighting this compound as a promising antivirulence agent. After reduction of their pathogenicity by such an agent, bacteria are supposed to be cleared by host defense mechanisms or with the help of conventional antibiotics.<sup>10,11</sup> In this context, our results highlight a direct restoration effect of the antibacterial activity of the host defense peptide AMP LL-37 *in vitro*. Indirectly, the activity of host defense mechanisms might be further improved by the inhibition of biofilm formation and eDNA release also observed for 7. The same holds true for the effectiveness of conventional antibiotics, which is seriously hampered by bacterial biofilms. Overall, these findings show hydroxamate 7 to be an interesting lead, which could pave the way for the rational development of selective protease inhibitors as potential new antibiotics.

## METHODS

**Chemistry.** All reagents were used from commercial suppliers without further purification. Procedures were not optimized regarding yield. NMR spectra were recorded on a Bruker Fourier 300 (300 MHz) spectrometer. Chemical shifts are given in parts per million (ppm) and referenced against the residual proton,  $^1\text{H}$ , or carbon,  $^{13}\text{C}$ , resonances of the >99% deuterated solvents as internal reference. Coupling constants ( $J$ ) are given in Hertz. Data are reported as follows: chemical shift, multiplicity ( $s$  = singlet,  $d$  = doublet,  $t$  = triplet,  $m$  = multiplet,  $br$  = broad and combinations of these) coupling constants and integration. Mass spectrometry was performed on a SpectraSystems-MSQ LCMS system (Thermo Fisher). Flash chromatography was performed on silica gel 60 M, 0.04–0.063 mm (Machery-Nagel) or using the automated flash chromatography system CombiFlash Rf+ (Teledyne Isco) equipped with RediSepRf silica columns (Axel Semrau) or Chromabond Flash C18 columns (Machery-Nagel). Purity of compounds synthesized by us was determined by LCMS using the area percentage method on the UV trace recorded at a wavelength of 254 nm and found to be >95%.

**Expression and Purification of LasB.** LasB was expressed and purified as described previously.<sup>32</sup>

***In vitro* Inhibition Assays.** The LasB *in vitro* inhibition assay and the MMP assay were performed as described previously.<sup>32</sup> The LasB concentration that was used in the assay was 0.3 nM. The  $K_m$  value for LasB was determined to be 363.0  $\mu\text{M}$ .

**ADAM Inhibition Assay.** ADAM-17 (TACE) Inhibitor Screening Assay Kit was purchased from Sigma-Aldrich. The assay was performed according to the guidelines of the manufacturer. Fluorescence signals were measured in a CLARIOstar plate reader (BMG Labtech).

**HDAC Inhibition Assay.** HDAC3 and HDAC8 inhibitor screening kits were purchased from Sigma-Aldrich. The assay was performed according to the guidelines of the manufacturer.

Fluorescence signals were measured in a CLARIOstar plate reader (BMG Labtech).

**Cytotoxicity Assays.** Hep G2 or HEK293 cells ( $2 \times 10^5$  cells per well) were seeded in 24-well, flat-bottomed plates. Culturing of cells, incubations, and OD measurements were performed as described previously<sup>61</sup> with small modifications. Twenty-four hours after seeding the cells, the incubation was started by the addition of compounds in a final DMSO concentration of 1%. The living cell mass was determined after 48 h. At least three independent measurements were performed for each compound.

**X-ray Crystallography and Image Preparation.** LasB was concentrated to 10–12 mg mL<sup>-1</sup> and mixed with inhibitor 7 at a final concentration of 1 mM. Complex crystals were grown by the sitting drop method using a reservoir solution containing 1.8 M AM<sub>2</sub>SO<sub>4</sub> and 0.1 M Tris-Cl, pH 8.8. Crystals were cryoprotected in glycerol, and diffraction data was collected from single crystals at 100 K at beamline ID29 (ESRF) at a wavelength of 1.738 Å. Data was processed using Xia2,<sup>62</sup> and the structure was solved using PHASER<sup>63</sup> molecular replacement with *Pseudomonas aeruginosa* LasB (PAE, PDB ID 1EZM) as a search model. The solution was manually rebuilt with COOT<sup>64</sup> and refined using PHENIX<sup>65</sup> and Refmac5.<sup>66</sup> The final refined structure of LasB in complex with compound 7 was deposited in the Protein Data Bank (PDB) as entry 6FZX. Structural superimposition of complex structures of human MMPs (1MMQ and 4G9L) and LasB (6FZX) was achieved through alignment of residues 201–205 of 4G9L (corresponds to residues 140–144 of 1MMQ) using the align atoms algorithm of YASARA structure (YASARA Biosciences GmbH).<sup>67</sup> Image was rendered using PovRay (<http://www.povray.org/>).

**Bacterial Growth Assay.** The assay was performed in 96 well plates (Greiner) with a final volume of 200  $\mu\text{L}$ . LL-37 was purchased from AnaSpec (Fremont) and diluted to a final concentration of 25  $\mu\text{g mL}^{-1}$  from 125  $\mu\text{g mL}^{-1}$  stocks in 18M $\Omega$  H<sub>2</sub>O. Prior to culture addition 7 was serially diluted in DMSO. A preculture of PA14 was adjusted to the final start OD<sub>600</sub> 0.02 in lysogeny broth medium. All samples contained 1% DMSO and 40% of 18M $\Omega$  H<sub>2</sub>O. OD<sub>600</sub> was measured using a FLUOstar Omega (BMG Labtech) after inoculation and after incubation for 16.5 h at 37 °C with 200 rpm. Given OD<sub>600</sub> values were obtained after subtraction of the respective start OD<sub>600</sub> and represent three independent measurements with at least two replicates each. One-way ANOVA was performed using GraphPad Prism 6 software.

**Biofilm and eDNA Assays.** The assays were performed as described previously.<sup>68</sup>

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acschembio.8b00257.

Supplementary Table 1, X-ray data collection and refinement statistics; Supplementary Table T2, residual activities of LasB in the *in vitro* assay; Supplementary



Figures 1–3, comparison of the LasB-1 complex to LasB-7 and structures of hydroxamate-based MMP inhibitors 9, 10, Batimastat, and HDAC inhibitor Trichostatin A; synthetic procedures; and spectral data for compounds 3–8 (PDF)

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: Rolf.Hartmann@helmholtz-hzi.de.

### ORCID

Andreas M. Kany: 0000-0001-7580-3658

Samir Yahiaoui: 0000-0001-5134-5007

Jörg Haupenthal: 0000-0003-3991-2800

Martin Empting: 0000-0002-0503-5830

Jesko Köhnke: 0000-0002-7153-1365

Rolf W. Hartmann: 0000-0002-5871-5231

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

J.K. acknowledges the DFG for an Emmy–Noether Fellowship (KO 4116/3-1). We thank J. Jung, S. Amann, and D. Jener for technical support and T. Röhrig for help with statistical analysis. The image of LL-37 used in the TOC graphic was created using PDB entry 2K6O.

## REFERENCES

- (1) Turk, B. (2006) Targeting proteases. Successes, failures and future prospects. *Nat. Rev. Drug Discovery* 5, 785–799.
- (2) Agbowuro, A. A., Huston, W. M., Gamble, A. B., and Tyndall, J. D. A. (2018) Proteases and protease inhibitors in infectious diseases. *Med. Res. Rev.* 38, 1295.
- (3) Culp, E., and Wright, G. D. (2017) Bacterial proteases, untapped antimicrobial drug targets. *J. Antibiot.* 70, 366–377.
- (4) Cooper, M. A., and Shlaes, D. (2011) Fix the antibiotics pipeline. *Nature* 472, 32.
- (5) Taubes, G. (2008) The bacteria fight back. *Science* 321, 356–361.
- (6) Coates, A. R. M., Halls, G., and Hu, Y. (2011) Novel classes of antibiotics or more of the same? *Br. J. Pharmacol.* 163, 184–194.
- (7) Payne, D. J., Gwynn, M. N., Holmes, D. J., and Pompliano, D. L. (2007) Drugs for bad bugs. Confronting the challenges of antibacterial discovery. *Nat. Rev. Drug Discovery* 6, 29–40.
- (8) Dickey, S. W., Cheung, G. Y. C., and Otto, M. (2017) Different drugs for bad bugs. Antivirulence strategies in the age of antibiotic resistance. *Nat. Rev. Drug Discovery* 16, 457–471.
- (9) Heras, B., Scanlon, M. J., and Martin, J. L. (2015) Targeting virulence not viability in the search for future antibacterials. *Br. J. Clin. Pharmacol.* 79, 208–215.
- (10) Rasko, D. A., and Sperandio, V. (2010) Anti-virulence strategies to combat bacteria-mediated disease. *Nat. Rev. Drug Discovery* 9, 117–128.
- (11) Clatworthy, A. E., Pierson, E., and Hung, D. T. (2007) Targeting virulence. A new paradigm for antimicrobial therapy. *Nat. Chem. Biol.* 3, 541–548.
- (12) Kamal, A. A. M., Maurer, C. K., Allegretta, G., Haupenthal, J., Empting, M., and Hartmann, R. W. (2017) Quorum Sensing Inhibitors as Pathoblockers for *Pseudomonas aeruginosa* Infections: A New Concept in Anti-Infective Drug Discovery. In *Topics in Medicinal Chemistry*, pp 1–26, Springer, Berlin, Heidelberg.
- (13) Travis, J., and Potempa, J. (2000) Bacterial proteinases as targets for the development of second-generation antibiotics. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* 1477, 35–50.
- (14) World Health Organization (2017) Antibacterial Agents in Clinical Development. *An Analysis of the Antibacterial Clinical Development Pipeline, Including Tuberculosis*, World Health Organization, Geneva.
- (15) Aloush, V., Navon-Venezia, S., Seigman-Igra, Y., Cabili, S., and Carmeli, Y. (2006) Multidrug-resistant *Pseudomonas aeruginosa*. Risk factors and clinical impact. *Antimicrob. Agents Chemother.* 50, 43–48.
- (16) Wagner, S., Sommer, R., Hinsberger, S., Lu, C., Hartmann, R. W., Empting, M., and Titz, A. (2016) Novel Strategies for the Treatment of *Pseudomonas aeruginosa* Infections. *J. Med. Chem.* 59, 5929–5969.
- (17) Sordé, R., Pahissa, A., and Rello, J. (2011) Management of refractory *Pseudomonas aeruginosa* infection in cystic fibrosis. *Infect. Drug Resist.* 4, 31–41.
- (18) Storz, M. P., Maurer, C. K., Zimmer, C., Wagner, N., Brengel, C., de Jong, J. C., Lucas, S., Müsken, M., Häussler, S., Steinbach, A., and Hartmann, R. W. (2012) Validation of PqsD as an anti-biofilm target in *Pseudomonas aeruginosa* by development of small-molecule inhibitors. *J. Am. Chem. Soc.* 134, 16143–16146.
- (19) Strateva, T., and Mitov, I. (2011) Contribution of an arsenal of virulence factors to pathogenesis of *Pseudomonas aeruginosa* infections. *Ann. Microbiol.* 61, 717–732.
- (20) Lu, C., Maurer, C. K., Kirsch, B., Steinbach, A., and Hartmann, R. W. (2014) Overcoming the unexpected functional inversion of a PqsR antagonist in *Pseudomonas aeruginosa*. An in vivo potent antivirulence agent targeting pqs quorum sensing. *Angew. Chem., Int. Ed.* 53, 1109–1112.
- (21) Morihara, K., Tsuzuki, H., Oka, T., Inoue, H., and Ebata, M. (1965) *Pseudomonas aeruginosa* elastase: isolation, crystallization, and preliminary characterization. *J. Biol. Chem.* 240, 3295–3304.
- (22) Wretling, B., and Pavlovskis, O. R. (1983) *Pseudomonas aeruginosa* elastase and its role in *Pseudomonas* infections. *Clin. Infect. Dis.* 5, S998–1004.
- (23) Schmidtchen, A., Frick, I.-M., Andersson, E., Tapper, H., and Björck, L. (2002) Proteinases of common pathogenic bacteria degrade and inactivate the antibacterial peptide LL-37. *Mol. Microbiol.* 46, 157–168.
- (24) Kamath, S., Kapatral, V., and Chakrabarty, A. M. (1998) Cellular function of elastase in *Pseudomonas aeruginosa*. Role in the cleavage of nucleoside diphosphate kinase and in alginate synthesis. *Mol. Microbiol.* 30, 933–941.
- (25) Yu, H., He, X., Xie, W., Xiong, J., Sheng, H., Guo, S., Huang, C., and Di Zhang, K. (2014) Elastase LasB of *Pseudomonas aeruginosa* promotes biofilm formation partly through rhamnolipid-mediated regulation. *Can. J. Microbiol.* 60, 227–235.
- (26) Høiby, N., Bjarnsholt, T., Givskov, M., Molin, S., and Ciofu, O. (2010) Antibiotic resistance of bacterial biofilms. *Int. J. Antimicrob. Agents* 35, 322–332.
- (27) Lambert, P. A. (2002) Mechanisms of antibiotic resistance in *Pseudomonas aeruginosa*. *J. R. Soc. Med.* 95 Suppl 41, 22–26.
- (28) Cathcart, G. R. A., Quinn, D., Greer, B., Harriott, P., Lynas, J. F., Gilmore, B. F., and Walker, B. (2011) Novel inhibitors of the *Pseudomonas aeruginosa* virulence factor LasB. A potential therapeutic approach for the attenuation of virulence mechanisms in pseudomonal infection. *Antimicrob. Agents Chemother.* 55, 2670–2678.
- (29) Zhu, J., Cai, X., Harris, T. L., Gooyit, M., Wood, M., Lardy, M., and Janda, K. D. (2015) Disarming *Pseudomonas aeruginosa* virulence factor LasB by leveraging a *Caenorhabditis elegans* infection model. *Chem. Biol.* 22, 483–491.
- (30) Burns, F. R., Paterson, C. A., Gray, R. D., and Wells, J. T. (1990) Inhibition of *Pseudomonas aeruginosa* elastase and *Pseudomonas keratitis* using a thiol-based peptide. *Antimicrob. Agents Chemother.* 34, 2065–2069.
- (31) Kessler, E., Israel, M., Landshman, N., Chechick, A., and Blumberg, S. (1982) In vitro inhibition of *Pseudomonas aeruginosa* elastase by metal-chelating peptide derivatives. *Infect. Immun.* 38, 716–723.
- (32) Kany, A. M., Sikandar, A., Haupenthal, J., Yahiaoui, S., Maurer, C. K., Proschak, E., Koehnke, J., and Hartmann, R. W. (2018) Binding Mode Characterization and Early in Vivo Evaluation of Fragment-

Like Thiols as Inhibitors of the Virulence Factor LasB from *Pseudomonas aeruginosa*. *ACS Infect. Dis.* 4, 988–997.

- (33) Schönauer, E., Kany, A. M., Haupenthal, J., Hüsecken, K., Hoppe, I. J., Voos, K., Yahiaoui, S., Elsässer, B., Ducho, C., Brandstetter, H., and Hartmann, R. W. (2017) Discovery of a Potent Inhibitor Class with High Selectivity toward Clostridial Collagenases. *J. Am. Chem. Soc.* 139, 12696–12703.
- (34) Pereillo, J.-M., Maftouh, M., Andrieu, A., Uzabiaga, M.-F., Fedeli, O., Savi, P., Pascal, M., Herbert, J.-M., Maffrand, J.-P., and Picard, C. (2002) Structure and stereochemistry of the active metabolite of clopidogrel. *Drug Metab. Dispos.* 30, 1288–1295.
- (35) Primi, M. P., Bueno, L., Baumer, P., Berard, H., and Lecomte, J. M. (1999) Racecadotril demonstrates intestinal antisecretory activity in vivo. *Aliment. Pharmacol. Ther.* 13, 3–7.
- (36) Cushman, D. W., and Ondetti, M. A. (1991) History of the design of captopril and related inhibitors of angiotensin converting enzyme. *Hypertension* 17, 589–592.
- (37) Pramari, Y., Das Gupta, V., and Bethea, C. (1992) Stability of captopril in some aqueous systems. *J. Clin. Pharm. Ther.* 17, 185–189.
- (38) Adekoya, O. A., Sjöli, S., Wuxiuer, Y., Bילו, I., Marques, S. M., Santos, M. A., Nuti, E., Cercignani, G., Rossello, A., Winberg, J.-O., and Sylte, I. (2015) Inhibition of pseudolysin and thermolysin by hydroxamate-based MMP inhibitors. *Eur. J. Med. Chem.* 89, 340–348.
- (39) Sjöli, S., Nuti, E., Camodeca, C., Bילו, I., Rossello, A., Winberg, J.-O., Sylte, I., and Adekoya, O. A. (2016) Synthesis, experimental evaluation and molecular modelling of hydroxamate derivatives as zinc metalloproteinase inhibitors. *Eur. J. Med. Chem.* 108, 141–153.
- (40) Jacobsen, J. A., Major Jourden, J. L., Miller, M. T., and Cohen, S. M. (2010) To bind zinc or not to bind zinc. An examination of innovative approaches to improved metalloproteinase inhibition. *Biochim. Biophys. Acta, Mol. Cell Res.* 1803, 72–94.
- (41) Nishino, N., and Powers, J. C. (1980) *Pseudomonas aeruginosa* elastase. Development of a new substrate, inhibitors, and an affinity ligand. *J. Biol. Chem.* 255, 3482–3486.
- (42) Browner, M. F., Smith, W. W., and Castelano, A. L. (1995) Matrilysin-inhibitor complexes. Common themes among metalloproteases. *Biochemistry* 34, 6602–6610.
- (43) Belviso, B. D., Caliendo, R., Siliqi, D., Calderone, V., Arnesano, F., and Natile, G. (2013) Structure of matrix metalloproteinase-3 with a platinum-based inhibitor. *Chem. Commun. (Cambridge, U. K.)* 49, 5492–5494.
- (44) Holmes, M. A., and Matthews, B. W. (1981) Binding of hydroxamic acid inhibitors to crystalline thermolysin suggests a pentacoordinate zinc intermediate in catalysis. *Biochemistry* 20, 6912–6920.
- (45) Holland, D. R., Tronrud, D. E., Pley, H. W., Flaherty, K. M., Stark, W., Jansonius, J. N., McKay, D. B., and Matthews, B. W. (1992) Structural comparison suggests that thermolysin and related neutral proteases undergo hinge-bending motion during catalysis. *Biochemistry* 31, 11310–11316.
- (46) Overall, C. M., and Kleifeld, O. (2006) Towards third generation matrix metalloproteinase inhibitors for cancer therapy. *Br. J. Cancer* 94, 941–946.
- (47) Fisher, J. F., and Mobashery, S. (2006) Recent advances in MMP inhibitor design. *Cancer Metastasis Rev.* 25, 115–136.
- (48) Vandenbroucke, R. E., and Libert, C. (2014) Is there new hope for therapeutic matrix metalloproteinase inhibition? *Nat. Rev. Drug Discovery* 13, 904–927.
- (49) Park, H. I., Jin, Y., Hurst, D. R., Monroe, C. A., Lee, S., Schwartz, M. A., and Sang, Q.-X. A. (2003) The intermediate S1' pocket of the endometase/matrilysin-2 active site revealed by enzyme inhibition kinetic studies, protein sequence analyses, and homology modeling. *J. Biol. Chem.* 278, 51646–51653.
- (50) Rasmussen, H. S., and McCann, P. P. (1997) Matrix metalloproteinase inhibition as a novel anticancer strategy. A review with special focus on batimastat and marimastat. *Pharmacol. Ther.* 75, 69–75.

(51) Overall, C. M., and Kleifeld, O. (2006) Tumour microenvironment - opinion. Validating matrix metalloproteinases as drug targets and anti-targets for cancer therapy. *Nat. Rev. Cancer* 6, 227–239.

(52) Shen, S., and Kozikowski, A. P. (2016) Why Hydroxamates May Not Be the Best Histone Deacetylase Inhibitors—What Some May Have Forgotten or Would Rather Forget? *ChemMedChem* 11, 15–21.

(53) Schrupp, D. S. (2009) Cytotoxicity mediated by histone deacetylase inhibitors in cancer cells. Mechanisms and potential clinical implications. *Clin. Cancer Res.* 15, 3947–3957.

(54) de Ruijter, A. J. M., van Gennip, A. H., Caron, H. N., Kemp, S., and van Kuilenburg, A. B. P. (2003) Histone deacetylases (HDACs). Characterization of the classical HDAC family. *Biochem. J.* 370, 737–749.

(55) West, A. C., and Johnstone, R. W. (2014) New and emerging HDAC inhibitors for cancer treatment. *J. Clin. Invest.* 124, 30–39.

(56) Drummond, D. C., Noble, C. O., Kirpotin, D. B., Guo, Z., Scott, G. K., and Benz, C. C. (2005) Clinical development of histone deacetylase inhibitors as anticancer agents. *Annu. Rev. Pharmacol. Toxicol.* 45, 495–528.

(57) Chen, C. I.-U., Schaller-Bals, S., Paul, K. P., Wahn, U., and Bals, R. (2004) Beta-defensins and LL-37 in bronchoalveolar lavage fluid of patients with cystic fibrosis. *J. Cystic Fibrosis* 3, 45–50.

(58) Flemming, H.-C., and Wingender, J. (2010) The biofilm matrix. *Nat. Rev. Microbiol.* 8, 623–633.

(59) Whitchurch, C. B., Tolker-Nielsen, T., Ragas, P. C., and Mattick, J. S. (2002) Extracellular DNA required for bacterial biofilm formation. *Science* 295, 1487.

(60) Lewenza, S. (2013) Extracellular DNA-induced antimicrobial peptide resistance mechanisms in *Pseudomonas aeruginosa*. *Front. Microbiol.* 4, 21.

(61) Haupenthal, J., Baehr, C., Zeuzem, S., and Piiper, A. (2007) RNase A-like enzymes in serum inhibit the anti-neoplastic activity of siRNA targeting polo-like kinase 1. *Int. J. Cancer* 121, 206–210.

(62) Winter, G. (2010) xia2. An expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* 43, 186–190.

(63) McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674.

(64) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 66, 486–501.

(65) Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 66, 213–221.

(66) Skubák, P., Murshudov, G. N., and Pannu, N. S. (2004) Direct incorporation of experimental phase information in model refinement. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 60, 2196–2201.

(67) Krieger, E., Koraimann, G., and Vriend, G. (2002) Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. *Proteins: Struct., Funct., Genet.* 47, 393–402.

(68) Thomann, A., de Mello Martins, A. G. G., Brengel, C., Empting, M., and Hartmann, R. W. (2016) Application of Dual Inhibition Concept within Looped Autoregulatory Systems toward Antivirulence Agents against *Pseudomonas aeruginosa* Infections. *ACS Chem. Biol.* 11, 1279–1286.



Cite this: DOI: 10.1039/c8np00064f

## The role of protein–protein interactions in the biosynthesis of ribosomally synthesized and post-translationally modified peptides

Asfandyar Sikandar  and Jesko Koehnke \*

Covering: up to 02/2019

This review covers the role of protein–protein complexes in the biosynthesis of selected ribosomally synthesized and post-translationally modified peptide (RiPP) classes. The genomic organization of RiPP systems usually allows the expression of each biosynthetic enzyme as an individual unit, which is in stark contrast to the giant assembly lines found in non-ribosomal peptide and polyketide synthesis systems. Evidence is mounting however that the formation of multi-enzyme complexes is critical for efficient RiPPs biosynthesis and that these complexes may be involved in substrate channeling or conformational sampling. In some pathways, polyfunctional enzymes have evolved, which can be viewed as perpetual protein complexes. We summarize what is currently known on enzyme complexes in RiPP systems for lasso peptides, cyanobactins, linear azolic peptides, thiopeptides, and lanthipeptides.

Received 12th July 2018

DOI: 10.1039/c8np00064f

rsc.li/npr

1. Introduction
2. Lasso peptides
3. Cyanobactins
4. Linear azolic peptides (LAPs)
5. Thiopeptides
6. Lanthipeptides
7. Conclusion
8. Conflicts of interest
9. Acknowledgements
10. Notes and references

### 1. Introduction

Natural products represent a rich source of new chemical scaffolds and have played a key role in advancing our understanding of biology and the development of medicine.<sup>1–3</sup> Many of these natural products are derived from polyketide synthase (PKS) or non-ribosomal peptide synthetase (NRPS) families including well-known antibiotics such as beta-lactams (*e.g.* carbenicillin), tetracyclines (*e.g.* tetracycline), macrolides (*e.g.* erythromycin), and glycopeptides (*e.g.* vancomycin) (Fig. 1A).<sup>4,5</sup> However, given the rising antibiotic resistance to these proven antibiotic classes, along with the limited biosynthetic malleability of NRPS and PKS systems, the synthetic biological

approaches have thus far been unsuccessful to turn the tide against antibiotics resistance.<sup>6,7</sup> Therefore, there is an urgent need to discover alternative sources of antibiotics. Recently, genome sequencing efforts have resulted in the identification of another major class of natural products produced in all three domains of life – ribosomally synthesized and post-translationally modified peptides (RiPPs).<sup>8–11</sup> RiPPs biosynthesis is initiated by expression of a precursor peptide, which undergoes (in some cases extensive) post-translational modifications to yield a mature RiPP natural product with a wide variety of sometimes unique structural features (Fig. 1B).<sup>12–15</sup> Many RiPPs have been shown to have potent antiviral, anti-tumor, and antibacterial properties with modes of action distinct from currently used drugs.<sup>16–18</sup> Therefore, RiPPs represent an untapped source of chemical scaffolds with vast therapeutic potential, which requires our attention.<sup>19–21</sup>

As the name implies, RiPP biosynthesis begins with the expression of a precursor peptide *via* the standard ribosomal route, which usually consists of an N-terminal leader sequence, that is important for recognition of the precursor peptide by the biosynthetic machinery, and one or more core peptides, which ultimately become the natural product (Fig. 2).<sup>13</sup> The biosynthetic enzymes bind their respective precursor peptide through specific recognition sequences located in the leader peptide or sometimes at the C-terminus, which is followed by modification(s) of the core peptide, sometimes extensively, to form for example azole heterocycles, lanthionine or sactonine cross-links, D-amino acids, macrocycles, and other modifications.<sup>22</sup>

Workgroup Structural Biology of Biosynthetic Enzymes, Helmholtz Institute for Pharmaceutical Research Saarland, Helmholtz Centre for Infection Research, Saarland University, Campus Geb. E8.1, 66123 Saarbrücken, Germany. E-mail: jesko.koehnke@helmholtz-hzi.de

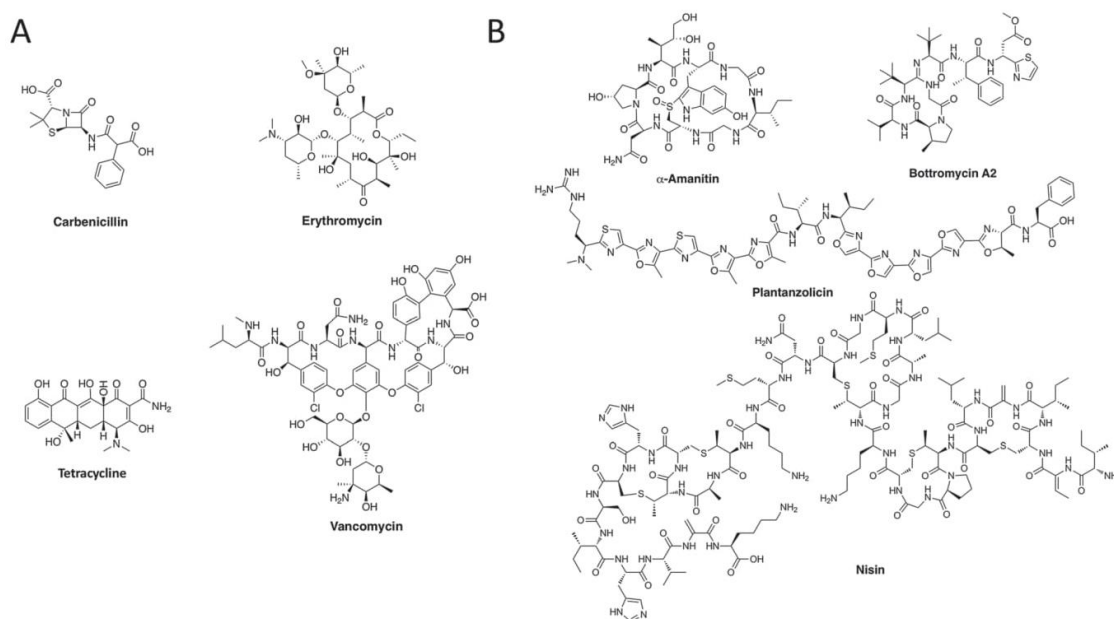


Fig. 1 Representative members of several NRPS/PKS (A) and RiPP (B) compounds highlighting natural product diversity.

The leader peptide is usually removed by proteolytic cleavage late in the maturation process or during export, to yield the final natural product.<sup>22</sup> In addition to these leader-dependent biosynthetic enzymes, RiPPs can be further modified by a range of 'tailoring' enzymes, which act on the core peptide independently of the leader peptide, thereby increasing the chemical diversity of RiPPs.<sup>15</sup>

The physical separation of the recognition sequence encoded in the leader peptide and the core peptide allows leader peptide-dependent tailoring enzymes to be substrate specific, while the

spatial separation between substrate recognition and catalysis allows the enzymes to process a wide variety of substrates.<sup>21,23–25</sup>

This extraordinary biosynthetic malleability sets RiPP apart from PKS/NRPS systems.<sup>26–30</sup> A key difference between RiPP and PKS/NRPS systems, which require large multimodular enzyme complexes to incorporate proteogenic and non-proteogenic amino acids into a peptide backbone, is that RiPP enzymes have thus far been largely studied as independent biosynthetic units.<sup>15,31</sup> Evidence is mounting however that at least some RiPP enzymes require larger, multi-protein, assemblies for efficient



Asfandiyar Sikandar obtained his BSc (Hons) in Biomolecular Science from the University of St. Andrews, UK. This was followed by an MSc in Biochemistry at the University of Calgary, Canada. Currently, he is pursuing his PhD at the Helmholtz Institute for infectious research Saarland (HIPS) under the supervision of Prof. Jesko Köhnke. His work revolves around the structural and biochemical characteriza-

tion of proteins involved in natural product synthesis and bacterial defense mechanisms.



Jesko Koehnke studied Biochemistry at Leibniz University Hannover, Germany, where he graduated in 2005. He pursued his PhD at Columbia University in the City of New York, USA, under the supervision of Prof. Lawrence Shapiro. After graduating in 2010, he joined the group of Prof. James Naismith at the University of St Andrews, UK, as a postdoctoral researcher. Since 2015 he has been a junior group leader at the Helmholtz Institute for Pharmaceutical Research Saarland, Germany. In 2018 he was also appointed as an associated Junior Professor at Saarland University. His work focuses on the structural biology and biochemistry of RiPPs biosynthesis.

tion of proteins involved in natural product synthesis and bacterial defense mechanisms.

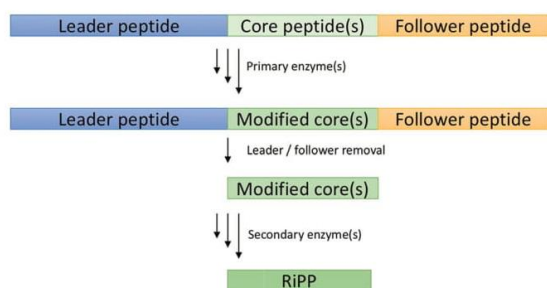


Fig. 2 General schematic of RiPPs biosynthesis. The precursor peptide is composed of a leader or follower peptide (blue/yellow) and core peptide(s) (light green). Frequently the precursor peptide contains multiple core peptides arranged as tandem repeats. Primary enzymes recognize and require the leader peptide for (full) activity. Secondary enzymes act leader-independent, usually after proteolytic removal of the leader peptide.

catalysis<sup>32,33</sup> and with more attention devoted to the study of RiPP enzymes, more examples will be reported in the near future. This review will discuss new insights into the role of protein–protein interactions in key post-translational modifications of the precursor peptide in comparatively well-studied classes of RiPPs: lasso peptides, cyanobactins, linear azolic peptides, thiopeptide, and lanthipeptides.

## 2. Lasso peptides

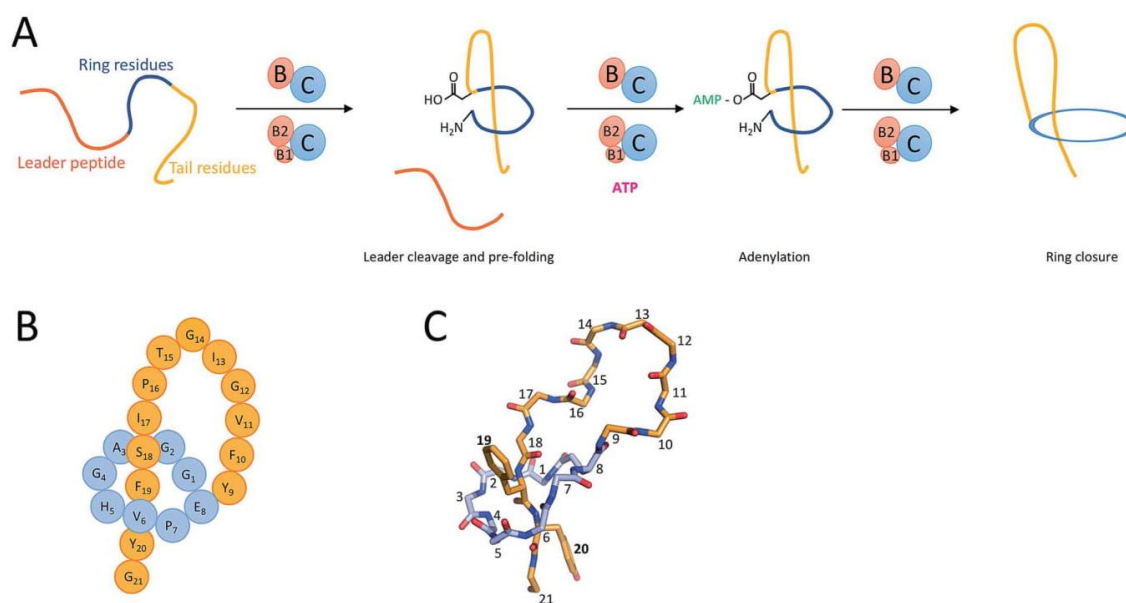
Lasso peptides were first described in 1991 and until 2017 more than 40 different lasso peptides along with some of their gene clusters had been identified from bioactivity screens or genome mining data.<sup>34,35</sup> In 2017 a comprehensive analysis of lasso peptide biosynthetic space was conducted using a novel genome mining approach. It revealed the existence of more than 1300 potential lasso peptide biosynthetic gene clusters, most of which have not been studied.<sup>36</sup> Recently, the same group updated the dataset, demonstrating that ~3000 lasso peptide biosynthetic gene clusters can be found using current genome mining tools.<sup>37</sup> Compared to other known RiPPs, lasso peptides are distinct due to their unique topology. They consist of a macrolactam ring that is composed of seven to nine residues and are threaded by a linear C-terminal peptide tail resulting in a special three-dimensional structure that resembles the knot of a lasso (Fig. 3A).<sup>38</sup> The knot is stabilized primarily by steric interactions between the ring and bulky side chains (so called plugs) present on either side of the ring.<sup>34,38,39</sup> This specific topology makes lasso peptides stable to chemical, proteolytic, and often thermal degradation.<sup>34</sup>

Like all RiPPs, lasso peptides are first ribosomally synthesized as a precursor peptide (A), which is then processed into its mature form by two enzymes: a two-domain cysteine protease B (expressed as individual B1 and B2 proteins in the majority of cases) and ATP-dependent lactam synthetase C.<sup>34,38,40</sup> Based on genome mining data, the overall organization of lasso peptide gene clusters is very diverse.<sup>36,40,41</sup> In addition, some gene clusters harbor downstream genes coding

for an ATP-binding cassette (ABC) transporter (Protein D) for efflux of the lasso peptide and a peptide-tailoring kinase (Protein K).<sup>41–43</sup> Interestingly, a large number of recently discovered lasso peptide biosynthetic gene clusters, which are responsible for the production of lasso peptides that evidently lack antimicrobial activity, can be found adjacent to a highly conserved gene encoding a lasso peptide specific isopeptidase (IsoP).<sup>40,44,45</sup> IsoP has been shown to hydrolyze the isopeptide bond of the macrolactam ring to yield a linear peptide.<sup>40,45</sup> Since astexin isopeptidases have been shown to be highly specific, it might be possible that the lasso peptides from such clusters have another yet unknown function. This argument is strengthened by the observation that the isopeptidases from the sphingopyxin I and benenodin-1 pathways possess the same specificity.<sup>46,47</sup>

In general, the production of lasso peptides requires at least three genes encoding a precursor peptide and the two maturation enzymes (BC).<sup>34</sup> *In vitro* and *in vivo* studies performed with the microcin J25 (MccJ25) biosynthesis system have shown that the precursor peptide (MccJ25) is processed by the cysteine protease MccJ25, which cleaves off the leader peptide.<sup>32,48</sup> This is thought to be followed by the closure of the macrolactam ring *via* isopeptide bond formation between the N-terminal glycine of the core peptide and the Glu8 side chain by the adenylate-forming enzyme, MccJ25, at a cost of one ATP.<sup>32</sup> Mutational studies of MccJ25 have shown that the catalytic triad (C150, H182 and D194) typical of cysteine proteases is present in the C-terminal domain of MccJ25.<sup>32,34</sup> The activities of MccJ25 and C are interdependent, as no production of MccJ25 was observed *in vitro* in the absence of MccJ25.<sup>32</sup> Moreover, the absence of active MccJ25 had a detrimental effect on the proteolytic activity of MccJ25.<sup>32</sup> Taken together, these findings strongly imply the presence of quaternary structure formation between MccJ25 and MccJ25 during MccJ25 maturation (Fig. 3B and C).<sup>32</sup> This interdependency does not appear to hold true for other systems.<sup>40,49</sup> Slight differences in the catalytic mechanisms of related RiPPs enzymes have been observed before, for example in the case of YcaO enzymes, which may phosphorylate or pyrophosphorylate their substrates to achieve the same chemical outcome.<sup>50,51</sup> Unfortunately, to date no crystal structure of either MccJ25 and MccJ25 or the multi-subunit complex has been reported which might shed light on the mechanistic interplay between the two proteins.

In many lasso peptide clusters, *e.g.* streptomycin, paenidin, lassomycin, and lariatin, protein B is organized as two ORFs, B1 and B2, which share homology with the N- and C-terminal domains of the intact B protein.<sup>34</sup> B1 has been shown to share homology with PqqD, a small protein involved in the biosynthesis of another RiPP, pyrroloquinoline quinone (PQQ).<sup>40,52</sup> This PqqD-like domain appears in over half of the known RiPP classes and is reported to be involved in binding to the precursor peptide.<sup>52</sup> This also holds true for lasso peptides as the B1 proteins from paenidin (PadeB1) and lariatin (LarB1) have been shown to interact with their leader peptides with nanomolar affinity.<sup>40</sup> Using hydrogen–deuterium exchange (HDX) mass spectrometry it has been shown that residues present all over PadeB1 are involved in binding to the leader peptide.<sup>40</sup> This indicates that the B1 protein undergoes



**Fig. 3** Biosynthesis and structure of lasso peptides. (A) Proposed mechanism of lasso peptides biosynthesis. Initially the B protein (cysteine protease, red), cleaves the leader peptide off the core peptide and pre-organizes the core peptide for formation of the macrolactam. After substrate activation via adenylation of a glutamic acid side-chain the macrolactam ring is formed. The B protein is often expressed as two individual proteins (B1 and B2). Some systems appear to require complex formation of B and C for protein function, while others appear not to have this restriction. Whether complex formation is still beneficial for catalysis in these systems requires further study. (B) Schematic of the lasso peptide MccJ25 to demonstrate general lasso peptide topology. Single-letter amino acid code and the color scheme of (A) were used. (C) Same as B, but shown as a 3-D structure (PDB ID 1PP5). Residues are numbered and only the side-chains of the two "plug" residues are shown.

a significant conformational rearrangement upon binding to the leader peptide, which might be important for the tight binding with the leader peptide or for interaction with the downstream maturation machinery. *In vitro*, B2 failed to hydrolyze the precursor peptide in the absence of the B1 protein, demonstrating that B1 and B2 need to form a complex for the hydrolysis reaction to take place.<sup>40</sup> In addition, PadeB1 residues D23 and Y38 are thought to be important for B1–B2 interaction, as mutations D23A and Y38A have been shown to have a detrimental effect on the hydrolysis of the precursor peptide, even though these mutants still bind the leader peptide with near wild type affinity.<sup>40</sup> Impairment of the PadeB1–B2 interface could explain the sharp decline in precursor peptide cleavage observed. Very recent work on the lasso peptide fusilassin also demonstrated the importance of the interaction between B1 and B2 (termed FusB and FusE for fusilassins).<sup>37</sup> Biochemical data and modeling studies have shown that the B1 protein, in addition to its expected role in leader binding, is also potentially involved in positioning the precursor peptide such that the scissile bond between the leader and the core peptide is solvent exposed.<sup>53</sup> This presumably positions it for cleavage by the B2 enzyme.<sup>54</sup> Taken together, these studies further strengthen the notion that B1–B2 need to form a complex for precursor peptide maturation to take place. Photo-crosslinking and docking studies in lariatin suggested that in addition to the leader peptide binding function of B1, the LarB1 protein is also interacting with the core peptide residues.<sup>53</sup> Very recent data on

paeninodin however demonstrates that PadeB1 only binds to the leader peptide.<sup>54</sup> This discrepancy will need to be resolved in the future.

Given the lack of structural data for each component of the maturation machinery, a number of key questions remain to be answered regarding the biosynthesis of lasso peptides. While the interactions between the precursor peptide, B1 and B2 proteins was recently investigated<sup>37</sup> using GREMLIN,<sup>55</sup> which can analyze large, diverse sets of sequences to identify sites of compensatory mutations between putative binding partners, the precise nature of the interaction between B and C proteins remains unexplored. Genome mining data suggests that B-enzymes expressed as two parts (B1 and B2) are the evolutionary precursor of fused B-enzymes.<sup>36</sup> This notion is somewhat supported by reports on artificial B1–B2 fusion proteins, which still support lasso peptide biosynthesis.<sup>43</sup> Whether fused B1 and B2 proteins offer advantages with regards to catalysis will require further study.

### 3. Cyanobactins

For cyanobactins we will focus our discussion on patellamides, as they are arguably the best-studied members of this family. Patellamides, found in marine ascidians, are a product of the symbiont cyanobacteria, *Prochloron didemni*.<sup>56</sup> They are pseudosymmetrical cyclic peptides containing oxazoline and thiazol(in)e rings (Fig. 4A).<sup>56</sup> The patellamide gene cluster encodes

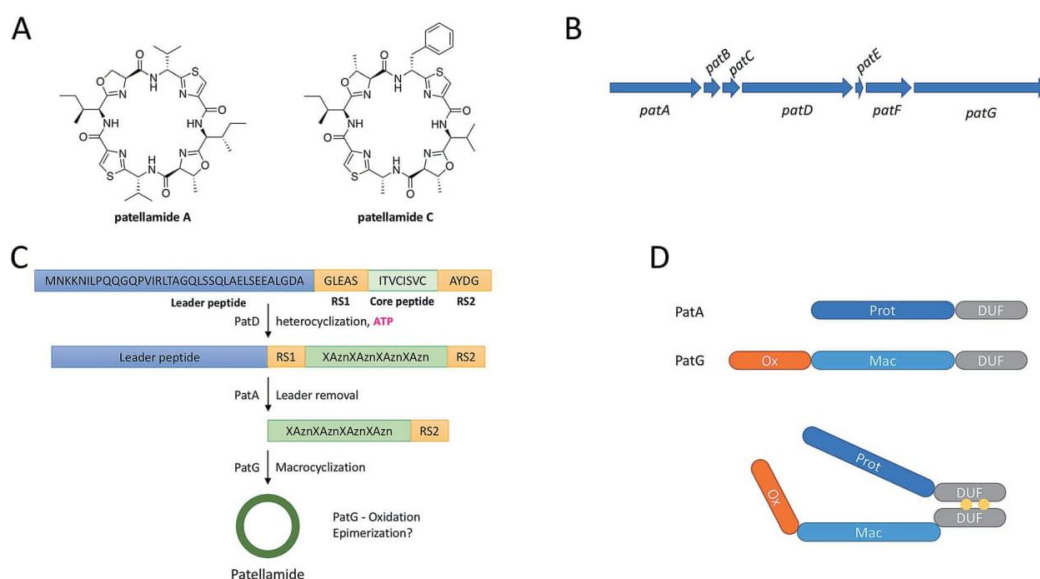


Fig. 4 Structure and biosynthesis of patellamides. (A) Structures of patellamides A and C. (B) The patellamide gene cluster contains genes *patA* to *patG*. Of these, only *patA*, *D*, *E* (precursor peptide), *F* and *G* are essential for patellamide biosynthesis. (C) Order of steps in patellamide biosynthesis. The precursor peptide is first processed by PatD, generating thiazoline and oxazoline heterocycles (Azn) flanked by hydrophobic residues (X). After removal of the leader peptide by PatA the modified core peptide is macrocyclized by PatG. The oxidation of thiazolines to thiazoles can occur at any point after heterocyclization. It is not clear how and when two core peptide amino acids epimerize, but the natural product contains two *D*-amino acids. (D) PatA and PatG are both multi-domain proteins. PatA encodes a protease and a DUF; whereas PatG in addition to DUF also possesses oxidase and macrocyclase domains (top). Possible  $Zn^{2+}$  (yellow circles)-mediated dimerization of the two proteins.

the precursor peptide (PatE), which is flanked by five genes (Pat A, B, C, D, F and G).<sup>57</sup> Of these, only Pat A, D, F and G are essential for patellamide biosynthesis (Fig. 4B and C).<sup>33,57–59</sup> The *patE* gene encodes a precursor peptide of 71 amino acids containing two core peptides, which give rise to two 8-residue macrocycles, for example patellamides A and C (Fig. 4A).<sup>57</sup> The first 37 residues of the precursor peptide serve as a leader peptide required for processing. The leader peptide is followed by two tandem repeats: 5-amino acid conserved protease cleavage sites consisting of the consensus G(L/V)E(A/P)S; two conserved 3-amino acid macrocyclization signatures AYD and the 8-amino acid core peptides, namely cassette I (VTACITFC) and II (ITVCISVC), which are converted to mature patellamides A and C, respectively (Fig. 4A).<sup>57,60</sup>

The hallmark of cyanobactins is the macrocycle, formed *via* the peptide backbone as an amide bond and not *via* amino acid side chains like many other macrocyclic natural products.<sup>58,61–63</sup> Macrocyclization occurs through a two-step process catalyzed by two multidomain proteins: PatA and PatG (Fig. 4C).<sup>61,62</sup> The first step involves the removal of the leader peptide and separation of the core peptides by PatA, which cleaves the precursor peptide after the N-terminal protease site, G(L/V)E(A/P)S.<sup>61</sup> The crystal structure of the PatA protease domain has been solved and it harbors a catalytic triad composed of serine, aspartate, and histidine-typical of subtilisin-like serine protease.<sup>61,62</sup> Interestingly, PatG also contains a subtilisin-like protease domain, PatG macrocyclase (PatGmac), which shares some sequence similarity with PatA

(~40% similarity).<sup>61,62</sup> However, PatGmac acts on a small, linear, non-activated peptide generated by PatA to form the macrocycle.<sup>62</sup> It has been shown that PatGmac cuts before the C-terminal macrocyclization motif, AYD (corresponding to positions P1'–P3'), and macrocyclizes the preceding eight residues, *via* an acyl-enzyme intermediate.<sup>61,62</sup> The crystal structure of PatGmac revealed it to be structurally homologous to PatA, but with an additional helix-loop-helix “macrocylation” insertion.<sup>61,62</sup> This insertion, present in all PatGmac homologs, is thought to protect the acyl-enzyme intermediate from water, which then reacts with the core peptide’s N-terminus as an internal amine nucleophile, rather than water, leading to transamidation and macrocyclization.<sup>61,62,64</sup> Unlike most other proteases, both PatA and G have slow rates of conversion.<sup>62</sup> It has been proposed that this apparent slow rate of conversion might play a role in regulation of biosynthesis.<sup>61,62</sup> However, it may also be due to the absence of protease domain interaction partners, which for example can act to stabilize protease-substrate binding, thereby increasing processivity. PatA and G both contain C-terminal domains of unknown function (DUF), which share high sequence similarity (56%) (Fig. 4D).<sup>33,57,65</sup> The conserved nature of this domain in all patellamide-like cyanobactin biosynthetic pathways characterized to date suggests that this DUF might have a functional role in biosynthesis. The X-ray crystal structure of PatG-DUF has been reported and it forms a dimer mediated by  $Zn^{+2}$  ions (Fig. 4D).<sup>65</sup> Given DUFs from PatA and G are homologous, it is attractive to consider that

these domains could lead to heterodimers of PatA-PatG, which might be important for the activity of these proteins *in vivo*. Such a complex may accelerate the slow conversion rates reported for both proteins, and would allow for substrate channeling, which would reduce the exposure of precursor peptides to cellular proteases and thereby their degradation. The PatG DUF dimer indeed forms a long channel where the precursor peptide could potentially bind.<sup>65</sup> However, binding was neither observed for PatG DUF with linear peptides or linear peptides with heterocycles by isothermal titration calorimetry (ITC).<sup>65</sup> This apparent lack of binding could be due to a number of reasons *e.g.* testing individual domain rather than full length protein and/or unsuitable precursor peptide length *etc.* Therefore, further experiments are needed to investigate the biological relevance of DUF dimerization and its potential significance in patellamide biosynthesis.

#### 4. Linear azolic peptides (LAPs)

Microcin B17 (MccB17) is an antimicrobial peptide produced by several *Escherichia coli* strains.<sup>66</sup> It contains up to nine oxazole and thiazole heterocycles that are introduced post-translationally and are required for its activity (Fig. 5A).<sup>66,67</sup> The biosynthetic gene cluster, which consists of a seven-gene operon (*mcbABCDEFGHI*), was first reported in 1985.<sup>68</sup> The *mcbA* gene codes for a precursor peptide, whereas the *mcbE* and *mcbF* gene products form the export pump to actively transport mature MccB17 out of producing cells.<sup>69–71</sup> The *mcbG* gene product is a pentapeptide repeat protein, which confers self-resistance; it is assumed to act in a similar fashion like its homolog Qnr, which binds to DNA gyrase and protects it from quinolones.<sup>72,73</sup> The defining feature of MccB17 is the presence of thiazole and oxazole heterocycles, which are formed by the

McbBCD synthetase complex: dehydrogenase (McbC) and cyclodehydratase (McbB and McbD) (Fig. 5A).<sup>69,70,74</sup> After a certain number of heterocycles have been formed (major form contains four oxazole and four thiazole heterocycles), the leader sequence is cleaved with the help of the products of two highly conserved genes in prokaryotes *tldD* and *tldE*, before being exported out of the cells.<sup>75,76</sup> The TldD and TldE proteolytic complex appears to process all unfolded peptides in a sequence-independent manner, provided the peptides sterically fit into the active-site. Their effect on microcin biosynthesis may therefore be a by-product of a broader cellular function.<sup>76</sup> *In vitro* reconstitution of MccB17 biosynthesis showed that the presence of all three enzymes, McbBCD, is essential for thiazole/oxazole ring formation.<sup>69,74</sup> However, it was the recent work on *Bacillus sp. Al Hakam* (Balh) LAP that helped shed light on the finer details of the underlying mechanism of the McbBCD synthetase complex.<sup>77</sup> Fortuitously, in Balh, the cyclodehydratase and dehydrogenase activity are not coupled thereby allowing to focus on each biochemical modification independently.<sup>77,78</sup> The Balh protein D was shown to be responsible for catalyzing the dehydratase activity using ATP to phosphorylate the amide backbone of the peptide substrate during the cyclodehydration reaction (Fig. 5B).<sup>50,77,78</sup> Interestingly, protein D could independently install azolines in the Balh precursor peptide (BalhA1) at a much reduced rate compared to the Balh cyclodehydratase DC complex (equivalent to McbBD).<sup>78</sup> While it has been shown that protein C recruits the precursor peptide, it is likely that additional conformational changes are required to promote the activity of protein D in the synthetase complex. Clearly, proteins McbB and D act cooperatively and given this fact it is not surprising that both these proteins are found fused into one polypeptide chain in roughly half of the known LAP clusters.

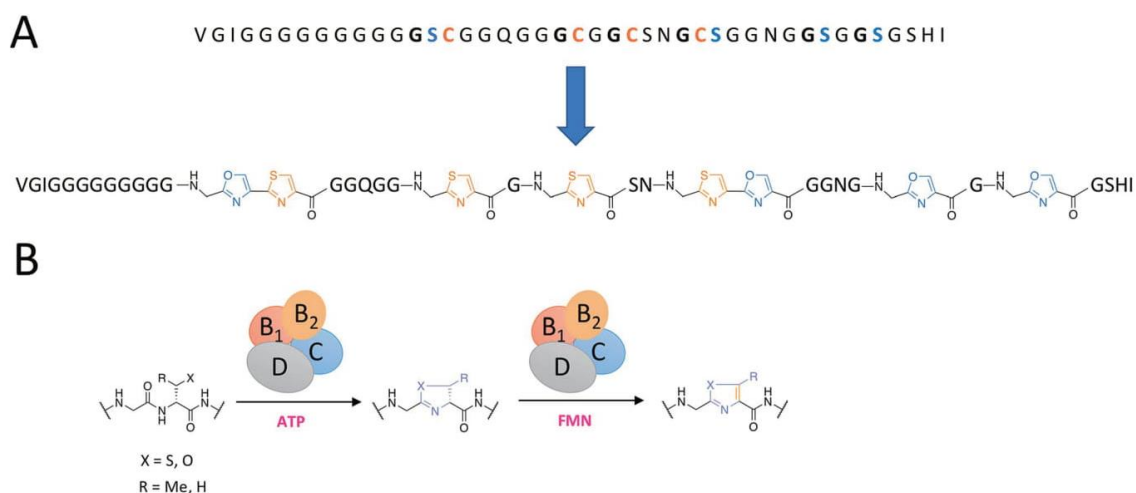


Fig. 5 Structure and biosynthesis of microcin B17. (A) Structure of microcin B17. (B) The maturation of the MccB17 precursor peptide (not shown) carried out by the McbBCD synthetase complex is a two-step process (one half of the  $B_4C_2D_2$  octamer is shown). First, the cyclodehydratase (B and D-proteins) cyclize serine/threonine or cysteine residues into the corresponding azolines. Subsequently, azoline oxidation to azoles is carried out by a FMN-dependent dehydrogenase (C protein).



McbA dehydrogenation is carried out by McbC, which is an FMN-dependent dehydrogenase (Fig. 5B)<sup>69,77</sup> and dehydrogenation is thought to immediately follow cyclization. While azoline intermediates could be verified,<sup>79</sup> the process is coupled and isolation of azoline intermediates has not been successful.<sup>74,77</sup> This is in contrast to cyanobactin and Balh dehydrogenases, which are dispensable for the formation of heterocycles. However, coimmunoprecipitations studies have demonstrated that formation of MccB17 requires proteins B, C and D to form a heterotrimeric complex (Fig. 5B), in which each component is responsible for carrying out a unique function.<sup>69</sup> *In vitro* studies conducted to test the malleability of the cyclodehydratase for noncognate dehydrogenases showed that the MccB17 cyclodehydratase is stringent in terms of its dehydrogenase partner, unlike other LAPs.<sup>77</sup> The very recent report of the crystal structure of the MccB17 synthetase complex sheds light on the observed stringency.<sup>80</sup> It revealed that the McbBCD complex had a very unexpected architecture: it is made up of two dimers, each of which is a tetramer with a B<sub>2</sub>CD composition, yielding an octameric B<sub>4</sub>C<sub>2</sub>D<sub>2</sub> complex. It revealed that the B protein is comprised of two domains, reminiscent of domains 1 and 2 of cyanobactin heterocyclases LynD and TruD,<sup>51,81</sup> which form an asymmetric dimer and binds the leader peptide of McbA. Both, McbC and McbD make contact with one McbB copy each, but the interface between the two tetramers is predominantly composed of McbD–McbD (oxidase) contacts. Surprisingly the active sites of heterocyclase (McbC) and oxidase (McbD) are not adjacent, so the substrate has to be shuttled between the two active centers during iterative heterocyclization and oxidation steps. This complex structure will be highly important for rationally designed biochemical experiments aimed at elucidating the finer details of MccB17 biosynthesis.

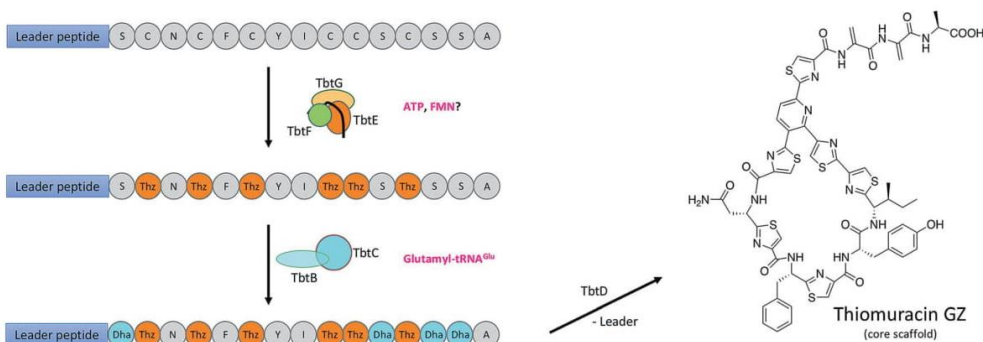
Given the new data available, it is now clear why all three proteins must be present for MccB17 biosynthesis and why the omission of any protein of the Mcb synthetase result in termination of all detectable enzymatic activity. Therefore, the

assembly of the synthetase complex precedes biosynthesis of MccB17.

## 5. Thiopeptides

Thiopeptides were originally discovered as peptidic antibiotics, first in the form of micrococin<sup>82</sup> and soon after the now very extensively studied thiostrepton was found in extracts of *Streptomyces azureus*.<sup>83</sup> After more than 60 years of research into thiopeptides, the first thiopeptide biosynthetic gene clusters were finally reported by several groups.<sup>84–86</sup> Their discovery was facilitated through the genome sequencing efforts of the early 21<sup>st</sup> century and they were shown to be RiPPs. Anecdotal evidence would suggest that thiopeptides were considered too complex to be of ribosomal origin by many in the community. Aided by new genome mining tools, to date over 100 thiopeptides have been discovered.<sup>87,88</sup> Several members of this RiPP family display impressive antibacterial activities, for example towards multi-drug resistant Gram-positive bacteria, while other family members have displayed antitumor and immunosuppressive activities.<sup>24,89,90</sup> Despite the huge structural variations found within this RiPPs family, they all share one common feature: A six-membered azacycle.<sup>91</sup> In addition, thiopeptides usually contain thiazole (and oxazole) heterocycles, which are derived from cysteine (and serine/threonine) residues and dehydrated serine/threonine residues in the form of dehydroalanine and dehydrobutyrine, respectively (Fig. 6).<sup>92</sup>

Two hallmark publications by the Mitchell and van der Donk groups gave unprecedented insights into the biosynthesis of thiopeptides by revealing the substrate specificity and biosynthetic timing of the biosynthesis of the thiomuracin core scaffold through *in vitro* reconstitution of the enzymatic activities.<sup>93,94</sup> Here it was shown that the first step in the biosynthesis of thiomuracin was the installation of all (six) thiazoles in an ordered but non-linear fashion. This step required the presence of three proteins: Ocin-ThiF-like protein TbtF, cyclodehydratase TbtG and dehydrogenase TbtE. It had



**Fig. 6** Conversion of the thiomuracin precursor peptide into the core scaffold of the natural product. In the first step, all cysteine residues are converted into thiazoles (Thz, orange) by the proteins TbtE–G (the precursor peptide is shown as a black line in the complex). Complex formation of leader peptide binding protein TbtF and heterocyclase (thiazoline formation) is likely. It is unclear if oxidase TbtE would also be part of this complex. In the next step, enzymes TbtB and TbtC convert all serines to dehydroalanines (Dha, cyan). It is unclear if these two proteins form a complex or not. A comparison with lanthipeptide biosynthesis (see below) would suggest complex formation. Finally, TbtD removes the leader peptide and triggers the [4 + 2] azo cycloaddition to complete construction of the core scaffold.

been shown previously that TbtF was able to bind the leader peptide of the thiomuracin precursor peptide TbtA with nanomolar affinity.<sup>52</sup> Combining TbtF with TbtG resulted in the ATP-dependent formation of six thiazolines. Addition of TbtE to this reaction lead to the formation and subsequent oxidation of all six thiazolines to thiazoles. Omission of the leader peptide binding protein TbtF (incubation with just TbtG and TbtE) abolished processing of the precursor peptide. While these data strongly imply complex formation of some sort, several points are currently still unclear. It is not known if the cyclodehydratase TbtG alone can act on TbtA (probably not), and affinities of the three components (TbtE/F/G) for each other in a possibly TbtA-dependent manner have not yet been reported.

Once the thiazoles have been installed in the thiomuracin precursor peptide, four serine residues are converted to dehydroalanines by the dehydratase enzymes TbtB and TbtC.<sup>93,94</sup> In contrast to heterocyclization, this process was reported as a C- to N-terminal process. Like heterocyclization (and possibly oxidation), this process is also dependent on the leader peptide. Akin to lanthipeptides (see next section), it was demonstrated that the first step in dehydroalanine formation is the activation of the side-chain's alcohol by glutamylation, using Glu-tRNA as a donor. In thiomuracin, this process is catalyzed by dehydratase TbtB. The Glu is then eliminated by the second enzyme, TbtC. When used on its own, TbtB, which also contains a leader-peptide binding domain, was able to glutamylate one serine, not multiple. Full activity was only observed when both proteins and appropriate co-factors were incubated together. It is currently unclear if TbtB and TbtC form a complex which is required for full activity, or if the observed reduced activity for TbtB is due to a lack of Glu elimination of the first modified residue. Further investigations will be required to delineate the biosynthesis of this fascinating class of RiPPs.

## 6. Lanthipeptides

Lanthipeptides are arguably the most intensively studied and largest family of RiPP compounds.<sup>9,95-97</sup> The founding member nisin (Fig. 1) was first described in 1928 as an inhibitor of bacterial growth.<sup>98</sup> Since then, newly discovered lanthipeptides have shown remarkable diversity in their bioactivities, ranging from anti-infective (anti-bacterial, -viral, and -fungal)<sup>99-101</sup> to anti-nociceptive.<sup>102</sup> Lanthipeptides with antibacterial activity are referred to as lantibiotics,<sup>101</sup> which were found through activity screening without the necessity of genomic information. With the advent of bacterial genome sequencing projects the discovery of new lanthipeptides was no longer restricted to bioactivity assays and has accelerated the discovery of new biosynthetic gene clusters exponentially.<sup>11,103-110</sup> The name lanthipeptides describes their hallmark biochemical transformation: the dehydration of serine residues to dehydroalanine and of threonines to dehydrobutyrine. These dehydroamino acids can then be cross-linked *via* a thioether connecting the C<sub>β</sub> atoms of the two amino acids, resulting in a lanthionine. The thioether is formed by the nucleophilic attack of a cysteine residue Fig. 7.<sup>15</sup>

Various efforts to classify lanthipeptides have been undertaken, but the sheer number of different biosynthetic gene clusters discovered to date has made this a very difficult task. The subdivision of lanthipeptides into four different classes based on differences in their biosynthesis does not make do without exceptions, but it has proven very useful.<sup>111</sup> Details of lanthipeptide biosynthesis and the classifications have been covered very thoroughly elsewhere.<sup>92,112-118</sup> Briefly, class I lanthipeptides use two separate enzymes to convert Ser/Thr to their respective dehydro form (dehydratase) and to then install thioether cross-links (cyclase).<sup>119,120</sup> Dehydration is achieved by glutamylation of the side-chain's alcohol or thiol function and subsequent elimination.<sup>121,122</sup> In class II lanthipeptides the two enzymatic functions are found in a two-domain protein performing both functions. Curiously, here the side-chains are phosphorylated rather than glutamylated to obtain the same final product.<sup>123-125</sup> Class III and IV lanthipeptides are the most recent additions to the family.<sup>103,126,127,128</sup> They carry out cyclization and dehydration reactions using three-domain proteins consisting of an N-terminal lyase domain (phosphate elimination), followed by a kinase domain (to phosphorylate Ser/Thr side-chains) and a C-terminal cyclase domain.<sup>129-134</sup> The difference between these two families lies in their cyclase domains<sup>134</sup> and enzymes belonging to class III are able to produce unprecedented cross-links (labionin), in which new carbon-carbon bonds are formed.<sup>135</sup>

Since the installation of lanthionine and labionin structures in a core peptide results in drastic conformational changes and restrictions, it may be intuitive to assume an ordered process – otherwise certain positions could no longer be accessible after a particular cross-link has been formed. While there is an overall N- to C-terminal directionality to the process in many examples, non-directionality can also be found.<sup>132,136-139</sup> When directionality is observed, it is by no means strict. These observations highlight the tremendous substrate tolerance of lanthipeptide enzymes and their great appeal for biotechnology applications – since pathways tend to have one dehydratase and one cyclase (or one multi-domain enzyme comprised of both), catalysis is carried out on a substrate undergoing multiple drastic changes and those changes do not have to be facilitated in a strict order.

The existence of multicomponent enzymatic complexes has been studied for class I and II lanthipeptides. In class I lanthipeptides, investigations into nisin biosynthesis using co-immunoprecipitation experiments implied a complex between the nisin dehydratase NisB, the corresponding cyclase NisC and the transporter NisT, which exports nisin from the cell (Fig. 7C).<sup>140</sup> It was demonstrated that NisB and NisC associate with the cytoplasmic membrane and interact with NisT.<sup>140</sup> Comparable results were reported for the enzymes and transporter from the subtilin biosynthetic pathway<sup>141</sup> and complex assembly has also been implied for enzymes involved in the biosynthesis of lantibiotic NAI-107.<sup>137</sup> In addition, the interaction between NisB and NisC, as well as NisC and NisT was demonstrated using yeast two-hybrid experiments<sup>140</sup> and the interaction between NisB and NisC appears to require binding of the precursor peptide.<sup>142,143</sup> In spite of these data, it was

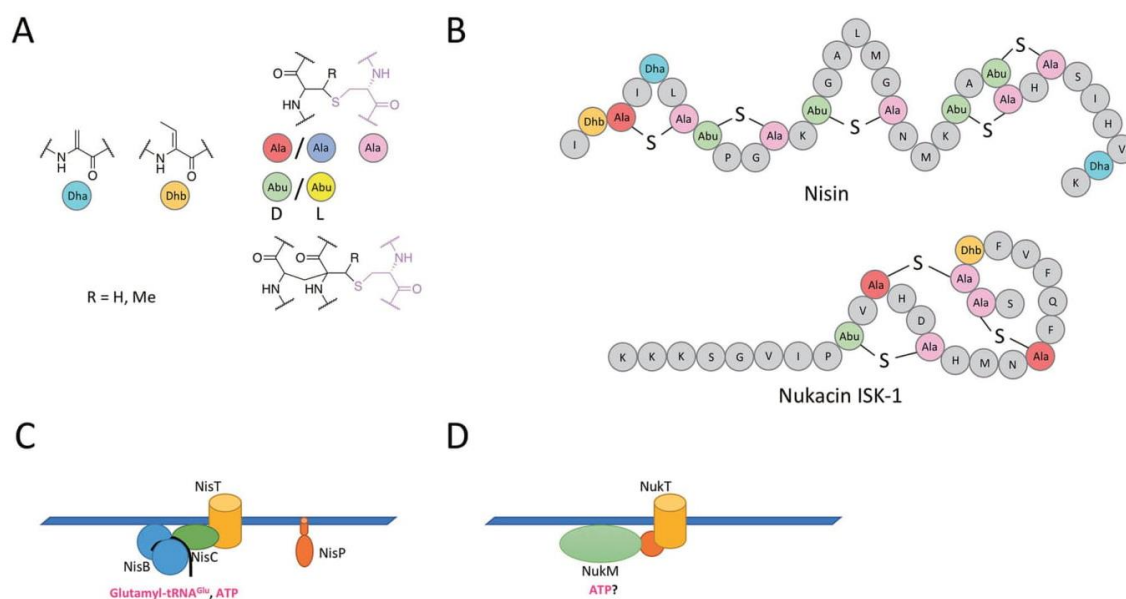


Fig. 7 (A) Serine-derived dehydroalanines (Dha, cyan) and threonine-derived dehydrobutyrine (Dhb, orange) react with cysteines to form the characteristic lanthionine cross-links. D and L refers to the stereochemistry at the  $\alpha$ -carbon of the former serine/threonine residue and the appropriate color-code is used in (B). In class III lanthipeptides an additional type of bond cross-linking two dehydro- and one cysteine residue to form a labionin structure can be found. Thus far, only  $LDL$ -labionine has been observed when two dehydroalanines react. The stereochemistry for Dha–Dhb–Cys labionines is unknown. It is also unclear which stereochemical combinations are possible, and if a Dhb–Dhb–Cys cross-link is possible. (B) Schematic representation of the lanthipeptides nisin and nukacin ISK-1. (C) Based on available data on nisin biosynthesis, NisC (blue) dimer forms a complex with one NisC (green) monomer and the transporter NisT (yellow). It is unclear if the leader peptidase NisP (orange) can also interact with this complex. (D) In the biosynthesis of nukacin ISK-1, the bifunctional protein NukM (green) forms a complex with the protease domain (orange) of the transporter NukT (yellow).

shown that each of the three proteins can perform their respective function independently of the other proteins; thus complex formation was not necessary for enzymatic activity.<sup>120–122,144</sup> In the absence of cyclization, individual dehydration events appear to be largely independent of one another and additional residues could be processed.<sup>145</sup> These positions are likely inaccessible after cyclization. Interestingly, pulse-labeling experiments investigating the activities of NisB and NisT in the absence of the cyclase NisC revealed impaired kinetics.<sup>136</sup> Wild-type-activity could be restored by supplying catalytically inactive NisC.<sup>145</sup> These data indicate that impaired kinetics were not the result of missing cyclization activity but rather of interference with complex formation.<sup>145</sup> A first glimpse of how the NisB–NisC complex may be arranged was recently reported based on *in vitro* work using small-angle X-ray scattering.<sup>146</sup> Here, one NisB dimer associated with one NisC monomer in a precursor peptide-dependent manner. The involvement of NisP, the membrane-anchored protease removing the leader peptide prior to export in nisin biosynthesis,<sup>147,148</sup> in the NisBCT biosynthetic complex has not been investigated. While the activities of NisP and NisT do not appear to be tightly coupled,<sup>149</sup> it appears likely that a four-component complex (NisBCPT) may form. In some class I and most class II lanthipeptide biosynthetic systems protease and exporter are contained in one polypeptide,<sup>150,151</sup> which is a perpetual

complex and supportive of above speculation. A more detailed view, including whether substrate channeling or conformational sampling or a combination of both is employed, will only be possible once more structural data is available. The authors are partial to conformational sampling, since the perpetual to and fro of channeling a substrate between two enzymes appears a rather laborious and complicated process. In contrast, conformational sampling offers a much more straight forward and elegant solution, but there is insufficient evidence to exclude either process.

Data on complex formation in class II lanthipeptide biosynthesis is limited. Since the dehydratase and cyclase activities are encoded in two-domain enzymes, investigations in these areas will be limited to two-component systems (dehydratase/cyclase and exporter). Data is available for enzymes involved in nukacin ISK-1 biosynthesis: here, the biosynthetic enzyme NukM appears to interact with the protease domain of the nukacin exporter NukT (Fig. 7D).<sup>152</sup>

No data is available yet for class III and class IV lanthipeptides.

## 7. Conclusion

In this review, we summarize the current knowledge on protein–protein complex formation in selected RiPP biosynthetic systems. We have also tried to highlight the strong need for

structural data of RiPP enzyme complexes. While some multi-domain RiPP enzyme structures have been reported,<sup>51,122</sup> only one protein–protein complex structure is available.<sup>80</sup> It is important to keep in mind that biochemical and binding assays alone do not reveal atomic level details of the protein–protein interactions. This void can only be filled by high-resolution structures of proteins and protein–protein complexes. A technique thus far not broadly applied to the study of RiPP enzyme complexes that may prove very valuable is HDX mass spectrometry, which offers an alternative to structure determination and has been used in the study of lasso peptide biosynthesis. Since it is an emerging field, the biochemical and structural biology focus thus far has largely been on individual enzymes performing unique transformations. The impressive diversity of structural scaffolds of RiPPs showcases the versatility of their biosynthetic machineries. Taking this into account it is difficult to comment on the importance and the prevalence of multi-modular proteins and protein–protein interactions across all RiPPs. However, all the classes of RiPPs studied to date reveal some basic biosynthetic commonality. In addition, work on the evolution of RiPP enzymes (such as lasso peptide B proteins) points to individual proteins that eventually become fused as one polypeptide. This implies an intimate interaction before the fusion event occurs. In the absence of structural data, the use of algorithms such as GREMLIN will aid the discovery of residues facilitating such interactions by providing testable hypotheses. Owing to this, it would not be surprising to find that protein–protein complexes will play a vital role in RiPP biosynthesis. We are convinced that as the field matures, a combination of *in vitro*, *in vivo* and *in silico* data will fill the gaps in our understanding of this important family of natural compounds.

## 8. Conflicts of interest

There are no conflicts to declare.

## 9. Acknowledgements

We thank Laura Franz, Bastien Schnell and Dr Chengzhang Fu for critical reading for the manuscript. This work was supported by an Emmy Noether fellowship of the DFG (KO 4116/3-1).

## 10. Notes and references

- 1 G. M. Cragg and D. J. Newman, *Biochim. Biophys. Acta*, 2013, **1830**, 3670–3695.
- 2 D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2007, **70**, 461–477.
- 3 J. Krause and G. Tobin, in *Using Old Solutions to New Problems*, ed. M. Kulka, IntechOpen, 2013, ch. 1, DOI: 10.5772/56424.
- 4 D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2016, **79**, 629–661.
- 5 L. Katz and R. H. Baltz, *J. Ind. Microbiol. Biotechnol.*, 2016, **43**, 155–176.
- 6 C. L. Ventola, *Pharm. Therapeut.*, 2015, **40**, 277–283.
- 7 C. R. Hutchinson, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 3010–3012.
- 8 A. C. Letzel, S. J. Pidot and C. Hertweck, *BMC Genomics*, 2014, **15**, 983.
- 9 M. A. Skinnider, C. W. Johnston, R. E. Edgar, C. A. Dejong, N. J. Merwin, P. N. Rees and N. A. Magarvey, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, E6343–E6351.
- 10 N. Ziemert, M. Alanjary and T. Weber, *Nat. Prod. Rep.*, 2016, **33**, 988–1005.
- 11 H. Mohimani, R. D. Kersten, W. T. Liu, M. Wang, S. O. Purvine, S. Wu, H. M. Brewer, L. Pasa-Tolic, N. Bandeira, B. S. Moore, P. A. Pevzner and P. C. Dorrestein, *ACS Chem. Biol.*, 2014, **9**, 1545–1551.
- 12 K. J. Hetrick and W. A. van der Donk, *Curr. Opin. Chem. Biol.*, 2017, **38**, 36–44.
- 13 M. A. Ortega and W. A. van der Donk, *Cell Chem. Biol.*, 2016, **23**, 31–44.
- 14 M. A. Funk and W. A. van der Donk, *Acc. Chem. Res.*, 2017, **50**, 1577–1586.
- 15 P. G. Arnison, M. J. Bibb, G. Bierbaum, A. A. Bowers, T. S. Bugni, G. Bulaj, J. A. Camarero, D. J. Campopiano, G. L. Challis, J. Clardy, P. D. Cotter, D. J. Craik, M. Dawson, E. Dittmann, S. Donadio, P. C. Dorrestein, K. D. Entian, M. A. Fischbach, J. S. Garavelli, U. Goransson, C. W. Gruber, D. H. Haft, T. K. Hemscheidt, C. Hertweck, C. Hill, A. R. Horswill, M. Jaspars, W. L. Kelly, J. P. Klinman, O. P. Kuipers, A. J. Link, W. Liu, M. A. Marahiel, D. A. Mitchell, G. N. Moll, B. S. Moore, R. Muller, S. K. Nair, I. F. Nes, G. E. Norris, B. M. Olivera, H. Onaka, M. L. Patchett, J. Piel, M. J. Reaney, S. Rebuffat, R. P. Ross, H. G. Sahl, E. W. Schmidt, M. E. Selsted, K. Severinov, B. Shen, K. Sivonen, L. Smith, T. Stein, R. D. Sussmuth, J. R. Tagg, G. L. Tang, A. W. Truman, J. C. Vederas, C. T. Walsh, J. D. Walton, S. C. Wenzel, J. M. Willey and W. A. van der Donk, *Nat. Prod. Rep.*, 2013, **30**, 108–160.
- 16 D. Destoumieux-Garzon, X. Thomas, M. Santamaria, C. Goulard, M. Barthelemy, B. Boscher, Y. Bessin, G. Molle, A. M. Pons, L. Letellier, J. Peduzzi and S. Rebuffat, *Mol. Microbiol.*, 2003, **49**, 1031–1041.
- 17 T. Dang and R. D. Sussmuth, *Acc. Chem. Res.*, 2017, **50**, 1566–1576.
- 18 T. Otaka and A. Kaji, *FEBS Lett.*, 1983, **153**, 53–59.
- 19 L. Kjaerulff, A. Sikandar, N. Zaburanyi, S. Adam, J. Herrmann, J. Koehnke and R. Muller, *ACS Chem. Biol.*, 2017, **12**, 2837–2841.
- 20 X. Just-Baringo, F. Albericio and M. Alvarez, *Mar. Drugs*, 2014, **12**, 317–351.
- 21 G. A. Hudson and D. A. Mitchell, *Curr. Opin. Microbiol.*, 2018, **45**, 61–69.
- 22 X. Yang and W. A. van der Donk, *Chemistry*, 2013, **19**, 7662–7677.
- 23 D. E. Ruffner, E. W. Schmidt and J. R. Heemstra, *ACS Synth. Biol.*, 2015, **4**, 482–492.
- 24 X. Just-Baringo, F. Albericio and M. Alvarez, *Angew. Chem., Int. Ed. Engl.*, 2014, **53**, 6602–6616.

- 25 B. J. Burkhart, N. Kakkar, G. A. Hudson, W. A. van der Donk and D. A. Mitchell, *ACS Cent. Sci.*, 2017, **3**, 629–638.
- 26 D. E. Cane and C. T. Walsh, *Chem. Biol.*, 1999, **6**, R319–R325.
- 27 D. E. Cane, C. T. Walsh and C. Khosla, *Science*, 1998, **282**, 63–68.
- 28 S. Smith, *Chem. Biol.*, 2002, **9**, 955–956.
- 29 H. D. Mootz, D. Schwarzer and M. A. Marahiel, *ChemBioChem*, 2002, **3**, 490–504.
- 30 C. M. Czekster, H. Ludewig, S. A. McMahon and J. H. Naismith, *Nat. Commun.*, 2017, **8**, 1045.
- 31 W. J. Crone, N. M. Vior, J. Santos-Aberturas, L. G. Schmitz, F. J. Leeper and A. W. Truman, *Angew. Chem., Int. Ed. Engl.*, 2016, **55**, 9639–9643.
- 32 K. P. Yan, Y. Li, S. Zirah, C. Goulard, T. A. Knappe, M. A. Marahiel and S. Rebuffat, *ChemBioChem*, 2012, **13**, 1046–1052.
- 33 J. Koehnke, A. F. Bent, W. E. Houssen, G. Mann, M. Jaspars and J. H. Naismith, *Curr. Opin. Struct. Biol.*, 2014, **29**, 112–121.
- 34 J. D. Hegemann, M. Zimmermann, X. Xie and M. A. Marahiel, *Acc. Chem. Res.*, 2015, **48**, 1909–1919.
- 35 M. O. Maksimov and A. J. Link, *J. Ind. Microbiol. Biotechnol.*, 2014, **41**, 333–344.
- 36 J. I. Tietz, C. J. Schwalen, P. S. Patel, T. Maxson, P. M. Blair, H. C. Tai, U. I. Zakai and D. A. Mitchell, *Nat. Chem. Biol.*, 2017, **13**, 470–478.
- 37 A. J. DiCaprio, A. Firouzbakht, G. A. Hudson and D. A. Mitchell, *J. Am. Chem. Soc.*, 2018, DOI: 10.1021/jacs.8b09928.
- 38 M. O. Maksimov, S. J. Pan and A. James Link, *Nat. Prod. Rep.*, 2012, **29**, 996–1006.
- 39 M. Zimmermann, J. D. Hegemann, X. Xie and M. A. Marahiel, *Chem. Biol.*, 2013, **20**, 558–569.
- 40 S. Zhu, C. D. Fage, J. D. Hegemann, A. Mielcarek, D. Yan, U. Linne and M. A. Marahiel, *Sci. Rep.*, 2016, **6**, 35604.
- 41 J. D. Hegemann, M. Zimmermann, S. Zhu, D. Klug and M. A. Marahiel, *Biopolymers*, 2013, **100**, 527–542.
- 42 H. G. Choudhury, Z. Tong, I. Mathavan, Y. Li, S. Iwata, S. Zirah, S. Rebuffat, H. W. van Veen and K. Beis, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 9145–9150.
- 43 S. Zhu, J. D. Hegemann, C. D. Fage, M. Zimmermann, X. Xie, U. Linne and M. A. Marahiel, *J. Biol. Chem.*, 2016, **291**, 13662–13678.
- 44 M. O. Maksimov, I. Pelczer and A. J. Link, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 15223–15228.
- 45 M. O. Maksimov and A. J. Link, *J. Am. Chem. Soc.*, 2013, **135**, 12038–12047.
- 46 C. D. Fage, J. D. Hegemann, A. J. Nebel, R. M. Steinbach, S. Zhu, U. Linne, K. Harms, G. Bange and M. A. Marahiel, *Angew. Chem., Int. Ed. Engl.*, 2016, **55**, 12717–12721.
- 47 C. Zong, M. J. Wu, J. Z. Qin and A. J. Link, *J. Am. Chem. Soc.*, 2017, **139**, 10403–10409.
- 48 W. L. Cheung, S. J. Pan and A. J. Link, *J. Am. Chem. Soc.*, 2010, **132**, 2514–2515.
- 49 C. Zong, M. O. Maksimov and A. J. Link, *ACS Chem. Biol.*, 2016, **11**, 61–68.
- 50 K. L. Dunbar, J. O. Melby and D. A. Mitchell, *Nat. Chem. Biol.*, 2012, **8**, 569–575.
- 51 J. Koehnke, G. Mann, A. F. Bent, H. Ludewig, S. Shirran, C. Botting, T. Lebl, W. Houssen, M. Jaspars and J. H. Naismith, *Nat. Chem. Biol.*, 2015, **11**, 558–563.
- 52 B. J. Burkhart, G. A. Hudson, K. L. Dunbar and D. A. Mitchell, *Nat. Chem. Biol.*, 2015, **11**, 564–570.
- 53 W. L. Cheung, M. Y. Chen, M. O. Maksimov and A. J. Link, *ACS Cent. Sci.*, 2016, **2**, 702–709.
- 54 J. D. Hegemann, C. J. Schwalen, D. A. Mitchell and W. A. van der Donk, *Chem. Commun.*, 2018, **54**, 9007–9010.
- 55 S. Ovchinnikov, H. Kamisetty and D. Baker, *eLife*, 2014, **3**, e02030.
- 56 C. M. Ireland, A. R. Durso, R. A. Newman and M. P. Hacker, *J. Org. Chem.*, 1982, **47**, 1807–1811.
- 57 E. W. Schmidt, J. T. Nelson, D. A. Rasko, S. Sudek, J. A. Eisen, M. G. Haygood and J. Ravel, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 7315–7320.
- 58 T. Katoh, Y. Goto, M. S. Reza and H. Suga, *Chem. Commun.*, 2011, **47**, 9946–9958.
- 59 M. S. Donia, B. J. Hathaway, S. Sudek, M. G. Haygood, M. J. Rosovitz, J. Ravel and E. W. Schmidt, *Nat. Chem. Biol.*, 2006, **2**, 729–735.
- 60 C. M. Czekster, Y. Ge and J. H. Naismith, *Curr. Opin. Chem. Biol.*, 2016, **35**, 80–88.
- 61 V. Agarwal, E. Pierce, J. McIntosh, E. W. Schmidt and S. K. Nair, *Chem. Biol.*, 2012, **19**, 1411–1422.
- 62 J. Koehnke, A. Bent, W. E. Houssen, D. Zollman, F. Morawitz, S. Shirran, J. Vendome, A. F. Nneoyiege, L. Trembleau, C. H. Botting, M. C. Smith, M. Jaspars and J. H. Naismith, *Nat. Struct. Mol. Biol.*, 2012, **19**, 767–772.
- 63 B. J. Burkhart, C. J. Schwalen, G. Mann, J. H. Naismith and D. A. Mitchell, *Chem. Rev.*, 2017, **117**, 5389–5456.
- 64 N. F. Bras, P. Ferreira, A. R. Calixto, M. Jaspars, W. Houssen, J. H. Naismith, P. A. Fernandes and M. J. Ramos, *Chemistry*, 2016, **22**, 13089–13097.
- 65 G. Mann, J. Koehnke, A. F. Bent, R. Graham, W. Houssen, M. Jaspars, U. Schwarz-Linek and J. H. Naismith, *Acta Crystallogr., Sect. F: Struct. Biol. Commun.*, 2014, **70**, 1597–1603.
- 66 I. A. Khmel, V. M. Bondarenko, I. M. Manokhina, E. I. Basyuk, A. Z. Metlitskaya, V. A. Lipasova and Y. M. Romanova, *FEMS Microbiol. Lett.*, 1993, **111**, 269–274.
- 67 R. Sinha Roy, N. L. Kelleher, J. C. Milne and C. T. Walsh, *Chem. Biol.*, 1999, **6**, 305–318.
- 68 J. L. San Millan, C. Hernandez-Chico, P. Pereda and F. Moreno, *J. Bacteriol.*, 1985, **163**, 275–281.
- 69 Y. M. Li, J. C. Milne, L. L. Madison, R. Kolter and C. T. Walsh, *Science*, 1996, **274**, 1188–1193.
- 70 M. V. Meterev and D. A. Giliarov, *J. Mol. Biol.*, 2014, **48**, 36–54.
- 71 J. Davagnino, M. Herrero, D. Furlong, F. Moreno and R. Kolter, *Proteins*, 1986, **1**, 230–238.
- 72 J. H. Tran, G. A. Jacoby and D. C. Hooper, *Antimicrob. Agents Chemother.*, 2005, **49**, 118–125.
- 73 I. Shkundina, M. Serebryakova and K. Severinov, *J. Bacteriol.*, 2014, **196**, 1759–1767.

- 74 J. C. Milne, R. S. Roy, A. C. Eliot, N. L. Kelleher, A. Wokhlu, B. Nickels and C. T. Walsh, *Biochemistry*, 1999, **38**, 4768–4781.
- 75 N. Allali, H. Afif, M. Couturier and L. Van Melderen, *J. Bacteriol.*, 2002, **184**, 3224–3231.
- 76 D. Ghilarov, M. Serebryakova, C. E. M. Stevenson, S. J. Hearnshaw, D. S. Volkov, A. Maxwell, D. M. Lawson and K. Severinov, *Structure*, 2017, **25**, 1549–1561.
- 77 J. O. Melby, X. Li and D. A. Mitchell, *Biochemistry*, 2014, **53**, 413–422.
- 78 J. O. Melby, K. L. Dunbar, N. Q. Trinh and D. A. Mitchell, *J. Am. Chem. Soc.*, 2012, **134**, 5309–5316.
- 79 K. L. Dunbar and D. A. Mitchell, *J. Am. Chem. Soc.*, 2013, **135**, 8692–8701.
- 80 D. Ghilarov, C. E. M. Stevenson, D. Y. Travin, J. Piskunova, M. Serebryakova, A. Maxwell, D. M. Lawson and K. Severinov, *Mol. Cell*, 2019, **73**, 749–762.
- 81 J. Koehnke, A. F. Bent, D. Zollman, K. Smith, W. E. Houssen, X. Zhu, G. Mann, T. Lebl, R. Scharff, S. Shirran, C. H. Botting, M. Jaspars, U. Schwarz-Linek and J. H. Naismith, *Angew. Chem., Int. Ed. Engl.*, 2013, **52**, 13991–13996.
- 82 T. L. Su, *Br. J. Exp. Pathol.*, 1948, **29**, 473–481.
- 83 J. D. Dutcher and J. Vandeputte, *Antibiot. Annu.*, 1955, **3**, 560–561.
- 84 R. Liao, L. Duan, C. Lei, H. Pan, Y. Ding, Q. Zhang, D. Chen, B. Shen, Y. Yu and W. Liu, *Chem. Biol.*, 2009, **16**, 141–147.
- 85 W. L. Kelly, L. Pan and C. Li, *J. Am. Chem. Soc.*, 2009, **131**, 4327–4334.
- 86 R. P. Morris, J. A. Leeds, H. U. Naegeli, L. Oberer, K. Memmert, E. Weber, M. J. LaMarche, C. N. Parker, N. Burren, S. Esterow, A. E. Hein, E. K. Schmitt and P. Krastel, *J. Am. Chem. Soc.*, 2009, **131**, 5946–5955.
- 87 Q. Zheng, H. Fang and W. Liu, *Org. Biomol. Chem.*, 2017, **15**, 3376–3390.
- 88 C. J. Schwalen, G. A. Hudson, B. Kille and D. A. Mitchell, *J. Am. Chem. Soc.*, 2018, **140**, 9494–9501.
- 89 C. Li and W. L. Kelly, *Nat. Prod. Rep.*, 2010, **27**, 153–164.
- 90 Q. Zhang and W. Liu, *Nat. Prod. Rep.*, 2013, **30**, 218–226.
- 91 M. C. Bagley, J. W. Dale, E. A. Merritt and X. Xiong, *Chem. Rev.*, 2005, **105**, 685–714.
- 92 L. M. Repka, J. R. Chekan, S. K. Nair and W. A. van der Donk, *Chem. Rev.*, 2017, **117**, 5457–5520.
- 93 Z. Zhang, G. A. Hudson, N. Mahanta, J. I. Tietz, W. A. van der Donk and D. A. Mitchell, *J. Am. Chem. Soc.*, 2016, **138**, 15511–15514.
- 94 G. A. Hudson, Z. Zhang, J. I. Tietz, D. A. Mitchell and W. A. van der Donk, *J. Am. Chem. Soc.*, 2015, **137**, 16012–16015.
- 95 J. R. Doroghazi, J. C. Albright, A. W. Goering, K. S. Ju, R. R. Haines, K. A. Tchalukov, D. P. Labeda, N. L. Kelleher and W. W. Metcalf, *Nat. Chem. Biol.*, 2014, **10**, 963–968.
- 96 P. Cimermancic, M. H. Medema, J. Claesen, K. Kurita, L. C. Wieland Brown, K. Mavrommatis, A. Pati, P. A. Godfrey, M. Koehrsen, J. Clardy, B. W. Birren, E. Takano, A. Sali, R. G. Linington and M. A. Fischbach, *Cell*, 2014, **158**, 412–421.
- 97 C. J. Walsh, C. M. Guinane, C. Hill, R. P. Ross, P. W. O'Toole and P. D. Cotter, *BMC Microbiol.*, 2015, **15**, 183.
- 98 L. A. Rogers, *J. Bacteriol.*, 1928, **16**, 321–325.
- 99 K. I. Mohr, C. Volz, R. Jansen, V. Wray, J. Hoffmann, S. Bernecker, J. Wink, K. Gerth, M. Stadler and R. Muller, *Angew. Chem., Int. Ed. Engl.*, 2015, **54**, 11254–11258.
- 100 G. Ferir, M. I. Petrova, G. Andrei, D. Huskens, B. Hoorelbeke, R. Snoeck, J. Vanderleyden, J. Balzarini, S. Bartoschek, M. Bronstrup, R. D. Sussmuth and D. Schols, *PLoS One*, 2013, **8**, e64010.
- 101 N. Schnell, K. D. Entian, U. Schneider, F. Gotz, H. Zahner, R. Kellner and G. Jung, *Nature*, 1988, **333**, 276–278.
- 102 M. Iorio, O. Sasso, S. I. Maffioli, R. Bertorelli, P. Monciardini, M. Sosio, F. Bonezzi, M. Summa, C. Brunati, R. Bordoni, G. Corti, G. Tarozzo, D. Piomelli, A. Reggiani and S. Donadio, *ACS Chem. Biol.*, 2014, **9**, 398–404.
- 103 Q. Zhang, J. R. Doroghazi, X. Zhao, M. C. Walker and W. A. van der Donk, *Appl. Environ. Microbiol.*, 2015, **81**, 4339–4350.
- 104 J. E. Velasquez and W. A. van der Donk, *Curr. Opin. Chem. Biol.*, 2011, **15**, 11–21.
- 105 A. de Jong, A. J. van Heel, J. Kok and O. P. Kuipers, *Nucleic Acids Res.*, 2010, **38**, W647–W651.
- 106 M. H. Medema, K. Blin, P. Cimermancic, V. de Jager, P. Zakrzewski, M. A. Fischbach, T. Weber, E. Takano and R. Breitling, *Nucleic Acids Res.*, 2011, **39**, W339–W346.
- 107 R. D. Kersten, Y. L. Yang, Y. Xu, P. Cimermancic, S. J. Nam, W. Fenical, M. A. Fischbach, B. S. Moore and P. C. Dorrestein, *Nat. Chem. Biol.*, 2011, **7**, 794–802.
- 108 A. J. van Heel, A. de Jong, M. Montalban-Lopez, J. Kok and O. P. Kuipers, *Nucleic Acids Res.*, 2013, **41**, W448–W453.
- 109 M. Singh and D. Sareen, *PLoS One*, 2014, **9**, e91352.
- 110 T. Weber, K. Blin, S. Duddela, D. Krug, H. U. Kim, R. Bruccoleri, S. Y. Lee, M. A. Fischbach, R. Muller, W. Wohlleben, R. Breitling, E. Takano and M. H. Medema, *Nucleic Acids Res.*, 2015, **43**, W237–W243.
- 111 U. Pag and H. G. Sahl, *Curr. Pharm. Des.*, 2002, **8**, 815–833.
- 112 R. J. Siezen, O. P. Kuipers and W. M. de Vos, *Antonie van Leeuwenhoek*, 1996, **69**, 171–184.
- 113 H. G. Sahl and G. Bierbaum, *Annu. Rev. Microbiol.*, 1998, **52**, 41–79.
- 114 C. van Kraaij, W. M. de Vos, R. J. Siezen and O. P. Kuipers, *Nat. Prod. Rep.*, 1999, **16**, 575–587.
- 115 O. McAuliffe, R. P. Ross and C. Hill, *FEMS Microbiol. Rev.*, 2001, **25**, 285–308.
- 116 C. Chatterjee, M. Paul, L. Xie and W. A. van der Donk, *Chem. Rev.*, 2005, **105**, 633–684.
- 117 P. D. Cotter, C. Hill and R. P. Ross, *Curr. Protein Pept. Sci.*, 2005, **6**, 61–75.
- 118 G. Jung, *Angew. Chem., Int. Ed. Engl.*, 1991, **30**, 1051–1068.
- 119 L. D. Kluskens, A. Kuipers, R. Rink, E. de Boef, S. Fekken, A. J. Driessen, O. P. Kuipers and G. N. Moll, *Biochemistry*, 2005, **44**, 12827–12834.
- 120 B. Li, J. P. Yu, J. S. Brunzelle, G. N. Moll, W. A. van der Donk and S. K. Nair, *Science*, 2006, **311**, 1464–1467.
- 121 N. Garg, L. M. Salazar-Ocampo and W. A. van der Donk, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 7258–7263.

- 122 M. A. Ortega, Y. Hao, Q. Zhang, M. C. Walker, W. A. van der Donk and S. K. Nair, *Nature*, 2015, **517**, 509–512.
- 123 L. Xie, L. M. Miller, C. Chatterjee, O. Averin, N. L. Kelleher and W. A. van der Donk, *Science*, 2004, **303**, 679–681.
- 124 M. S. Gilmore, R. A. Segarra, M. C. Booth, C. P. Bogie, L. R. Hall and D. B. Clewell, *J. Bacteriol.*, 1994, **176**, 7335–7344.
- 125 C. Chatterjee, L. M. Miller, Y. L. Leung, L. Xie, M. Yi, N. L. Kelleher and W. A. van der Donk, *J. Am. Chem. Soc.*, 2005, **127**, 15332–15333.
- 126 S. Kodani, M. E. Hudson, M. C. Durrant, M. J. Buttner, J. R. Nodwell and J. M. Willey, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 11448–11453.
- 127 V. Wiebach, A. Mainz, M. J. Siegert, N. A. Jungmann, G. Lesquame, S. Tirat, A. Dreux-Zigha, J. Aszodi, D. Le Beller and R. D. Sussmuth, *Nat. Chem. Biol.*, 2018, **14**, 652–654.
- 128 J. D. Hegemann and W. A. van der Donk, *J. Am. Chem. Soc.*, 2018, **140**, 5743–5754.
- 129 W. M. Muller, T. Schmiederer, P. Ensle and R. D. Sussmuth, *Angew. Chem., Int. Ed. Engl.*, 2010, **49**, 2436–2440.
- 130 Y. Goto, A. Okesli and W. A. van der Donk, *Biochemistry*, 2011, **50**, 891–898.
- 131 W. M. Muller, P. Ensle, B. Krawczyk and R. D. Sussmuth, *Biochemistry*, 2011, **50**, 8362–8373.
- 132 N. A. Jungmann, B. Krawczyk, M. Tietzmann, P. Ensle and R. D. Sussmuth, *J. Am. Chem. Soc.*, 2014, **136**, 15222–15228.
- 133 B. Krawczyk, G. H. Voller, J. Voller, P. Ensle and R. D. Sussmuth, *ChemBioChem*, 2012, **13**, 2065–2071.
- 134 Y. Goto, B. Li, J. Claesen, Y. Shi, M. J. Bibb and W. A. van der Donk, *PLoS Biol.*, 2010, **8**, e1000339.
- 135 K. Meindl, T. Schmiederer, K. Schneider, A. Reicke, D. Butz, S. Keller, H. Guhring, L. Vertesy, J. Wink, H. Hoffmann, M. Bronstrup, G. M. Sheldrick and R. D. Sussmuth, *Angew. Chem., Int. Ed. Engl.*, 2010, **49**, 1151–1154.
- 136 H. B. van den Berg van Saparoea, P. J. Bakkes, G. N. Moll and A. J. Driessen, *Appl. Environ. Microbiol.*, 2008, **74**, 5541–5548.
- 137 M. A. Ortega, Y. Hao, M. C. Walker, S. Donadio, M. Sosio, S. K. Nair and W. A. van der Donk, *Cell Chem. Biol.*, 2016, **23**, 370–380.
- 138 Q. Zhang, M. Ortega, Y. Shi, H. Wang, J. O. Melby, W. Tang, D. A. Mitchell and W. A. van der Donk, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 12031–12036.
- 139 S. Mukherjee and W. A. van der Donk, *J. Am. Chem. Soc.*, 2014, **136**, 10450–10459.
- 140 K. Siegers, S. Heinzmann and K. D. Entian, *J. Biol. Chem.*, 1996, **271**, 12294–12301.
- 141 P. Kiesau, U. Eikmanns, Z. Gutowski-Eckel, S. Weber, M. Hammelmann and K. D. Entian, *J. Bacteriol.*, 1997, **179**, 1475–1481.
- 142 R. Khusainov and O. P. Kuipers, *ChemBioChem*, 2012, **13**, 2433–2438.
- 143 R. Khusainov and O. P. Kuipers, *PLoS One*, 2013, **8**, e74890.
- 144 A. Kuipers, J. Wierenga, R. Rink, L. D. Kluskens, A. J. Driessen, O. P. Kuipers and G. N. Moll, *Appl. Environ. Microbiol.*, 2006, **72**, 7626–7633.
- 145 J. Lubelski, R. Khusainov and O. P. Kuipers, *J. Biol. Chem.*, 2009, **284**, 25962–25972.
- 146 J. Reiners, A. Abts, R. Clemens, S. H. Smits and L. Schmitt, *Sci. Rep.*, 2017, **7**, 42163.
- 147 S. K. Mazmanian, H. Ton-That and O. Schneewind, *Mol. Microbiol.*, 2001, **40**, 1049–1057.
- 148 J. R. van der Meer, J. Polman, M. M. Beerthuyzen, R. J. Siezen, O. P. Kuipers and W. M. De Vos, *J. Bacteriol.*, 1993, **175**, 2578–2588.
- 149 A. Kuipers, E. de Boef, R. Rink, S. Fekken, L. D. Kluskens, A. J. Driessen, K. Leenhouts, O. P. Kuipers and G. N. Moll, *J. Biol. Chem.*, 2004, **279**, 22176–22182.
- 150 L. S. Havarstein, D. B. Diep and I. F. Nes, *Mol. Microbiol.*, 1995, **16**, 229–240.
- 151 C. M. Franke, J. Tiemersma, G. Venema and J. Kok, *J. Biol. Chem.*, 1999, **274**, 8484–8490.
- 152 J. Nagao, Y. Aso, T. Sashihara, K. Shioya, A. Adachi, J. Nakayama and K. Sonomoto, *Biosci., Biotechnol., Biochem.*, 2005, **69**, 1341–1347.

## Tutuillamides A–C: Vinyl-Chloride-Containing Cyclodepsipeptides from Marine Cyanobacteria with Potent Elastase Inhibitory Properties

Lena Keller, Kirley Marques Canuto, Chenxi Liu, Brian M. Suzuki, Jehad Almaliti, Asfandyar Sikandar, C. Benjamin Naman, Evgenia Glukhov, Danmeng Luo, Brendan M. Duggan, Hendrik Luesch, Jesko Koehnke, Anthony J. O'Donoghue, and William H. Gerwick\*

Cite This: <https://dx.doi.org/10.1021/acscchembio.9b00992>

Read Online

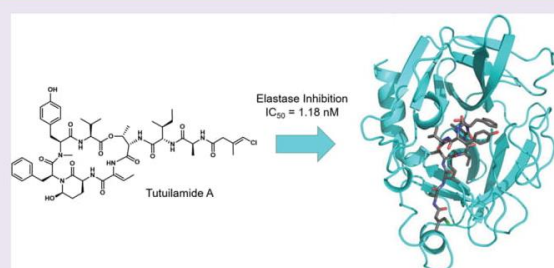
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** Marine cyanobacteria (blue-green algae) have been shown to possess an enormous capacity to produce structurally diverse natural products that exhibit a broad spectrum of potent biological activities, including cytotoxic, antifungal, antiparasitic, antiviral, and antibacterial activities. Using mass-spectrometry-guided fractionation together with molecular networking, cyanobacterial field collections from American Samoa and Palmyra Atoll yielded three new cyclic peptides, tutuillamides A–C. Their structures were established by spectroscopic techniques including 1D and 2D NMR, HR-MS, and chemical derivatization. Structure elucidation was facilitated by employing advanced NMR techniques including nonuniform sampling in combination with the 1,1-ADEQUATE experiment. These cyclic peptides are characterized by the presence of several unusual residues including 3-amino-6-hydroxy-2-piperidone and 2-amino-2-butenic acid, together with a novel vinyl chloride-containing residue. Tutuillamides A–C show potent elastase inhibitory activity together with moderate potency in H-460 lung cancer cell cytotoxicity assays. The binding mode to elastase was analyzed by X-ray crystallography revealing a reversible binding mode similar to the natural product lyngbyastatin 7. The presence of an additional hydrogen bond with the amino acid backbone of the flexible side chain of tutuillamide A, compared to lyngbyastatin 7, facilitates its stabilization in the elastase binding pocket and possibly explains its enhanced inhibitory potency.



Cyanobacteria produce a wide range of cyclic depsipeptides that contain an unusual 3-amino-6-hydroxy-2-piperidone (Ahp) moiety; these peptides are generally referred to as cyanopeptolin-like peptides or Ahp-cyclodepsipeptides.<sup>1–3</sup> The general structure includes an amino acid chain of variable length of which six amino acid residues form a macrocyclic ring. Most cyanopeptolin-like peptides contain the Ahp residue and an ester linkage between the  $\beta$ -OH group of L-threonine and the carboxyl group of the C-terminal amino acid. Additionally, among these Ahp-containing peptides, there are more than 20 compounds that contain a highly conserved hexadepsipeptide core bearing an additional 2-amino-2-butenic acid (Abu) moiety adjacent to the Ahp ring together with a highly variable side chain (Table 1).

Ahp-containing peptides have frequently been observed from both marine and freshwater cyanobacteria as well as other bacteria, and over 200 are now known.<sup>3</sup> However, peptides containing the Abu moiety are generally found in marine bacteria, with the exception being stigonemapeptin, which derives from a freshwater cyanobacterium (Table 1).<sup>4</sup> Interestingly, the Abu moiety in stigonemapeptin is (*E*)-configured, whereas the Abu moieties derived from marine

sources are consistently (*Z*)-configured. X-ray crystallography of the elastase complex formed with lyngbyastatin 7 (4) revealed that these Abu-cyclodepsipeptides act as substrate mimics.<sup>13</sup> The Abu moiety was shown to occupy the S1 substrate binding pocket and engage in a noncovalent interaction. Additionally, a study of the binding mode of scyptolin showed that the Ahp moiety occupies a crucial part of the active site pocket and thereby prevents hydrolysis.<sup>15</sup> It was suggested that the Abu moiety increases the potency of these elastase inhibitors and that the pendant side chain is responsible for modulating their activity and selectivity.<sup>13</sup>

In the current work, we report the isolation of three new members of this Ahp and Abu-containing family of peptides, named tutuillamides A (1) and B (2) based on the location of collection of the source cyanobacterium, *Schizothrix* sp., along

Received: December 11, 2019

Accepted: January 14, 2020

Published: January 14, 2020



**Table 1.** Abu and Ahp-Containing Peptides Isolated from Cyanobacteria

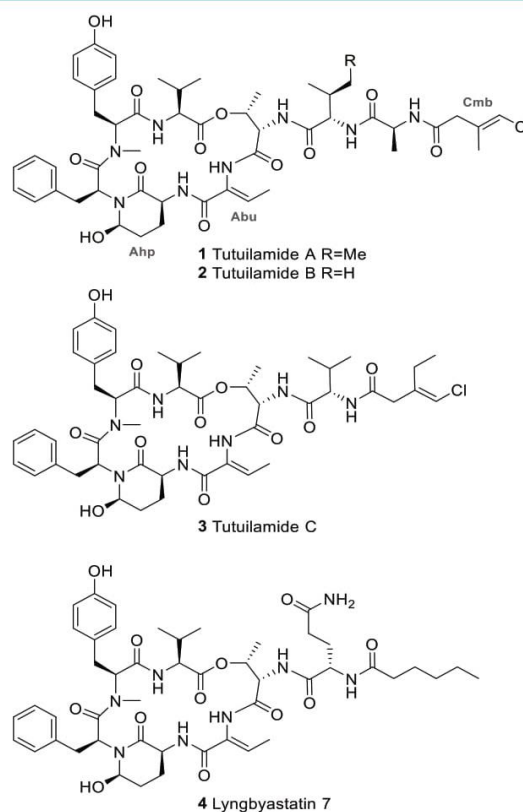
compound name	year	collected from	isolated organism
dolastatin 13 <sup>5</sup>	1989	East Africa, Indian Ocean	<i>Dolabella auricularia</i> (sea hare)
symplostatin 2 <sup>6</sup>	1999	Guam, Pago Bay	<i>Symploca hydnooides</i>
somamide A+B <sup>7</sup>	2001	Fiji, Somo Somo	<i>Lyngbya majuscula</i> / <i>Schizothrix</i> sp.
lyngbyastatin 4 <sup>8</sup>	2007	South Florida, Atlantic coast	<i>Lyngbya confervoides</i>
lyngbyastatin 5–7 <sup>9</sup>	2007	South Florida, Fort Lauderdale/Florida Keys, Summerland Key	<i>Lyngbya confervoides</i>
lyngbyastatin 8–10 <sup>10</sup> /bouillomide A+B <sup>11</sup>	2009	Guam, Tumon Bay	<i>Lyngbya semiplena</i>
molassamide <sup>12</sup>	2010	Florida, Molasses Reef	<i>Dichothrix utahensis</i>
stigonemapeptin <sup>4</sup>	2012	Wisconsin, North Nokomis Lake	<i>Stigonema</i> sp. (freshwater)
symplostatin 5–10 <sup>13</sup>	2013	Guam, Cetti Bay	red <i>Symploca</i> sp.
kurahamide <sup>14</sup>	2014	Japan, Kuraha	<i>Lyngbya</i> sp. assembly

with tutuilamide C (3) isolated from a *Coleofasciculus* sp. These structures were assembled by a combination of NMR, mass spectrometry (MS), and chromatographic analysis following acid hydrolysis and chiral derivatization. They feature an unusual vinyl-chloride-containing residue never previously observed in this structure class. The new natural products, as well as two semisynthetic derivatives, were evaluated for serine protease inhibition, and all of the cyclic species were found to be highly potent. The crystal structure of elastase in complex with tutuilamide A revealed extensive binding interactions in the substrate binding pocket, as has been shown previously with lyngbyastatin 7 (4). However, we identified additional hydrogen bond interactions between tutuilamide A and elastase that did not occur in the lyngbyastatin 7 cocrystal structure. These may be responsible for the increased potency of tutuilamide A compared to lyngbyastatin 7.

## RESULTS AND DISCUSSION

Our discovery strategy to locate natural products with novel structural frameworks includes MS<sup>2</sup>-based metabolomics (Molecular Networking) for strain selection and dereplication as well as chromatographic methods for isolation driven by structural features. Cyanobacterial colonies of the genus *Schizothrix* sp. and *Coleofasciculus* sp. were collected by hand from the main island of Tutuila in American Samoa in 2016 and Palmyra Atoll in 2008, respectively, using SCUBA gear. The crude CH<sub>2</sub>Cl<sub>2</sub>–MeOH (2:1) extract was initially fractionated using vacuum-liquid chromatography (VLC) as well as solid phase extraction (SPE) for further analysis by MS and NMR. This approach revealed the presence of peptides with unusual features and led to the HPLC isolation of tutuilamides A and B from *Schizothrix* sp. and tutuilamide C from *Coleofasciculus* sp. In addition, we identified related peptides such as symplostatin 2 together with several derivatives of dolastatin 13 that have yet to be characterized. Tutuilamide A (1) and B (2) share the same cyclic peptide

core and possess the unusual Ahp and Abu moieties. They differ by a single side chain residue wherein isoleucine is replaced by valine (Figure 1). Meanwhile, tutuilamide C (3) is

**Figure 1.** Chemical structure of tutuilamide A (1), B (2), and C (3) together with the related lyngbyastatin 7 (4).<sup>9</sup>

an analog of 2 that lacks an alanine moiety but bears an additional methylene unit in the aliphatic side-chain. Moreover, they are the first cyanopeptolins to possess a vinyl chloride residue in the side chain. Vinyl chloride functionalities in cyanobacteria have been shown to biosynthetically result from a unique cassette of enzymes that involve polyketide synthase (PKS) beta branch formation along with radical-based chlorination of an intermediate, such as in jamaicamide A.<sup>16</sup>

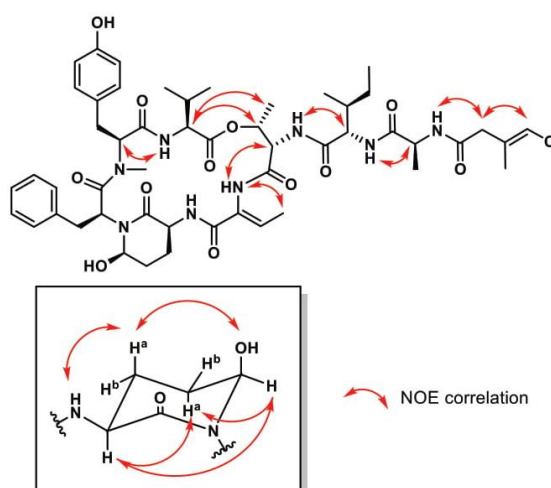
High-resolution electrospray ionization mass spectrometry (HR-ESI-MS) of tutuilamide A (1) displayed an ion peak at  $m/z$  1043.4616  $[M + Na]^+$  (calculated for C<sub>51</sub>H<sub>69</sub>ClN<sub>8</sub>O<sub>12</sub>Na, 1043.4616,  $\Delta = 0.0$  ppm), consistent with the molecular formula C<sub>51</sub>H<sub>69</sub>ClN<sub>8</sub>O<sub>12</sub> containing 21 double-bond equivalents. The isotope pattern for the molecular ion cluster indicated the clear presence of one chlorine atom. The <sup>1</sup>H NMR spectrum of 1 in DMSO-*d*<sub>6</sub> exhibited signals characteristic of a peptide including seven  $\alpha$ -proton signals at  $\delta$  4.65 (overlap), 4.89 (1H, dd,  $J = 11.5, 2.2$  Hz), 4.73 (1H, dd,  $J = 11.4, 4.0$  Hz), 3.79 (1H, m), 4.67 (overlap), 4.40 (1H, t,  $J = 7.8$  Hz), and 4.36 (1H, p,  $J = 7.2$  Hz) along with six amide NH signals at  $\delta$  7.53 (1H, t,  $J = 8.4$  Hz), 7.22 (1H, broad), 9.23 (1H, broad), 7.91 (1H, broad), and 8.21 (1H, d,  $J = 7.4$  Hz; Table S1). Additionally, a downfield pair of triplets at  $\delta$  7.18

(2H, t,  $J = 7.2$  Hz) and 7.15 (1H, t,  $J = 7.2$  Hz) and a doublet at  $\delta$  6.83 (2H, d,  $J = 7.2$  Hz) were characteristic of a phenyl group; similarly, two doublets at  $\delta$  7.00 (2H, d,  $J = 8.3$  Hz) and 6.78 (2H, d,  $J = 8.3$  Hz) were characteristic for a *para*-substituted phenol group. The proton NMR spectrum also showed a singlet at  $\delta$  2.77 (3H, s), indicative of an *N*-methyl amide, together with eight partially overlapping methyl signals at  $\delta$  0.75 (3H, d,  $J = 6.8$  Hz),  $\delta$  0.80 (3H, t,  $J = 7.5$  Hz),  $\delta$  0.84 (3H, d,  $J = 6.8$  Hz),  $\delta$  0.88 (3H, d,  $J = 6.8$  Hz),  $\delta$  1.19 (3H, d,  $J = 7.0$  Hz),  $\delta$  1.21 (3H, d,  $J = 6.6$  Hz),  $\delta$  1.48 (3H, d,  $J = 7.1$  Hz), and  $\delta$  1.72 (3H, d,  $J = 1.2$  Hz).

The HSQC spectrum revealed the presence of six methylene groups ( $\delta$ C-3<sub>Tyr</sub> 32.5,  $\delta$ H-3<sub>Tyr</sub> 3.11/2.69;  $\delta$ C-3<sub>Phe</sub> 35.2,  $\delta$ H-3<sub>Phe</sub> 2.88/1.83;  $\delta$ C-3<sub>Ahp</sub> 21.9,  $\delta$ H-3<sub>Ahp</sub> 2.42/1.58;  $\delta$ C-4<sub>Ahp</sub> 29.2,  $\delta$ H-4<sub>Ahp</sub> 1.72/1.58;  $\delta$ C-4<sub>Ile</sub> 23.9,  $\delta$ H-4<sub>Ile</sub> 1.43/1.08;  $\delta$ C-2<sub>Cmb</sub> 42.5,  $\delta$ H-2<sub>Cmb</sub> 2.97), two methines ( $\delta$ C-3<sub>Val</sub> 30.5,  $\delta$ H-3<sub>Val</sub> 2.07;  $\delta$ C-3<sub>Ile</sub> 36.5,  $\delta$ H-3<sub>Ile</sub> 1.81), two oxygenated methines ( $\delta$ C-5<sub>Ahp</sub> 73.7,  $\delta$ H-5<sub>Ahp</sub> 5.08;  $\delta$ C-3<sub>Thr</sub> 71.7,  $\delta$ H-3<sub>Thr</sub> 5.53), and two vinylic methines ( $\delta$ C-3<sub>Abu</sub> 131.5,  $\delta$ H-3<sub>Abu</sub> 6.52;  $\delta$ C-4<sub>Cmb</sub> 114.3,  $\delta$ H-4<sub>Cmb</sub> 6.10).

A detailed analysis of the 2D NMR data (HSQC, HMBC, and DQF-COSY) established the presence of valine, *N*-methyltyrosine, phenylalanine, threonine, isoleucine, and alanine together with an Ahp, an Abu, and a 4-chloro-3-methylbut-3-enoic acid (Cmb) residue. The Ahp residue was identified by COSY correlations following the sequential spin system starting at the  $\alpha$ -proton at  $\delta$  3.79 (H-2<sub>Ahp</sub>) followed by two methylene protons at  $\delta$  2.42/1.58 (H-3<sub>Ahp</sub>), another set of methylene protons at  $\delta$  1.72/1.58 (H-4<sub>Ahp</sub>), and an oxygenated methine at  $\delta$  5.08 (H-5<sub>Ahp</sub>). The ring closure was based on HMBC correlations from the  $\alpha$ -proton at  $\delta$  3.79 (H-2<sub>Ahp</sub>), the methylene protons at  $\delta$  2.42 and 1.58 (2H-3<sub>Ahp</sub>), and the oxygenated methine proton at  $\delta$  5.08 (H-5<sub>Ahp</sub>) to the carbonyl carbon at  $\delta$  168.5 (C-1<sub>Ahp</sub>).

The Abu structure was based on a COSY correlation between the methyl protons at  $\delta$  1.48 (Me-4<sub>Abu</sub>) and a vinylic methine proton at  $\delta$  6.52 (H-3<sub>Abu</sub>) as well as HMBC correlations from both proton signals to carbon resonances at  $\delta$  129.8 (C-2<sub>Abu</sub>) and 162.6 (C-1<sub>Abu</sub>). The structure of the remaining C<sub>5</sub>H<sub>6</sub>ClO residue was deduced as follows. Long-range COSY correlations between the vinylic methine proton at  $\delta$  6.10 (H-4<sub>Cmb</sub>), the methyl protons at  $\delta$  1.72 (Me-5<sub>Cmb</sub>), and the methylene protons at  $\delta$  2.97 (2H-2<sub>Cmb</sub>) gave an unclear picture of the residue structure. Additionally, HMBC correlations from all three of these proton signals to a quaternary carbon at  $\delta$  133.9 (C-3<sub>Cmb</sub>) and a carbonyl signal at  $\delta$  168.4 (C-1<sub>Cmb</sub>) left the position of the double bond unresolved as four-bond ( $^4J_{\text{CH}}$ ) interactions are occasionally observed in the proximity of double bonds. To differentiate between a 4-chloro-3-methylbut-3-enoic acid and a 4-chloro-3-methyl-2-enoic acid residue, a 1,1-ADEQUATE spectrum with nonuniform sampling (NUS) was recorded which showed a two bond correlation from the methylene protons at  $\delta$  2.97 (2H-2<sub>Cmb</sub>) to the carbonyl signal at  $\delta$  168.4 (C-1<sub>Cmb</sub>), thereby confirming this as a 4-chloro-3-methylbut-3-enoic acid residue. Finally, an NOE correlation between the associated amide proton at  $\delta$  8.21 (NH<sub>Ala</sub>) and the methylene protons at  $\delta$  2.97 (2H-2<sub>Cmb</sub>) helped to confirm the identity and location of the residue. The sequence of residues was established as Val-N-MeTyr-Phe-Ahp-Abu-Thr-Ile-Ala-Cmb through HMBC correlations between consecutive  $\alpha$ -protons to carbonyl carbons of adjacent residues together with NOE correlations between  $\alpha$ -protons and amide protons (Figure 2). The ring closure



**Figure 2.** Selected NOE correlations for tutuilamide A (**1**). Inset shows selected correlations observed for the Ahp residue.

between the C-terminal valine and the  $\beta$ -hydroxy group of threonine was established based on NOE correlations from the valine  $\alpha$ -proton at  $\delta$  4.65 (H-2<sub>Val</sub>) to the threonine methine and methyl protons at  $\delta$  5.53 (H-3<sub>Thr</sub>) and  $\delta$  1.21 (Me-3<sub>Thr</sub>), respectively.

The second and considerably less abundant new compound isolated from this extract, tutuilamide B (**2**), showed a HR-ESI-MS molecular ion peak at  $m/z$  1029.4445 [ $M + \text{Na}$ ]<sup>+</sup> (calcd for C<sub>50</sub>H<sub>67</sub>ClN<sub>8</sub>O<sub>12</sub>Na, 1029.4459,  $\Delta = 1.36$  ppm), thus having one less methylene unit than **1**. Comparison of NMR features for **2** (Table S2) showed it to be highly similar in all regards to **1**, with the primary difference being that the triplet methyl from the Ile residue at  $\delta$  0.80 (Me-5<sub>Ile</sub>) in **1** was replaced by an additional doublet methyl at  $\delta$  0.82 (Me-5<sub>Val2</sub>) in **2**, thus revealing **2** to be the valine analog of **1**.

The absolute configurations of the amino acid residues in both natural products were determined by LC-MS analysis of the 1-fluoro-2,4-dinitro-phenyl-5-D-alanineamide (D-FDAA) derivatives derived from the acid hydrolyzate of **1** and **2** (Marfey's method, Table S3).<sup>17</sup> This revealed that all of the amino acid residues in both compounds were of *L*-configuration, as is the case for all other cyanopeptolin-like peptides. In addition, PDC oxidation followed by acid hydrolysis liberated *L*-glutamic acid from the Ahp residue, as determined by Marfey's analysis, therefore establishing the configuration of C-2<sub>Ahp</sub> as *S*. To determine the relative stereochemistry between C-2<sub>Ahp</sub> and C-5<sub>Ahp</sub> in **1**, a combination of proton–proton coupling constants and NOE correlations was used (Figure 2). The relative conformation of the Ahp ring was determined based on the H-2<sub>Ahp</sub> to H-3<sub>Ahp</sub> coupling constant; a large value was observed ( $^3J_{\text{H-2, H-3a}} = 12.3$  Hz), indicating that these two deshielded protons were axially oriented.<sup>18</sup> NOE correlations between H-2<sub>Ahp</sub>, H-4<sub>Ahp</sub>, and H-5<sub>Ahp</sub> as well as between H-3<sub>Ahp</sub>, OH<sub>Ahp</sub>, and NH<sub>Ahp</sub> revealed the relative configuration of the two stereocenters and therefore established the configuration of C-5<sub>Ahp</sub> as *R*. The highly comparable NMR data sets for this Ahp residue for compounds **1–3** (see Supporting Information, Tables 1 and 2) strongly suggest that they are all of the same relative configuration, and likely of the same absolute configuration.

C

<https://dx.doi.org/10.1021/acschembio.9b00992>  
ACS Chem. Biol. XXXX, XXX, XXX–XXX

Table 2. Potency of Compounds for Select Serine Proteases Following 30 min Incubations

name (analog number)	IC <sub>50</sub> values (nM) with 95% confidence interval values			
	elastase	chymotrypsin	trypsin	proteinase K
tutuillamide A (1)	1.18 (0.33–1.97)	1014 (853.0–1227)	>20 000	103.7 (84.6–129.2)
tutuillamide B (2)	2.05 (1.68–2.46)	576.6 (447.0–767.3)	>20 000	87.6 (65.2–122.2)
tutuillamide C (3)	4.93 (2.72–7.48)	542.0 (445.0–671.2)	>20 000	>5000
lyngbyastatin 7 (4)	11.50 (9.25–14.28)	not done	not done	1574 (806.0 to >5000)
symplostatin 2	5.41 (4.30–6.69)	not done	not done	216.5 (155.2–326.0)
tutuillamide A-linear (5)	3994 (2363–11 735)	not done	not done	>5000
tutuillamide A-methyl (6)	1.83 (1.56–2.12)	not done	not done	98.6 (94.4–100.1)

The geometry of the Abu olefinic bond was determined as *Z* based on an NOE correlation in DMSO-*d*<sub>6</sub> between the methyl protons at  $\delta$  1.48 (Me-4<sub>Abu</sub>) and the amide proton at  $\delta$  9.23 (NH<sub>Abu</sub>). Finally, a 1D NOE experiment with selective irradiation of the vinylic proton at  $\delta$  6.07 (in methanol-*d*<sub>4</sub> to avoid overlapping signals with OH<sub>Ahp</sub>) established the *E* geometry of the Cmb residue due to a correlation to the methylene protons at  $\delta$  3.03 (2H-2<sub>Cmb</sub>).

Subsequently, a third tutuillamide was isolated from a *Coleofasciculus* sp. extract (3). The HR-ESI-MS of tutuillamide C showed a precursor ion at *m/z* 972.4252 [M + Na]<sup>+</sup> (calcd for C<sub>48</sub>H<sub>64</sub>ClN<sub>7</sub>O<sub>11</sub>Na, 972.4255,  $\Delta$  = 0.31 ppm), and the NMR data were highly similar to that of 2, except for the lack of the alanine moiety resonances and possessing an additional methylene in the side chain (Table S2). HMBC, TOCSY, and ROESY correlations corroborated the same connectivities observed between the amino acid residues as for 1 and 2. Because the NMR data of 2 and 3 were highly similar, we propose that they are of the same relative configuration at comparable stereocenters. This hypothesis is additionally supported by the observation that all Abu and Ahp-containing peptides reported to date are composed mainly of L-amino acids.<sup>7–14</sup>

**Inhibition of Serine Proteases with Tutuillamide Analogs.** Studies have shown that Ahp-Abu-containing cyclic hexadepsipeptides inhibit serine proteases such as elastase by binding to the active site in a substrate-like manner. Therefore, 5  $\mu$ M to 85  $\mu$ M of 1, 2, and 3 were incubated with porcine pancreatic elastase, and potency was directly compared to lyngbyastatin 7 (4) and symplostatin 2. Compounds 1 and 2 were found to be the most potent in this group, inhibiting elastase with IC<sub>50</sub>s of 1–2 nM (Table 2, Figures S2 and S3). Compound 3 and symplostatin 2 had IC<sub>50</sub> values of  $\sim$ 5 nM, while 4 was the weakest inhibitor with an IC<sub>50</sub> of 11.5 nM.

Elastase is a member of the S1 family of serine proteases that forms a double  $\beta$ -barrel structure. Trypsin and chymotrypsin are also members of the S1 family, and therefore we determined if tutuillamide A, B, and C can also inhibit these structurally related enzymes. When bovine trypsin was incubated with up to 20  $\mu$ M of these compounds, no inhibition was detected. This enzyme has a strong preference for binding to peptide inhibitors containing basic amino acids such as arginine and lysine, both of which are absent from these natural products. When incubated with chymotrypsin the IC<sub>50</sub> values ranged from 542 nM to 1014 nM. Although the potency of these compounds is considerably weaker for chymotrypsin than for elastase, the relative potency of the three tutuillamide analogs differs between enzymes. Chymotrypsin appears to have a preference for valine as found in 2 and 3 over isoleucine found in compound 1, resulting in a 2-fold reduction in potency. This inhibition profile is consistent

with several other Ahp-Abu-containing cyclic hexadepsipeptides that preferentially inhibit elastase over chymotrypsin and trypsin.<sup>19</sup>

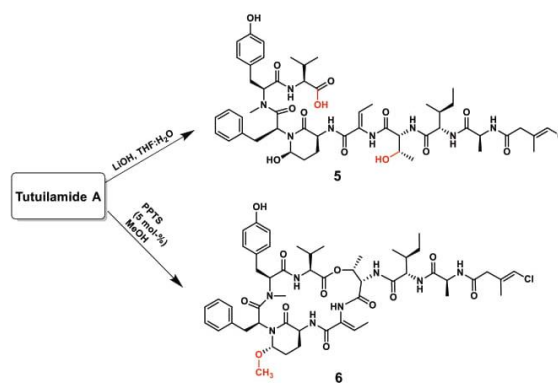
Outside of the S1 family, commonly studied serine proteases are from the S8 family. These enzymes have no sequence or structural homology with S1 family serine proteases and include the bacterial and fungal subtilisin enzymes commonly used in laundry detergents and the mammalian kexin family of enzymes.<sup>20</sup> When 1 and 2 were incubated with proteinase K, a representative S8 protease isolated from the soil fungus *Engyodontium album*, we observed IC<sub>50</sub> values of 103.7 and 87.6 nM, respectively.

However, 3 had more than a 50-fold weaker inhibition for this enzyme (IC<sub>50</sub> > 5  $\mu$ M), suggesting that longer peptides have higher potency for this enzyme. When comparing the potency of symplostatin 2, an 8-mer cyclic hexadepsipeptide, to the 7-mer peptide 4, the IC<sub>50</sub> value decreased 7.3-fold for the shorter peptide inhibitor. These data show that changes in peptide length and amino acid composition of these cyclic hexadepsipeptides result in differences in potency and selectivity for serine proteases.

In addition, we determined the cytotoxic activity of compounds 1–3 toward the H-460 human lung cancer cell line. The compounds show moderate activity with IC<sub>50</sub> values of 0.53  $\pm$  0.04  $\mu$ M, 1.27  $\pm$  0.21  $\mu$ M, and 4.78  $\pm$  0.45  $\mu$ M, respectively (Figure S1).

**Synthesis of Semisynthetic Analogs of 1.** A limited structure–activity relationship study on tutuillamide A was performed to explore the structural features related to the elastase and proteinase K inhibitory activity (Scheme 1). In the first semisynthetic analog of 1, the depsipeptide ester was hydrolyzed using LiOH/H<sub>2</sub>O to obtain the acyclic analog 5.

#### Scheme 1. Synthesis of Semisynthetic Analogs 5 and 6 of Tutuillamide A (1)



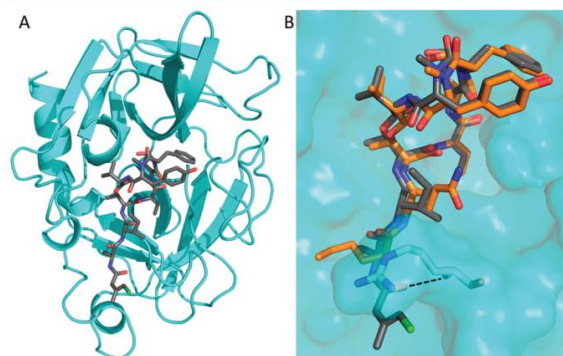
D

<https://dx.doi.org/10.1021/acschembio.9b00992>  
ACS Chem. Biol. XXXX, XXX, XXX–XXX

This analog failed to inhibit proteinase K activity at a concentration of 5  $\mu\text{M}$ , while potency for elastase was decreased 2600-fold when compared to the cyclic tutuilamide A (**1**). These data demonstrate the importance of the cyclic structure to inhibit these serine proteases.

In addition, the hemiaminal moiety in **1** was methylated with methanol in the presence of pyridinium *p*-toluenesulfonate (PPTS) to generate analog **6**. This modification is predicted to improve the metabolic stability of the Abu moiety without affecting the potency of the compound; indeed, analog **6** showed excellent single digit nanomolar potency to elastase (Table 2).

**Rationalizing the Binding of Tutuilamide A to Elastase.** We next determined the crystal structure of porcine pancreatic elastase in complex with **1** in order to compare the binding mode of the depsipeptide with related molecules. The complex crystallized in space group  $P2_12_12_1$ , and data were collected to 2.2 Å. The structure was determined by molecular replacement using the published elastase apo structure (PDB ID: 1LVY) as a search model (Figure 3A). Full details of the



**Figure 3.** (A) Cartoon representation of PPE (cyan) in complex with **1** (shown as gray sticks). (B) Comparison of **1** (gray) and **4** (PDB: 4GVU; orange) binding to PPE (cyan surface). The side chains of PPE residues Ser216 and Arg217 are shown as sticks. The additional hydrogen bond found in the PPE–**1** complex structure is indicated by a dashed line.

data collection and refinement statistics can be found in Table S4. Compound **1** was found to occupy the same binding pocket as lynbyastatin **7** (**4**), explaining the observed identical mode of action of tutuilamide A compared to other Abu-bearing cyclic depsipeptides, with the Abu moiety and the N-terminal residues occupying the S1–S4 pockets (Figure 3B).<sup>13,15</sup> However, compared to **4**, the carbonyl group of the flexible Cmb moiety of **1** forms an additional hydrogen bond with the backbone amide of Arg217 (Figure 3B and Figure S4). This additional interaction could explain the slightly more potent inhibitory activity of tutuilamides compared to lynbyastatin **7** in that this interaction would further stabilize the binding of tutuilamide A to the elastase binding pocket.

The potency of semisynthetic analogs **5** and **6** was evaluated in the context of our cocrystal structure. As shown for **4**, the cyclic core of **1** also forms a network of direct and water mediated inter- and intramolecular hydrogen bonds (Figure S5). Therefore, one would expect any changes to the cyclic nature of these compounds to have a negative impact on their potency. This is indeed the case as shown by the observed nearly 3400-fold decrease in potency of acyclic analog **5**. The

hemiaminal moiety of **1** is facing a hydrophobic pocket formed by Leu 63 and Phe 65 (Figure S5). This pocket is large enough to accommodate methylated analog **6**, without resulting in a steric clash with the binding pocket (Figure S5). Given the size of the pocket, even larger hydrophobic functional groups at this position may be accommodated. This could result in an enhancement of hydrophobic interactions, thereby further stabilizing the compound in the active site. However, further optimization of this group must be pursued cautiously, as the oxygen atom of the hydroxy or methoxy group is involved in intramolecular hydrogen bonding with the neighboring residues and thus facilitates the formation of a *cis* peptide bond (Figure S5). In turn, this contributes to the rigidity of the cyclic core, making these compounds less sensitive to protease activity.<sup>13,15,21,22</sup>

## CONCLUSION

Tutuilamides A–C are new representatives of the family of Ahp-cyclodepsipeptides (cyanopeptolin-like peptides) isolated from cyanobacteria. They share a cyclic peptide core, including the distinctive 2-amino-2-butenic acid (Abu) residue, with a group of cyclic depsipeptides mainly isolated from marine cyanobacteria; however, they also contain an unprecedented 4-chloro-but-3-enoic acid moiety. The predicted biosynthesis of these new cyclic lipopeptides suggests initiation of the pathway by a polyketide synthase which produces an intermediate  $\beta$ -ketobutyrate moiety. This is expected to undergo transformation by a  $\beta$ -branch cassette of genes that also introduces a chlorine atom via radical chemistry, as shown in the case of jamaicamide A biosynthesis.<sup>16</sup>

The new compounds are potent inhibitors of porcine pancreatic elastase and fungal proteinase K. Direct comparison to the potent elastase inhibitor lynbyastatin **7**<sup>23</sup> revealed that the new compounds have 2- to 4-fold increased potency. Structural analysis of tutuilamide A in complex with the porcine pancreatic elastase confirms the same binding mechanism of lynbyastatin **7** with an additional strong hydrogen bond (2.8 Å) between the Cmb carbonyl group and the backbone amide group of elastase residue R226; this additional hydrogen bond appears to stabilize the ligand in the binding pocket and may explain the increased potency of tutuilamides A over lynbyastatin **7**.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acscchembio.9b00992>.

General experimental procedures; extraction and isolation; pyridinium dichromate (PDC) oxidation; advanced Marfey's method; protease activity assays; cytotoxicity to cancer cells; semisynthesis of tutuilamide analog **5**; semisynthesis of tutuilamide analog **6**; crystallization and structure determination; tabulated <sup>1</sup>H and <sup>13</sup>C NMR data for compound **1** in methanol-*d*<sub>4</sub>; <sup>1</sup>H NMR, <sup>13</sup>C NMR, <sup>1</sup>H–<sup>1</sup>H COSY, <sup>1</sup>H–<sup>1</sup>H NOESY, <sup>1</sup>H–<sup>13</sup>C HSQC, <sup>1</sup>H–<sup>13</sup>C HMBC, and 1,1-ADEQUATE spectra in DMSO-*d*<sub>6</sub> as well as 1D-NOE in methanol-*d*<sub>4</sub> for compound **1**; <sup>1</sup>H NMR, <sup>1</sup>H–<sup>1</sup>H COSY, <sup>1</sup>H–<sup>13</sup>C HSQC, and <sup>1</sup>H–<sup>13</sup>C HMBC spectra in DMSO-*d*<sub>6</sub> for compounds **2** and **3**; dose response curves for elastase and H-460 cancer cell assays (PDF)

E

<https://dx.doi.org/10.1021/acscchembio.9b00992>  
ACS Chem. Biol. XXXX, XXX, XXX–XXX

**Accession Codes**

Coordinates were deposited in the Protein Databank with accession number 6TH7.

**AUTHOR INFORMATION****Corresponding Author**

**William H. Gerwick** – Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography and Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, California 92093, United States; [orcid.org/0000-0003-1403-4458](https://orcid.org/0000-0003-1403-4458); Phone: (858)-534-0576; Email: [wgerwick@ucsd.edu](mailto:wgerwick@ucsd.edu)

**Authors**

**Lena Keller** – Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093, United States

**Kirley Marques Canuto** – Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093, United States; Embrapa Agroindústria Tropical, Fortaleza 60511-110, Brazil

**Chenxi Liu** – Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, California 92093, United States

**Brian M. Suzuki** – Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, California 92093, United States

**Jehad Almaliti** – Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093, United States; Department of Pharmaceutical Sciences, Faculty of Pharmacy, University of Jordan, Amman 11942, Jordan

**Asfandiyar Sikandar** – Workgroup Structural Biology of Biosynthetic Enzymes, Helmholtz Institute for Pharmaceutical Research Saarland, Helmholtz Centre for Infection Research, Saarland University, Saarbrücken 66123, Germany

**C. Benjamin Naman** – Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093, United States; Li Dak Sum Yip Yio Chin Kenneth Li Marine Biopharmaceutical Research Center, Department of Marine Pharmacy, College of Food and Pharmaceutical Sciences, Ningbo University, Ningbo 315800, People's Republic of China

**Evgenia Glukhov** – Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093, United States

**Danmeng Luo** – Department of Medicinal Chemistry and Center for Natural Products, Drug Discovery and Development (CNP3), University of Florida, Gainesville, Florida 32610, United States

**Brendan M. Duggan** – Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, California 92093, United States

**Hendrik Luesch** – Department of Medicinal Chemistry and Center for Natural Products, Drug Discovery and Development (CNP3), University of Florida, Gainesville, Florida 32610, United States; [orcid.org/0000-0002-4091-7492](https://orcid.org/0000-0002-4091-7492)

**Jesko Koehnke** – Workgroup Structural Biology of Biosynthetic Enzymes, Helmholtz Institute for Pharmaceutical Research Saarland, Helmholtz Centre for Infection Research, Saarland University, Saarbrücken 66123, Germany

**Anthony J. O'Donoghue** – Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, California 92093, United States; [orcid.org/0000-0001-5695-0409](https://orcid.org/0000-0001-5695-0409)

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acscchembio.9b00992>

**Author Contributions**

All authors have given approval to the final version of the manuscript.

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

We acknowledge the Territory of American Samoa for permitting the collection of this strain of *Schizothrix* sp. from Fagasa Bay, Tutuila in 2014, and the Department of Fish and Wildlife for permits to make collections in Palmyra Atoll. We acknowledge GM107550 and CA100851 for financial support of this project. L.K. would like to thank the German Research Foundation (DFG) for a Research Fellowship (KE 2172/3-1 and KE 2172/4-1). H.L. thanks R01CA172310 for support of the chemical synthesis of lyngbyastatin 7. The authors are grateful to Y. Su for recording the HRMS spectra at the UCSD Molecular Mass Spectrometry Facility. We acknowledge use of ESRF beamline ID30A-3. J.K. thanks the German Research Foundation for an Emmy Noether Fellowship (KO 4116/3-2).

**REFERENCES**

- (1) Weckesser, J., Martin, C., and Jakobi, C. (1996) Cyanopeptolins, Depsipeptides from Cyanobacteria. *Syst. Appl. Microbiol.* 19, 133–138.
- (2) Welker, M., and von Döhren, H. (2006) Cyanobacterial Peptides - Nature's Own Combinatorial Biosynthesis. *FEMS Microbiol. Rev.* 30, 530–563.
- (3) Köcher, S., Resch, S., Kessenbrock, T., Schräpp, L., Ehrmann, M., and Kaiser, M. (2020) From Dolastatin 13 to Cyanopeptolins, Micropeptins, and Lyngbyastatins: The Chemical Biology of Ahp-Cyclodepsipeptides. *Nat. Prod. Rep.*, DOI: 10.1039/C9NP00033J.
- (4) Kang, H.-S., Kronic, A., and Orjala, J. (2012) Stigonemapeptin, an Ahp-Containing Depsipeptide with Elastase Inhibitory Activity from the Bloom-Forming Freshwater Cyanobacterium *Stigonema* Sp. *J. Nat. Prod.* 75, 807–811.
- (5) Pettit, G. R., Kaman, Y., Herald, C. L., Dufresne, C., Cerny, R. L., Herald, D. L., Schmidt, J. M., and Kizu, H. (1989) Antineoplastic Agent. 174. Isolation and Structure of the Cytostatic Depsipeptide Dolastatin 13 from the Sea Hare *Dolabella Auricularia*. *J. Am. Chem. Soc.* 111, 5015–5017.
- (6) Harrigan, G. G., Luesch, H., Yoshida, W. Y., Moore, R. E., Nagle, D. G., and Paul, V. J. (1999) Symplostatins 2: A Dolastatin 13 Analogue from the Marine Cyanobacterium *Symploca Hydnoidea*. *J. Nat. Prod.* 62, 655–658.
- (7) Nogle, L. M., Williamson, R. T., and Gerwick, W. H. (2001) Somamides A and B, Two New Depsipeptide Analogues of Dolastatin 13 from a Fijian Cyanobacterial Assemblage of *Lyngbya Majuscula* and *Schizothrix* Species. *J. Nat. Prod.* 64, 716–719.
- (8) Matthew, S., Ross, C., Rocca, J. R., Paul, V. J., and Luesch, H. (2007) Lyngbyastatin 4, a Dolastatin 13 Analogue with Elastase and Chymotrypsin Inhibitory Activity from the Marine Cyanobacterium *Lyngbya Confervoides*. *J. Nat. Prod.* 70, 124–127.
- (9) Taori, K., Matthew, S., Rocca, J. R., Paul, V. J., and Luesch, H. (2007) Lyngbyastatins 5–7, Potent Elastase Inhibitors from Floridian Marine Cyanobacteria, *Lyngbya* spp. *J. Nat. Prod.* 70, 1593–1600.
- (10) Kwan, J. C., Taori, K., Paul, V. J., and Luesch, H. (2009) Lyngbyastatins 8–10, Elastase Inhibitors with Cyclic Depsipeptide

F

<https://dx.doi.org/10.1021/acscchembio.9b00992>  
ACS Chem. Biol. XXXX, XXX, XXX–XXX

Scaffolds Isolated from the Marine Cyanobacterium *Lyngbya Semiplena*. *Mar. Drugs* 7, 528–538.

(11) Rubio, B. K., Parrish, S. M., Yoshida, W., Schupp, P. J., Schils, T., and Williams, P. G. (2010) Depsipeptides from a Guamanian Marine Cyanobacterium, *Lyngbya Bouillonii*, with Selective Inhibition of Serine Proteases. *Tetrahedron Lett.* 51, 6718–6721.

(12) Gunasekera, S. P., Miller, M. W., Kwan, J. C., Luesch, H., and Paul, V. J. (2010) Molassamide, a Depsipeptide Serine Protease Inhibitor from the Marine Cyanobacterium *Dichothrix Utahensis*. *J. Nat. Prod.* 73, 459–462.

(13) Salvador, L. A., Taori, K., Biggs, J. S., Jakoncic, J., Ostrov, D. A., Paul, V. J., and Luesch, H. (2013) Potent Elastase Inhibitors from Cyanobacteria: Structural Basis and Mechanisms Mediating Cytoprotective and Anti-Inflammatory Effects in Bronchial Epithelial Cells. *J. Med. Chem.* 56, 1276–1290.

(14) Iwasaki, A., Sumimoto, S., Ohno, O., Suda, S., and Suenaga, K. (2014) Kurahamide, a Cyclic Depsipeptide Analog of Dolastatin 13 from a Marine Cyanobacterial Assemblage of *Lyngbya* Sp. *Bull. Chem. Soc. Jpn.* 87, 609–613.

(15) Matern, U., Schleberger, C., Jelakovic, S., Weckesser, J., and Schulz, G. E. (2003) Binding Structure of Elastase Inhibitor Scyptolin A. *Chem. Biol.* 10, 997–1001.

(16) Gu, L., Wang, B., Kulkarni, A., Geders, T. W., Grindberg, R. V., Gerwick, L., Håkansson, K., Wipf, P., Smith, J. L., Gerwick, W. H., and Sherman, D. H. (2009) Metamorphic Enzyme Assembly in Polyketide Diversification. *Nature* 459, 731–735.

(17) Harada, K., Fujii, K., Mayumi, T., Hibino, Y., Suzuki, M., Ikai, Y., and Oka, H. (1995) A Method Using LC/MS for Determination of Absolute Configuration of Constituent Amino Acids in Peptide - Advanced Marfey's Method. *Tetrahedron Lett.* 36, 1515–1518.

(18) Minch, M. J. (1994) Orientational Dependence of Vicinal Proton-Proton NMR Coupling Constants: The Karplus Relationship. *Concepts Magn. Reson.* 6, 41–56.

(19) Liu, L., and Rein, K. S. (2010) New Peptides Isolated from *Lyngbya* Species: A Review. *Mar. Drugs* 8, 1817–1837.

(20) Siezen, R. J., and Leunissen, J. A. (1997) Subtilases: The Superfamily of Subtilisin-like Serine Proteases. *Protein Sci.* 6, 501–523.

(21) Lee, A. Y., Smitka, T. A., Bonjouklian, R., and Clardy, J. (1994) Atomic Structure of the Trypsin-A90720A Complex: A Unified Approach to Structure and Function. *Chem. Biol.* 1, 113–117.

(22) Pelay-Gimeno, M., Tulla-Puche, J., and Albericio, F. (2013) Head-to-Side-Chain" Cyclodepsipeptides of Marine Origin. *Mar. Drugs* 11, 1693–1717.

(23) Luo, D., Chen, Q. Y., and Luesch, H. (2016) Total Synthesis of the Potent Marine-Derived Elastase Inhibitor *Lyngbyastatin 7* and in Vitro Biological Evaluation in Model Systems for Pulmonary Diseases. *J. Org. Chem.* 81, 532–544.