

# Structural biology and *in vitro* biosynthetic studies of Bottromycin and Crocagin – opportunities for exploration and optimization of natural products

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

von  
**Sebastian Adam**

Saarbrücken

2019

Tag des Kolloquiums: 12.12.2019

Dekan: Prof. Dr. Guido Kickelbick

Berichterstatter: Jun. Prof. Dr. Jesko Köhnke

Prof. Dr. Rolf Hartmann

Vorsitz: Prof. Dr. Andriy Luzhetskyy

Akademischer Mitarbeiter: Dr. Daniel Krug



Diese Arbeit entstand unter der Anleitung von Dr. Jesko Köhnke am Helmholtz-Institut für pharmazeutische Forschung Saarland, zugehörig den Naturwissenschaftlich-Technischen Fakultäten der Universität des Saarlandes von April 2015 bis April 2019.

**Den Wissenschaftlern geht es wie den Chaoten.  
Es ist alles da, man muss es nur suchen.**

(Franz Kern)

## Danksagung

Zuallererst gilt ein großer Dank meinem Doktorvater Jun.-Prof. Dr. Jesko Köhnke, der mir im Frühjahr 2015 die große Verantwortung übergab, als erster Doktorand und Mitarbeiter seiner neu gegründeten Arbeitsgruppe an einem neuen Standort das Labor aufzubauen und die Arbeitsabläufe zu etablieren. Durch das entgegengebrachte Vertrauen durfte ich sehr viel Eigeninitiative ergreifen, wodurch ich mich persönlich und beruflich weiterentwickeln konnte. Weiterhin habe ich mich durch die stets exzellente und nahe wissenschaftliche Betreuung immer gut aufgehoben, aber auch stets herausgefordert gefühlt, wodurch ich mich über den gesamten Zeitraum für die wissenschaftlichen Fragestellungen begeistern konnte.

Ich danke meinem Promotionskommittee, bestehend aus Jun.-Prof. Dr. Jesko Köhnke, Prof. Dr. Rolf Hartmann, Dr. Tadeja Lukežič und zuletzt Dr. Chengzhang Fu, für ihre Bereitschaft zur wissenschaftlichen Betreuung über den Zeitraum meiner Promotion und die vielen hilfreichen Ratschläge, die maßgeblich zum Gelingen dieser Arbeit beigetragen haben.

Ich möchte mich bei meinen Freunden und Kollegen bedanken, die mich während meiner Doktorandenzeit in Saarbrücken begleitet und diese positiv geprägt haben. Hierbei gilt ein Dank Asfandiyar Sikandar, Laura Franz und Andreas Klein, die mir als meine Laborkollegen immer hilfreich zur Seite standen und gerne bereit waren in vielen Situationen einzuspringen.

Ich bedanke mich beim Sekretariat, den technischen Assistenten und anderen Mitarbeitern der Arbeitsgruppe Mikrobielle Naturstoffe sowie der IT-Abteilung des HIPS für die stetige Hilfsbereitschaft. Ein besonderer Dank gilt Kris Hippler, Markus Neuber, Joachim Hug, Nikolas Eckert, Nicolas Frank, Jan Schlemmer, Stephanie Schmidt und Rebekka Christmann für die vielen tollen Momente und Erlebnisse der letzten 4 Jahre.

Zuletzt möchte ich mich bei Janina, meiner Familie und allen voran meinen Eltern für die konstante Unterstützung bedanken, ohne euch wäre diese Arbeit nicht möglich gewesen.

## Vorveröffentlichungen aus dieser Dissertation

Teile dieser Dissertation wurden mit Genehmigung der entsprechenden Verlage sowie der naturwissenschaftlich-technischen Fakultäten, vertreten durch den Mentor der Arbeit, in folgenden Journalbeiträgen veröffentlicht oder sind in Vorbereitung zur Veröffentlichung.

### Publikationen

Greg Mann<sup>1</sup>, Liujie Huo<sup>1</sup>, **Sebastian Adam**, Brunello Nardone, Jeremie Vendome, Nicholas James Westwood, Rolf Müller and Jesko Koehnke\*: Structure and substrate recognition of the Bottromycin maturation enzyme BotP. **Chembiochem**. 2016 Dec 2;**17**(23):2286-2292.

**DOI: 10.1002/cbic.201600406**

Laura Franz<sup>1</sup>, **Sebastian Adam**<sup>1</sup>, Javier Santos-Aberturas, Andrew W. Truman and Jesko Koehnke\*: Macroamidine formation in Bottromycins is catalyzed by a divergent YcaO enzyme. **J Am Chem Soc**. 2017 Dec 20;**139**(50):18158-18161.

**DOI: 10.1021/jacs.7b09898**

Konrad Viehrig, Frank Surup, Carsten Volz, Jennifer Herrmann, Antoine Abou Fayad, **Sebastian Adam**, Jesko Koehnke, Dirk Trauner and Rolf Müller\*: Structure and biosynthesis of Crocagins: Polycyclic posttranslationally modified peptides from *Chondromyces crocatus*. **Angew Chem Int Ed Engl**. 2017 Jun 19;**56**(26):7407-7410.

**DOI: 10.1002/anie.201612640**

**Sebastian Adam**, Andreas Klein, Frank Surup and Jesko Koehnke\*: The structure of CgnJ, a domain of unknown function protein from the crocagin gene cluster. **Acta Crystallogr F Struct Biol Commun**. 2019 Mar 1;**75**(Pt 3):205-211.

**DOI: 10.1107/S2053230X19000712**

---

<sup>1</sup> These authors contributed equally to the manuscript

\* Corresponding author

## Publikationen, die nicht Teil dieser Arbeit sind

Louise Kjaerulff, Asfandiyar Sikandar, Nestor Zaburannyi, **Sebastian Adam**, Jennifer Herrmann, Jesko Koehnke and Rolf Müller\*: Thioholgamides: Thioamide-containing cytotoxic RiPP natural products. **ACS Chem Biol.** 2017 Nov 17;**12**(11):2837-2841.

DOI: 10.1021/acscchembio.7b00676

## Tagungsbeiträge

**Sebastian Adam**, Greg Mann, Liujie Huo, Brunello Nardone, Jeremie Vendome, Nicholas James Westwood, Rolf Müller and Jesko Koehnke **(2015)**: Structure and substrate recognition of the Bottromycin maturation enzyme BotP. (Poster) **5<sup>th</sup> Helmholtz-Institut für Pharmazeutische Forschung Saarland (HIPS) Symposium, Saarbrücken, Germany**

**Sebastian Adam**, Greg Mann, Liujie Huo, Brunello Nardone, Jeremie Vendome, Nicholas James Westwood, Rolf Müller and Jesko Koehnke **(2016)**: Structural and functional investigation of the Bottromycin maturation enzyme BotP. (Oral communication) **6<sup>th</sup> Helmholtz-Institut für Pharmazeutische Forschung Saarland (HIPS) Symposium, Saarbrücken, Germany**

**Sebastian Adam**, Konrad Viehrig, Frank Surup, Rolf Müller and Jesko Köhnke **(2016)**: Structural and biochemical investigation of Bottromycin and Crocagin biosynthetic enzymes. (Oral communication) **5<sup>th</sup> Symposium of the interdisciplinary graduate school of natural product research, Saarbrücken, Germany**

**Sebastian Adam**, Laura Franz, Asfandiyar Sikandar and Jesko Koehnke **(2017)**: Investigation of the Bottromycin P450 maturation enzyme: Structure and active site. (Poster) **7<sup>th</sup> Helmholtz-Institut für Pharmazeutische Forschung Saarland (HIPS) Symposium, Saarbrücken, Germany**

**Sebastian Adam**, Laura Franz, Javier Santos-Aberturas, Andrew W. Truman and Jesko Koehnke **(2018)**: Macroamidine formation in Bottromycins is catalyzed by a divergent YcaO enzyme. (Oral communication) **8<sup>th</sup> Helmholtz-Institut für Pharmazeutische Forschung Saarland (HIPS) Symposium, Saarbrücken, Germany**

## Zusammenfassung

Ribosomal synthetisierte und post-translational modifizierte Peptide (RiPPs) sind eine relativ neue Klasse der Naturstoffe, welche für den Menschen interessante biologische Aktivitäten aufweisen. Dadurch stellen sie potentiell spannende Alternativen für die Behandlung der größten Gesundheitsrisiken im Zeitalter der modernen Medizin dar. Obwohl einige dieser interessanten Wirkstoffkandidaten bereits vor mehr als 50 Jahren entdeckt wurden, wurde ihr Potential für die klinische Anwendung oftmals nicht erforscht. Da die Biosynthese der RiPPs komplexe enzymatische Modifikationen des ribosomal produzierten Vorläuferpeptids beinhaltet, ist eine totalsynthetische Herstellung dieser Verbindungen häufig kompliziert und kostspielig. Diese Modifikationen werden im Fall der RiPPs von Enzymen katalysiert, für welche sich die codierenden Gene innerhalb des für Naturstoffe üblichen Genclusters befinden. Im Laufe dieser Arbeit wurden *in vitro* Studien zur Biosynthese zweier interessanter Naturstoffe durchgeführt. Die Charakterisierung der einzelnen Biosyntheseschritte trägt elementar zur Erweiterung des Wissens über die noch zum Großteil wenig verstandene Biosynthese innerhalb des Spektrums der RiPPs bei. Weiterhin stellt sie den ersten Schritt zu einer möglichen enzymatischen Syntheseroute dieser chemisch komplexen Naturstoffe dar, welche eine Derivatisierung beziehungsweise semi-synthetische Herstellung dieser potentiellen Wirkstoffkandidaten deutlich erleichtern würde.

## Abstract

Ribosomally synthesized and post-translationally modified peptides (RiPPs) are a relatively new class of natural products, which can exhibit interesting biological activities. As such, they could represent exciting alternatives for treatment of some of the biggest health concerns of modern medicine, which are multidrug-resistant pathogens and cancer. Even though some of these interesting drug candidates were discovered more than 50 years ago, their clinical potential has often not been explored. RiPPs biosynthesis involves the heavy enzymatic modification of the ribosomally produced precursor peptide. The total synthesis of these compounds is frequently as complex as it is expensive. These modifications are usually catalyzed by enzymes for which the corresponding genes are encoded together with the precursor peptide within the biosynthetic gene cluster of the natural product. During the course of this work, the biosynthetic pathways of two interesting RiPP natural products have been investigated using an *in vitro* approach. The full characterization of the enzyme-substrate relationship of the biosynthetic steps broadens our knowledge of the still poorly understood general properties of RiPPs biosynthesis. Furthermore, this work represents the first step towards a possible enzymatically-driven synthesis of these natural products, which in turn allows for facilitated derivatization as well as a semisynthetic route for these potential clinical drug candidates.

# Table of Contents

<b>Chapter 1.....</b>	<b>12</b>
<b>Introduction.....</b>	<b>12</b>
1.1 Antimicrobial resistance in pathogens – a global challenge .....	12
1.2 Natural products - a source of novelty.....	19
1.3 Ribosomally synthesized and post-translationally modified peptides (RiPPs) – natural products with great potential .....	28
1.4 Bottromycins.....	32
1.5 Crocagins .....	37
1.6 Outline of this work .....	37
1.7 References.....	39
<b>Chapter 2.....</b>	<b>46</b>
<b>Structure and substrate recognition of the Bottromycin maturation enzyme BotP .....</b>	<b>46</b>
Supporting Information .....	63
<b>Chapter 3.....</b>	<b>70</b>
<b>Macroamidine formation in Bottromycins is catalyzed by a divergent YcaO enzyme .....</b>	<b>70</b>
Supporting Information .....	82
<b>Chapter 4.....</b>	<b>128</b>
<b>Structural analysis of IpoC, an unusual YcaO cyclodehydratase from the Bottromycin pathway ...</b>	<b>128</b>
Supporting Information .....	148
<b>Chapter 5.....</b>	<b>159</b>
<b>Characterization of the stereoselective P450 enzyme BotCYP enables the <i>in vitro</i> biosynthesis of the Bottromycin core scaffold .....</b>	<b>159</b>
Supporting Information .....	171
<b>Chapter 6.....</b>	<b>197</b>
<b>Structure and Biosynthesis of Crocagins: Polycyclic Postranslationally Modified Ribosomal Peptides from <i>Chondromyces crocatus</i> .....</b>	<b>197</b>
Supporting Information .....	208
<b>Chapter 7.....</b>	<b>234</b>
<b>The structure of CgnJ, a domain of unknown function protein from the Crocagin gene cluster .....</b>	<b>234</b>
Supporting Information .....	250
<b>Chapter 8.....</b>	<b>255</b>
<b>Initiation of crocagin biosynthesis – implications for a macromolecular biosynthetic complex .....</b>	<b>255</b>



---

<i>Supporting Information</i> .....	281
<b>Chapter 9</b> .....	<b>295</b>
<b>Discussion and Outlook</b> .....	<b>295</b>
<i>9.1 The importance of understanding natural product biosynthesis and its implications for industrial biotechnology</i> .....	295
<i>9.2 The potential of RiPP engineering for the optimization of natural product scaffolds</i> .....	304
<i>9.3 References</i> .....	307
<b>Chapter 10 Appendix</b> .....	<b>309</b>

# Chapter 1

## Introduction

### 1.1 Antimicrobial resistance in pathogens – a global challenge

#### The evolution of antimicrobial resistance

The development of modern medicine in the last 150 years in combination with better nutritional possibilities has increased the human life expectancy at birth by more than 100% since 1875.<sup>[1]</sup> Given the enormous amount of medicinal research and development progress that has been made within this time in the field of antibiotics, infectious diseases have been surpassed by cancer and heart disease as a leading cause of death in the United States already in 1940.<sup>[2]</sup> However, the use of antibiotics in a widespread fashion throughout the industrialized nations has brought humanity a new challenge, as antibiotic resistance started developing in common pathogens. In fact, the World Health Organization has recently named antibiotic resistance a top public health threat,<sup>[3]</sup> with one report on antimicrobial resistance suggesting it will surpass cancer as the leading cause of death by 2050.<sup>[4]</sup>

Given the challenge we are facing, it is crucial that we understand the principles underlying antimicrobial resistance development. Looking at the sources of clinically approved antibiotics, traditionally there have been two major approaches to discover new antibiotic leads – natural products and synthetic compounds.<sup>[5]</sup> Of the nine major antibiotic classes that are commonly used today, six originate from a natural product, while three were developed mainly through synthetic chemistry.<sup>[6]</sup> The majority of antibiotics from natural sources are produced by the order of *Actinomycetes*, which are gram-positive bacteria commonly found in soil.<sup>[7]</sup> An absolute quantification of bacterial cells utilizing modern methodologies that account for non-culturable organisms estimated the bacterial cell count to range between  $1 \times 10^8$  and  $1 \times 10^{10}$  cells/g of dry soil, depending on which method has been applied.<sup>[8]</sup> As they were observing at least nine different phyla of bacteria at two different locations, we can imagine a constant strife for nutrients happening in microbial environments, which in turn can be related to the production of antibacterial agents.

Considering the fact that a large portion of antibacterial producers are bacteria themselves, it appears only natural that they had to develop mechanisms of defense against these substances over the course of evolution. Depending on the compound, these so-called self-

resistance mechanisms can include chemical modification(s) of the compound, expression of efflux pumps and/or target alteration in the producer.<sup>[9]</sup> It is generally believed that microbial ecosystems continuously undergo changes depending on their physical and chemical properties, resulting in either temporary or permanent adaptation.<sup>[10]</sup> Therefore, we can assume the evolutionary pressure resulting from bacterial co-existence as the main driving force of resistance gene development.<sup>[11][12]</sup> The latter arises mainly through random chromosomal point mutations, even of non-producing organisms.<sup>[11][12]</sup> In conclusion, we can consider antimicrobial resistance an ancient phenomenon, which is in fact what has been suggested by a study estimating the age of the biosynthetic genes of certain antibiotics.<sup>[13]</sup> Though, if we are dealing with an ever-present event, under which circumstances is it posing a threat to us today?

There are several crucial differences, which lead to the naturally occurring resistance development in ecosystems and the resistance development in pathogens through human application being virtually non-comparable.<sup>[14]</sup> First, we have to consider the mechanisms whereby bacteria transfer resistance genes within their population, but also to nearby co-existing bacteria. Possibilities of genetic exchange between bacteria include direct uptake of DNA (transformation), phage-mediated transduction, through direct intercellular contact (conjugation) or the mobilization of genomic DNA elements (transposition).<sup>[15]</sup> In a microbial ecosystem, these horizontal gene transfer (HGT) mechanisms can lead to the spread of antibiotic resistance genes located on plasmids, most commonly by means of conjugation.<sup>[16]</sup> However, this effect is somewhat self-regulatory in nature because the maintenance of a mobile genetic element or likewise the stable integration of a DNA element are often correlated with fitness costs for the organism.<sup>[17][18]</sup> Due to the fact that many antibiotics target important cellular processes, resistance to them disrupts the natural processes or imposes large energetic burdens on the cell.<sup>[18]</sup> In combination with low overall selection pressure, a plasmid-located, transferred resistance gene in the environment is likely to be kept in a confined space or even lost after a certain time, thereby controlling the overall distribution of resistance among a population. In fact, a study surveying the incompatibility of plasmids between modern resistance (R) plasmids and plasmids isolated before the medical use of antibiotics from similar strains demonstrated that antimicrobial resistance in the last 50 years was mostly acquired by the insertion of new genes into existing plasmids.<sup>[19]</sup>

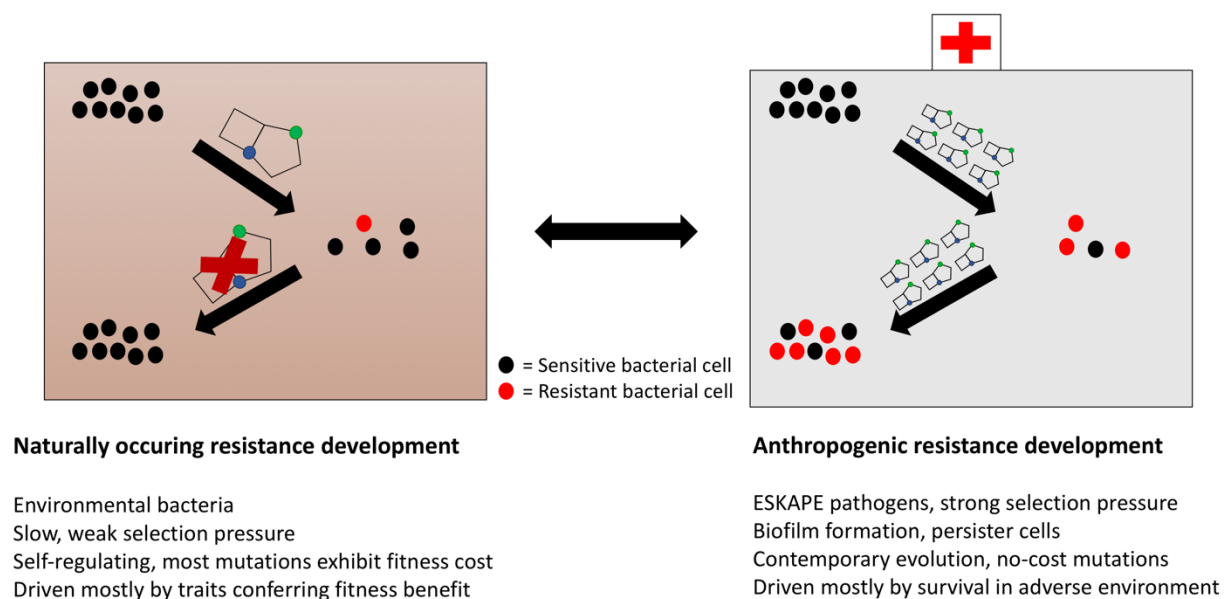
Furthermore, it displayed the rarity of resistance genes on plasmids isolated from today's medically relevant pathogens in the timeframe from 1917-1954.<sup>[19]</sup>

### **The development of anthropogenic resistance development**

In a clinical setting, however, the self-regulatory mechanism of fitness deficit related to most resistance mutations ceases to be effective. This is believed to be a consequence of an enormous selection pressure being applied on clinical isolates due to human activity, e. g. prolonged exposure to a certain antibiotic class over decades.<sup>[11]</sup> One of the reasons for this is a phenomenon called compensatory evolution, a fitness restoring mechanism by which resistant bacteria select for mutations which compensate for the original fitness cost without any significant loss of resistance.<sup>[20]</sup> Additionally, it has been shown that there is a substantial variation in fitness costs concerning mutations related to different species and drugs.<sup>[18]</sup> Consequently, studies have suggested the emergence of no-cost resistances and even resistances with a fitness benefit, depending on the species, the drug and the environmental setting.<sup>[21][22]</sup> A combination of these factors is believed to be responsible for the persistence of antibiotic resistance in clinical bacterial populations as the natural mechanisms of reversibility are suspended.<sup>[23]</sup> In conclusion, the contemporary selection pressure resulting from antibiotic use cannot be compared to the evolution of microbes over several billions of years as selection occurs mostly for survival in adverse environments, counteracting the aforementioned regulation processes.<sup>[14]</sup>

Looking at both settings, we also have to take the general differences in characteristics of the organisms acquiring resistance into consideration. In comparison to environmental soil bacteria, the anthropogenic resistance development triggered increased antimicrobial resistance in human pathogens. These are organisms, which are especially adapted to hostile conditions as they are in constant battle with the immune system. Consequently, an evolutionary arms-race has been proposed to continuously happen as a result of the pathogen-host interaction, causing human pathogenic strains to innately exhibit specific features making them exceptionally dangerous in infection.<sup>[24]</sup> In particular, the so-called ESKAPE pathogens are a panel of microorganisms with the most clinical relevance due to their ability to successfully escape the biocidal action of antibiotics.<sup>[12][25]</sup> As such, their clinical isolates are often resistant to multiple antibiotic classes, leading to them being defined as either multidrug-resistant (MDR), extensively drug-resistant (XDR) or even pandrug-resistant

(PDR) strains.<sup>[26]</sup> Additionally, epidemiological data suggest the incidence of infection with ESKAPE organisms and the associated mortality rate is significantly increased in comparison with susceptible strains.<sup>[27]</sup> Concerns connected to ESKAPE pathogens also arise from their ability to form biofilms, which are complex microbial communities embedded into an extracellular layer of polymeric substances.<sup>[28]</sup> Biofilm communities have been associated with different properties compared to planktonic bacterial organization. These include an increased mutation rate, increased rate of horizontal gene transfer and stationary-phase physiology, resulting in an increase of minimal inhibitory concentration (MIC) of clinical antibiotics by 100-1000 fold.<sup>[29]</sup> Furthermore, their extracellular matrix can function as a shield for persister cells, which are specialized survivor cells able to reinitiate infection after antibiotic therapy has been terminated detected in all ESKAPE pathogens.<sup>[30]</sup> A summary of the different resistance development processes and their specific characteristics is shown as a schematic in **Figure 1**.



**Figure 1:** Schematic of the fundamentally different processes underlying naturally occurring and anthropogenic resistance development as well as their associated traits.

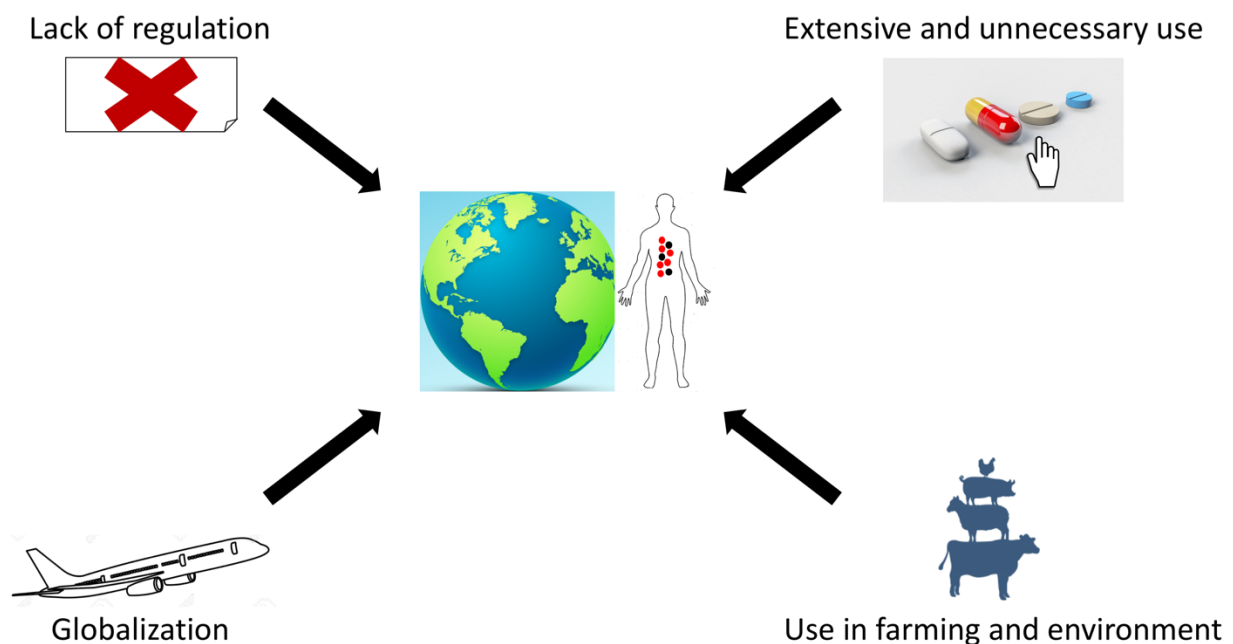
## The influence of different human activities on resistance development

Now that we have discussed the principles and characteristics of anthropogenic resistance development, we can take a look at the various ways in which human activity contributes to the modern situation. Antibiotics represent a large pharmaceutical market with an ever-increasing demand, given that their total consumption has increased by 35% between the years of 2000 and 2010.<sup>[31]</sup> However, therapeutic human use accounts for less than half of all applications of commercially produced antibiotics.<sup>[14]</sup> It should be noted that the BRICS

countries (Brazil, Russia, India, China and South Africa) are responsible for 76% of this increase.<sup>[31]</sup> These countries have shown massive economic growth and therefore rising incomes, which has led to a rising demand for animal protein.<sup>[32]</sup> Consequently, the countries have answered this demand by largely investing into cost-efficient livestock production systems.<sup>[32]</sup> These systems often necessitate the use of antibiotics to maintain productivity and keep the animals healthy, but also for non-therapeutic purposes including growth promotion.<sup>[33]</sup> In combination with a lack of regulation (e.g. no required veterinary prescription), the use of antibiotics in livestock dramatically increases the global consumption and is expected to further rise by 67% until 2030 (from 2010).<sup>[34]</sup> There are several potential routes of transmission of antimicrobial resistance through extensive application in farming, with the consumption of meat being the most apparent one. An indirect route of transmission is emerging from the excretion of animal waste, which is often used as manure for crop fertilization, as it has been shown that 75 to 90% of compound is not metabolized by the animals.<sup>[33]</sup> Therefore, the intact compound (degradation dependent on the antibiotic class) is having a direct impact on environmental ecosystems and can possibly re-enter the human food chain by means of crop consumption or even ground water.<sup>[35]</sup> The subsequent distribution and dilution of compound concentration can result in subinhibitory concentrations in the environment, which may facilitate resistance development.<sup>[36]</sup> This can partly be correlated to a direct link between an activation of the SOS response (a regulatory network involved in bacterial adaptation to stress response) and an upregulation of integron integrases that are crucial for the spread of antibiotic resistance determinants in gram-negative bacteria.<sup>[37]</sup>

Concerning therapeutic human application of antibiotics, we are dealing with the problem of extensive and even unnecessary use, especially but not restricted to countries with a lack of regulation. The overuse can be exemplified by a study calculating the number of antibiotic prescriptions in the US exceeding the number of people in certain states, implying that every person would be theoretically in need of more than one treatment in a given year.<sup>[38]</sup> This implication is contradictory to the observation that life-threatening infectious diseases requiring antibiotic treatment occur mostly in elderly and immunocompromised patients.<sup>[39]</sup> Especially in developing countries, the availability of antibiotic cures without prescription can lead to self-diagnostics and therefore self-medication, which is especially problematic in the context of people not being educated about the potentially dangerous consequences.<sup>[40]</sup>

However, even when antibiotic therapy cannot be obtained without a prescription, inappropriate prescribing of antibiotics in primary care with no significant beneficial or even detrimental results has been reported in different studies (up to 50% in the US and 20% in the UK).<sup>[41][42]</sup> In a clinical setting, we are in need of enhanced diagnostic procedures and a great interdisciplinary effort to optimize antibiotic therapy, while simultaneously improving education on antimicrobial resistance through advanced training for medical professionals and the general public.<sup>[43][44]</sup> Finally, we have to consider the effects of globalization on the epidemiology of antimicrobial resistance. The increased amount of travel leading to the dissemination of evolutionary successful resistance genes among the globe has been exemplified by the global spread of CTX-M extended spectrum  $\beta$ -lactamase (ESBL).<sup>[45]</sup> A schematic highlighting the major causes of anthropogenic resistance development, which can be related to different human activities, is shown in **Figure 2**.



**Figure 2:** Schematic of the main causes of anthropogenic resistance development in bacteria that can be related to various human activities.

Dealing with these various factors driving anthropogenic resistance development will require an increasing global effort in regulation and management of antibiotic use. However, reversibility of occurrence of antimicrobial resistance in pathogens can be expected to be a rather slow process. So, what about the problematic clinical situation at hand? Alternative strategies to combat antimicrobial resistance especially in ESKAPE pathogens include

bacteriophage therapy,<sup>[46]</sup> photodynamic light therapy<sup>[47]</sup> and nanomedicine.<sup>[48]</sup> However, as these technologies are likely to require further research and development efforts concerning both their application specificity and efficacy, a prompter solution demands novel antibiotic classes. Traditionally, compounds derived from a natural origin offered considerable diversity in both their chemical scaffolds as well as their mode of action, making their exploration a promising endeavor to this day.



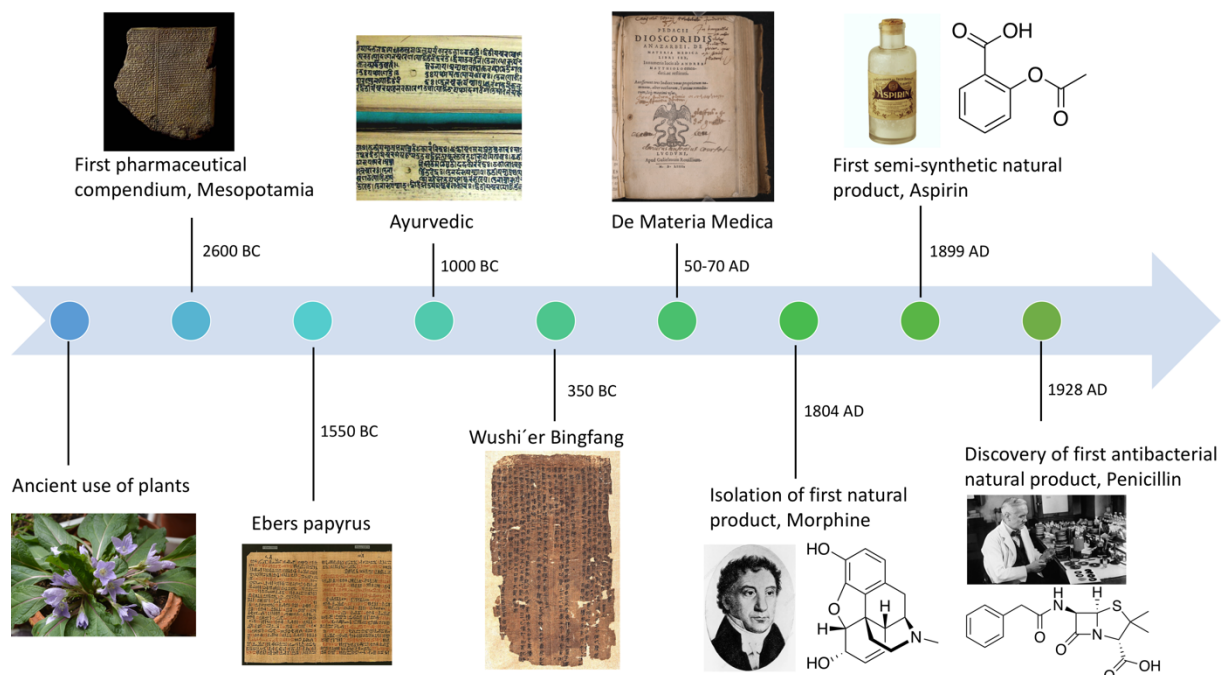
## 1.2 Natural products - a source of novelty

### A brief history of natural products

The potential harbored by compounds of a natural origin has been known for thousands of years, with the first application of leaves for relieving pain by our ancestors preceding any form of what can be considered medicinal use.<sup>[49]</sup> This process can be considered somewhat evolutionary-based as there are many examples of invertebrates, birds and mammals, which have been observed to ingest or apply special plants containing bioactive ingredients in response to a stimulus such as infection.<sup>[50]</sup> Therefore, it is probable that humans have undergone a learning process including trial and error procedures by observing animals over the course of evolution, adapting their behavior and teaching their offspring about the potential use of pharmaceutically active plants. The passing on of this gathered knowledge on medicine over generations has ultimately led to the first records of a pharmaceutical compendium of ~1000 plants containing bioactive substances, which dates back to 2600 BC and was found written on clay tables in Mesopotamia.<sup>[51]</sup> Interestingly, many of these formulations are still in use today.<sup>[51]</sup> Though, arguably the best known ancient pharmaceutical documents on the use of natural products originate from early advanced civilisations like the Egyptians (Ebers Papyrus, >700 formulations of natural extracts, 1550 BC), the Indian (Ayurvedic, >300 formulations of natural extracts, 1000 BC) and Chinese (Wushi'er Bingfang, >240 natural substances mentioned, 350 BC).<sup>[52][53]</sup> Likewise, roman and greek physicians and pharmacists subsequently built on this extensive knowledge and added their own insights and experiences, laying the foundation of modern pharmacology in Europe (e.g. *De Materia Medica* written 50-70 AD by Pedanius Dioscorides).<sup>[49][54]</sup> In 1618, the publication *Pharmacopoeia Londinensis* by the Royal College of Physicians depicted the first standard list of medicines and their ingredients to be used in England.

However, it took until the early 1800's until specific pharmacological actions were related to isolated chemical structures found in the natural extract rather than the extracts themselves. In 1804, the isolation of morphine from *Papaver somniferum* was the first isolated natural product by Friedrich Sertürner, which also became the first pure natural product being sold commercially by Merck in 1826.<sup>[51][55]</sup> Additional milestones in natural product research were the introduction of Aspirin as the first synthetic analogue of a natural product in 1899 as well as the discovery of Penicillin as the first antibacterial substance in 1928, starting the golden

age of antibiotic discovery.<sup>[51]</sup> A brief history of natural product discovery and development is displayed in the timeline of **Figure 3**.



**Figure 3:** Timeline of the milestones of natural product research up to the discovery of penicillin in 1928.

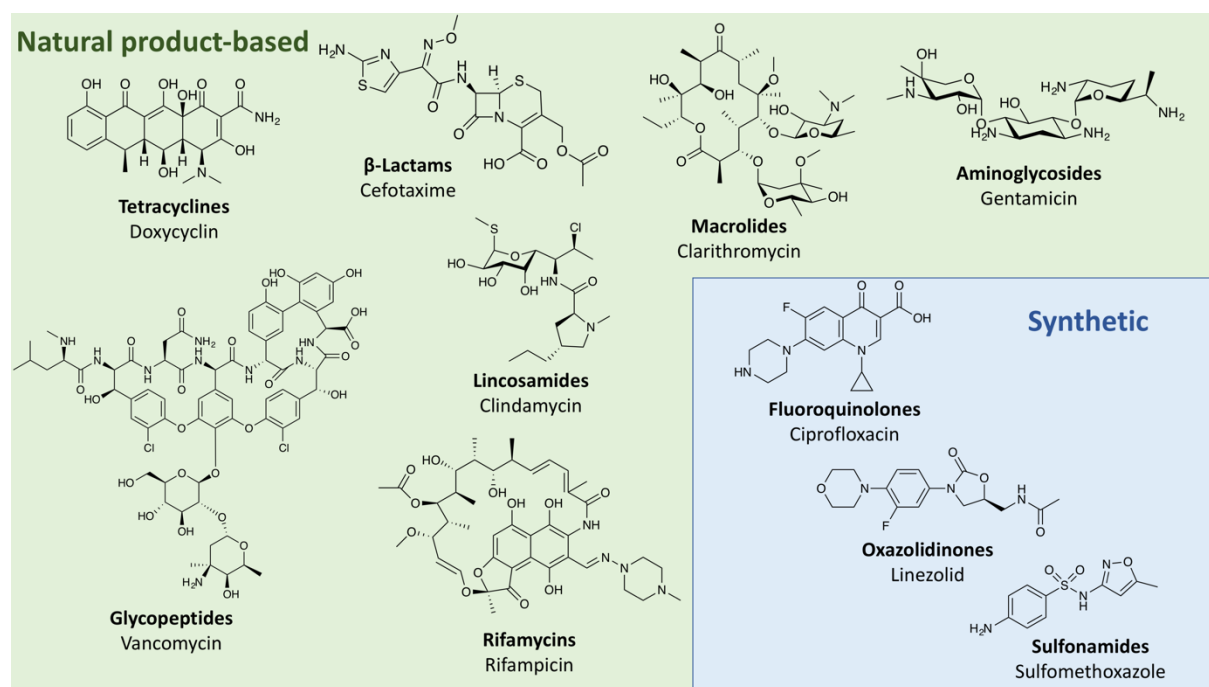
## The rise and fall of microbial natural product research

The discovery of penicillin successfully demonstrated the biosynthetic potential of microorganisms and revolutionized the way natural product research was approached thereafter, also being awarded the nobel prize for physiology and medicine in 1945. As microorganisms are assumed to comprise an immense (largely uncharacterized) biological diversity with e.g. only  $\sim\frac{1}{4}$  of bacterial phyla being isolated in the laboratory as of today, they also represent a widely divergent chemical space to be explored by researchers.<sup>[56][57]</sup> Going hand in hand with the isolation of Streptomycin from *Streptomyces griseus* in 1944, the following period has been named the golden age of antibiotic discovery as it featured a worldwide search for novel microbial natural products.<sup>[58]</sup> Consequently, these efforts also resulted in countless success stories of important discoveries for other therapeutic areas such as the versatile rapamycin from *Streptomyces hygroscopicus*, which is suggested to have various indications due to its target mTOR being involved in different cellular processes.<sup>[59]</sup> In contrast, many pharmaceutical companies have scaled back their investments into natural product research towards the end of the 20<sup>th</sup> century, shifting towards synthetic libraries

containing millions of compounds due to advances in both high-throughput screening (HTS) and combinatorial synthesis.<sup>[60]</sup> In the time from 1981 to 2010, however, only 1 new chemical entity (NCE) was approved using combinatorial chemistry even though it has been used as a discovery source for 70% of this timeframe.<sup>[61]</sup> This is especially critical for the field of anti-infectives as new chemical scaffolds are expected to have a higher chance of lacking cross-resistance and a lower potential of one-step resistance.<sup>[62]</sup> Thus, the methodology has failed to deliver a sufficient amount of new leads for further pharmaceutical development, possibly leading to a resurgence of interest into natural products.<sup>[63]</sup> To understand the reasons of this outcome, we have to consider the main differences between the two approaches.

### **Differences between natural product-based and synthetic scaffolds**

One of the big advantages of natural products in this context is that they have been defined as being privileged compounds in their therapeutic space. Originally introduced as a concept in 1988, it refers to substructures that show high affinity binding to more than one target and selective modification of the privileged structure on just one position can provide opportunities for high selectivity.<sup>[64]</sup> For natural product anti-infectives, it is relating to their specific physicochemical properties, such as efficient bacterial cell penetration and high molecular selectivity, which is believed to be the result of millions of years of natural selection.<sup>[65]</sup> They cover much wider and larger chemical space than combinatorial or synthetic compounds, which is reflected by a large number of chiral centers, ring fusions and a higher density of functional groups.<sup>[66]</sup> An interesting study has directly linked three-dimensional properties granted by carbon saturation and chiral centers to a higher success of drug candidates in early development.<sup>[67]</sup> As such, natural product based privileged substructures can be very useful leads for novel bioactive compounds.<sup>[68]</sup> Therefore, it is not surprising that natural products and inspirations thereof constituted about 50% of all small molecule approved drugs in the timeframe from 1981-2006.<sup>[68]</sup> Concerning the major clinical antibiotic classes of the 20<sup>th</sup> century, we can clearly observe a higher structural diversity in the core scaffolds of natural product-based agents compared to those from a synthetic origin (**Figure 4**).



**Figure 4:** Overview of the main clinical antibiotic classes of the 20<sup>th</sup> century, highlighting their origin as either natural product-based or synthetic. We observe a lot more chemical diversity in the core structures of natural product-based compounds.

In contrast, screening libraries such as corporate compound collections have often been designed according to Lipinski's "rule of five" and are therefore unsuitable for anti-infective research, as known antibacterial agents generally do not follow these rules.<sup>[69]</sup> Additionally, discovery efforts of screening libraries have often been steered towards achiral and aromatic molecules, which can easily be synthesized, but represent a limited chemical space.<sup>[67]</sup> However, potential concepts to overcome these deficits and to combine the best of both worlds have been proposed, such as Diversity-Oriented Synthesis (DOS). Compared to classical Targeted-Oriented Synthesis (TOS) and combinatorial chemistry methodologies, DOS aims at establishing synthetic strategies of introducing substitutional, skeletal and stereochemical diversity into a small molecule compound library.<sup>[70]</sup> To that end, an optimal DOS strategy employs building blocks in a 3- to 5-step synthetic process, which allow for regio- and stereoselective introduction of functional groups in the process of expanding the molecular architecture.<sup>[71]</sup> Intriguingly, a DOS approach can also be applied to privileged natural product scaffolds, which is directed towards identifying novel biological properties different from the original privileged compound.<sup>[70]</sup> A similar concept is depicted by the so-called Biology-Oriented Synthesis (BIOS), which is utilizing a structural classification of natural products (SCONP) as guiding scaffolds for the design and synthesis of compound collections.<sup>[72]</sup> The

application of these novel technologies might increase the hit rate and success of synthetic libraries in terms of discovering new drug leads, especially for anti-infectives. In addition to the development of novel antibiotics, the further exploration of compounds that were dismissed during the golden age of antibiotic discovery due to various reasons such as toxicity, efficacy, stability *etc.* should also be considered. As such, only 0.7% of the ~28,000 natural product antibiotics discovered over the past decades are in clinical use today.<sup>[65]</sup> A reinvestigation of Daptomycin, which was once discarded due to its narrow spectrum and toxicity issues, led to one of the few newly approved antibiotics of the 21<sup>st</sup> century highly valuable for use in multidrug-resistant infections today.<sup>[65]</sup> The advancements in genomic technologies like heterologous expression and strain development should also help overcome issues of compounds with low yields in the original producer.

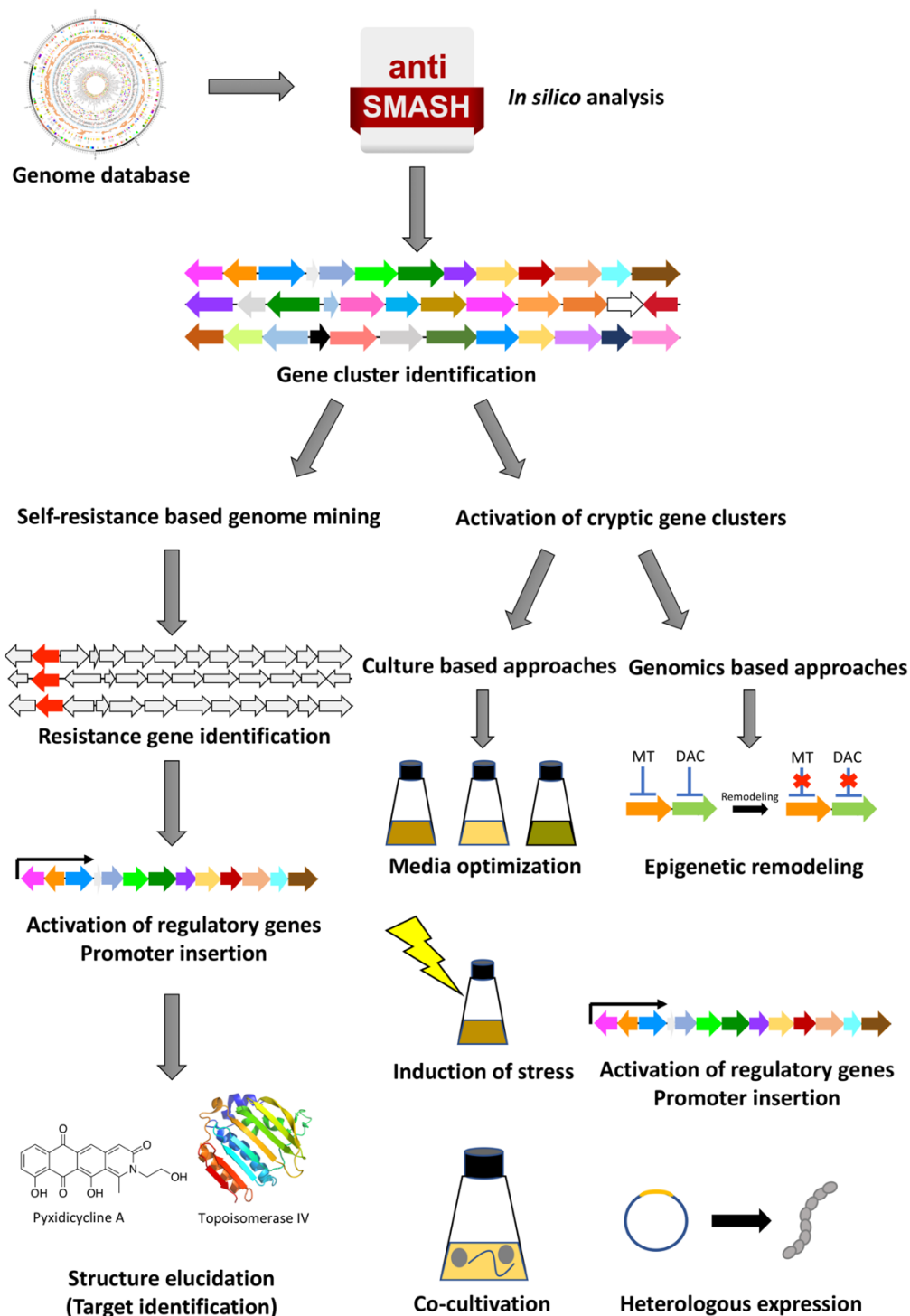
### **Modern natural product research – the impact of the genomic age**

The beginning of the genomic era marked a new milestone in natural product research, as the broadened knowledge about the biosynthetic origin of natural products enabled a completely new methodic scope of research and development. The development of Next-Generation Sequencing technologies has led to a dramatic increase in sequenced bacterial genomes (by 100-fold since 2005), with the total number deposited in the NCBI exceeding 30,000.<sup>[73]</sup> Bacteria devote > 7.5% of their genome to secondary metabolite biosynthesis, which was shown by a global analysis of prokaryotic genomes.<sup>[74]</sup> The genetic information responsible for biosynthesis of a specific secondary metabolite is typically clustered within a certain part of the genome and is therefore referred to as a biosynthetic gene cluster (BGC). To date, the mechanisms driving gene clustering remain unclear. Though, one prevailing hypothesis is that gene clustering enables tight regulatory control of the biosynthetic enzymes and therefore tends to be evolutionary selected for, as the simultaneous expression of these genes allows for their direct interaction.<sup>[75]</sup> Historically, natural product drug discovery in the golden age was constrained to the conventional top-down approach, which employed the biological screening of samples, followed by compound isolation and characterization.<sup>[76]</sup> However, this process has several limitations, which include a restriction to isolatable and culturable organisms and the fraction of secondary metabolites that are expressed in a given laboratory environment.<sup>[76]</sup> As such, the shutdown of a lot of natural product-related departments in the

pharmaceutical industry at the end of the 20<sup>th</sup> century can at least partly be related to high rediscovery rates.<sup>[76]</sup>

New discovery strategies envisioned the use of genomic data for the exploration of hidden or overlooked secondary metabolites, consequently called the bottom-up approaches.<sup>[77]</sup> However, the amount of data necessitated bioinformatic strategies to connect biosynthetic gene clusters to the vast amount of already known molecules and predict novel BGCs that encode for the most promising compounds.<sup>[78]</sup> This led to the development of one of today's invaluable bioinformatic tools for *in silico* natural product gene cluster analysis, the antibiotics & Secondary Metabolite Analysis Shell (antiSMASH), first released in 2011.<sup>[79]</sup> An early software-guided approach was the so-called classical genome mining, which makes use of the existing similarity of the underlying biosynthetic principles of secondary metabolites, which in turn allows for the detection of putative biosynthetic enzymes based on sequence homology. Examples of conserved biosynthetic machinery can be found throughout the whole spectrum of secondary metabolite classes such as polyketide synthases (PKS), non-ribosomal peptide synthetases (NRPS), ribosomally synthesized and post-translationally modified peptides (RiPPs), aminoglycosides and many more.<sup>[78]</sup> Nowadays, genome mining technologies are widely used in natural product research and have undergone advancements based on various biological principles. One of the most important concepts for antibacterial drug discovery is arguably the idea of resistance-based genome mining, which relies on the fact that many antibiotic-encoding BGCs will also encode one or more genes that confer self-resistance to the product.<sup>[80]</sup> As self-resistance genes often show high similarity to the resistance genes acquired as protection mechanisms by other bacteria, the presence of a resistance gene within a BGC can indicate its biosynthetic function as the production of an antibacterial substance.<sup>[80]</sup> If the resistance is conferred through point mutation of a housekeeping gene, the technology can even be extended to target identification, which has been successfully demonstrated in the case of Telomycin.<sup>[81]</sup> Furthermore, an analysis of microbial genomes quickly revealed that the number of annotated BGCs largely exceeds the number of detected secondary metabolites in a standard laboratory setting. Therefore, another important approach revolves around the exploration of the biosynthetic space which is not accessible by the conventional top-down approach, the so called silent gene loci usually referred to as cryptic or orphan pathways.<sup>[82]</sup> Typical methodologies for the systematic investigation of the hidden microbial secondary metabolome include rather classic optimization of cultivation

conditions (e.g. media, induction of stress and co-cultivation triggering a natural response) and novel genomics-based approaches (epigenetic remodeling, activation of regulatory genes in the native host, heterologous expression of the BGC).<sup>[82]</sup> An overview of the modern bottom-up approaches intended to extend the accessible microbial biosynthetic space is exemplified with some examples in **Figure 5**.



**Figure 5:** Overview of a typical bottom-up workflow for the discovery of novel natural products. Pixidicycline was discovered with a similar approach<sup>[83]</sup>, the protein structure was taken from PDB ID 3FV5.

## Mechanisms of microbial secondary metabolite biosynthesis

Concerning their biosynthesis, microbial secondary metabolites can be assigned to a specific class based on certain characteristics, with the most prominent natural product classes being PKS, NRPS, terpenes, RiPPs and hybrids thereof. Consisting of mostly macromolecular enzyme complexes with a defined modular architecture, the so-called assembly line, biosynthetic enzymes of the first 3 classes usually contain conserved sequences facilitating their detection:

1. The general biosynthetic paradigm of polyketide synthases is based on the incorporation of one building block onto the growing polyketide chain by each module.<sup>[84]</sup>

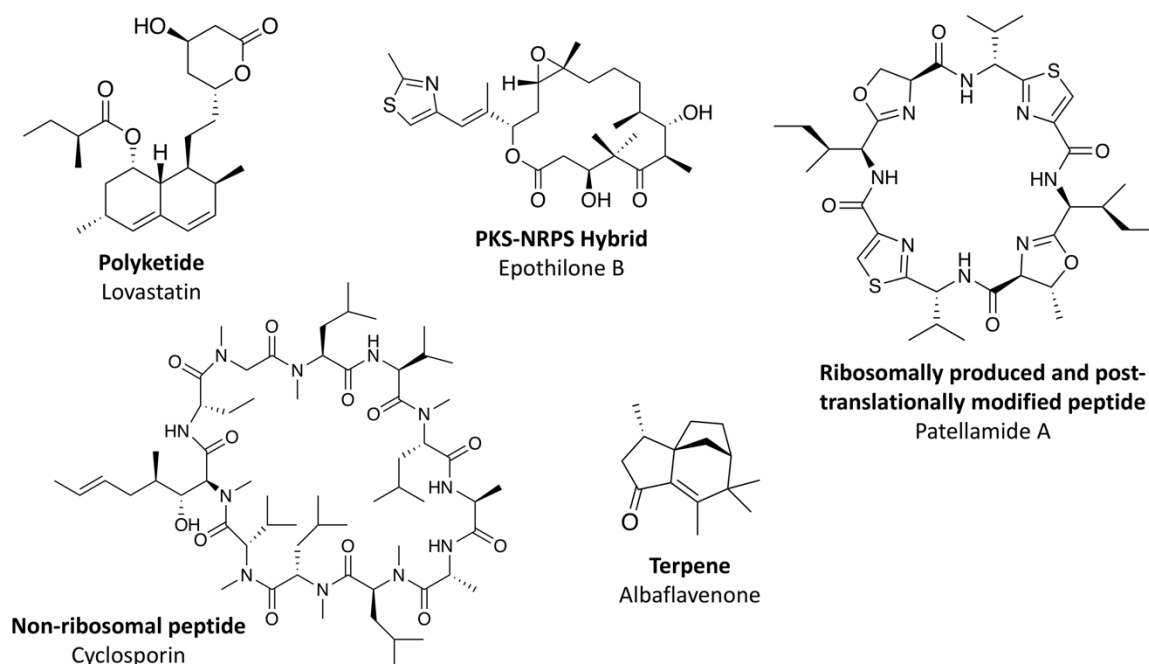
The minimal domain architecture of a *cis*-AT PKS consists of a ketosynthase (KS), an acyltransferase (AT) and an acyl carrier protein (ACP) domain. The biosynthetic logic of PKSs is based on ACP domains carrying a free thiol group derived from coenzyme A, which is attached to a conserved serine residue.<sup>[84]</sup> In a linear process, the AT domain selects and transfers acyl-CoA building blocks to the ACP that are connected in a condensation reaction catalyzed by the KS domain.<sup>[84]</sup> However, PKS systems are frequently found to be much more complex depending on the scaffold of the natural product, with additional functional domains [Ketoreductase (KR), Dehydratase (DH), Thioesterase (TE), Aromatase/Cyclase (ARO/CYC)] incorporated into modules of the assembly line, to just name a few.<sup>[85]</sup>

2. Non-ribosomal peptide synthetases (NRPS) are large, multimeric enzyme complexes capable of synthesizing oligopeptides independent of ribosome action, which extends the scope of possible incorporated amino acids beyond the available tRNAs used in translation. Like PKSs, NRPS systems also require at least three domains for peptide chain elongation, which are defined as a module that inserts one amino acid.<sup>[86]</sup> The A-domain selects and activates a specific amino acid as an amino acyl adenylate, which is transferred to the peptidyl carrier protein (PCP) that transports it to the C-domain responsible for the formation of the peptide bond.<sup>[86]</sup> In most cases, the C-terminal module contains a thioesterase (TE) domain, which is responsible for the release of the matured peptide.<sup>[87]</sup> Intriguingly, a lot of NRPS systems are also coupled with tailoring enzymes, which catalyze e.g. heterocyclization, epimerization or  $\beta$ -hydroxylation and can act during all stages of the synthetic process.<sup>[88]</sup> As such, NRPS products can be structurally very diverse, but lack the several ribosomal proofreading mechanisms during their synthesis.<sup>[86]</sup>



3. Bacterial terpene synthesis starts out from building blocks of either dimethylallyl diphosphate (DMAPP) or isopentenyl diphosphate (IPP), which undergo condensation to universal, acyclic precursors that are further cyclized by a common terpene cyclization reaction.<sup>[89]</sup> Even though bacterial terpene synthases show limited sequence homology to each other and to respective synthases from fungi or plants, they typically contain a combination of two highly conserved metal-binding motifs, the acidic amino acid-rich motif and a triad of residues.<sup>[89]</sup>

In contrast, the striking features of ribosomally synthesized and post-translationally modified peptides (RiPPs) are their origins as ribosomally synthesized precursor peptides and the absence of a required multimeric modular architecture for their synthesis, as their tailoring has been shown to usually be carried out by single enzymatic reactions. They comprise the youngest of all presented classes and are featured in more detail in the next section. As such, the number of known RiPPs BGCs in a prokaryotic, global analysis of 1154 genomes has been determined to be the second lowest of any class, highlighting the remaining biosynthetic space which can be explored.<sup>[74]</sup> A compilation of prominent representatives of every natural product compound class is displayed in **Figure 6**.



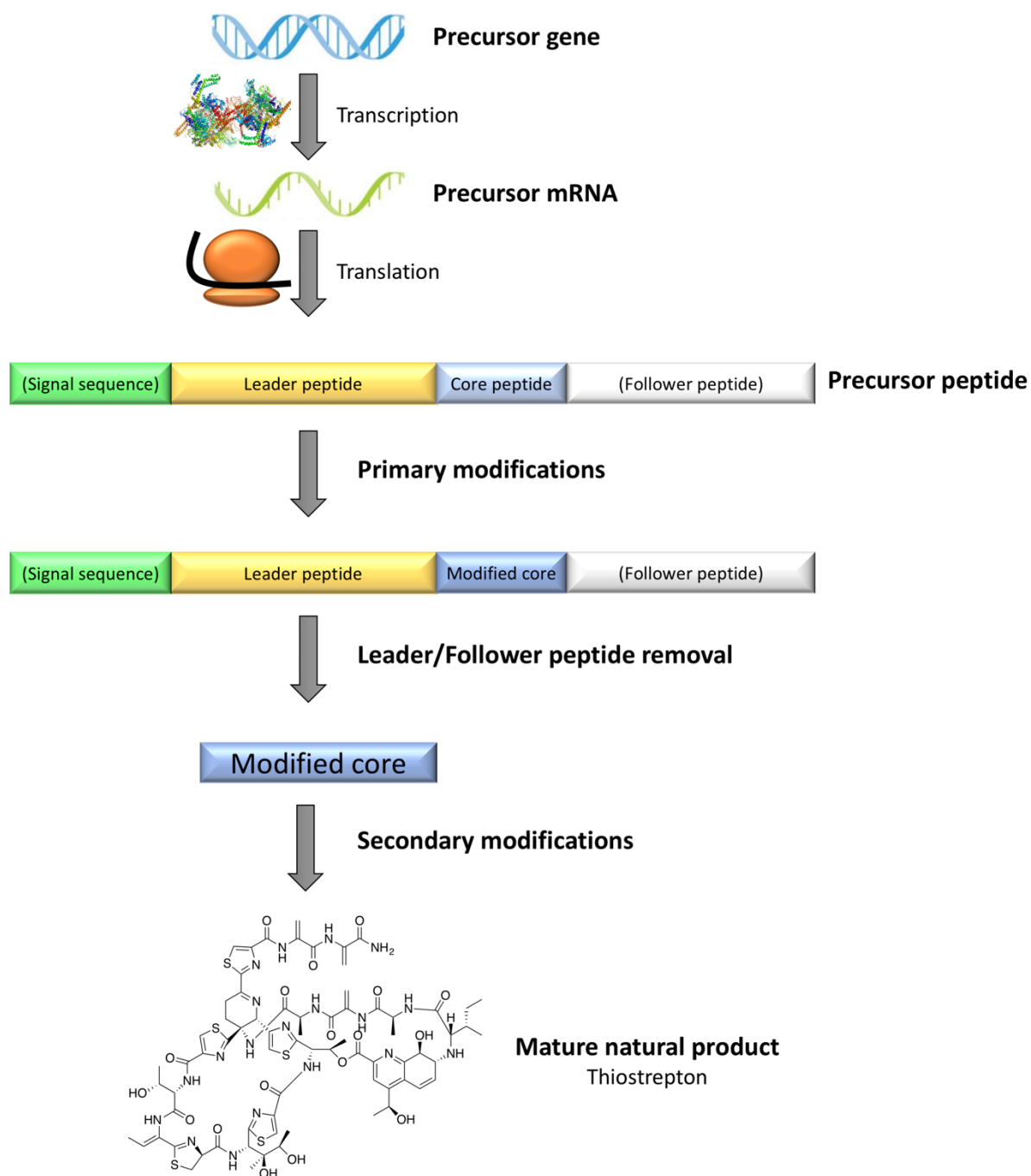
**Figure 6:** Compilation of prominent natural product representatives of every major biosynthetic class.

### 1.3 Ribosomally synthesized and post-translationally modified peptides (RiPPs) – natural products with great potential

The existence of peptide-based secondary metabolites in nature has been known for a very long time, but a clear distinction between a ribosomal or non-ribosomal origin was often only possible by investigating the underlying mechanisms of biosynthesis, e. g. initial NRPS research on tyrocidine in the 1960's.<sup>[90]</sup> As NRPS systems are capable of synthesizing products of enormous chemical diversity due to a combination of different domains and tailoring enzymes, it was an obvious notion for many years that heavily tailored peptidic natural products are originating from NRPS. However, the emerging link between specific BGCs and their products after the respective genomes had been sequenced, led to the observation of open reading frames in the BGC fitting to peptide sequences incorporated into the natural product scaffold. This class of natural products has consequently been defined as being distinct from NRPS and received increasing attention in the 21<sup>st</sup> century, to the point where a universal nomenclature guideline has been proposed in 2013, naming them ribosomally synthesized and post-translationally modified peptides (RiPPs).<sup>[91]</sup> Subsequently, once established information on peptide-based natural products had to be revised, leading to the reinvestigation and determination of many "classic" natural products as RiPPs, such as Thiostrepton and Botromycin (see next section).<sup>[92][93]</sup> To date, RiPP biosynthetic genes and natural products have been found in all three domains of life, with a respective vast structural diversity.<sup>[91]</sup>

The first evidence for a ribosomal origin of today's important bioactive lantibiotics dates back to 1988, when a group examined the biosynthesis of epidermin and proposed a ribosomal synthesis followed by an enzymatic maturation procedure.<sup>[94]</sup> This work already suggested the cleavage of a so-called pre-peptide to be an important tailoring step, which was later confirmed and re-named to the leader peptide in studies on the prominent antimicrobial peptide nisin produced by *Lactococcus lactis*.<sup>[95]</sup> The nisin leader peptide was subsequently found to be crucial for recognition of both modifying and transport proteins in a following publication.<sup>[96]</sup> Today, the biosynthetic paradigm of RiPPs has been established by various investigations on different pathways. Nearly all RiPPs originate from a ribosomally synthesized precursor peptide with a length of ~20-110 residues, which is encoded by a structural gene usually found within the BGC of the natural product.<sup>[91]</sup> The precursor peptide usually consists of an N-terminal leader peptide followed by a C-terminal core peptide, with the latter being

tailored to become the natural product.<sup>[97]</sup> However, both positions are swapped in specific biosynthetic pathways to an N-terminal core peptide followed by a C-terminal follower peptide, with arguably the best studied example being Bottromycins (see next section).<sup>[98]</sup> Eukaryotic RiPPs, the conopeptides, have been found to comprise a highly conserved N-terminal signal sequence, which directs the precursor peptide to the endoplasmic reticulum, and can be assumed to be absent in prokaryotic RiPPs.<sup>[99]</sup> The core peptide is subjected to primary post-translational modifications, which usually require the leader (or follower) peptide for substrate recognition.<sup>[97]</sup> After a (defined) order of early phase tailoring steps has taken place, the leader (or follower) peptide is proteolytically removed. It should be noted that proteolysis is often a highly regulated step in RiPPs biosynthesis as it needs to select for completely modified core peptides, since primary enzymes are no longer active after the leader (or follower) peptide has been removed. Oftentimes, but not always, secondary enzymes then act on the modified core peptide after which the matured RiPP is released.<sup>[100]</sup> An overview of a typical RiPP biosynthesis is shown in **Figure 7**.



**Figure 7:** Scheme of a typical RiPP biosynthesis. The genetically encoded precursor peptide is ribosomally synthesized and is subjected to primary modifications, typically requiring a leader, or in some cases, follower peptide, for substrate recognition. After proteolytic removal of the leader (or follower) peptide, the modified core peptide is sometimes further tailored to be released as a mature natural product. An N-terminal signal sequence has been observed in eukaryotic RiPPs, but is optional.

Since the emergence of RiPPs as a novel natural product class, various studies have discussed their unique biosynthetic scheme. It has been shown for several biosynthetic pathways that RiPP tailoring enzymes possess catalytic promiscuity towards changes in the core peptide region.<sup>[101]</sup> Consequently, a physical separation of recognition and catalysis has been

envisioned, to the extent that certain enzymes catalyzing primary modifications in RiPP pathways only selectively recognize the leader peptide, therefore tolerating changes in the core peptide.<sup>[100]</sup> This notion is also backed up by reports on enzymes capable of acting on hypervariable core peptides attached to highly conserved leader peptides.<sup>[102]</sup> Indeed, a conserved peptide-binding domain has been detected in different modifying enzymes of the majority of prokaryotic RiPP BGCs and defined as the RiPP recognition element (RRE).<sup>[103]</sup> In accordance with structural studies, RREs are believed to play an important role in binding the leader and guiding the core peptide into the active site of primary tailoring enzymes.<sup>[103]</sup> The leader peptide-guided biosynthetic strategy has been proposed to result in highly evolvable pathways with a relatively small number of involved enzymes, which can access a high chemical diversity at a low genetic cost.<sup>[91]</sup>

Opposed to multimeric PKS and NRPS systems, RiPP biosynthesis features mostly single post-translational modifications, making them a prime target for *in vitro* studies and peptide engineering starting out from the linear precursor peptide. However, their biosynthesis is by no means simple, as diverse regulation mechanisms and complex formation based on protein-protein interactions of biosynthetic enzymes are regularly observed.<sup>[104]</sup> Furthermore, chemical scaffolds of RiPP natural products can be extensively modified, even leading to an absence of peptide bonds in the final molecule as in the case of the bacterial cofactor pyrroloquinoline quinone (PQQ).<sup>[91]</sup> To date, observed post-translational modifications include cyclodehydration reactions forming different types of azoline heterocycles, macrocyclizations, *D* amino acids, sactonine and lanthionine crosslinks and cycloadditions.<sup>[101]</sup> Given their biosynthetic logic, especially complex chemical reactions with low yields in a synthetic scheme towards bioactive scaffolds can possibly be substituted by enzymatic reactions. Additionally, a reconstituted pathway can also enable facilitated derivatization studies of bioactive natural products by simply submitting the derivatized precursor peptide to the enzymatic machinery.

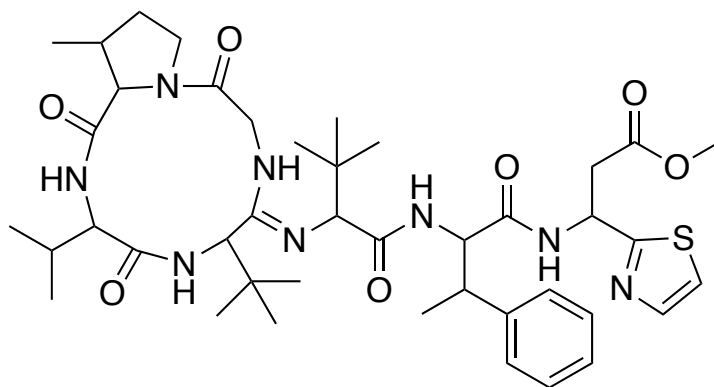
However, it is reasonable to assume that these are only the most apparent concepts of what is possible in RiPP engineering, as a lot of innovative research ideas are published based on our broadening understanding of RiPP biosynthesis each year. A study based on an extremely promiscuous biosynthetic enzyme capable of generating a library of  $10^6$  different lanthipeptides in *E. coli* has impressively demonstrated the scope of RiPP enzyme engineering.<sup>[105]</sup> Another intriguing approach relies on examining the structurally important

features of leader peptides from different classes for substrate recognition of interesting primary enzymes, as this information could be applied to generate chimeric leader peptides.<sup>[100]</sup> Bearing in mind the substrate scope and reaction conditions of every enzyme, these chimeric leader peptides could be used to rationally design and synthesize a library of new-to-nature hybrid RiPPs with potentially diverse chemical scaffolds.<sup>[101]</sup>

It is perhaps fortunate that the emergence of scientific interest in RiPP research happened simultaneously to advancements in bioinformatics, opening up novel possibilities in drug discovery. Therefore, gathered knowledge on mechanisms of post-translational modification by *in vivo* or *in vitro* studies can directly be used to model the structures of novel RiPPs. A corresponding approach using hypothetical structure enumeration and evaluation to analyze tandem MS data for RiPP structure elucidation has been presented.<sup>[106]</sup> Though it cannot account for stereochemistry, the sites and connectivity of post-translational modification were correctly identified as a proof of concept.<sup>[106]</sup> Similarly, peptide topologies of the diverse class of lasso peptides compared to their branched-cyclic analogues could be identified using a specialized mass spectrometry method.<sup>[107]</sup> It will be fascinating to explore the future potential of RiPPs for various human applications and possibly extend it beyond the panel of the important food preservative nisin and the topical veterinary application of thiostrepton. The current portfolio of RiPP natural products and derivatives thereof in clinical trials looks promising. Some examples are portrayed by the semisynthetic thiopeptide derivative LFF571 showing activity against *C. difficile* infections<sup>[108]</sup> and the lanthipeptide duramycin combating cystic fibrosis,<sup>[109]</sup> both having successfully completed phase II clinical trials.

## 1.4 Bottromycins

Bottromycins are natural products that were first isolated long before the emergence of RiPPs as a class of natural products in extracts of the terrestrial bacterium *Streptomyces bottropensis* during the golden age of antibiotic discovery in 1957.<sup>[110]</sup> The compound family received initial scientific interest after its first discovery, with the chemical structure and biological activity studies revealing an inhibition of protein synthesis in *E. coli* both being reported in 1966.<sup>[111][112]</sup> However, the original structure elucidation depicted a linear octapeptide, which was revised in 1983 and now correctly contained the characteristic macroamidine function (**Figure 8**).<sup>[113]</sup>

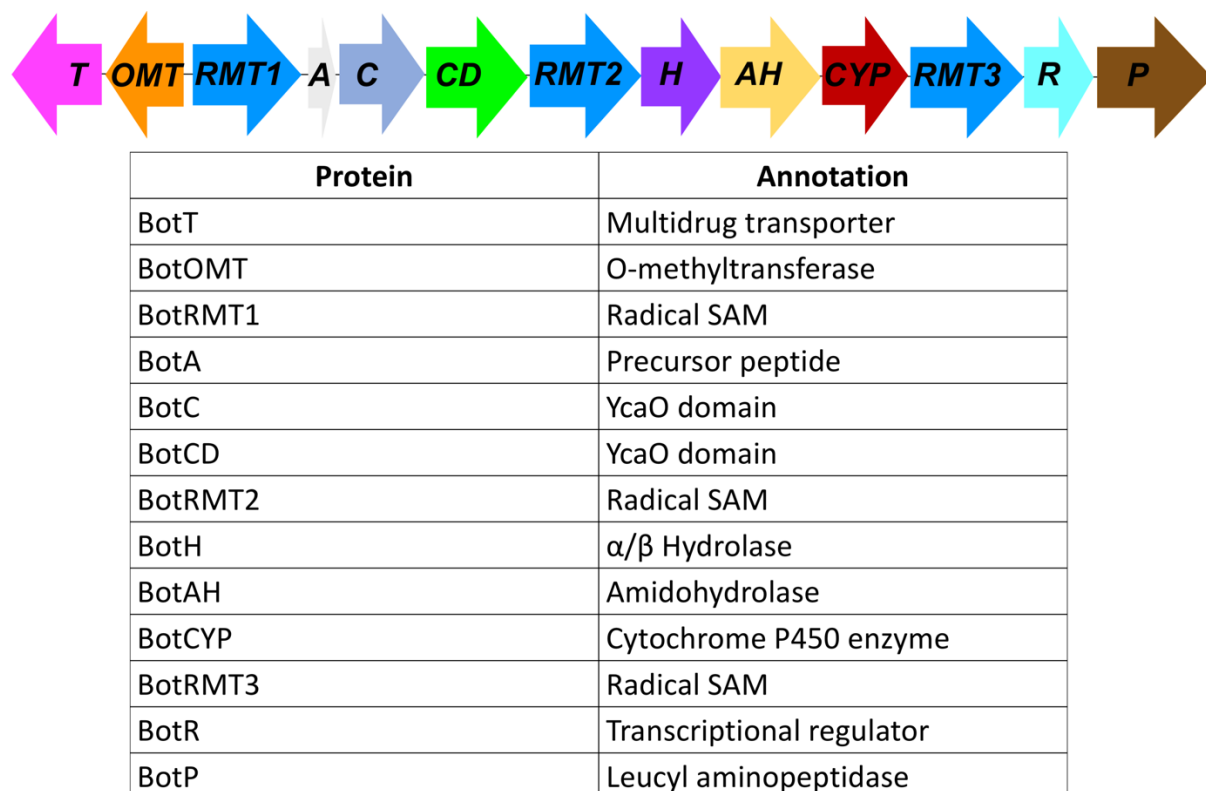


**Figure 8:** The revised chemical scaffold of Bottromycin A2 released in 1983, adapted from <sup>[113]</sup>.

Early studies on the biological activity of Bottromycin A were quite promising, as it showed higher *in vitro* activity than the well-known macrolide antibiotics against *Mycoplasma* species.<sup>[114]</sup> Follow-up research on pharmacokinetic properties of Bottromycin A and its hydrazide derivative using an *in vivo* mouse model indicated good tissue distribution and low toxicity.<sup>[115]</sup> Though, they also pointed out the rapid degradation of Bottromycin A in the animal body compared to its derivative, possibly by esterase activity on the O-methylated aspartate residue.<sup>[115]</sup> In 1976, a more detailed investigation on the inhibition of protein biosynthesis by Bottromycins determined the mode of action as the release of aminoacyl- or peptidyl-tRNA from the A site of the 50S bacterial ribosome.<sup>[116]</sup> This original proposal was later confirmed and narrowed down to a specific inhibition of binding to, but also a release of aminoacyl-tRNA from the A site in a study using isolated polysomes.<sup>[117]</sup> As such, it does not share a common binding site (and resistance mechanism) with the clinical antibiotic classes of tetracyclines and aminoglycosides, which have been shown to bind in the 30S subunit.<sup>[118]</sup> Amphenicoles, Macrolides and Lincosamides comprise some of the most prominent protein biosynthesis inhibitors targeting the 50S subunit. However, complex crystal structures have shown that there are various effective binding sites of these molecules within the 50S subunit, e. g. the polypeptide exit tunnel and the peptidyl transferase center.<sup>[119][120]</sup> Therefore, we can assume Bottromycins to exhibit a unique binding mode within the ribosome, also taken into account its new chemical entity compared to other 50S binders. Still, there was limited research and development progress towards a potential clinical use of these compounds until 2000.

With the start of the 21<sup>st</sup> century, there had been a renewed scientific interest in Bottromycin for several reasons. The rising emergence of antimicrobial resistance in pathogens had consequently started an investigation of “overlooked” compounds with an activity against multidrug-resistant strains. To that end, Bottromycin A2 was tested and confirmed to be

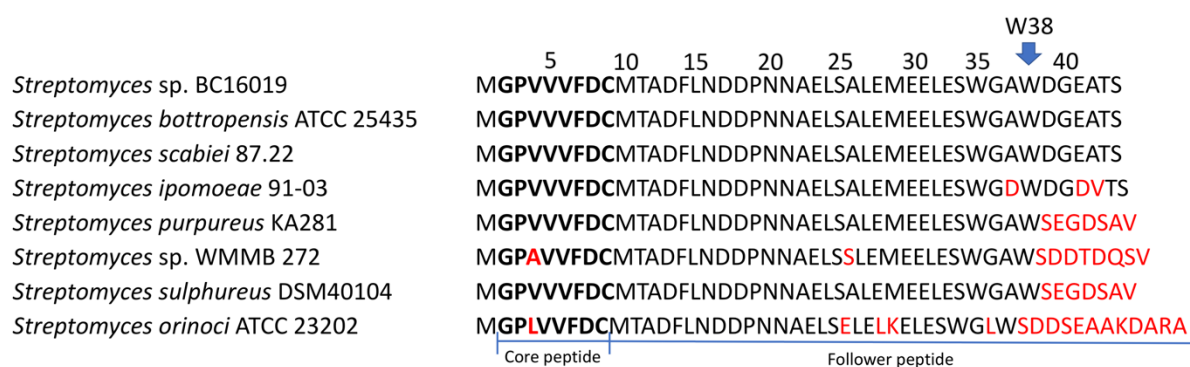
effective against multidrug-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant *Enterococci* (VRE).<sup>[121][122]</sup> Given their final chemical structure being heavily modified, the first total synthesis published in 2009 including the determination of the absolute configuration of Bottromycin A2 required 17 steps with a macrolactamization as the last step yielding only 51%.<sup>[121]</sup> As a first *in vivo* study depicted rapid degradation of Bottromycin A2 related to its aspartate methyl ester, derivatization of this position was shown to yield stable compounds in mouse plasma while still retaining activity.<sup>[123]</sup> Though, arguably one of the biggest advances in the field was the identification of the biosynthetic gene cluster of Bottromycins in different organisms, revealing its biosynthetic origin as a RiPP in 2012.<sup>[93][98][124][125]</sup> To this day, several *Streptomyces* strains carrying a Bottromycin biosynthetic gene cluster with different gene nomenclatures have been reported, but for this work we will focus on the nomenclature designated for *Streptomyces* sp. BC16019. The respective gene cluster as well as a table of every protein together with its most likely function based on sequence homology is shown in **Figure 9**.<sup>[93]</sup>



**Figure 9:** Schematic of the Bottromycin biosynthetic gene cluster found in *Streptomyces* sp. BC16019. The table shows every protein of the biosynthetic gene cluster together with its annotated function based on sequence homology, adapted from <sup>[93]</sup>.

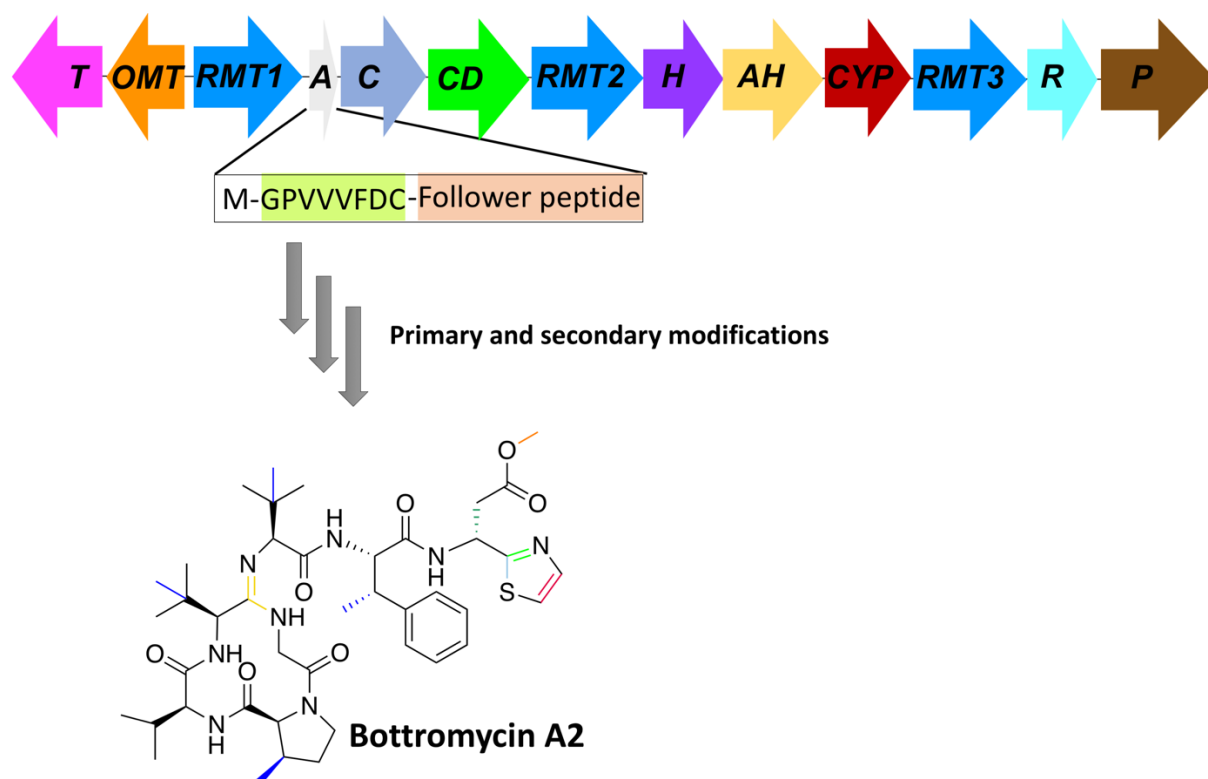


One of the most striking features of the Bottromycin biosynthetic gene cluster is the composition of the precursor peptide BotA. It does not show the typical RiPP architecture of a leader and a core peptide, but instead contains a N-terminal core peptide followed by a C-terminal follower sequence.<sup>[98]</sup> In a similar fashion as it was described for leader peptides, this follower peptide was proposed to be important for substrate recognition by the maturation enzymes found in the biosynthetic gene cluster.<sup>[98]</sup> The precursor peptide has been found to be highly conserved among the different Bottromycin producers, with major sequence variations only observed at the C-terminus after W38 (**Figure 10**).<sup>[126]</sup> Additionally, the only core peptide residue that appears to be variable is V4, although only minor variations to other hydrophobic, aliphatic amino acids (Alanine and Leucine) have been observed in two different strains thus far (**Figure 10**).



**Figure 10:** Comparison of the different Bottromycin precursor peptide sequences found in various *Streptomyces* strains, adapted from <sup>[126]</sup>.

Initial studies on *in vivo* knockout models have determined the function of the three radical SAM methyltransferases found within the cluster, as the methylation on different non-activated C<sub>β</sub>-atoms of the core peptide valines, proline and phenylalanine residues was missing.<sup>[93]</sup> However, the assignment of other post-translational modifications to the remaining enzymes of the cluster as well as their biosynthetic order had not been elucidated at the beginning of this work in early 2015. These include interesting biochemical reactions as the formation of the N-terminal macroamidine and the C-terminal thiazole moieties, the removal of the follower peptide as well as a possible epimerization leading to the observation of a *D*-aspartate methyl ester in the final natural product. As such, a biosynthetic scheme had been proposed and functions assigned to the enzymes based on sequence homology, which have been fitted and colour coded to the final natural product in **Figure 11**.



**Figure 11:** Proposed assignment of the post-translational modifications required for the final chemical scaffold of Botromycin A2. Modifications in the structure have been colour coded and assigned to specific biosynthetic enzymes based on sequence homology.

Concerning Botromycin biosynthesis, we presumed the enzymes responsible for the catalysis of the Botromycin core scaffold to be BotC, BotCD, BotH, BotAH and BotCYP. We expected either one or the combination of two YcaO domain enzymes BotC and BotCD to be involved in the formation of a C-terminal thiazoline, given the insights gained on YcaO domains involved in heterocycle formation of microcins.<sup>[127]</sup> The resulting thiazoline moiety was suggested to then be oxidized to become a thiazole by the Cytochrome P450 enzyme BotCYP. However, there was no available data on macroamide formation, so this function was assigned to BotAH. The epimerization of the aspartate residue was presumed to be spontaneous (no colour code in **Figure 11**), as there was a similar hypothesis for the patellamides before the start of this work.<sup>[128]</sup> With this data, we set out to explore the biosynthesis of this interesting natural product. The application of structural biology, biochemical and biophysical methods on every single biosynthetic step should allow us to gain a deep understanding of the underlying mechanisms and shed light on the generation of a new chemical entity starting its life as a ribosomally synthesized, linear precursor peptide.

## 1.5 Crocagins

As the crocagins were a completely novel class of RiPPs at the start of this work, all original research concerning their exploration is featured later in this thesis (**Chapters 6, 7 and 8**).

## 1.6 Outline of this work

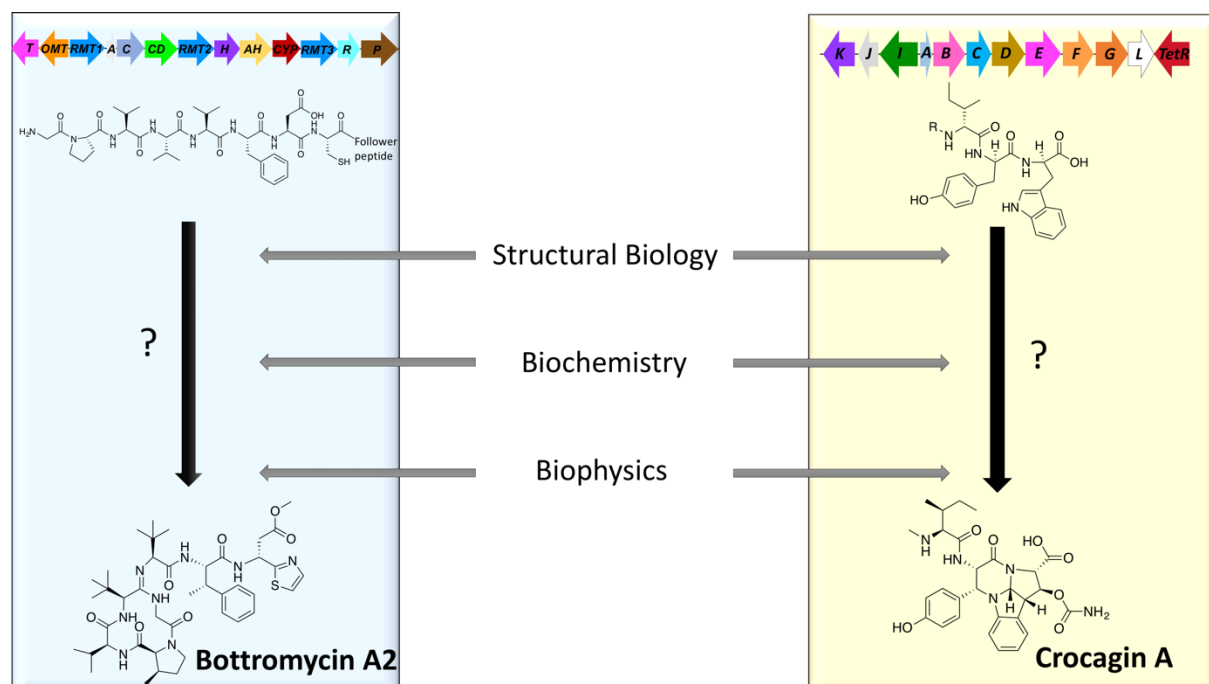
This work presents the investigation of the biosynthetic pathways of the interesting natural products Bottromycin and Crocagin using an *in vitro* approach. With limited experimental data available at the start of this project, the first task was the expression and purification of the potential biosynthetic enzymes of both clusters, which had previously been determined by sequence homology. Given their ribosomal origin as precursor peptides, their final chemical scaffold could be related to a specific set of post-translational modifications, although the respective order of those was completely unknown.

Starting out with bottromycin, a look at the matured natural product Bottromycin A2 revealed a missing N-terminal methionine residue, which implied the involvement of a biosynthetic aminopeptidase. The corresponding hypothesis could be confirmed by structural and biochemical data on BotP and is discussed in **Chapter 2**. Two of the most striking structural features of bottromycin A2 are arguably the C-terminal thiazole and the N-terminal macroamidine moieties. A study unraveling the underlying biochemistry and biosynthetic logic of these modifications suggests a similar mechanism of both responsible YcaO domain enzymes and is reported in **Chapter 3**. The gathered knowledge about the YcaO protein catalyzing the thiazoline formation, IpoC, is backed up by a structural analysis of the crystal structure presented in **Chapter 4**. Finally, the completion of the Bottromycin *in vitro* core scaffold by an oxidative decarboxylation, which is catalyzed by the cytochrome P450 enzyme BotCYP, could be achieved by a combined effort of our laboratory and is depicted in **Chapter 5** together with the respective crystal structure of a close homologue. The detailed understanding and *in vitro* reconstitution of every important biosynthetic step opens up a potential semi-synthetic route for the production of Bottromycin precursors and derivatives with enhanced pharmacokinetic properties. Consequently, these data are enabling the scientific community to further explore Bottromycins for possible clinical applications.

As a secondary project, this thesis examines the first reported RiPPs from myxobacteria, the Crocagins. They were discovered by means of secondary metabolome mining in extracts of

*Chondromyces crocatus* Cm c5 and represent a novel chemical scaffold. The elucidation of their structure in combination with the identification of their biosynthetic gene cluster and an *in vivo* knockout model confirmed their ribosomal origin and is discussed in **Chapter 6**. No biological activity for crocagins could be determined to date, but they have recently been linked to a potential function in quorum sensing and genetic competence as both the natural product and a protein of the gene cluster (CgnJ) show structural homology to their respective counterparts in *Bacillus subtilis* (**Chapter 7**). Lastly, parts of crocagin biosynthesis have been reconstituted *in vitro* and revealed implications for the formation of a possible macromolecular biosynthetic complex. Furthermore, the structural analyses of different biosynthetic enzymes from the cluster demonstrated a likely candidate for leader peptide hydrolysis (**Chapter 8**). In the context of horizontal gene transfer (HGT) mechanisms being the most prominent ways of spreading antimicrobial resistance genes, there is a heightened interest in understanding and possibly disrupting the underlying biology, which would in turn control the transmission. Consequently, exploring molecules that are possibly involved in these mechanisms broadens our general knowledge of the field and enables further research on the inhibition of HGT in clinically relevant pathogens.

A scheme of the summarized methodology of this work is displayed in **Figure 12**.



**Figure 12:** Outline of this work, highlighting the two main questions. The biosyntheses of the interesting natural products Bottromycin and Crocagin have been investigated starting out from the respective precursor peptides by means of structural biology, biochemical and biophysical methodologies.

## 1.7 References

- [1] Max Roser (2019) - "Life Expectancy". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/life-expectancy' [Online Resource]
- [2] Hannah Ritchie and Max Roser (2019) - "Causes of Death". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/causes-of-death' [Online Resource]
- [3] World Health Organization (2019): "Ten threats to global health in 2019". Published online at who.int Retrieved from 'https://www.who.int/emergencies/ten-threats-to-global-health-in-2019' [Online Resource]
- [4] O'Neill, J (2014): Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations. London: Review on Antimicrobial Resistance. 2014. Available from: <https://amr-review.org/Publications.html>
- [5] Singh SB, Barrett JF *Biochem Pharmacol.* **2006** 71 (7):1006-15
- [6] Rossiter SE, Fletcher MH, Wuest WM *Chem Rev.* **2017** 117 (19):12415-12474
- [7] Barka EA, Vatsa P, Sanchez L, Gaveau-Vaillant N, Jacquard C, Meier-Kolthoff JP, Klenk HP, Clément C, Ouhdouch Y, van Wezel GP *Microbiol Mol Biol Rev.* **2015** 80 (1):1-43
- [8] Zhang Z, Qu Y, Li S, Feng K, Wang S, Cai W, Liang Y, Li H, Xu M, Yin H, Deng Y *Sci Rep.* **2017** 7 (1):4837
- [9] Almabruk KH, Dinh LK, Philmus B *ACS Chem Biol.* **2018** 13 (6):1426-1437
- [10] Jahn LJ, Munck C, Ellabaan MMH, Sommer MOA *Front Microbiol.* **2017** 8:816
- [11] Peterson E, Kaur P *Front Microbiol.* **2018** 9:2928
- [12] Pendleton JN, Gorman SP, Gilmore BF *Expert Rev Anti Infect Ther.* **2013** 11 (3):297-308
- [13] Baltz, R. H. *SIM News* **2005** 55:186–196
- [14] Davies J, Davies D *Microbiol Mol Biol Rev.* **2010** 74 (3):417-33
- [15] Wright GD *Nat Rev Microbiol.* **2007** 5 (3):175-86
- [16] Volkova VV, Lu Z, Besser T, Gröhn YT *Appl Environ Microbiol.* **2014** 80 (14):4350-62
- [17] Bengtsson-Palme J, Kristiansson E, Larsson DGJ *FEMS Microbiol Rev.* **2018** 42 (1)
- [18] Melnyk AH, Wong A, Kassen R *Evol Appl.* **2015** 8 (3):273-83
- [19] Datta N, Hughes VM *Nature.* **1983** 306 (5943):616-7
- [20] Andersson DI, Hughes D *Nat Rev Microbiol.* **2010** 8 (4):260-71
- [21] Enne VI, Delsol AA, Davis GR, Hayward SL, Roe JM, Bennett PM *J Antimicrob Chemother.* **2005** 56 (3):544-51
- [22] Marcusson LL, Frimodt-Møller N, Hughes D *PLoS Pathog.* **2009** 5 (8):e1000541.

- [23] Andersson DI, Hughes D *FEMS Microbiol Rev.* **2011** 35 (5):901-11
- [24] Bliven KA, Maurelli AT *Microbiol Spectr.* **2016** 4 (1)
- [25] Rice LB *J Infect Dis.* **2008** 197 (8):1079-81
- [26] Magiorakos AP, Srinivasan A, Carey RB, Carmeli Y, Falagas ME, Giske CG, Harbarth S, Hindler JF, Kahlmeter G, Olsson-Liljequist B, Paterson DL, Rice LB, Stelling J, Struelens MJ, Vatopoulos A, Weber JT, Monnet DL *Clin Microbiol Infect.* **2012** 18 (3):268-81
- [27] Pogue JM, Kaye KS, Cohen DA, Marchaim D *Clin Microbiol Infect.* **2015** 21 (4):302-12
- [28] Santajit S, Indrawattana N *Biomed Res Int.* **2016** 2016:2475067
- [29] Høiby N, Bjarnsholt T, Givskov M, Molin S, Ciofu O *Int J Antimicrob Agents.* **2010** 35 (4):322-32
- [30] Michiels JE, Van den Bergh B, Verstraeten N, Fauvart M, Michiels J *Antimicrob Agents Chemother.* **2016** 60 (8):4630-7
- [31] Van Boeckel TP, Gandra S, Ashok A, Caudron Q, Grenfell BT, Levin SA, Laxminarayan R *Lancet Infect Dis.* **2014** 14 (8):742-750
- [32] Van Boeckel TP, Brower C, Gilbert M, Grenfell BT, Levin SA, Robinson TP, Teillant A, Laxminarayan R *Proc Natl Acad Sci U S A.* **2015** 112 (18):5649-54
- [33] Marshall BM, Levy SB *Clin Microbiol Rev.* **2011** 24 (4):718-33
- [34] Maron DF, Smith TJ, Nachman KE *Global Health.* **2013** 9:48
- [35] Harbarth S, Balkhy HH, Goossens H, Jarlier V, Kluytmans J, Laxminarayan R, Saam M, Van Belkum A, Pittet D *Antimicrob Resist Infect Control.* **2015** 4:49
- [36] Davies J, Spiegelman GB, Yim G *Curr Opin Microbiol.* **2006** 9 (5):445-53
- [37] Guerin E, Cambray G, Sanchez-Alberola N, Campoy S, Erill I, Da Re S, Gonzalez-Zorn B, Barbé J, Ploy MC, Mazel D *Science.* **2009** 324 (5930):1034
- [38] Gross M *Curr Biol.* **2013** 23 (24):R1063-5
- [39] Yoshikawa TT *Clin Infect Dis.* **2000** 30 (6):931-3
- [40] Rather IA, Kim BC2, Bajpai VK1, Park YH *Saudi J Biol Sci.* **2017** 24 (4):808-812
- [41] Centers for Disease Control and Prevention, Office of Infectious Disease Antibiotic resistance threats in the United States (2013) Available from: [https://www.cdc.gov/drugresistance/biggest\\_threats.html](https://www.cdc.gov/drugresistance/biggest_threats.html)
- [42] Davies SC *J Antimicrob Chemother.* **2018** 73 (4):833-834
- [43] Messacar K, Parker SK, Todd JK, Dominguez SR *J Clin Microbiol.* **2017** 55 (3):715-723

- [44] Hill-Cawthorne G, Negin J, Capon T, Gilbert GL, Nind L, Nunn M, Ridgway P, Schipp M, Firman J, Sorrell TC, Marais BJ *BMJ Glob Health*. **2019** 4 (2):e001283
- [45] Morgan D1, Okeke IN, Laxminarayan R, Perencevich EN, Weisenberg S *Lancet Infect Dis*. **2011** 11 (9):692-701
- [46] Lin DM, Koskella B, Lin HC *World J Gastrointest Pharmacol Ther*. **2017** 8 (3):162-173
- [47] Barra F, Roscetto E, Soriano AA, Vollaro A, Postiglione I, Pierantoni GM, Palumbo G, Catania MR *Int J Mol Sci*. **2015** 16 (9):20417-30
- [48] Colino CI, Millán CG, Lanao JM *Int J Mol Sci*. **2018** 19 (6)
- [49] Ji HF, Li XJ, Zhang HY *EMBO Rep*. **2009** 10 (3):194-200
- [50] Clayton DH, Wolfe ND *Trends Ecol Evol*. **1993** 8 (2):60-3
- [51] Newman DJ, Cragg GM, Snader KM *Nat Prod Rep*. **2000** 17 (3):215-34
- [52] Otvos RA, Still KBM, Somsen GW, Smit AB, Kool J *SLAS Discov*. **2019** 24 (3):362-385
- [53] Wan F, Zhong GS *Acta Med Sin* **1990** 5: 55–58
- [54] De Vos P *J Ethnopharmacol*. **2010** 132 (1):28-47
- [55] Krishnamurti C, Rao SC *Indian J Anaesth*. **2016** 60 (11):861-862
- [56] Dudek NK, Sun CL, Burstein D, Kantor RS, Aliaga Goltsman DS, Bik EM, Thomas BC, Banfield JF, Relman DA *Curr Biol*. **2017** 27 (24):3752-3762.e6
- [57] Woodruff HB *Science*. **1980** 208 (4449):1225-9
- [58] Gould K *J Antimicrob Chemother*. **2016** 71 (3):572-5
- [59] Sabatini DM *Proc Natl Acad Sci U S A*. **2017** 114 (45):11818-11825
- [60] Shen B *Cell*. **2015** 163 (6):1297-300
- [61] Newman DJ, Cragg GM *J Nat Prod*. **2012** 75 (3):311-35
- [62] Theuretzbacher U *Clin Microbiol Infect*. **2017** 23( 10):713-717
- [63] Challis GL *J Med Chem*. **2008** 51 (9):2618-28
- [64] Evans BE, Rittle KE, Bock MG, DiPardo RM, Freidinger RM, Whitter WL, Lundell GF, Veber DF, Anderson PS, Chang RS, et al. *J Med Chem*. **1988** 31 (12):2235-46
- [65] Wright GD *Nat Prod Rep*. **2017** 34 (7):694-701
- [66] D. Buss, Antony & Butler, Mark. (2010). Natural Product Chemistry for Drug Discovery. Chapter 2: Chemical Space and the difference between natural products and synthetics. DOI:10.1039/9781847559890

- [67] Lovering F, Bikker J, Humblet C *J Med Chem.* **2009** 52 (21):6752-6
- [68] Newman DJ *J Med Chem.* **2008** 51 (9):2589-99
- [69] Payne DJ, Gwynn MN, Holmes DJ, Pompliano DL *Nat Rev Drug Discov.* **2007** 6 (1):29-40
- [70] Spandl RJ, Bender A, Spring DR *Org Biomol Chem.* **2008** 6 (7):1149-58
- [71] Burke M, Schreiber S *Angew.Chem.* **2004** 116:48-60
- [72] Nören-Müller A, Reis-Corrêa I Jr, Prinz H, Rosenbaum C, Saxena K, Schwalbe HJ, Vestweber D, Cagna G, Schunk S, Schwarz O, Schiewe H, Waldmann H *Proc Natl Acad Sci U S A.* **2006** 103 (28):10606-11
- [73] Land M, Hauser L, Jun SR, Nookaew I, Leuze MR, Ahn TH, Karpinets T, Lund O, Kora G, Wassenaar T, Poudel S, Ussery DW *Funct Integr Genomics.* **2015** 15 (2):141-61
- [74] Cimermancic P, Medema MH, Claesen J, Kurita K, Wieland Brown LC, Mavrommatis K, Pati A, Godfrey PA, Koehrsen M, Clardy J, Birren BW, Takano E, Sali A, Lington RG, Fischbach MA *Cell.* **2014** 158 (2):412-421
- [75] Miller IJ, Chevrette MG, Kwan JC *Mar Drugs.* **2017** 15 (6)
- [76] Zhang MM, Qiao Y, Ang EL, Zhao H *Expert Opin Drug Discov.* **2017** 12 (5):475-487
- [77] Luo Y, Cobb RE1, Zhao H *Curr Opin Biotechnol.* **2014** 30:230-7
- [78] Ziemert N, Alanjary M, Weber T *Nat Prod Rep.* **2016** 33 (8):988-1005
- [79] Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R *Nucleic Acids Res.* **2011** 39(Web Server issue):W339-46
- [80] Tracanna V, de Jong A, Medema MH, Kuipers OP *FEMS Microbiol Rev.* **2017** 41 (3):417-429
- [81] Johnston CW, Skinnider MA, Dejong CA, Rees PN, Chen GM, Walker CG, French S, Brown ED, Bérday J, Liu DY, Magarvey NA *Nat Chem Biol.* **2016** 12 (4):233-9
- [82] Scherlach K, Hertweck C *Org Biomol Chem.* **2009** 7 (9):1753-60
- [83] Panter F, Krug D, Baumann S, Müller R *Chem Sci.* **2018** 9 (21):4898-4908
- [84] Piel J *Nat Prod Rep.* **2010** 27 (7):996-1047
- [85] Tsai SC, Ames BD *Methods Enzymol.* **2009** 459:17-47
- [86] Finking R, Marahiel MA *Annu Rev Microbiol.* **2004** 58:453-88
- [87] Sieber SA, Marahiel MA *J Bacteriol.* **2003** 185 (24):7036-43
- [88] Walsh CT, Chen H, Keating TA, Hubbard BK, Losey HC, Luo L, Marshall CG, Miller DA, Patel HM *Curr Opin Chem Biol.* **2001** 5 (5):525-34
- [89] Yamada Y, Cane DE, Ikeda H *Methods Enzymol.* **2012** 515:123-62



- [90] Mach B, Reich E, Tatum EL *Proc Natl Acad Sci U S A*. **1963** 50 (1):175-81
- [91] Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, Camarero JA, Campopiano DJ, Challis GL, Clardy J, Cotter PD, Craik DJ, Dawson M, Dittmann E, Donadio S, Dorrestein PC, Entian KD, Fischbach MA, Garavelli JS, Göransson U, Gruber CW, Haft DH, Hemscheidt TK, Hertweck C, Hill C, Horswill AR, Jaspars M, Kelly WL, Klinman JP, Kuipers OP, Link AJ, Liu W, Marahiel MA, Mitchell DA, Moll GN, Moore BS, Müller R, Nair SK, Nes IF, Norris GE, Olivera BM, Onaka H, Patchett ML, Piel J, Reaney MJ, Rebuffat S, Ross RP, Sahl HG, Schmidt EW, Selsted ME, Severinov K, Shen B, Sivonen K, Smith L, Stein T, Süssmuth RD, Tagg JR, Tang GL, Truman AW, Vederas JC, Walsh CT, Walton JD, Wenzel SC, Willey JM, van der Donk WA *Nat Prod Rep*. **2013** 30 (1):108-60
- [92] Liao R, Duan L, Lei C, Pan H, Ding Y, Zhang Q, Chen D, Shen B, Yu Y, Liu W *Chem Biol*. **2009** 16 (2):141-7
- [93] Huo L, Rachid S, Stadler M, Wenzel SC, Müller R *Chem Biol*. **2012** 19 (10):1278-87
- [94] Schnell N, Entian KD, Schneider U, Götz F, Zähner H, Kellner R, Jung G *Nature*. **1988** 333 (6170):276-8
- [95] Kuipers OP, Beerthuyzen MM, Siezen RJ, De Vos WM *Eur J Biochem*. **1993** 216 (1):281-91
- [96] Lubelski J, Rink R, Khusainov R, Moll GN, Kuipers OP *Cell Mol Life Sci*. **2008** 65 (3):455-76
- [97] Oman TJ, van der Donk WA *Nat Chem Biol*. **2010** 6 (1):9-18
- [98] Gomez-Escribano J, Song L, Bibb M, Challis G *Chem. Sci*. **2012** 3:3522-3525
- [99] Craig AG, Bandyopadhyay P, Olivera BM *Eur J Biochem*. **1999** 264 (2):271-5
- [100] Burkhart BJ, Kakkar N, Hudson GA, van der Donk WA, Mitchell DA *ACS Cent Sci*. **2017** 3 (6):629-638
- [101] Hudson GA, Mitchell DA *Curr Opin Microbiol*. **2018** 45:61-69
- [102] Li B, Sher D, Kelly L, Shi Y, Huang K, Knerr PJ, Joewono I, Rusch D, Chisholm SW, van der Donk WA *Proc Natl Acad Sci U S A*. **2010** 107 (23):10430-5
- [103] Burkhart BJ, Hudson GA, Dunbar KL, Mitchell DA *Nat Chem Biol*. **2015** 11 (8):564-70
- [104] Sikandar A, Koehnke J *Nat Prod Rep*. **2019** DOI: 10.1039/c8np00064f. [Epub ahead of print]
- [105] Yang X, Lennard KR, He C, Walker MC, Ball AT, Doigneaux C, Tavassoli A, van der Donk WA *Nat Chem Biol*. **2018** 14 (4):375-380
- [106] Zhang Q, Ortega M, Shi Y, Wang H, Melby JO, Tang W, Mitchell DA, van der Donk WA *Proc Natl Acad Sci U S A*. **2014** 111 (33):12031-6.
- [107] Dit Fouque KJ, Moreno J, Hegemann JD, Zirah S, Rebuffat S, Fernandez-Lima F *Anal Chem*. **2018** 90 (8):5139-5146.
- [108] Mullane K, Lee C, Bressler A, Buitrago M, Weiss K, Dabovic K, Praestgaard J, Leeds JA, Blais J, Pertel P *Antimicrob Agents Chemother*. **2015** 59 (3):1435-40
- [109] Sandiford SK *Expert Opin Drug Discov*. **2015** 10 (4):315-20

- [110] Waisvisz JM, van der Hoeven MG, van Peppen J, Zwennis WCM, *J Am Chem Soc.* **1957** 79 (16):4520–4521
- [111] Nakamura S, Tanaka N, Umezawa H *J Antibiot (Tokyo)*. **1966** 19 (1):10-2
- [112] Tanaka N, Sashikata K, Yamaguchi H, Umezawa H *J Biochem.* **1966** 60 (4):405-10
- [113] Schipper D *J Antibiot (Tokyo)*. **1983** 36 (8):1076-7
- [114] Tanaka N, Nishimura T, Nakamura S, Umezawa H *J Antibiot (Tokyo)*. **1968** 21 (1):75-6
- [115] Tanaka N, Nishimura T, Nakamura S, Umezawa H *J Antibiot (Tokyo)*. **1966** 19 (4):149-54
- [116] Otaka T, Kaji A *J Biol Chem.* **1976** 251 (8):2299-306
- [117] Otaka T, Kaji A *FEBS Lett.* **1983** 153 (1):53-9
- [118] Franceschi F *Future Microbiol.* **2007** 2 (6):571-4
- [119] Dunkle JA, Xiong L, Mankin AS, Cate JH *Proc Natl Acad Sci U S A.* **2010** 107 (40):17152-7
- [120] Svetlov MS, Plessa E, Chen CW, Bougas A, Krokidis MG, Dinos GP, Polikanov YS *RNA.* **2019** 25 (5):600-606
- [121] Shimamura H, Gouda H, Nagai K, Hirose T, Ichioka M, Furuya Y, Kobayashi Y, Hirono S, Sunazuka T, Omura S *Angew Chem Int Ed Engl.* **2009** 48 (5):914-7
- [122] Sowa S, Masumi N, Inouye Y, Nakamura S, Takesue Y, Yokoyama T *Hiroshima J Med Sci.* **1991** 40 (4):137-44
- [123] Kobayashi Y, Ichioka M, Hirose T, Nagai K, Matsumoto A, Matsui H, Hanaki H, Masuma R, Takahashi Y, Omura S, Sunazuka T *Bioorg Med Chem Lett.* **2010** 20 (20):6116-20
- [124] Hou Y, Tianero MD, Kwan JC, Wyche TP, Michel CR, Ellis GA, Vazquez-Rivera E, Braun DR, Rose WE, Schmidt EW, Bugni TS *Org Lett.* **2012** 14 (19):5050-3
- [125] Crone WJK, Leeper FJ, Truman AW *Chem. Sci.* **2012** 3:3516-3521
- [126] Schwalen CJ, Hudson GA, Kosol S, Mahanta N, Challis GL, Mitchell DA *J Am Chem Soc.* **2017** 139 (50):18154-18157
- [127] Dunbar KL, Melby JO, Mitchell DA *Nat Chem Biol.* **2012** 8 (6):569-75
- [126] Koehnke J, Bent AF, Houssen WE, Mann G, Jaspars M, Naismith JH *Curr Opin Struct Biol.* **2014** 29:112-121

## Figure references

**Figure 2:** Farmantibiotics.org; Pixabay.com; 123rf.com; Freepik.com; Elementhealing.com; all accessed on 30.06.2019.

**Figure 3:** Spicegarden.eu; Ancientpages.com; Ayurvedainnepal.com; vie.ens.fr; gohighbrow.com; Chinesemedicalclassics.wordpress.com; exhibits.hsl.virginia.edu; sciencephoto.com; planet-vie.ens.fr; all accessed on 19.07.2019.

**Figure 5:** Circos.ca, Antismash.secondarymetabolites.org; both accessed on 26.07.2019.

**Figure 7:** Onlinesciencenotes.com; accessed on 28.07.2019; shown protein structure is RNA polymerase holoenzyme from *E. coli*; PDB ID 4MEY.

## Chapter 2

# Structure and substrate recognition of the Bottromycin maturation enzyme BotP

Previously published in:

Greg Mann<sup>[a]1</sup>, Liujie Huo<sup>[b]1</sup>, **Sebastian Adam**<sup>[c]</sup>, Brunello Nardone<sup>[a]</sup>, Jeremie Vendome<sup>[d],[e]</sup>,  
Nicholas James Westwood<sup>[a]</sup>, Rolf Müller<sup>f</sup> and Jesko Koehnke<sup>[c]\*</sup>

**Chembiochem.** 2016 Dec 2;**17**(23):2286-2292.

**DOI: 10.1002/cbic.201600406**

### Affiliation

<sup>[a]</sup> School of Chemistry and Biomedical Sciences Research Centre, University of St. Andrews, North Haugh, St. Andrews, KY16 9ST, UK.

<sup>[b]</sup> Roger Adams Laboratory 156, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, IL, 61801, USA.

<sup>[c]</sup> Workgroup Structural Biology of Biosynthetic Enzymes, Helmholtz Institute for Pharmaceutical Research, Helmholtz Centre for Infection Research, Saarland University, Universitätscampus E8 1, 66123, Saarbrücken, Germany.

<sup>[d]</sup> Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY, 10032, USA.

<sup>[e]</sup> Howard Hughes Medical Institute, Columbia University, New York, NY, 10032, USA.

<sup>[f]</sup> Department of Microbial Natural Products, Helmholtz Institute for Pharmaceutical Research, Helmholtz Centre for Infection Research, Saarland University, 66123, Saarbrücken, Germany.

<sup>1</sup> These authors contributed equally to the manuscript

## **Contributions and Acknowledgements**

### **Author's effort:**

The author contributed significantly to the experimental section of this study by designing and performing experiments as well as evaluating the resulting data. The author re-established the protein expression and purification in our laboratory. Furthermore, the author performed all biochemical experiments featured in the final paper, interpreted the resulting data and contributed to the structural biology work by refining the BotP structure with bound manganese. Lastly, the author contributed to writing and editing the final manuscript.

### **Other's effort:**

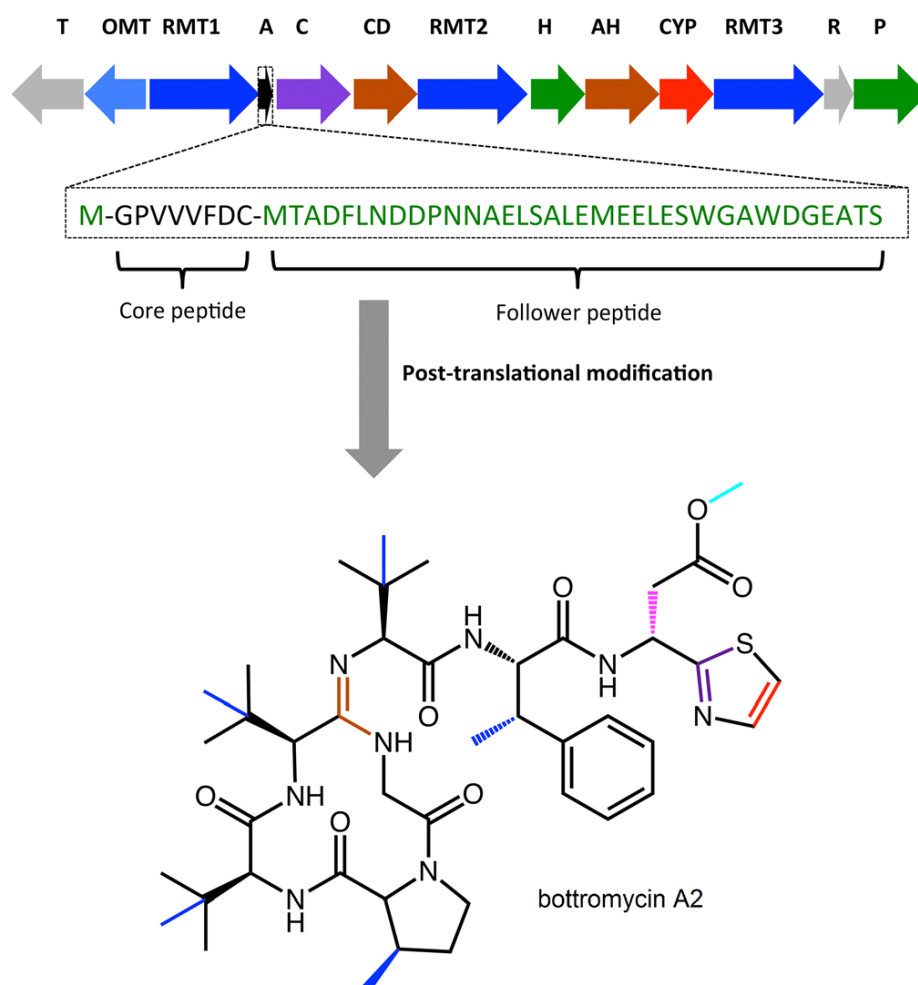
Liujie Huo contributed by doing the cloning as well as small scale protein expression and purification tests for BotP. Greg Mann contributed by performing the first crystal trials for the apo protein as well as collecting the first dataset for the apo protein leading to the molecular replacement solution. Brunello Nardone contributed by synthesizing the pentapeptides used in the biochemical experiments. Jeremie Vendome contributed by modeling the natural pentapeptide substrate into the active site.

## 2.1 Abstract

The bottromycins are a family of highly modified peptide natural products, which display potent antimicrobial activity against Gram-positive bacteria, including methicillin-resistant *Staphylococcus aureus*. Bottromycins have recently been shown to be ribosomally synthesized and post-translationally modified peptides (RiPPs). Unique amongst RiPPs, the precursor peptide BotA contains a C-terminal “follower” sequence, rather than the canonical N-terminal “leader” sequence. We report herein the structural and biochemical characterization of BotP, a leucyl-aminopeptidase-like enzyme from the bottromycin pathway. We demonstrate that BotP is responsible for the removal of the N-terminal methionine from the precursor peptide. Determining the crystal structures of both apo BotP and BotP in complex with  $Mn^{2+}$  allowed us to model a BotP/substrate complex and to rationalize substrate recognition. Our data represent the first step towards targeted compound modification to unlock the full antibiotic potential of bottromycin.

## 2.2 Introduction

Bottromycins have recently been identified as ribosomally synthesized and post-translationally modified peptides (RiPPs) - a growing class of natural products.<sup>[1]</sup> Like other RiPPs, bottromycins are derived from a core peptide that is part of a larger precursor peptide (BotA). BotA is encoded for by a structural gene, which undergoes extensive post-translational modifications (PTMs) by a series of enzymes to become the final product<sup>[1]</sup> (**Figure 1**). Unlike other RiPPs, BotA contains a C-terminal “follower” sequence, in place of the normal N-terminal “leader” sequence, which is thought to be important for stability and recognition of PTM enzymes.<sup>[1,2]</sup> The genes responsible for bottromycin biosynthesis, including that of the precursor peptide (*botA*) and the necessary tailoring enzymes, are localized within a single biosynthetic gene cluster<sup>[3,4]</sup> (**Figure 1**). Bottromycin maturation involves multiple methylation and proteolysis steps, amidine macrocycle and thiazole heterocycle formation and C-terminal decarboxylation<sup>[4-6]</sup> (**Figure 1**).



**Figure 1.** Bottromycin A2 is derived from a ribosomally synthesised precursor peptide BotA (black), containing a core peptide (black) and a C-terminal follower sequence (green). The precursor peptide undergoes a series of post-translational modifications to become the natural product Bottromycin A2. The genes within the gene cluster are coloured according to the modifications their gene products are proposed to catalyse.<sup>[5]</sup> The order of the modifications have been reported and include proteolysis on either side of the core peptide (green), heterocyclisation of the C-terminal cysteine within the core peptide (purple) oxidation of the resultant thiazoline to thiazole (red), C $\beta$ -methylation (dark blue), O-methylation (light blue), and macrocyclisation (dark red). Epimerisation (pink) is likely spontaneous.<sup>[5]</sup>

Previously, by using gene deletion experiments and subsequent analysis of the products formed, only the unusual set of methyl transferases had been assigned to their function in bottromycin biosynthesis.<sup>[3,4,7]</sup> Now, an order for these transformations as well as functions for the modifying enzymes have very recently been proposed based on metabolomics data of an engineered bottromycin pathway.<sup>[5]</sup> In that analysis, BotP appeared to be the first enzyme of the pathway to act on BotA.

BotP shares sequence homology with a number of M17 leucine aminopeptidases (LAPs) from various organisms<sup>[3,6]</sup> (**Figure S1** in the Supporting Information). These metal-dependent proteases hydrolyse peptide substrates with N-terminal leucine and methionine residues.<sup>[8]</sup> The function of LAPs in nature has evolved beyond ordinary roles in cell maintenance: degrading peptides released by endoproteases and the proteasome into amino acids. They are implicated in various additional biological functions including the processing of bioactive peptides, vesicular trafficking, transcriptional regulation, mediating site-specific recombination and as viral and toxin receptors.<sup>[8,9]</sup>

As a putative LAP, it was predicted that BotP facilitates the removal of methionine from the N-terminus of BotA.<sup>[3]</sup> Here we report the first biochemical and structural characterization of a PTM enzyme from the bottromycin biosynthetic pathway. The crystal structures of both apo BotP and BotP in complex with  $Mn^{2+}$  allowed us to build a BotP/substrate peptide model and determine the BotP residues involved in substrate recognition. Incubation of BotP with recombinantly produced BotA shows that this enzyme is responsible for the removal of the N-terminal methionine of BotA. The predictions based on our enzyme/substrate model were then tested in biochemical assays using peptide substrates.

## 2.3 Results and Discussion

### Overall structure of BotP

BotP was purified as described in the Experimental Section and subjected to crystallisation trials with commercially available sparse-matrix screens. Crystals grew after three days in a number of conditions but were too small to yield good diffraction patterns. Following optimization, the best crystals were selected and a data set was collected at the Diamond light source that diffracted to 1.76 Å (**PDB ID: 5LHJ**). The structure was determined using molecular replacement (ensemble of multiple search models). The data collection and refinement statistics can be found in **Table 1**.

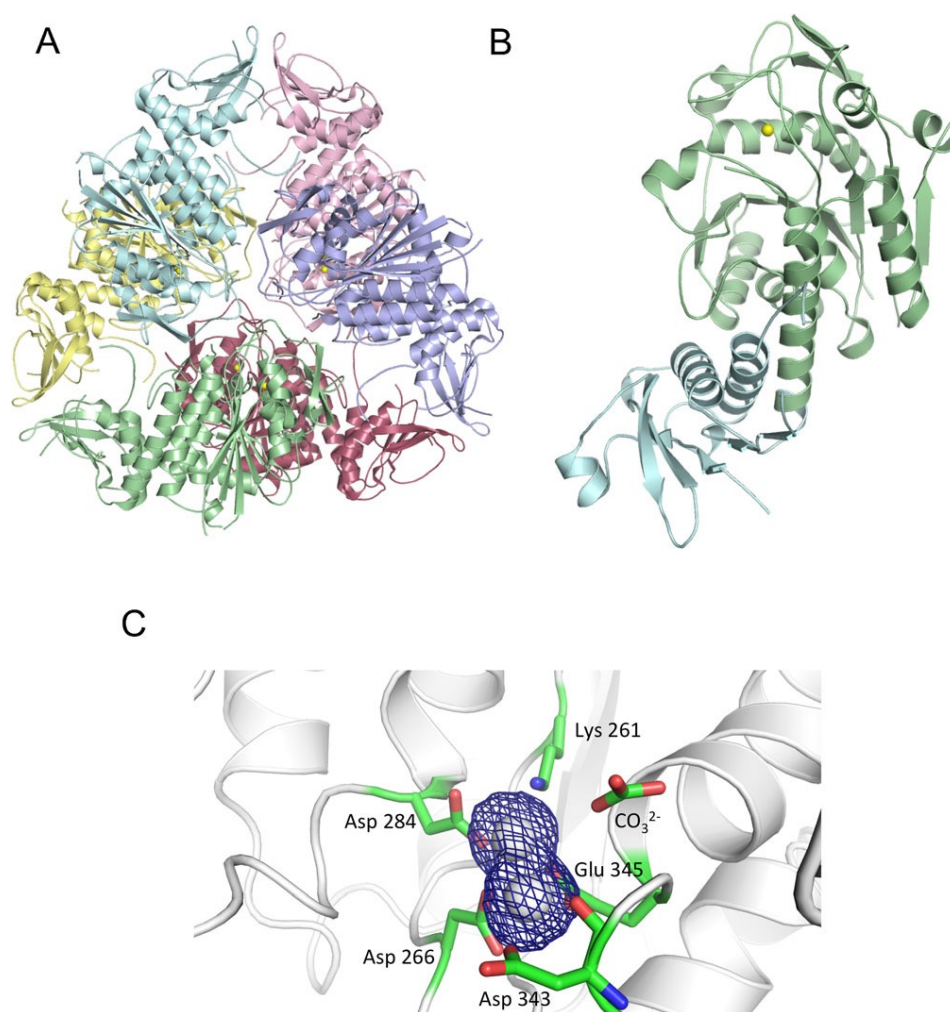


**Table 1: Data collection and refinement statistics**

	BotP apo	BotP-Mn <sup>2+</sup>
Resolution range [Å]	24.22-1.76 (1.82-1.76)	47.21-2.32 (2.38-2.32)
Space group	P 6 <sub>3</sub> 2 2	P 6 <sub>3</sub> 2 2
Unit cell		
$a, b, c$ [Å]	152.42, 152.42, 100.90	151.51, 151.51, 101.19
$\alpha, \beta, \gamma$ [°]	90.00, 90.00, 120.00	90.00, 90.00, 120.00
Total reflections	638 232 (97 638)	1 139 364 (77 380)
Unique reflections	68 447 (6724)	30 149 (2166)
Multiplicity	9.3 (8.9)	37.8 (35.7)
Wavelength [Å]	0.92	1.89
Completeness (%)	99.96 (99.96)	100.00 (99.90)
Mean $I/\sigma(I)$	17.0 (3.3)	26.8 (5.9)
Wilson B-factor	22.16	21.79
$R_{\text{merge}}$	0.067 (0.543)	0.142 (0.736)
$R_{\text{work}}$	0.1816 (0.2444)	0.1762 (0.1965)
$R_{\text{free}}$	0.2000 (0.2674)	0.1897 (0.1983)
Number of non-hydrogen atoms	3821	3640
Macromolecules	3358	3304
Ligands	11	6
Water	452	330
Protein residues	457	455
RMS (bond lengths)	0.008	0.008
RMS (angles)	1.32	1.37
Average B-factor	27.00	31.00
Macromolecules	25.00	30.26
Ligands	49.20	23.17
Solvent	41.40	38.61

One crystal was used per dataset. Values in parenthesis are for the highest resolution shell.  $R_{\text{merge}} = \sum hkl \sum i |I(hkl) - [I(hkl)]| / \sum hkl \sum i |I(hkl)|$ , where  $I(hkl)$  is the intensity of an individual measurement and  $[I(hkl)]$  is the average intensity from multiple observations.  $R_{\text{work}} = \sum hkl ||F_{\text{obs}} - k|F_{\text{calcd}}|| / \sum hkl |F_{\text{obs}}|$  ( $R_{\text{free}}$  was calculated in the same manner by using 5% of the reflection data chosen randomly and omitted from the start of refinement).

The crystal contained one monomer in the asymmetric unit, which, like other M17 LAPs, is one subunit of a hexamer (**Figure 2A**). Analysis with *PISA* (software that enables the exploration of macromolecular interfaces)<sup>[10]</sup> suggests this hexameric assembly to be biologically relevant, which is consistent with previous observations in homologous structures.<sup>[9,11,12]</sup> A BotP monomer consists of a small N-terminal domain (residues 35–163), and a larger pseudospherical C-terminal catalytic domain (residues 164–499), creating a characteristic “comma” shape (**Figure 2B**). The overall structure of BotP is highly similar to other M17 LAPs, and structural comparison with the recently characterized XoLAP<sup>[11]</sup> (**PDB ID: 3JRU**) shows a  $C_{\alpha}$  root-mean-square deviation (RMSD) of 1.61 Å over 339 residues. Superimposing the two structures in PyMOL reveals that while the two C-terminal domains are highly similar, the smaller N-terminal domains show slightly more deviation, which is a common observation within this protein family (**Figure S2**).



**Figure 2.** Cartoon representation of the crystal structure of BotP. A) BotP hexamer. For clarity individual monomers are coloured green, yellow, cyan, pink, red and purple. B) BotP monomer depicting the characteristic “comma” shape. The variable N-terminal domain is shown in blue and the catalytic C-terminal domain is shown in green. A chloride ion (in place of the bicarbonate ion in other LAPs) is shown as a yellow sphere. C) The active site of the BotP-Mn<sup>2+</sup> complex. Conserved catalytically important residues are shown as green sticks, Mn<sup>2+</sup>-ions as grey spheres and bicarbonate as green 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 metal ions, is shown as a blue isomesh.

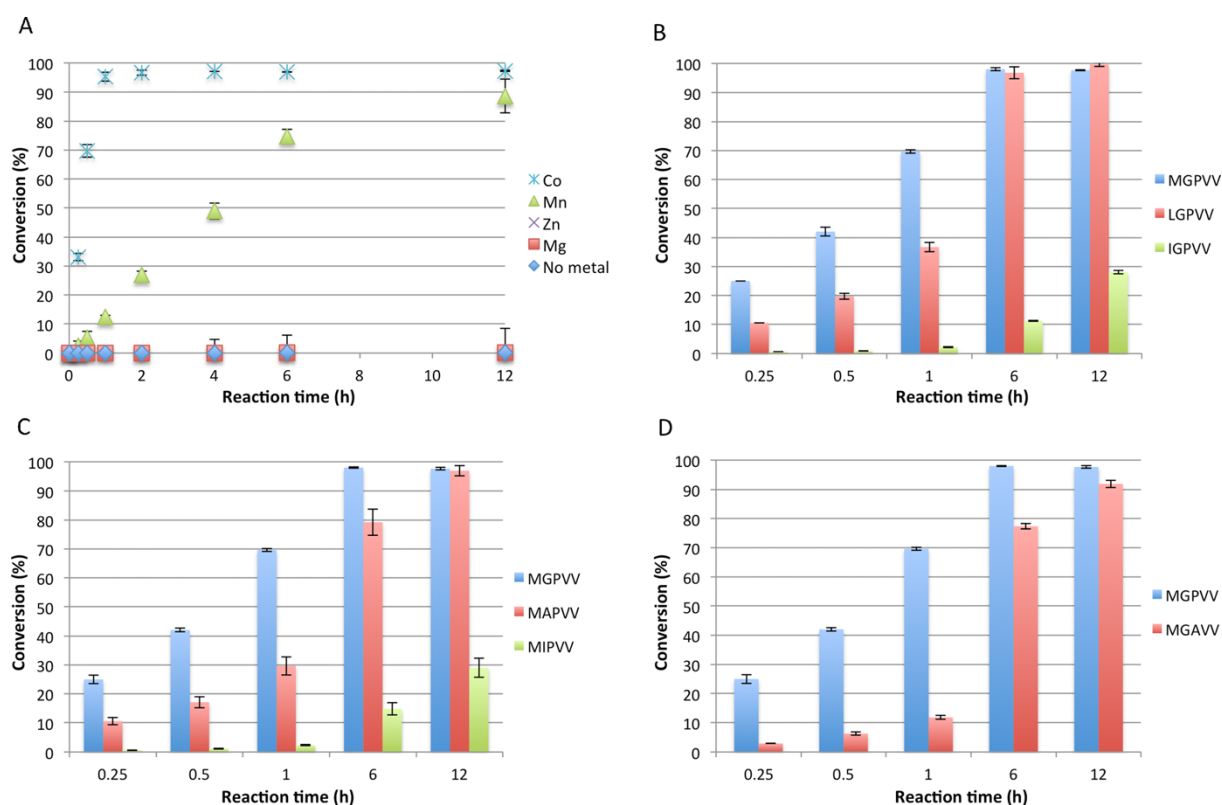
Putative catalytically important residues of BotP (based on alignments with other LAPs) are completely conserved in both sequence and structure (**Figure S1**). Within this cluster of active site residues, a chloride ion is in the place of the bicarbonate ion commonly found in LAPs. However, the two divalent metal ions observed in other LAPs are absent in the BotP structure, despite conservation of all the ligating residues (**Figure S3**).

### Biochemical characterization of BotP

Recombinantly produced precursor peptide BotA (100  $\mu$ M) was incubated with BotP (5  $\mu$ M) at 37 °C overnight and proteolysis was monitored by LC-MS (**Figure S4**). No activity could be detected, which we believed to be a result of the absence of metal ions (as seen in the structure). Typical metals found in active LAPs include  $\text{Zn}^{2+}$ ,  $\text{Mn}^{2+}$ ,  $\text{Co}^{2+}$  and  $\text{Mg}^{2+}$ -ions.<sup>[8,13]</sup> The proteolysis reaction was repeated with a truncated version of BotA in the presence of 100  $\mu$ M of each metal ion and reaction progress was monitored over time using LC-MS (**Figure 3A**). Addition of  $\text{Co}^{2+}$ -ions conferred the highest activity with complete substrate turnover observed after just 1 h (**Figure 3A** and **Figure S4**). BotP was less active in the presence of  $\text{Mn}^{2+}$ -ions, processing approximately 90 % of the substrate in 12 h, while almost no activity was detected with  $\text{Mg}^{2+}$  and  $\text{Zn}^{2+}$ -ions (**Figure 3A**).

### Metal binding of BotP

Given the partial and full restoration of activity in the presence of  $\text{Mn}^{2+}$  and  $\text{Co}^{2+}$ -ions respectively, we sought to determine the structures of BotP in complex with these ions. Soaking the optimized BotP apo crystals in the presence of  $\text{MnCl}_2$  was detrimental to the diffraction quality of the crystals. Consequently, BotP was pre-incubated with  $\text{MnCl}_2$  overnight on ice, prior to crystal screening. Crystals of the putative BotP- $\text{Mn}^{2+}$  complex grew under different conditions than the apo crystals. All of our attempts to obtain a BotP- $\text{Co}^{2+}$  complex, both by way of crystal soaking and pre-incubation, were unsuccessful. The putative BotP- $\text{Mn}^{2+}$  crystals were backsoaked prior to freezing to remove any unbound  $\text{Mn}^{2+}$ -ions. These crystals were then used in an X-ray fluorescence scan to confirm the presence of  $\text{Mn}^{2+}$  in the crystals. The recorded spectrum (**Figure S5**) is congruent with the presence of  $\text{Mn}^{2+}$  in the BotP- $\text{Mn}^{2+}$  crystals. Data for these crystals were collected at the SLS at the Mn-edge and the complex structure solved using Mn-SAD (manganese single-wavelength anomalous diffraction, **PDB ID: 5LHK**). As expected, both putative metal binding sites were occupied with  $\text{Mn}^{2+}$  (**Figure 2C**). This time a bicarbonate ion was found in place of the chloride ion in the apo structure.



**Figure 3.** A) Relative proteolysis rate of a truncated version of BotA (MGPVV) by BotP in the presence of different metal ions. Using BotP- $\text{Co}^{2+}$  we tested different residues in the B) P1, C) P1' and D) P2' positions. The reactions were monitored at specific time points by LC-MS. For each time point, the amount of substrate and product were determined, and the percent conversion was calculated. Results for full-length BotA are analogous (**Figure S3**).

### Substrate recognition by BotP

A common property of enzymes involved in RiPP biosynthesis is their broad tolerance of different substrates, which enables the same enzymes to synthesise multiple products.<sup>[14,15]</sup> Because co-crystallization of BotP with substrate peptides or full-length BotA proved to be intractable owing to slow crystal growth and fast enzymatic turnover, we used modelling to provide insights into the substrate specificity of BotP. The hexameric assembly of BotP prevents accurate modelling of BotA beyond residue 4, which led us to model the complex of BotP with the peptide sequence MGPV (P1–P3' in protease nomenclature; **Figure 4A**, **Figure S6**; the residue of the enzyme pocket binding the P1 residue is referred to as S1 and so forth.). We did not find any indication that residues of the substrate beyond MGPV are involved in BotP binding *in silico*, and the C-terminal Val residue only contributes marginally to the binding. To test this experimentally, we performed a competition assay using BotA and the truncated form of BotA (MGPVV). BotA gets processed in the same manner but approximately

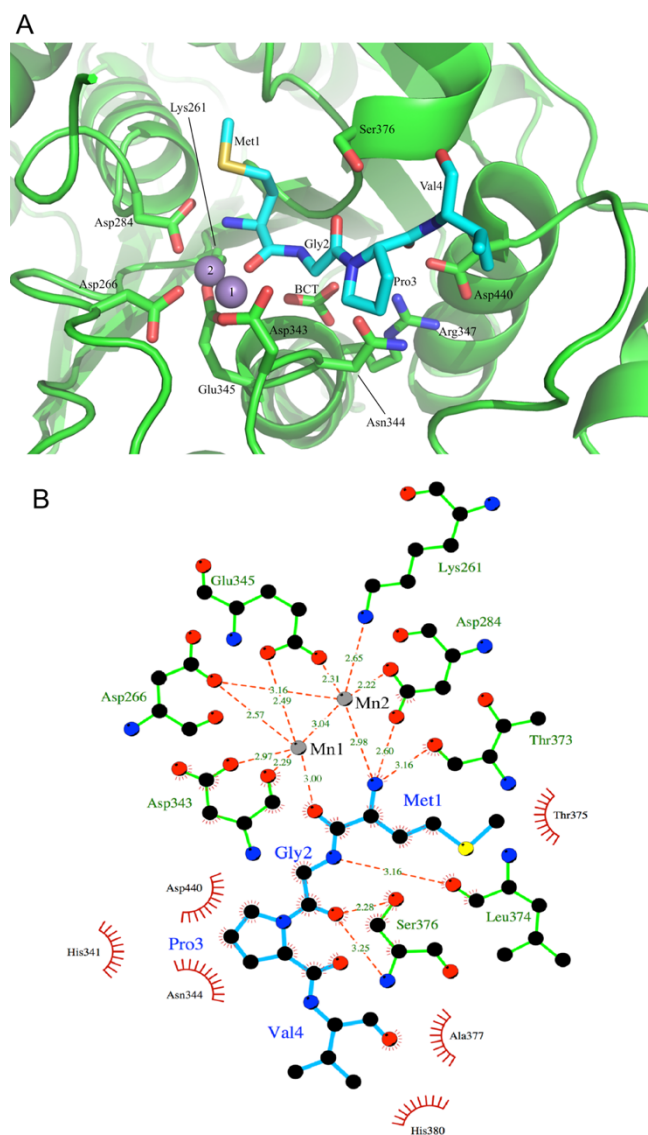
fourfold faster than the truncated version (**Figure S4 C**). It could not be determined whether this difference in rate is the result of (non)specific substrate–enzyme interactions or simply a thermodynamic effect of the substrate size. Nevertheless, our model adequately addresses the substrate specificity, which seems to primarily depend on the first three residues of the core peptide.

We verified this claim by using subsequent biochemical experiments. The long and narrow S1 pocket is very well suited to accommodate methionine residues with an extended side chain. The P1 methionine residue makes hydrophobic contact with Thr375 and forms a hydrogen bond with Thr373 (**Figure 4B**). Additionally, it partakes in the extensive hydrogen-bonding network involving the two  $Mn^{2+}$ -ions (**Figure 4B**). The S1 pocket could also allow the binding of Gly, Ala and  $C_{\gamma}$ - branched amino acids. Residues branched at the  $C_{\beta}$  would not fit into the pocket. To test this hypothesis, we incubated BotP with  $Co^{2+}$ -ions and peptides MGPVV (wild-type), LGPVV and IGPVV. The LGPVV peptide was processed slowly (compared to wild-type) and complete turnover was observed after 6 h (**Figure 3B**). However, the IGPVV peptide was only partially cleaved by BotP after 12 h, indicating it is a very poor substrate and, thus, confirming our hypothesis.

The S1' pocket is particularly narrow and the glycine forms hydrogen bonds with the carbonyl of Leu374 and the side- and main-chain of Ser376. At first glance, our model would only allow for glycine residues. However, assuming a slight rotation of the BotP Asn301 side-chain, larger residues could also be accommodated including Val or Ile. We sought to test this hypothesis by attempting to cleave MAPVV and MIPVV (**Figure 3C**). Both changes were accepted, but cleavage progressed much faster with the Ala substitution. To probe substrate tolerance at this position further an additional 16 peptide substrates were tested (**Figure S7**). As expected, Ser was processed as well as Ala and bulkier amino acids and those with longer side-chains were processed more slowly than Ala. One surprise was that when a Lys residue is in the P1' position, the peptide still gets processed half as fast as Ala/Ser at this position. We believe this to be the result of the extraordinary flexibility of the Lys side chain.

The S2' and S3' pockets provide an interesting insight into substrate recognition. BotP residues Asn344 and Asp440 block most of the S3' pocket (**Figure 4B**). These residues are usually Ala and Gly residues, respectively, in other LAPs. This is consistent with P2' being a proline residue,

which introduces an elbow into the substrate peptide, pointing it away from the surface of the enzyme. The Pro residue makes hydrophobic interactions with His341, Asn344 and Asp440. Substitution of P2' from Pro to Ala (MGAVV) was tolerated by BotP but at a rate slower compared with the wild-type sequence (**Figure 3D**).



**Figure 4.** Model of substrate peptide MGPV (cyan) with BotP-Mn<sup>2+</sup> (green). A) Cartoon representation. B) Ligplot representation. Hydrogen bonds and ligand bonds are shown as red dashes. Mn<sup>2+</sup>-ions are shown as purple (A) or grey (B) spheres, the bicarbonate ion (BCT) is only shown in (A), because it does not interact directly with the substrate or metal ions.

## 2.4 Conclusions

The bottromycins are members of the ever-growing ribosomal and post-translationally modified peptide synthesis family of natural products. They are synthesized first as a precursor peptide, which undergoes extensive enzymatic (and possibly non-enzymatic) post-translational modifications to give the final product. At present, much of the details of these enzymatic transformations remain unknown. One essential modification in bottromycin biosynthesis is the removal of the N-terminal methionine from the precursor peptide BotA. Sequence homology with a number of known leucyl aminopeptidases implicated BotP as the enzyme responsible for this modification.

Recombinant BotP, isolated from *E. coli*, was initially found to be inactive owing to the absence of divalent metal ions as cofactors, but addition of  $Mn^{2+}$  or  $Co^{2+}$ -ions restored the activity. The enzymatic turnover of BotP was faster with  $Co^{2+}$  than with  $Mn^{2+}$ -ions. There has been a long debate regarding the identity of the divalent metal ions bound at the active sites of LAPs.<sup>[8]</sup> While this study identifies  $Co^{2+}$  as the ion giving the highest relative level of activity for BotP, we cannot discern which metal ions (or combinations of metal ions) are preferred in the natural producer of bottromycins. Nevertheless, it would seem that  $Co^{2+}$  would be a good choice for future in vitro studies.

Most PTM enzymes from RiPP pathways act on the leader peptide or protease recognition elements outside the core peptide. This separation of recognition from catalysis results in a large tolerance for substrate changes by the respective enzymes. In fact, the sequence of several core peptides can essentially be hypervariable. In the case of bottromycins, the absence of an N-terminal protease site outside the core peptide could be interpreted as a severe restriction on the combinatorial possibilities for engineered substrates. However, it appears that, perhaps in an effort to limit this restriction, BotP has evolved to recognise and bind to just three residues, with binding pockets S3' partially and S4' and S5' completely inaccessible, compared with other LAPs. Even for the binding pockets that are available, BotP is able to at least tolerate drastic single amino acid changes, although the processing progresses more slowly. This may bode well for future engineering of the pathway, because BotP activity is essential to liberate the N-terminus of BotA, thus making it available for macrocyclisation. We would like to note that BotP processes its substrate significantly faster than other RiPP proteases. We believe this to be a reflection of its relative position in the

biochemical cascade—normally RiPPs proteases remove the leader and it has been suggested that they are deliberately slow to avoid cleavage before the PTMs are completed. BotP has no such restrictions. In fact, by being the first enzyme of the pathway, short processing times are essential for efficient bottromycin biosynthesis.

Bottromycins exhibit promising antimicrobial activity, owing to a unique molecular scaffold and a novel biological target. Their biosynthesis requires diverse chemical transformations. The better understanding of BotP as a PTM enzyme of this pathway is a first step towards structure-guided and targeted compound modification *in vitro* and *in vivo* to unlock the full potential of bottromycins.

## 2.5 Experimental section

**Protein cloning, expression and purification:** Full-length *botP* was cloned into pEHISMBPTEV (provided by Dr. Huanting Liu, St. Andrews University) and expressed in *E. coli* BL21(DE3) cells grown in auto-induction medium at 20 °C for 48 h.<sup>[16,17]</sup> Cells were harvested by centrifugation (4000 g, 4 °C, 15 min) and resuspended in lysis buffer (NaCl (500 mM), imidazole (20 mM), Tris (20 mM, pH 8.0),  $\beta$ -mercaptoethanol (BME, 3 mM)) supplemented with cOmplete EDTA-free protease inhibitor tablets (Roche) and DNase (0.4 mg g<sup>-1</sup> wet cells, Sigma). The cell suspension was lysed via passage through a cell disruptor (30 kpsi, Constant Systems Ltd), and the cell debris was removed by centrifugation (40000 g, 4 °C, 20 min). The supernatant was filtered through a 0.45 mm filter and applied to a HF-Ni-NTA column (GE Healthcare) pre-equilibrated in lysis buffer. The protein was eluted in lysis buffer supplemented with extra imidazole (250 mM). The protein eluent was passed over a desalting column (16/10 GE Healthcare) into “low-imidazole, low-salt” buffer (NaCl (150mM), imidazole (20mM), Tris (20mM, pH 8.0), BME (3 mM)) and subsequently incubated with TEV protease at room temperature for 2 h to remove the His<sub>6</sub>-tagged-MBP. Digested protein was loaded onto a HF-Ni-NTA column, and the flow through was collected. A homogenous and pure sample was obtained via passage down a Superdex 200 gel filtration column (GE Healthcare) that was pre-equilibrated with gel filtration buffer (NaCl (150 mM), HEPES (10 mM, pH 7.4), TCEP (1 mM)). Protein purity was assessed by SDS-PAGE and integrity confirmed by MS.

Full-length, codon-optimized *botA* was constructed from synthetic primers using PCR and cloned into the pEHISSUMOTEV (provided by Dr. David Owen, St. Andrews University) plasmid to introduce a TEV cleavable SUMO tag. Expression and purification were performed as described for BotP.



**X-ray crystallography:** Crystals of apo-BotP were obtained in PEG 3350 (20 %, w/v) and sodium malonate (0.24 M, pH 7.0). Crystals of BotP-Mn<sup>2+</sup> complex were obtained in PEG 4000 (10 % w/v), MES (0.1 M, pH 6.5). Diffraction data for apo-BotP and BotP-Mn<sup>2+</sup> crystals were collected at Beamlines I04–1 (Diamond light source) and X06DA (Swiss Light Source), respectively. Data was processed using Xia2 and the apo structure solved using PHASER Molecular Replacement, using an ensemble of multiple search models generated in Ensemblr.<sup>[18-20]</sup> The BotP-Mn<sup>2+</sup> structure was solved using PHENIX autosol with autobuild.<sup>[21,22]</sup> The models were manually rebuilt with COOT and refined using PHENIX<sup>[21,22]</sup> and Refmac5.<sup>[23]</sup> The structure was validated using MolProbity and all images were presented using PyMOL.<sup>[24,25]</sup> Interaction diagrams were created by using Ligplot.<sup>[26]</sup>

**Modelling of substrate complex:** The peptide MGPV was modelled in the catalytic pocket of the reported structure of BotP by using the crystal structure of the small peptidomimetic molecule Ubenimex (a.k.a. Bestatin) bound in the catalytic site of the zinc-dependent leucine aminopeptidase from *Pseudomonas putida* ppLAP (PDB ID: 3H8G, chain A) as a template. Specifically, the structure of *P. putida* ppLAP bound to Ubenimex was aligned to our structure of BotP, and the coordinates of chosen atoms of Ubenimex were used as the initial position for certain P1, P1' and P2' backbone atoms. Initial coordinates for the remaining atoms of P1, P1', P2' and P3' were generated using the build internal coordinates function in CHARMM (version c32a2 ; <https://www.charmm.org/charmm/>). The structure of the peptide was then minimized in the context of the BotP catalytic site, using 50 steps of steepest descent followed by 1000 steps of conjugate gradient minimization, with a position constraint of 10 kcal mol<sup>-1</sup> on all BotP heavy atoms. Because two peptidomimetic inhibitors of leucine aminopeptidase, namely Ubenimex and leucine p-nitroaminal, were previously reported to bind LAP with their N-terminus in a deprotonated form (NH), both the protonated (NH<sup>+</sup>) and deprotonated (NH) forms 232 of the MGPV peptide were modelled following the protocol described above. Consistent with what was observed for the two peptidomimetic inhibitors, the deprotonated form bound much more favourably in the BotP site, the NH<sub>2</sub> group notably coordinating the second zinc ion. The model with a deprotonated N-terminus is presented in the article.

**Peptide synthesis:** The solid-phase synthesis of the peptides was carried out on a Rink-amide PEG resin (0.45 mmol g<sup>-1</sup>) purchased from Merck-Millipore. Fmoc-protected amino acids and all other reagents were purchased from Sigma–Aldrich and were used without any further purification. Syntheses were performed manually on a 0.023 mmol scale. Amino acids were coupled using a fivefold excess (0.12 mmol). A solution of HBTU (0.5 M, 0.12 mmol, 46 mg) in DMF (1 mL) was added to the Fmoc-protected amino acid, followed by DIPEA (0.12 mmol, 21 mL) and the amino acid was activated for 30 s prior to coupling. The resin was bubbled with N<sub>2</sub>. A single treatment of 1 h was used for all amino acids. The Fmoc group was deprotected by using 20 % piperidine/DMF (v/v, 1 mL) for 10 min. The extent of

coupling was checked using the ninhydrin test. The resin was washed with DMF (6 x 1 mL) at the end of each coupling. The peptidyl-resin was dried under vacuum and the peptide was cleaved from the resin by using 2.5 mL of the reagent cocktail: trifluoroacetic acid (TFA, 81 %), phenol (5 %), thioanisol (5 %), water (5.5 %), dimethylsulfide (2 %) and ammonium iodide (1.5 %) for 2 h. The TFA solution containing the peptide was concentrated under vacuum to a minimal volume, added over cold Et<sub>2</sub>O and precipitated by centrifugation. The supernatant was removed and the residue was dissolved in 20 mL of acetonitrile/H<sub>2</sub>O (1:1) 0.1 % TFA (HPLC buffer) and lyophilized. Peptides were used for the assays without further purification.

**BotP proteolysis reactions:** Substrate peptides (100  $\mu$ M) were incubated with M<sup>2+</sup> (100  $\mu$ M) and BotP (5  $\mu$ M) at 37 °C, shaking at 200 rpm. The reactions were analysed by LC-MS at the time intervals specified.

**LC-MS:** HPLC-MS (Thermo Ultimate 3000 RSLC, coupled to a Bruker Daltonics amaZon ESI-MS ion trap instrument) data were obtained in positive ionization mode. Compounds were separated on a Waters Acquity BEH C18 column (50 x 2.1 mm ; 1.7 mm particle diameter) at a flow rate of 600  $\mu$ L min<sup>-1</sup> and 45 °C by a linear gradient using two buffers : A) H<sub>2</sub>O + 0.1 % formic acid (FA) and B) acetonitrile (ACN) + 0.1 % FA. The gradient was initiated by a 0.33 min isocratic step at 5 % B, followed by an increase to 95 % B over 9 min and ending with a 1 min flush step at 95 % B before re-equilibration to the initial conditions. Detection was carried out by both diode array (DAD) and ESI-MS. Extracted ion chromatograms were generated for the expected mass  $\pm$  0.5 m/z. The amount of substrate and the product were analysed by using peak area.

## 2.6 Acknowledgements

We acknowledge use of the Diamond (beamline I04–1) and SLS (beamline X06DA) synchrotrons. J.K. would like to thank the University of St. Andrews, which is supported by a Wellcome Trust Capital Award (086036), the Deutsche Forschungsgemeinschaft for an Emmy Noether fellowship (KO4116/3–1) and Daniel Sauer, Dr. Hilda Sucipto and Eva Luxenburger for help with the biochemical assays and MS analysis. B.N. would like to thank the European Research Council (339367).

## 2.7 References

- [1] 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, et al., *Nat. Prod. Rep.* **2013**, *30*, 108–160.
- [2] T. J. Oman, W. A. van der Donk, *Nat. Chem. Biol.* **2010**, *6*, 9 – 18.
- [3] L. Huo, S. Rachid, M. Stadler, S. C. Wenzel, R. Müller, *Chem. Biol.* **2012**, *19*, 1278–1287.
- [4] W. J. K. Crone, F. J. Leeper, A. W. Truman, *Chem. Sci.* **2012**, *3*, 3516 – 3521.
- [5] W. J. K. Crone, N. M. Vior, J. Santos-Aberturas, L. G. Schmitz, F. J. Leeper, A.W. Truman, *Angew. Chem. Int. Ed.* **2016**, *55*, 9639–9643; *Angew. Chem.* **2016**, *128*, 9791–9795.
- [6] Y. Hou, M. D. B. Tianero, J. C. Kwan, T. P. Wyche, C. R. Michel, G. A. Ellis, E. Vazquez-Rivera, D. R. Braun, W. E. Rose, E. W. Schmidt, et al., *Org. Lett.* **2012**, *14*, 5050–5053.
- [7] J. P. Gomez-Escribano, L. Song, M. J. Bibb, G. L. Challis, *Chem. Sci.* **2012**, *3*, 3522–3525.
- [8] M. Matsui, J. H. Fowler, L. L. Walling, *Biol. Chem.* **2006**, *387*, 1535 – 1544.
- [9] S. Dorus, E. C. Wilkin, T. L. Karr, *BMC Genomics* **2011**, *12*, 177.
- [10] E. Krissinel, K. Henrick, *J. Mol. Biol.* **2007**, *372*, 774–797.
- [11] J.-K. Kim, S. Natarajan, H. Park, K.-H. Huynh, S. H. Lee, J.-G. Kim, Y.-J. Ahn, L.-W. Kang, *J. Microbiol.* **2013**, *51*, 627 – 632.
- [12] H. Kim, W. N. Lipscomb, *Biochemistry* **1993**, *32*, 8465 – 8478.
- [13] A. Kale, T. Pijning, T. Sonke, B. W. Dijkstra, A.-M. W. H. Thunnissen, *J. Mol. Biol.* **2010**, *398*, 703 – 714.
- [14] W. E. Houssen, A. F. Bent, A. R. McEwan, N. Pieiller, J. Tabudravu, J. Koehnke, G. Mann, R. I. Adaba, L. Thomas, U. W. Hawas, et al., *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 14171 – 14174 ; *Angew. Chem.* **2014**, *126*, 14395 – 14398.
- [15] D. E. Ruffner, E. W. Schmidt, J. R. Heemstra, *ACS Synth. Biol.* **2015**, *4*, 482 – 492.
- [16] H. Liu, J. H. Naismith, *Protein Expression Purif.* **2009**, *63*, 102 – 111.
- [17] F. W. Studier, *Protein Expression Purif.* **2005**, *41*, 207 – 234.
- [18] G. Winter, *J. Appl. Crystallogr.* **2010**, *43*, 186–190.
- [19] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J. Appl. Crystallogr.* **2007**, *40*, 658 – 674.
- [20] X. Wang, J. Snoeyink, *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2008**, *5*, 525 – 533.

- [21] P. Emsley, K. Cowtan, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, 60, 2126 – 2132.
- [22] P. D. Adams, P. V. Afonine, G. Bunkóczy, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, et al., *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, 66, 213 – 221.
- [23] G. N. Murshudov, P. Skubník, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, 67, 355–367.
- [24] V. B. Chen, W. B. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, 66, 12–21.
- [25] The PyMOL Molecular Graphics System, Version 1.5.0.4, Schrödinger, LLC, n.d..
- [26] A. C. Wallace, R. A. Laskowski, J. M. Thornton, *Protein Eng.* **1995**, 8, 127 – 134.

## Supporting Information

# Structure and substrate recognition of the Bottromycin maturation enzyme BotP

Previously published in:

Greg Mann<sup>[a]1</sup>, Liujie Huo<sup>[b]1</sup>, **Sebastian Adam**<sup>[c]</sup>, Brunello Nardone<sup>[a]</sup>, Jeremie Vendome<sup>[d],[e]</sup>,  
Nicholas James Westwood<sup>[a]</sup>, Rolf Müller<sup>[f]</sup> and Jesko Koehnke<sup>[c]\*</sup>

**Chembiochem.** 2016 Dec 2;**17**(23):2286-2292.

**DOI: 10.1002/cbic.201600406**

### Affiliation

<sup>[a]</sup> School of Chemistry and Biomedical Sciences Research Centre, University of St. Andrews, North Haugh, St. Andrews, KY16 9ST, UK.

<sup>[b]</sup> Roger Adams Laboratory 156, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, IL, 61801, USA.

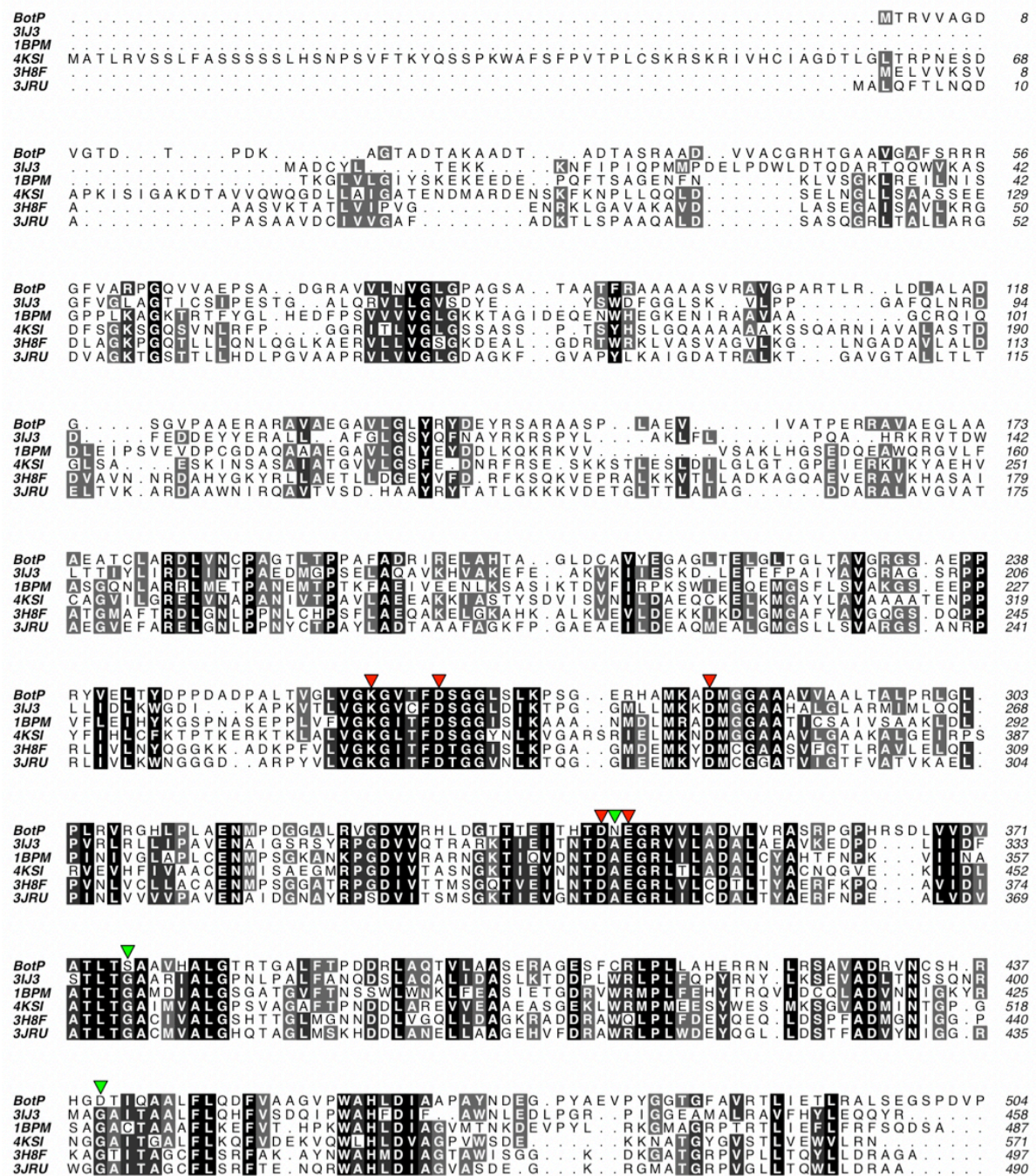
<sup>[c]</sup> Workgroup Structural Biology of Biosynthetic Enzymes, Helmholtz Institute for Pharmaceutical Research, Helmholtz Centre for Infection Research, Saarland University, Universitätscampus E8 1, 66123, Saarbrücken, Germany.

<sup>[d]</sup> Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY, 10032, USA.

<sup>[e]</sup> Howard Hughes Medical Institute, Columbia University, New York, NY, 10032, USA.

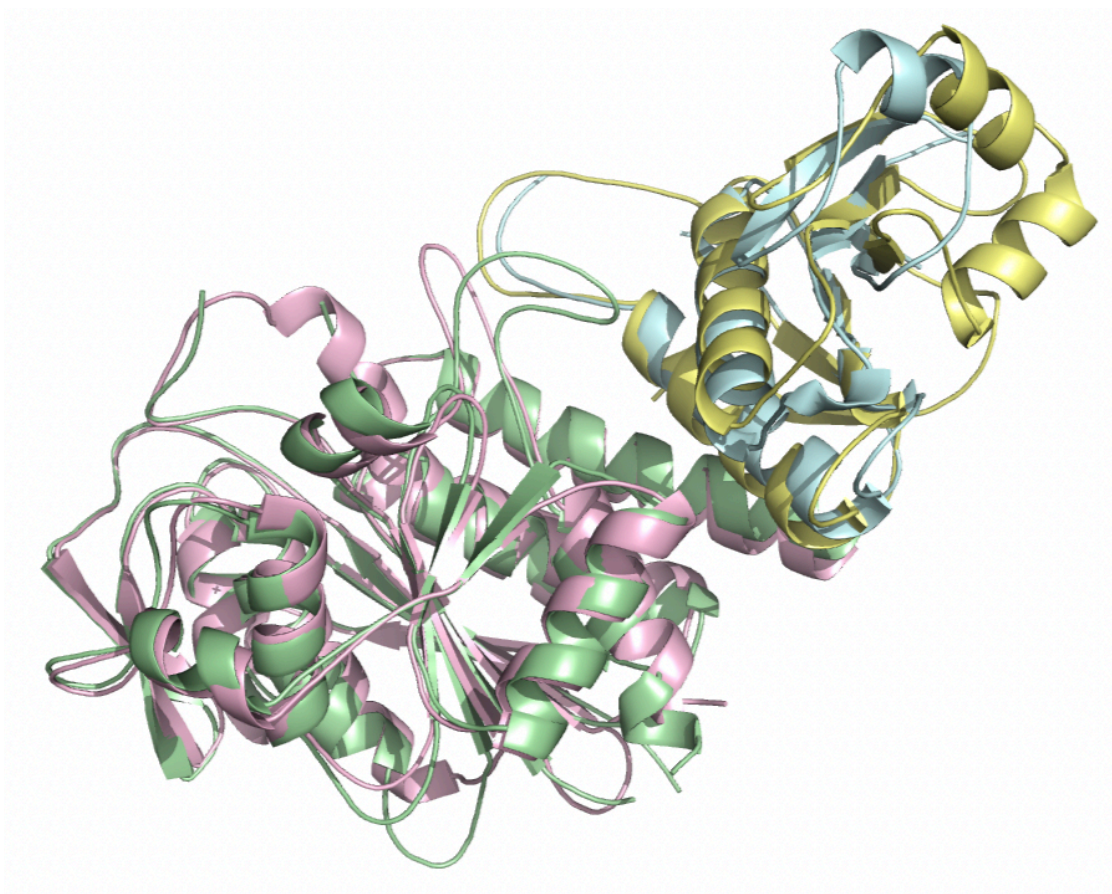
<sup>[f]</sup> Department of Microbial Natural Products, Helmholtz Institute for Pharmaceutical Research, Helmholtz Centre for Infection Research, Saarland University, 66123, Saarbrücken, Germany.

<sup>1</sup> These authors contributed equally to the manuscript

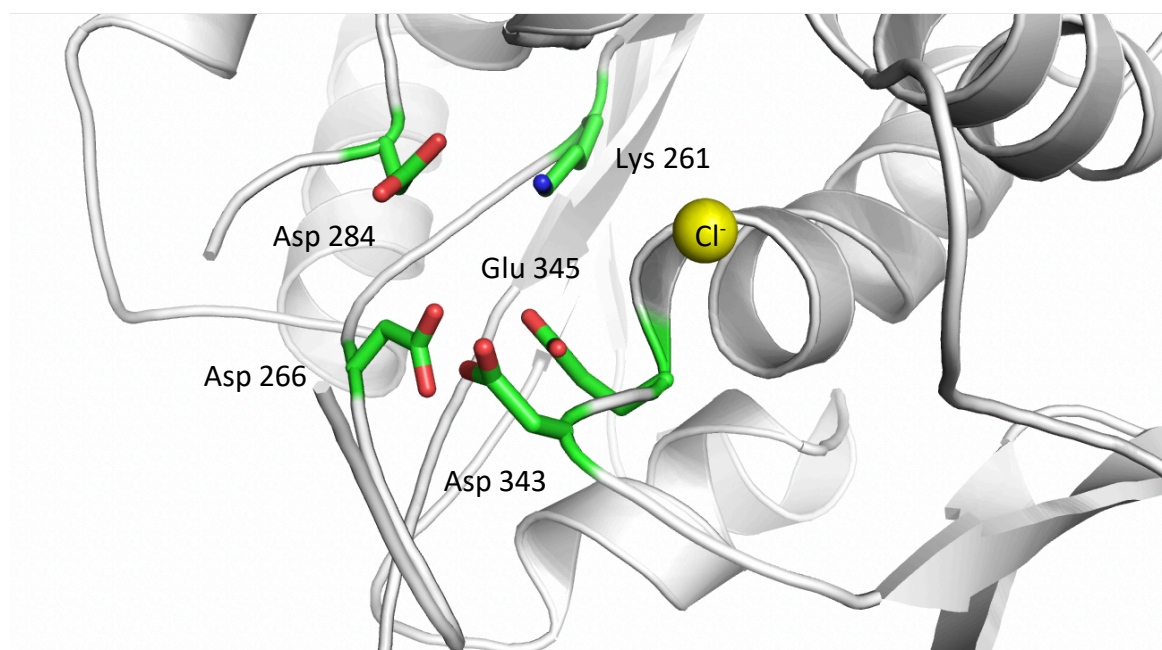


**Figure S1** Sequence alignment of BotP with other leucyl-amino peptidases from the PDB. Conserved catalytically important residues involved in metal ion coordination are highlighted with red triangles. Residues thought to result in a narrow, more restricted substrate binding pocket in BotP compared with other LAPs are highlighted with green triangles. The PDB codes for each of the proteins used in the alignment are shown.

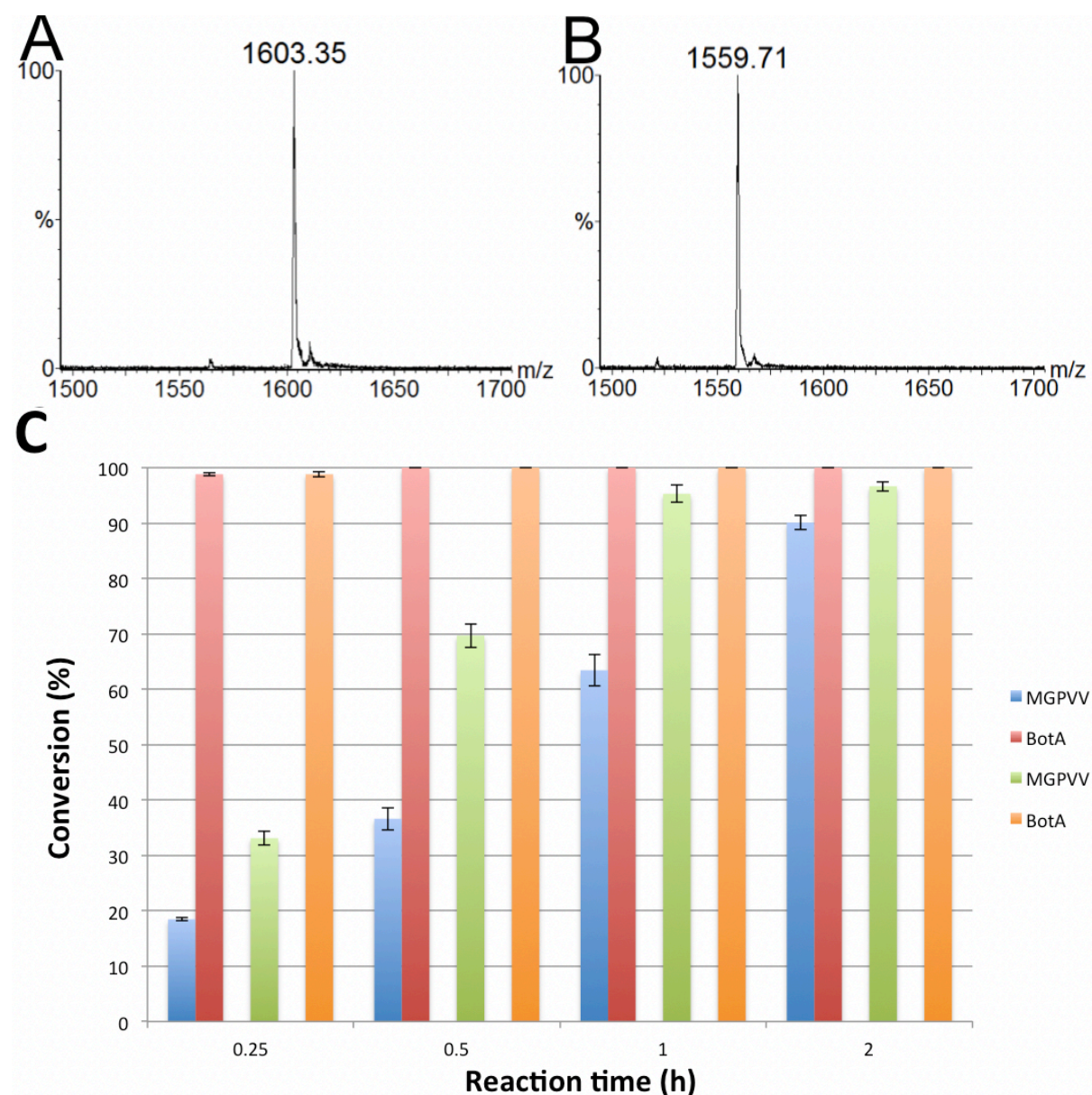




**Figure S2** Superimposition of BotP with XoLAP (PDB ID 3JRU) giving a  $C_{\alpha}$  root mean square deviation of (RMSD) of 1.61 Å over 339 residues.

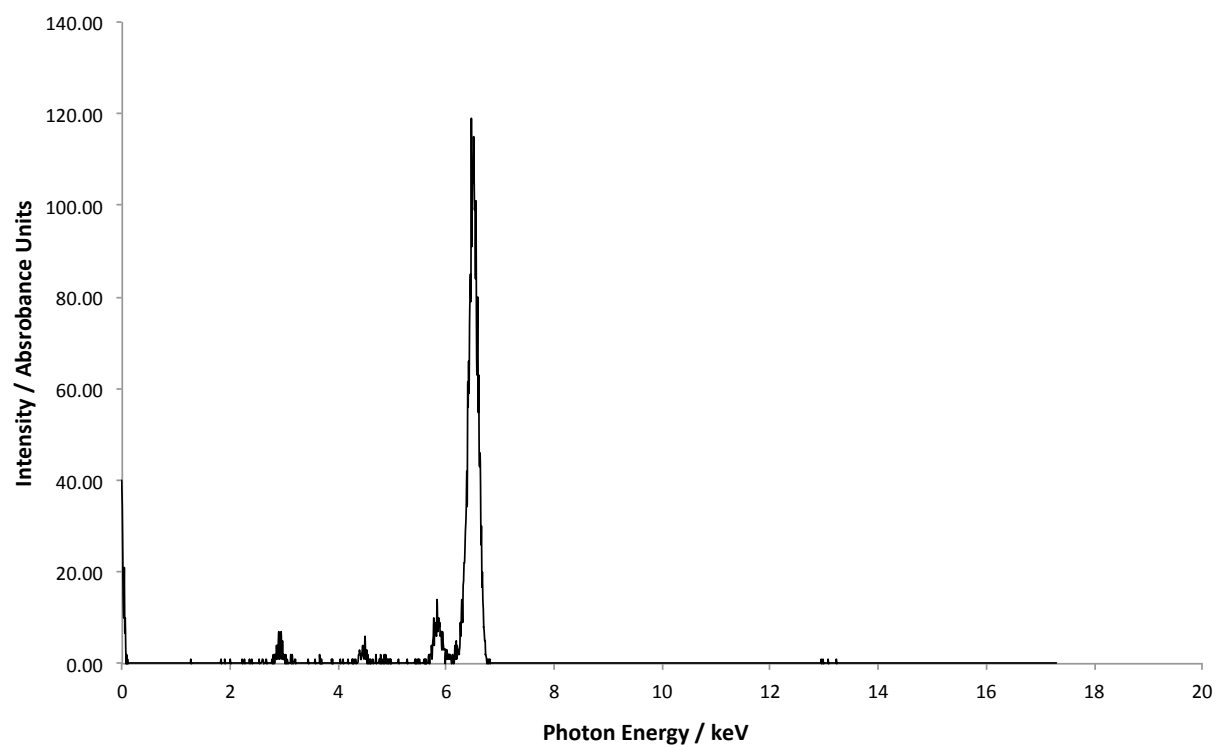


**Figure S3** Apo BotP active site. Conserved metal binding residues are shown as green sticks. BotP is shown as a white cartoon. A chloride ion in place of the carbonate ion is shown as a yellow sphere.

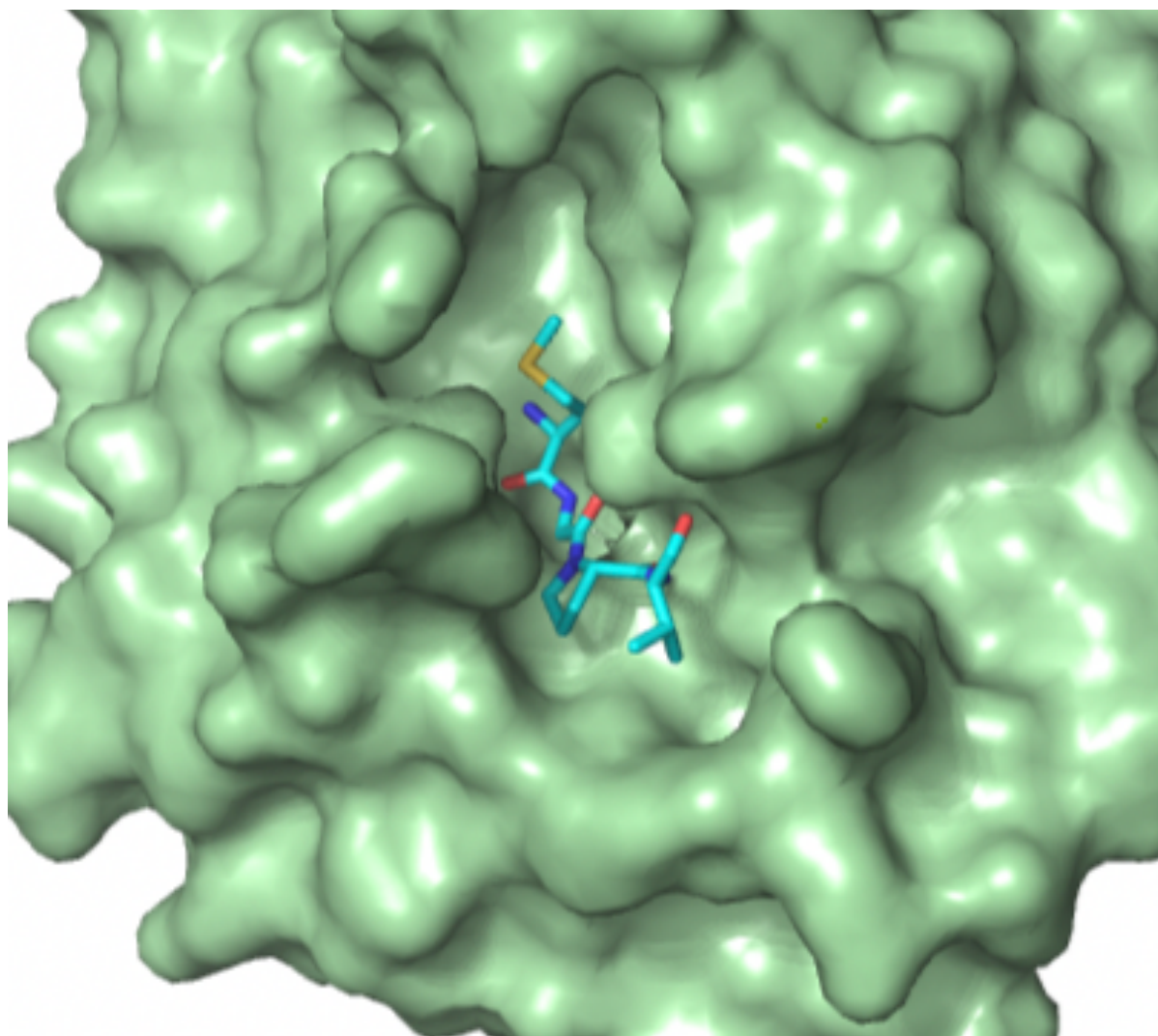


**Figure S4** LC-MS analysis of BotA incubated with A) BotP, B) BotP +  $\text{Co}^{2+}$ . Full-length BotA is detected as a triply ionized species with  $m/z = 1603.35$  (MW = 4808.2 Da). Hydrolysis of the N-terminal methionine by BotP results in a triply ionized species with  $m/z = 1559.71$  (MW = 4677.0 Da). C) Competition assay using BotP- $\text{Co}^{2+}$ , BotA and peptide MGPVV. For each time-point the amount of substrate and product were determined and the % conversion was calculated. Green and orange: Substrates in isolation, blue and red: Substrates co-incubated with enzyme.

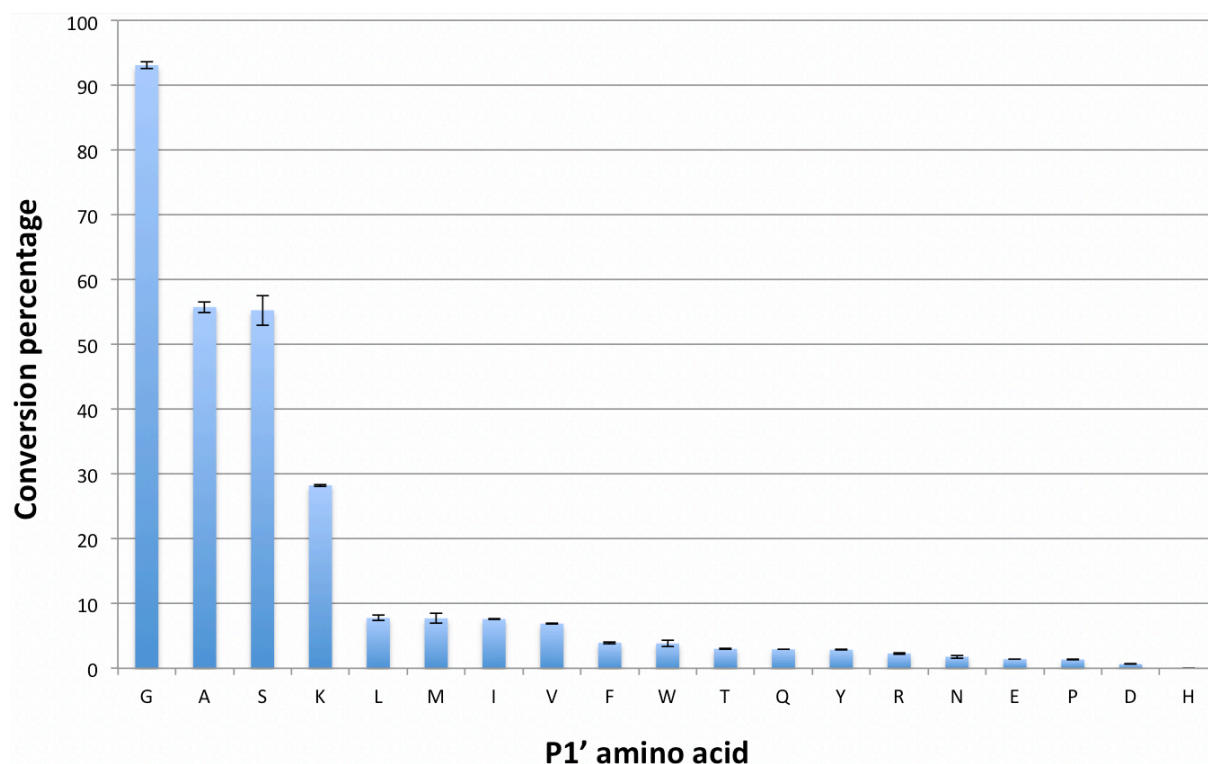




**Figure S5** Fluorescence scan of BotP-Mn<sup>2+</sup> crystal. A peak at 6.54 keV confirms the presence of Mn in the crystal.



**Figure S6** BotP-MGPV model. MGPV peptide is shown in as cyan sticks; BotP is shown as a green surface. A proline residue at position 3 is necessary to elbow the substrate peptide away from clashes with the surface of the protein. The figure was generated using PyMol.



**Figure S7** Relative rate of various test peptides with different amino acids in the P1' position (M-X-PVV) incubated with BotP-Co<sup>2+</sup> to test the substrate specificity of BotP. The reaction was stopped after 3 h at 37 °C and analyzed by LC-MS. For each amino acid the amount of substrate and product were determined the % conversion was calculated.

## Chapter 3

# Macroamidine formation in Bottromycins is catalyzed by a divergent YcaO enzyme

Previously published in:

Laura Franz<sup>[a]1</sup>, **Sebastian Adam**<sup>[a]1</sup>, Javier Santos-Aberturas<sup>[b]</sup>, Andrew W. Truman<sup>[b]</sup> and  
Jesko Koehnke<sup>[a]\*</sup>

**J Am Chem Soc.** 2017 Dec 20;**139**(50):18158-18161.

**DOI: 10.1021/jacs.7b09898**

### Affiliation

<sup>[a]</sup> Workgroup Structural Biology of Biosynthetic Enzymes, Helmholtz Institute for Pharmaceutical Research, Helmholtz Centre for Infection Research, Saarland University, Universitätscampus E8 1, 66123, Saarbrücken, Germany.

<sup>[b]</sup> Department of Molecular Microbiology, John Innes Centre, Colney Lane, Norwich NR4 7UH, United Kingdom.

<sup>1</sup> These authors contributed equally to the manuscript

## Contributions and Acknowledgements

### Author's effort:

The author contributed significantly to the manuscript by establishing the first reliable protein expression and purification protocols for the two featured YcaO proteins which had not been expressed in a soluble form beforehand. The author performed the cloning, expression and purification for the IpoC and PurCD point mutants and helped with expression and purification of the BotA point mutants. Furthermore, the author evaluated biochemical data, designed figures and edited the final manuscript.

### Other's effort:

Laura Franz contributed significantly by performing biochemical experiments and evaluating the resulting data qualitatively as well as quantitatively. She performed the thermal shift assays and helped with the cloning, expression and purification of the BotA point mutants. Furthermore, she designed figures and edited the final manuscript. Javier Santos-Aberturas contributed by designing, performing and evaluating the featured *in vivo* experiments as well as the MS/MS data.

### 3.1 Abstract

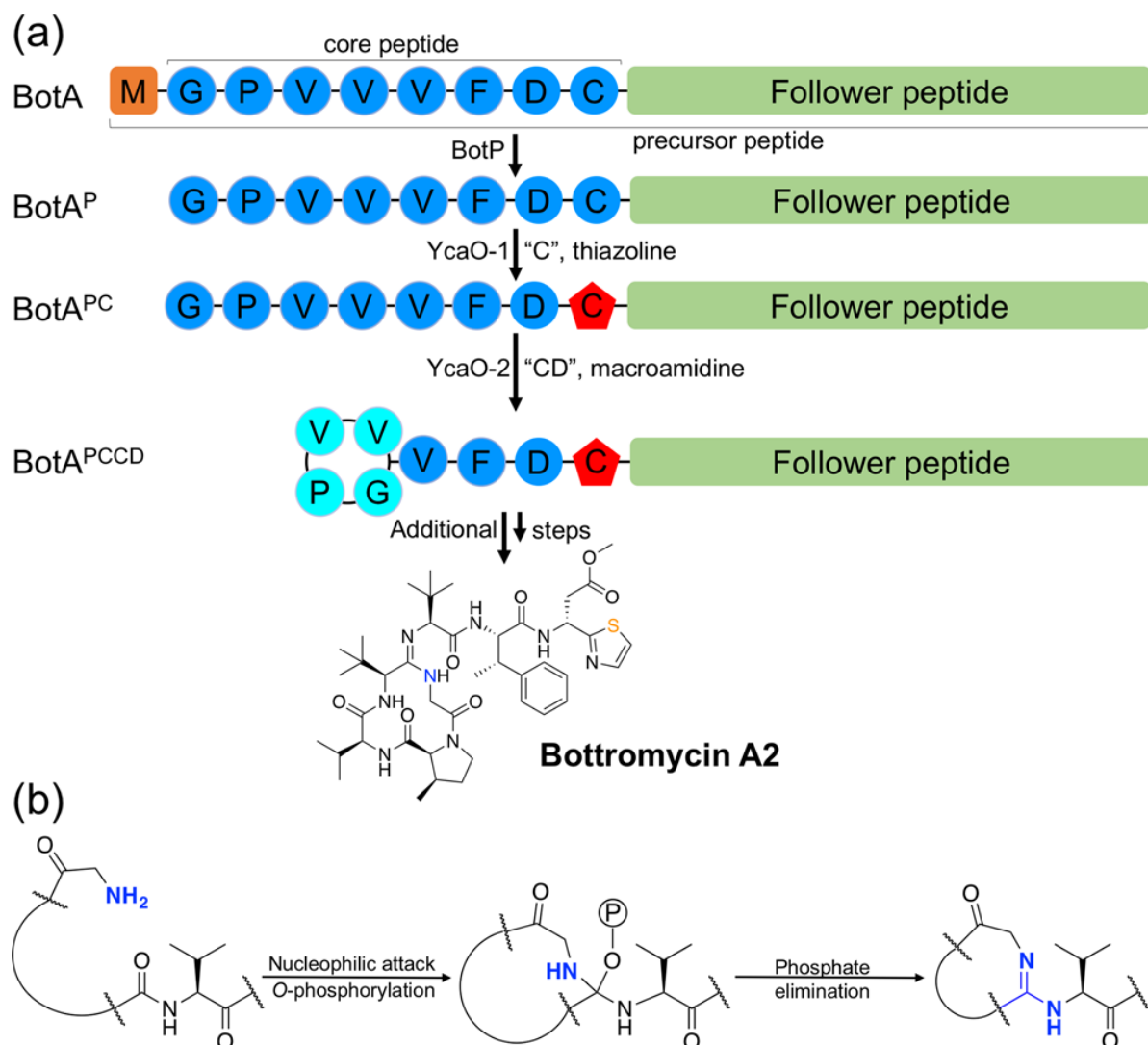
The YcaO superfamily of proteins catalyzes the phosphorylation of peptide backbone amide bonds, which leads to the formation of azolines and azoles in ribosomally synthesized and post-translationally modified peptides (RiPPs). Bottromycins are RiPPs with potent antimicrobial activity, and their biosynthetic pathway contains two divergent, stand-alone YcaO enzymes, IpoC and PurCD. From an untargeted metabolomics approach, it had been suggested that PurCD acts with a partner protein to form the 12-membered macroamidine unique to bottromycins. Here we report the biochemical characterization of IpoC and PurCD. We demonstrate that IpoC installs a cysteine-derived thiazoline, whereas PurCD alone is sufficient to create the macroamidine structure. Both enzymes are catalytically promiscuous, and we generated 10 different macroamidines. Our data provide important insights into the versatility of YcaO enzymes, their ability to utilize different nucleophiles and provide a framework for the creation of novel bottromycin derivatives with enhanced bioactivity.

### 3.2 Main text

Bottromycins (**Scheme 1**) are ribosomally synthesized and post-translationally modified peptide (RiPP) natural products that bind to the A site of the 50S ribosome, thus inhibiting prokaryotic protein synthesis and represent a novel class of antibiotics.<sup>[1]</sup> They are derived from the precursor peptide BotA, which is unique among the RiPPs family in that it has a C-terminal “follower” rather than the canonical N-terminal “leader” peptide.<sup>[2-5]</sup> BotA is tailored by a series of enzymatic (and possibly non-enzymatic) chemical transformations and their order has recently been proposed based on data from an untargeted metabolomics approach (**Scheme 1**).<sup>[6]</sup> The bottromycin biosynthetic gene cluster contains two unusual, stand-alone YcaO-domain enzymes, BotC and BotCD (sequence identity of 26%, **Figure S2**).

YcaO-domain proteins were enigmatic until they were linked to the formation of azol(in)e heterocycles in the biosynthesis of microcin B17.<sup>[7,8]</sup> Biosynthetic pathways responsible for the production of RiPPs frequently contain enzymes with a YcaO domain, which catalyze azoline formation in an ATP-dependent cyclodehydration reaction.<sup>[9]</sup>

**Scheme 1** (a) Bottromycin Biosynthesis with the Proposed Functions of the Two YcaO Enzymes; (b) Use of a Terminal Amine by a YcaO Enzyme as a Nucleophile to Achieve Macroamide Formation.

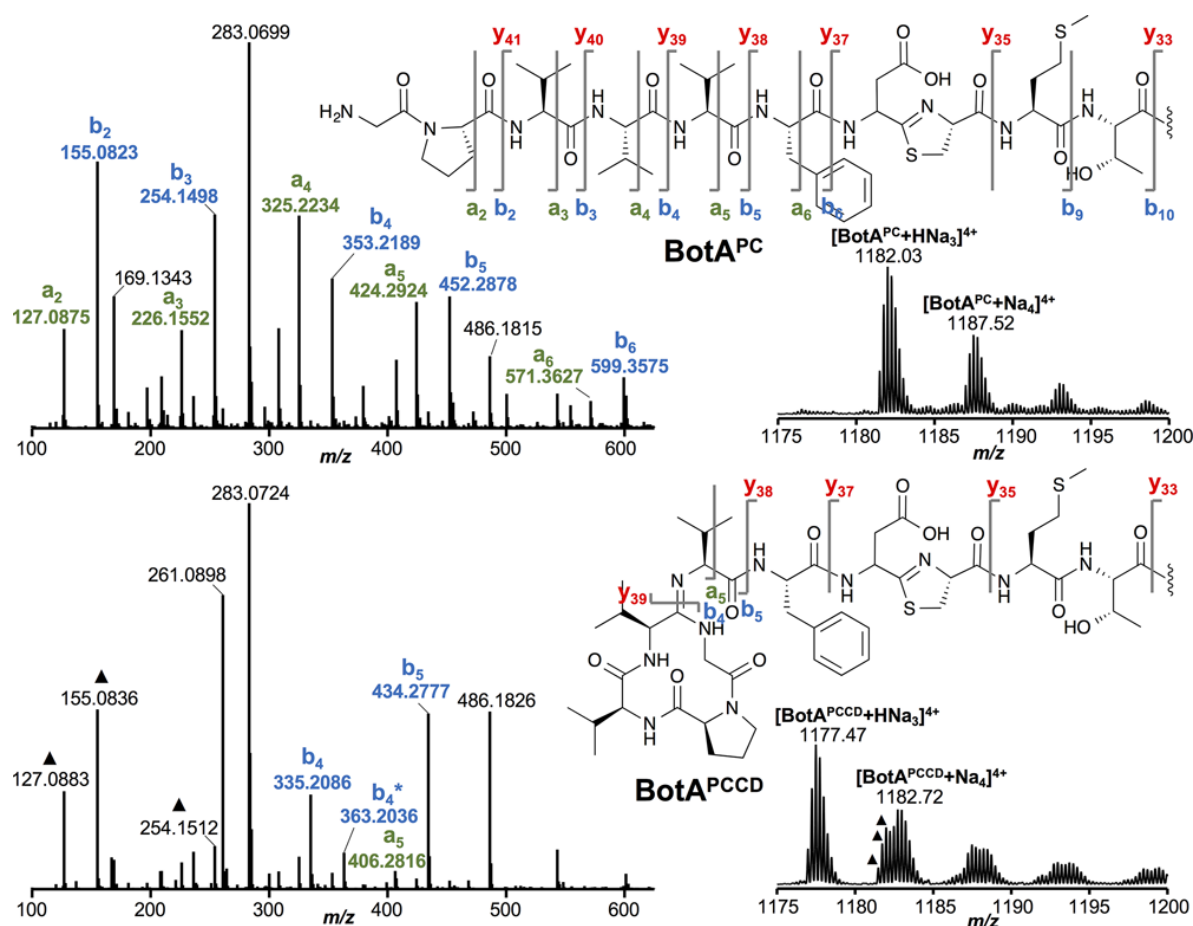


Vital for this process in all RiPPs systems studied to date is the presence of RiPP recognition element (RRE) domains.<sup>[10]</sup> Without the RRE domain of a specific system, which can either be present on a stand-alone E1 ubiquitin activating-like superfamily protein or part of a multidomain heterocyclase, the heterocyclase activity is minimal.<sup>[11]</sup> The YcaO domain of the responsible enzyme catalyzes the nucleophilic attack of a Cys/Ser/Thr side-chain on the preceding backbone carbonyl to form a hemiorthoamide.<sup>[12]</sup> This intermediate is then O-phosphorylated<sup>[12]</sup> (or pyrophosphorylated<sup>[13]</sup>) followed by an elimination reaction to yield the azoline (**Figure S1a**). It has been proposed (and one example reported<sup>[14]</sup>) that other nucleophiles may also be used by this enzyme family. Despite the presence of two YcaO enzymes, bottromycins contain only one azole, neither YcaO possesses a RRE and the gene cluster encodes no E1-like protein. An untargeted metabolomic approach using mass spectral

networking identified the order of reactions in bottromycin biosynthesis.<sup>[6]</sup> This study revealed that BotC was responsible for the formation of the thiazoline, whereas BotCD (in cooperation with a hydrolase) catalyzed the macrocyclization reaction to yield the macroamidine.<sup>[6]</sup> Macroamidine formation can be viewed as analogous to azoline formation, with the amino-terminus of the precursor peptide BotA functioning as the nucleophile (**Scheme 1b** and **Figure S1b**).

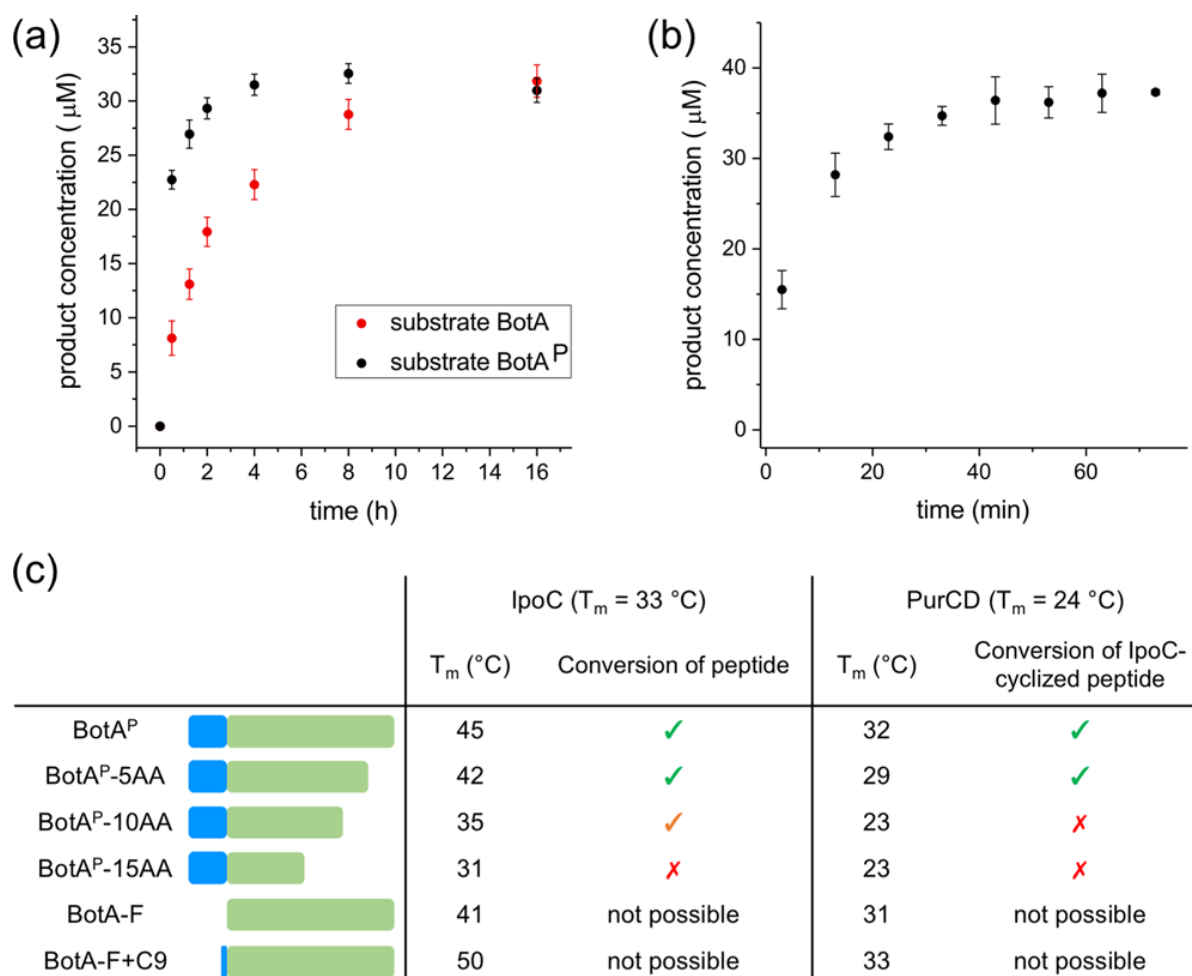
To investigate this potentially novel YcaO-domain catalysis, we sought to characterize BotC and BotCD in vitro, but neither protein (from *Streptomyces* sp. BC16019) could be expressed in a soluble form. In contrast, the homologues IpoC (76% sequence identity to BotC) from *Streptomyces ipomoeae* and PurCD (81% sequence identity to BotCD) from *Streptomyces purpureus* were tractable. Both proteins originate from gene clusters that are homologous and highly similar to those reported to produce bottromycins (**Figure S3**). When BotA was incubated with IpoC and ATP/Mg<sup>2+</sup>, we observed a loss of 18 Da consistent with cyclodehydration (**Figure S4a**), and the conversion of Cys9 to thiazoline was confirmed by tandem MS (MS<sup>2+</sup>) (**Figure S4b** and **Table S2**). We treated BotA with BotP (BotA<sup>P</sup>) to make the glycine amino group available as a nucleophile for macrocyclization and repeated the experiments.<sup>[15]</sup> The reaction with IpoC also resulted in thiazoline formation at Cys9 (**Figures 1** and **S5, Table S4**). For PurCD incubated with BotA<sup>P</sup>, we also observed a loss of 18 Da, but the yield was too low for MS<sup>2</sup> analysis (**Figure S5a**). On the basis of metabolomic data, we expected the heterocyclized intermediate (BotA<sup>PC</sup>) to be a better PurCD substrate.<sup>[6]</sup> When we reacted purified BotA<sup>PC</sup> with PurCD and ATP/Mg<sup>2+</sup>, we observed macroamidine formation (BotA<sup>PCCD</sup>), which was confirmed by MS<sup>2</sup>, where characteristic<sup>[6]</sup> macroamidine fragments of *m/z* 335.21, 363.20, and 434.28 were observed (**Figures 1** and **S6, Table S5**). Neither BotA nor heterocyclized BotA<sup>C</sup> were substrates for PurCD. These data established IpoC as the heterocyclase, whereas PurCD functions as the macrocyclase. PurCD is the first reported case of a YcaO enzyme able to (a) utilize nitrogen as a nucleophile and (b) form larger ring systems. Our data indicates that for efficient catalysis removal of the N-terminal methionine precedes heterocyclization, which is followed by macrocyclization.





**Figure 1.** MS<sup>2</sup> characterization of BotA<sup>P</sup> after incubation with IpoC and BotA<sup>PC</sup> after incubation with PurCD. Full details of the MS<sup>2</sup> fragmentation are provided in **Figures S5** and **S6**, **Tables S4** and **S5**. Signals in the BotA<sup>PCCD</sup> spectra that likely come from residual BotA<sup>PC</sup> are labeled with triangles.

A detailed analysis of the IpoC-catalyzed heterocyclization showed that 50  $\mu$ M BotA<sup>P</sup> required 8 h to reach completion when it was incubated with 5  $\mu$ M IpoC and 5 mM ATP/Mg<sup>2+</sup> at 37 °C (**Figure 2a**). In contrast, the reaction using unprocessed precursor peptide BotA required 16 h (**Figure 2a**). The heterocyclization rate was calculated to be 0.1 heterocycles per enzyme per minute with BotA<sup>P</sup>, whereas the reaction using BotA was roughly half this rate (**Figure S7**), indicating that although IpoC can process both versions of the precursor peptide, BotA<sup>P</sup> is preferred. Changes in the reaction conditions (temperature, pH, salt concentration) did not lead to accelerated heterocycle formation (data not shown).



**Figure 2.** (a) Time-course of IpoC-catalyzed heterocyclization of BotA (red) and BotA<sup>P</sup> (black) at pH 7.4. (b) Time-course of PurCD-catalyzed macroamidine formation at pH 9.5. (c) Effects of BotA<sup>P</sup> truncations and follower peptide on IpoC and PurCD stability (pH 8.5) and turnover (pH 7.4 and 9.5, respectively).

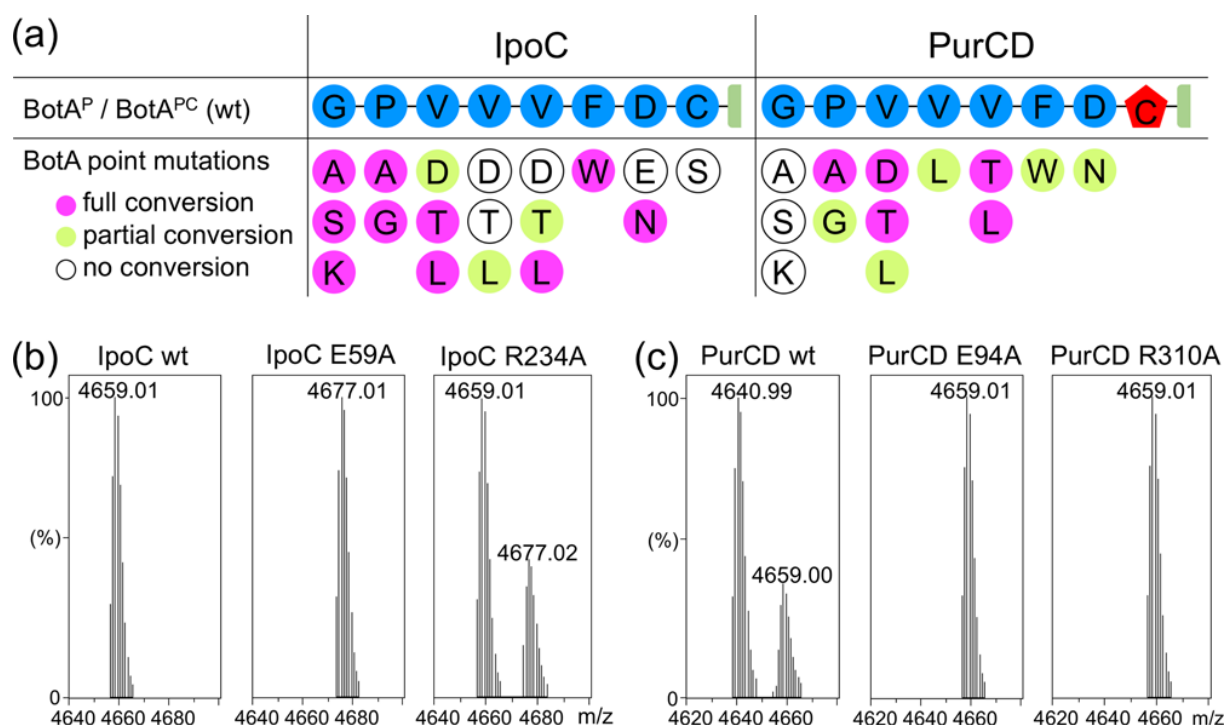
The macrocyclization reaction of PurCD (10 μM) with BotA<sup>PC</sup> (50 μM) was approximately 15 % complete after 90 min at pH 7.5 (**Figure S8a**). We hypothesized that the reaction may be accelerated by increasing the pH of the reaction as increasing the pH results in a larger proportion of deprotonated substrate N-terminus, which is then available as the nucleophile for macroamidine formation. We analyzed PurCD activity between pH 7 and 11, which is more than 1 pH unit above the  $pK_a$  of the amino group of glycine ( $pK_a$  9.6), so most of the substrate should be available in its neutral (nucleophilic) form. We observed a steady increase in product formation with a maximum at pH 9.5 (**Figure S8b**), with ~85 % turnover after 1 h (**Figures 2b** and **S8c**). We wondered if the decrease in enzyme activity above pH 9.5 was due to enzyme degradation. When we analyzed PurCD stability by measuring its melting temperature ( $T_m$ ) in thermal shift assays (TSAs), we observed a steady decrease of enzyme stability with increasing pH and were unable to record a  $T_m$  at pH 11 (**Figure S9a**). We therefore

postulate that pH 9.5 represents the best balance between substrate availability and enzyme stability. Interestingly, we observed the enzyme and ATP-dependent reopening of the macroamidine after extended incubations of the reactions, which was accelerated at lower pH (**Figure S8**). After 12 h, no BotA<sup>PCD</sup> could be observed in the samples. It appears possible that the hydrolase identified as essential for macroamidine formation in vivo<sup>[6]</sup> influences this reaction's equilibrium and thus prevents reopening of the ring. IpoC and PurCD seem to be tractable for the use in in vitro systems, but the surprising enzyme-induced hydrolysis of the macroamidine by PurCD must be taken into account in such systems to accomplish satisfactory yields.

A hallmark of YcaO domains characterized to date is their requirement to bind their substrate's leader peptide for efficient catalysis via an RRE domain. IpoC and PurCD do not have domains with homology to known RREs, so we investigated whether they were able to bind the follower peptide of BotA. Based on published precursor peptide–enzyme structures, binding to the leader appears to stabilize the corresponding enzyme.<sup>[13,16]</sup> This feature would allow TSAs to be used as a convenient tool to probe these critical interactions. In TSAs, one can measure the  $T_m$  of a protein in the absence and presence of ligands and stabilizing interactions between protein and ligand lead to an increase in  $T_m$ . IpoC and PurCD had  $T_m$  values of 33 and 24 °C, respectively, which increased dramatically after the addition of the substrate BotA<sup>P</sup>/BotA<sup>PC</sup> (**Figure S9b** and **Table S6**). To verify that this effect was not exerted by the core peptide, we repeated the experiment using follower peptide BotA-F. The effect of the follower on IpoC was almost as pronounced as for BotA<sup>P</sup> (**Figure S10a**). In the case of PurCD, the differences between BotA<sup>PC</sup> and BotA-F were within the experimental error (**Figures 2c** and **S10b**, **Table S6**). We asked if the presence of the substrate cysteine had an effect on IpoC stability and observed a drastic increase in  $T_m$  for IpoC + BotA-F with an additional N-terminal cysteine (**Figures 2c** and **S10a**, **Table S6**). To identify the follower region responsible for binding, we produced a series of systematic truncations (**Figures 2c** and **S10**, **Tables S7** and **S8**). BotA<sup>P</sup> with the 5 C-terminal residues removed (BotA<sup>P</sup>-5AA) had a slightly less stabilizing effect on IpoC and PurCD when compared to BotA<sup>P</sup> and was processed with similar efficiency. Removal of the next 5 residues (BotA<sup>P</sup>-10AA) abolished the stabilizing effect on both enzymes, severely hampered processing by IpoC and abolished PurCD activity. The next truncation, BotA<sup>P</sup>-15AA, had no stabilizing effect on either enzyme and was not processed. These data indicated that the region between residues 30 and 39 is critical for the

interaction between BotA and both enzymes. We produced a series of point mutants covering this area (**Figure S11**, **Tables S7** and **S8**), which all showed a reduced stabilizing effect in TSAs, with E31R and W35A reducing the effect to within experimental error. However, all point mutants could be processed by both enzymes (**Figure S12**). Only BotA<sup>P</sup>E28R, a mutation slightly outside the suggested range, abolished stabilization and processing, highlighting the importance of this residue. The observation that impairment of substrate binding (implied via  $T_m$  values) does not abolish processing has been observed for other YcaO enzymes.<sup>[13]</sup> A deeper understanding will require the determination of C and CD crystal structures in complex with BotA-F. However, we have demonstrated that the stabilizing effect of the follower or leader peptides observed in TSAs may be used to group tailoring enzymes into early (requiring leader/follower) and late (acting after leader/follower removal) enzymes, to reduce the complexity of reconstituting RiPPs pathways in vitro.

The YcaO enzymes studied to date have shown a great tolerance for changes in their respective substrates<sup>[9]</sup> and we wondered if IpoC and PurCD would also display catalytic promiscuity. To this end we tested a set of 18 core peptide mutants (**Figure S13**, **Tables S7** and **S8**). IpoC was unable to produce oxazoline from a C9S mutant, but processed 13 other point mutants (**Figures 3a** and **S14**). The most sensitive position appeared to be V5. The 13 heterocyclized peptides were tested for reactivity with PurCD. We observed macro- amidine formation for 10 of these (**Figures 3a** and **S15**, **Table S8**) with varying conversion rates. Our data indicates that Gly2 cannot be modified, and the requirement of a glycine at this position was likely, because aside from D-amino acids<sup>[17]</sup> it introduces the least strain into the ring system and is the nucleophile in macrocyclization.



**Figure 3.** (a) Tolerance of IpoC and PurCD to changes in the core peptide of BotA. Conversion rate given in relation to wt substrate. Effect of (b) IpoC and (c) PurCD ATP/Mg<sup>2+</sup> binding residue mutations on enzyme activity.

To date, only two YcaO domain proteins have been crystallized in complex with a nucleotide cofactor. Both are very distantly related to IpoC and PurCD (pairwise protein sequence identity <20%), but sequence alignment nevertheless revealed that key ATP and Mg<sup>2+</sup> binding residues are mostly conserved (**Figure S16**). To explore IpoC and PurCD ATP binding residues, we designed a series of point mutants of both enzymes based on the sequence alignments (**Figure S16**, **Tables S7** and **S8**). In IpoC, we mutated the putative Mg<sup>2+</sup> binding residue E59 to A or R, whereas we swapped the putative ATP (phosphate) binding residue R234 to A or E. All four mutants could be expressed, purified and stabilized by BotA<sup>P</sup> (TSAs), indicating correct protein folding (**Figure S17** and **Table S6**). The stabilizing effect of ATP/Mg<sup>2+</sup> remained as observed for wild-type IpoC, within experimental error. Mutations of E59 reduced activity to less than 5% of wt enzyme, while mutations of R234 reduced enzymatic activity (**Figure 3b** and **S18**). Equivalent *in vivo* mutants (**Table S9**) were consistent with this *in vitro* data. In PurCD we selected E94 (to A or R) and R310 (to A or E) for mutations. Here, only the conservative alanine substitutions could be expressed. Both mutant proteins were stabilized by BotA<sup>P</sup> and ATP/Mg<sup>2+</sup> as observed for wt PurCD, but were unable to process substrate (**Figures 3c**, **S19** and **S20**, **Tables S7** and **S8**). These deleterious *in vitro* results for PurCD were confirmed when the same mutations were tested *in vivo* (**Table S9**), which all provided the same metabolite

profile as a gene deletion. The conservation of key ATP/Mg<sup>2+</sup> binding residues in highly divergent enzymes from this family may be used to validate predicted YcaO domain enzymes *in vitro* or *in vivo* and offers a way to selectively abolish enzyme function without disrupting protein folding. We have demonstrated that the catalytic repertoire of YcaO enzymes also includes the formation of macroamidines. This enzyme family has also been implied in the formation of thioamide bonds,<sup>[18]</sup> and it will be fascinating to discover their true chemical scope. The promiscuity of these two key biosynthetic enzymes *in vitro* is an important cornerstone toward the development of an efficient system to generate bottromycin variants with enhanced stability and activity.

### 3.3 Acknowledgements

J.K. thanks the DFG for an Emmy Noether Fellowship (KO 4116/3-1). This work was also supported by BBSRC grant BB/ M003140/1 (A.W.T. and J.S.-A.) and a Royal Society University Research Fellowship (A.W.T.). We thank Dr. David Auerbach, Dr. Gerhard Saalbach and Daniel Sauer for assistance with the LC-MS measurements. We are grateful to Dr. Rafael Guimaraes da Silva and Prof. Rolf Müller for helpful discussions.

## 3.4 References

- [1] Otaka, T.; Kaji, A. *J. Biol. Chem.* **1976**, *251* (8), 2299–2306.
- [2] Crone, W. J. K.; Leeper, F. J.; Truman, A. W. *Chemical Science* **2012**, *3* (12), 3516–3521.
- [3] Gomez-Escribano, J. P.; Song, L.; Bibb, M. J.; Challis, G. L. *Chemical Science* **2012**, *3* (12), 3522–3525.
- [4] Hou, Y.; Tianero, M. D.; Kwan, J. C.; Wyche, T. P.; Michel, C. R.; Ellis, G. A.; Vazquez-Rivera, E.; Braun, D. R.; Rose, W. E.; Schmidt, E. W.; Bugni, T. S. *Org. Lett.* **2012**, *14* (19), 5050–3.
- [5] Huo, L.; Rachid, S.; Stadler, M.; Wenzel, S. C.; Muller, R. *Chem. Biol.* **2012**, *19* (10), 1278–87.
- [6] Crone, W. J.; Vior, N. M.; Santos-Aberturas, J.; Schmitz, L. G.; Leeper, F. J.; Truman, A. W. *Angew. Chem., Int. Ed.* **2016**, *55* (33), 9639–43.
- [7] Li, Y. M.; Milne, J. C.; Madison, L. L.; Kolter, R.; Walsh, C. T. *Science* **1996**, *274* (5290), 1188–93.
- [8] Bayer, A.; Freund, S.; Nicholson, G.; Jung, G. *Angew. Chem., Int. Ed. Engl.* **1993**, *32* (9), 1336–1339.
- [9] Burkhardt, B. J.; Schwalen, C. J.; Mann, G.; Naismith, J. H.; Mitchell, D. A. *Chem. Rev.* **2017**, *117* (8), 5389–5456.
- [10] Dunbar, K. L.; Tietz, J. I.; Cox, C. L.; Burkhardt, B. J.; Mitchell, D. A. *J. Am. Chem. Soc.* **2015**, *137* (24), 7672–7.
- [11] Dunbar, K. L.; Chekan, J. R.; Cox, C. L.; Burkhardt, B. J.; Nair, S. K.; Mitchell, D. A. *Nat. Chem. Biol.* **2014**, *10* (10), 823–9.
- [12] Dunbar, K. L.; Melby, J. O.; Mitchell, D. A. *Nat. Chem. Biol.* **2012**, *8* (6), 569–75.
- [13] Koehnke, J.; Mann, G.; Bent, A. F.; Ludewig, H.; Shirran, S.; Botting, C.; Lebl, T.; Houssen, W.; Jaspars, M.; Naismith, J. H. *Nat. Chem. Biol.* **2015**, *11* (8), 558–563.
- [14] Koehnke, J.; Morawitz, F.; Bent, A. F.; Houssen, W. E.; Shirran, S. L.; Fuszard, M. A.; Smellie, I. A.; Botting, C. H.; Smith, M. C.; Jaspars, M.; Naismith, J. H. *ChemBioChem* **2013**, *14* (5), 564–7.
- [15] Mann, G.; Huo, L.; Adam, S.; Nardone, B.; Vendome, J.; Westwood, N. J.; Muller, R.; Koehnke, J. *ChemBioChem* **2016**, *17* (23), 2286–2292.
- [16] Ortega, M. A.; Hao, Y.; Zhang, Q.; Walker, M. C.; van der Donk, W. A.; Nair, S. K. *Nature* **2015**, *517* (7535), 509–512.
- [17] Yongye, A. B.; Li, Y.; Giulianotti, M. A.; Yu, Y.; Houghten, R. A.; Martinez-Mayorga, K. J. *Comput.-Aided Mol. Des.* **2009**, *23* (9), 677–89.
- [18] Izawa, M.; Kawasaki, T.; Hayakawa, Y. *Appl. Environ. Microbiol.* **2013**, *79* (22), 7110–3.

## Supporting Information

# Macroamidine formation in Bottromycins is catalyzed by a divergent YcaO enzyme

Previously published in:

Laura Franz<sup>[a]1</sup>, **Sebastian Adam**<sup>[a]1</sup>, Javier Santos-Aberturas<sup>[b]</sup>, Andrew W. Truman<sup>[b]</sup> and  
Jesko Koehnke<sup>[a]\*</sup>

**J Am Chem Soc.** 2017 Dec 20;**139**(50):18158-18161.

**DOI: 10.1021/jacs.7b09898**

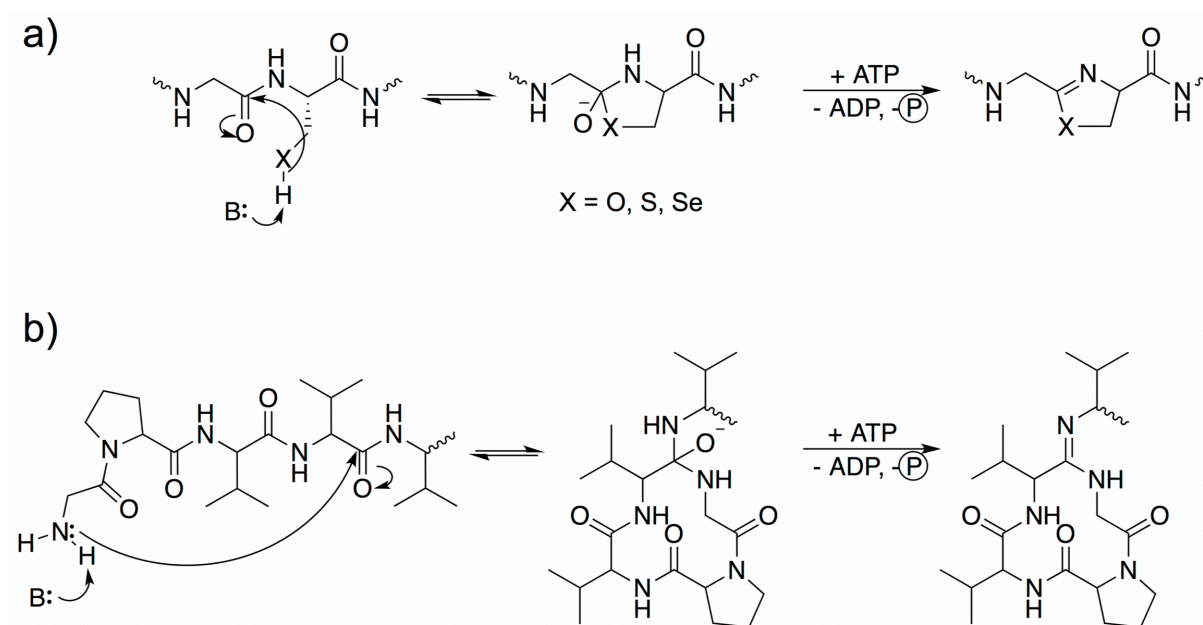
### Affiliation

<sup>[a]</sup> Workgroup Structural Biology of Biosynthetic Enzymes, Helmholtz Institute for Pharmaceutical Research, Helmholtz Centre for Infection Research, Saarland University, Universitäts-campus E8 1, 66123, Saarbrücken, Germany.

<sup>[b]</sup> Department of Molecular Microbiology, John Innes Centre, Colney Lane, Norwich NR4 7UH, United Kingdom.

<sup>1</sup> These authors contributed equally to the manuscript

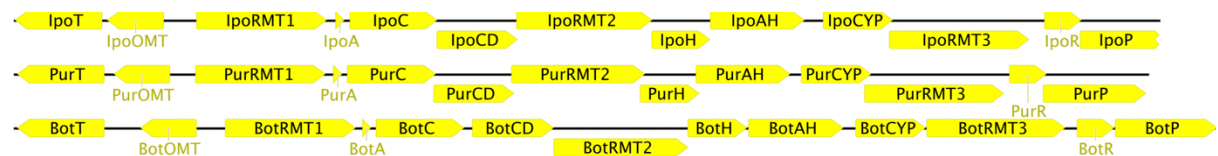




**Figure S1:** a) Proposed mechanism of heterocyclization by YcaO domain enzymes. ATP is used to phosphorylate the amide oxygen. Phosphate is eliminated from the hemiorthoamide intermediate afterwards and the azoline is formed. Nucleophiles reported to be utilized to date are O, S and Se. b) Possible route to macroamidine formation using N as a nucleophile in a reaction analogous to a).

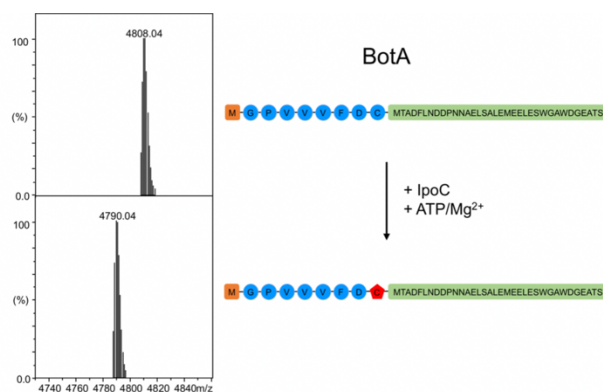
BotC	--MLEATATECELR---EVVHRSY-----PSDRTVTVRCTVRPAEGTAQ-ADGYGTATTE	49
BotCD	MRLTDAEGALCSRSSGKGLGHRSMASALFESAHEYHLDWRRDPRASEAEFLPGREIAG-Q	59
	: * . : * . : * * * * . : * . * : * * :	
BotC	AVARAKALSEAVERLVACTPFATV-ARPPTTPAPGS-GSGP-----VPPFP-AAGVRTA	100
BotCD	DTARRSALLRRLGGIMPDAFVLTRTYRPVEETLTGRADHGPAGRPHRHPVFLRDGGYRNW	119
	. * * . * * . : : : : * . * * * * . * * * * . * * .	
BotC	PDGCASREYRPLSGDGPRRVPLYWSSPWAGEELRAGTLSAAEARLSSTVGWAVAPTPEA	160
BotCD	PHPDDDR-----SFQPLWHYT-----SSAGYAAGATVHE	148
	* . . * . * . * . * . * . * . * . * . * .	
BotC	ALRGALLELTELDYGVFLHRRLAGPAR----PRSAGD-----ET	196
BotCD	ALVHAVNELIERDAWSYQLARSYFGLPGEGPELRVVDHDTLPaelRELtGTVEEVREAPV	208
	** * : ** * : . * * * * * . . . .	
BotC	LVVPLGGAVRTPTVLAVAYGRGRIMPATGLGCGASRAEATDRALLELAQAETMWRSNPTA	256
BotCD	LIVDITCDTGVPAYVCDARSREEVRLIGSGASPVRTYALQRALLEYLQVRTMFEHGPVD	268
	* : * : . . * : : . : * * . . * : * : * * * * * * . * * : . * .	
BotC	EPAERF-FLRRFERWPLLTRCATLDFELSDLSGLDGLSGLDGLSRLDGLSRLDHPSDLND	315
BotCD	ADTEAGQIATALARYPRHLAAAR--FDVGQLPHER-----RAFDADDGLPAAASPEQL--	319
	: * : : * : * . * * : : * . . . * * * * . : * :	
BotC	PAGRPAPYDDGPHPASPLEELEAGGITVWADSGAVDISGPDIPRTRLCFAHVSDPQPLL	375
BotCD	-----LRHLVDRLRDLGVDVHH-----RVLTTPPGSVTVVDVVAPGLEMF	358
	: : * . * : * * . * : . . . * * : : :	
BotC	GLVRAGIPVFDTGVEVRTLDPSRRDAHRDTPRGAPRDAARDSADRGPAGRGRQGRSA	434
BotCD	DKARAGYPVLPTGRLGERLRPRKDHE-----RGHGVHDH-----GKDRGGAPR-----	400
	. . * * * * : * * : . * * : . * * * . * * * * *	

**Figure S2:** Alignment of BotC (top) and BotCD (bottom) created using CLUSTAL Omega.

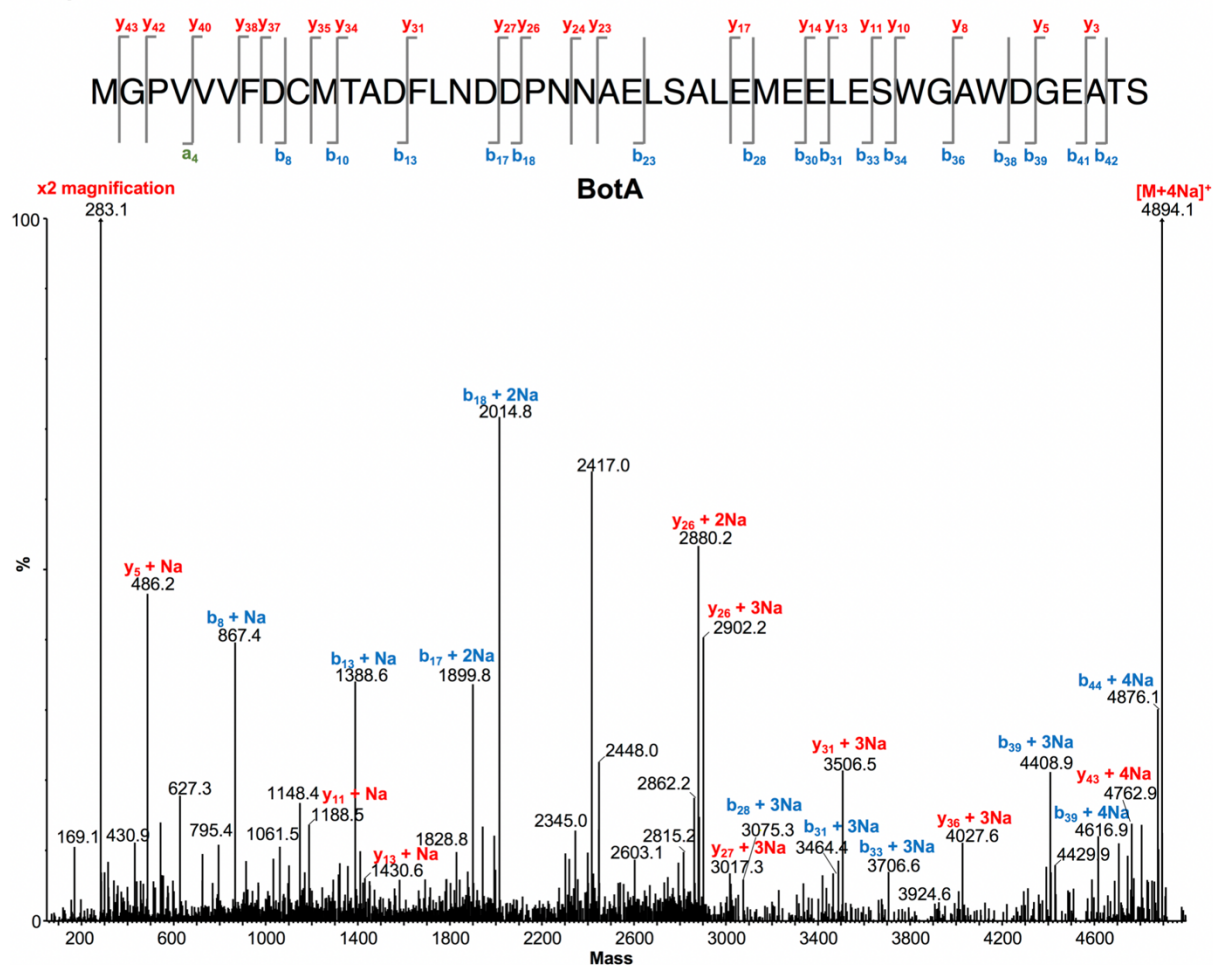


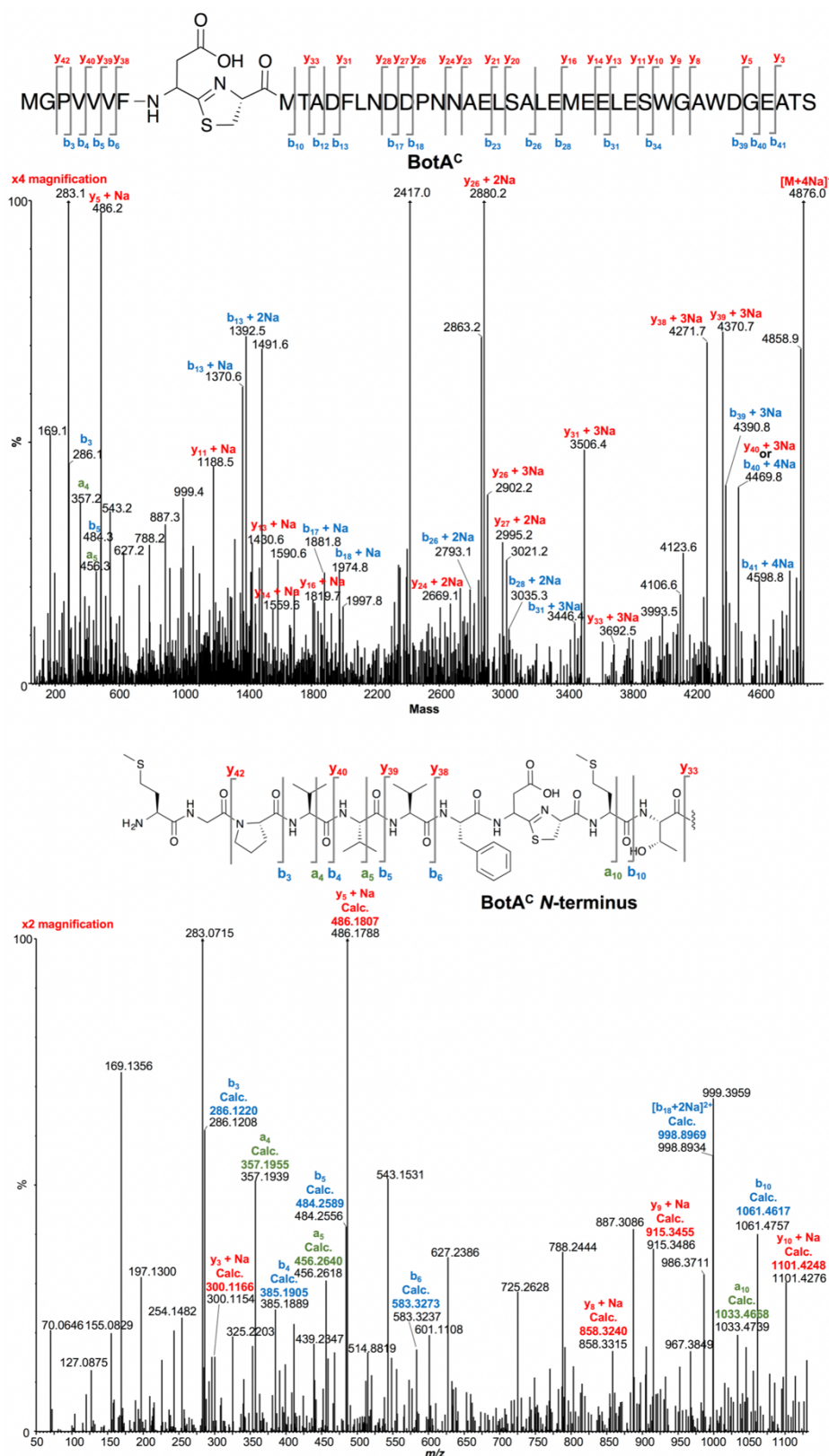
**Figure S3:** Gene clusters highly homologous in sequence and organisation to the bottromycin biosynthetic gene cluster identified in *Streptomyces* sp. BC16019 were identified in *Streptomyces ipomoeae* and *Streptomyces purpureus*. BotC homologue IpoC and BotCD homologue PurCD were used for this study.

a)



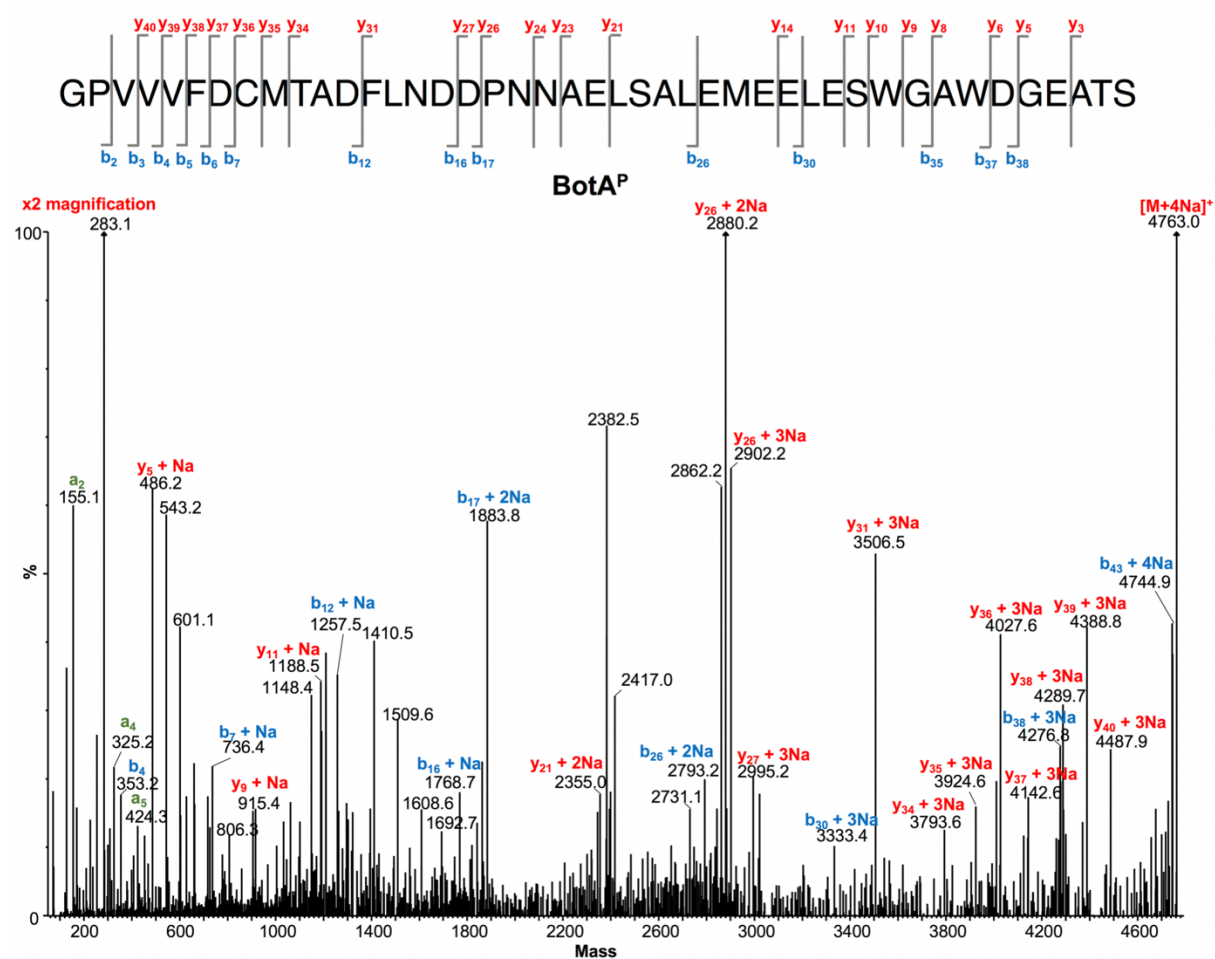
b)

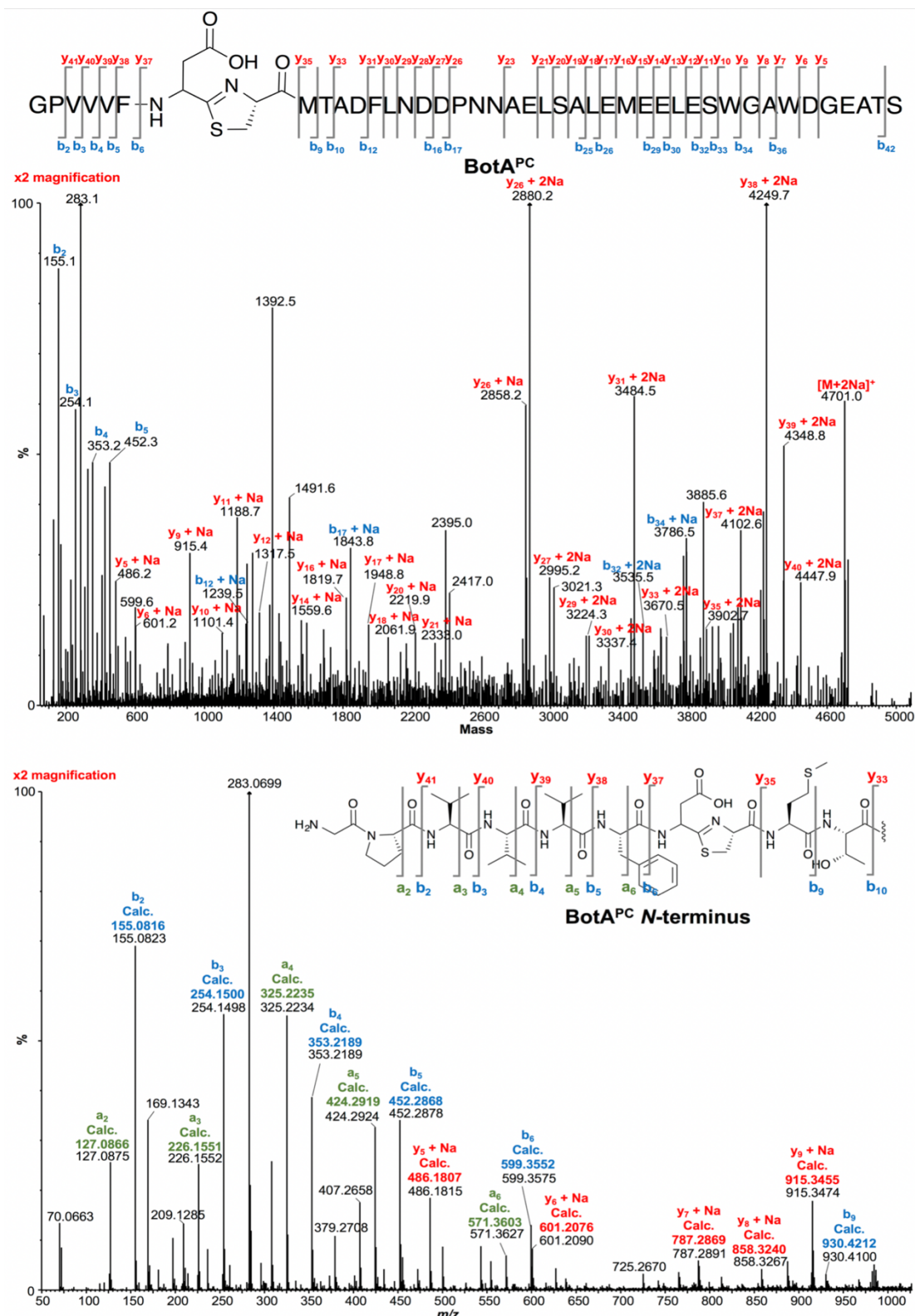




**Figure S4:** a) MS data showing BotA before and after incubation with IpoC and ATP/MgCl<sub>2</sub>. The product, BotA<sup>C</sup>, shows the characteristic loss of 18 Da, indicating loss of one water molecule b) MS<sup>2</sup> analysis of unmodified BotA and BotA<sup>C</sup>. The fragments visible in the MS<sup>2</sup> are indicated on the peptide sequence and these are detailed in Tables S1 and S2. The deconvoluted mass spectrum is displayed and is annotated with peptide fragments where possible. The presence of sodium salts meant that most fragments contained at least one sodium atom.

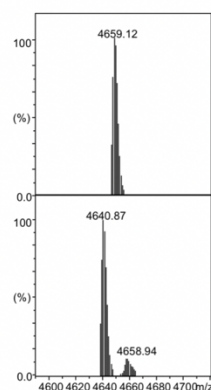




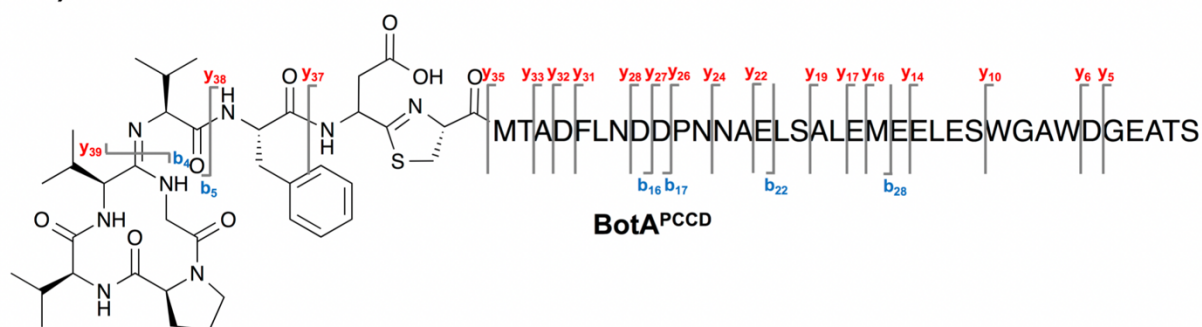


**Figure S5:** a) MS data showing BotA<sup>P</sup> before and after incubation with IpoC/PurCD and ATP/MgCl<sub>2</sub>. The products, BotA<sup>PC</sup> and BotA<sup>PCD</sup>, show the characteristic loss of 18 Da, indicating loss of one water molecule b) MS<sup>2</sup> analysis of BotA treated with aminopeptidase BotP (BotA<sup>P</sup>) and of BotA<sup>P</sup> treated with YcaO-domain enzyme IpoC (BotA<sup>PC</sup>). The fragments visible in the MS<sup>2</sup> are indicated on the peptide sequence and these are detailed in Tables S3 and S4. The deconvoluted mass spectrum is displayed and is annotated with peptide fragments where possible. The presence of sodium salts meant that most fragments contained at least one sodium atom.

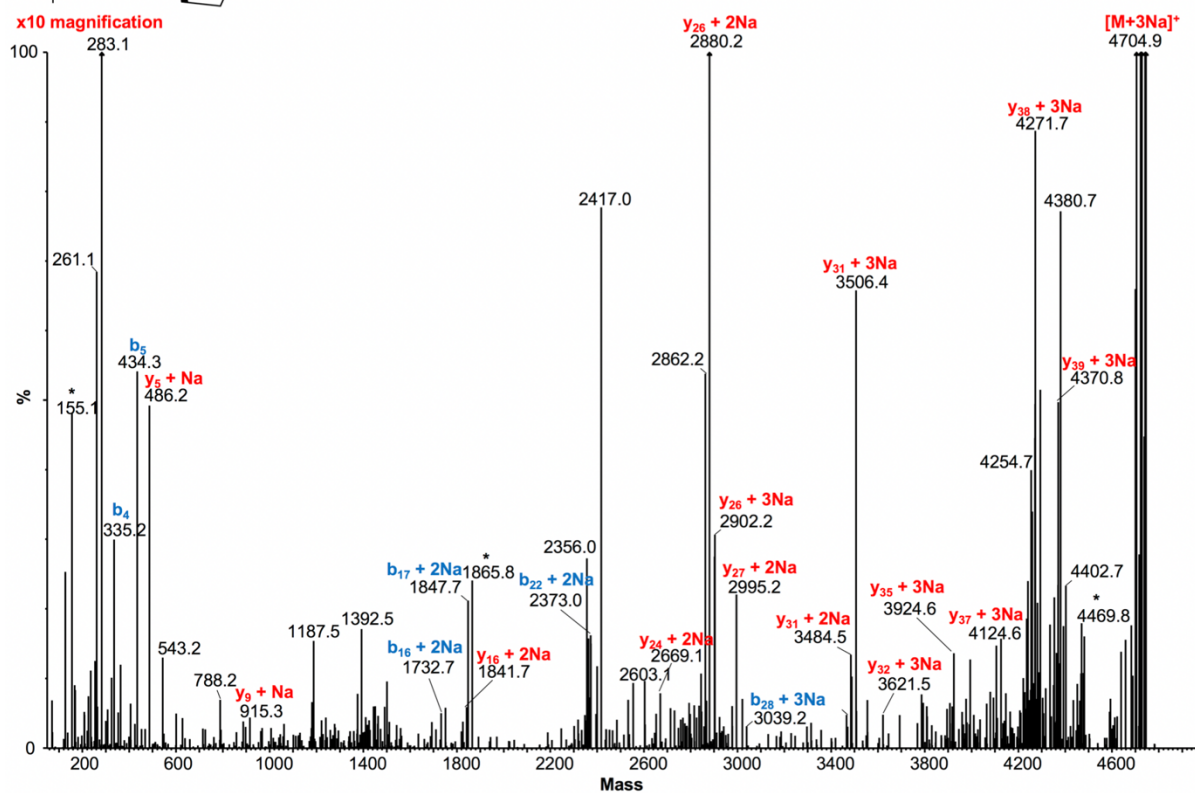
a)

BotA<sup>PC</sup>+ PurCD  
+ ATP/Mg<sup>2+</sup>

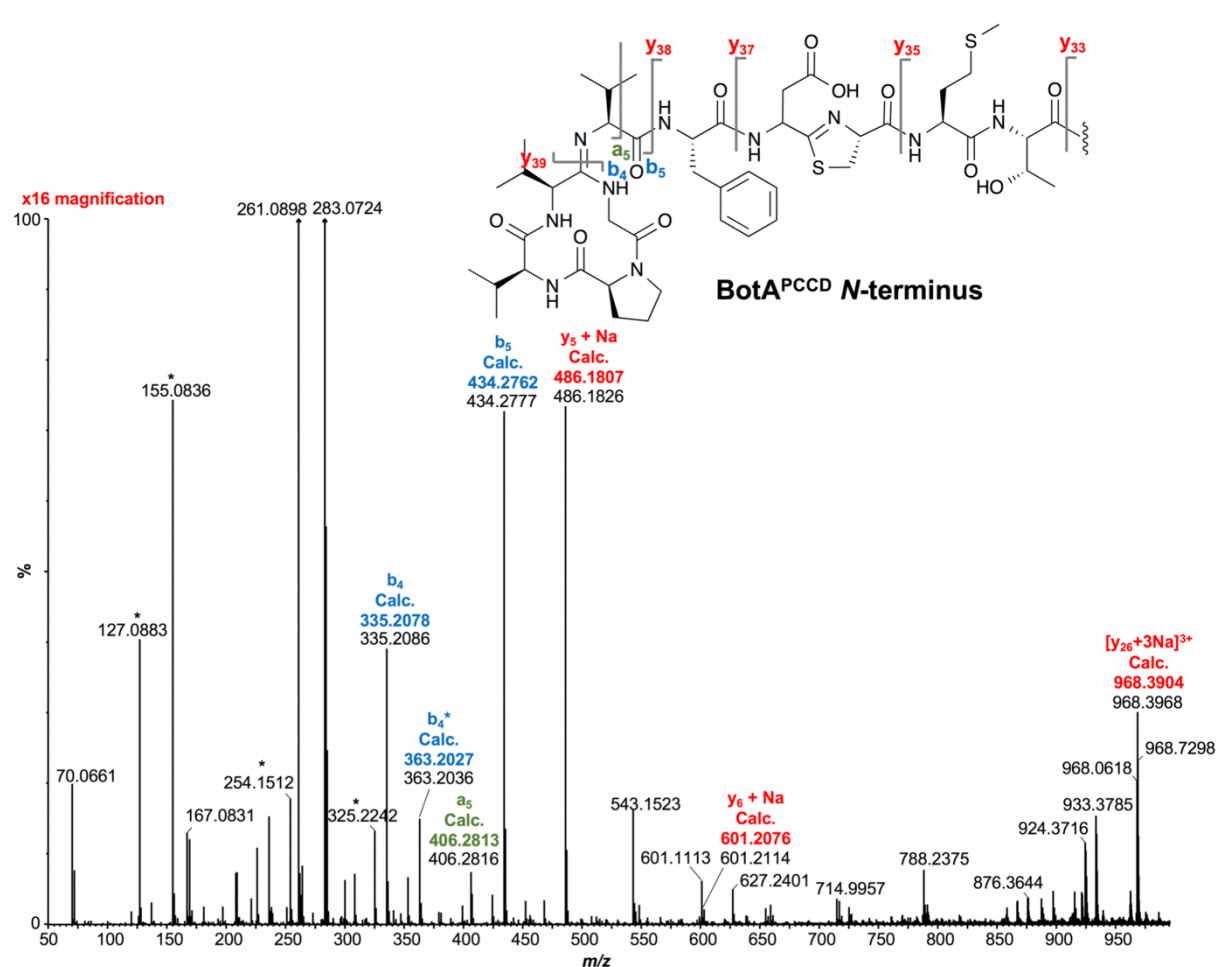
b)



**BotA<sup>PCCD</sup>**

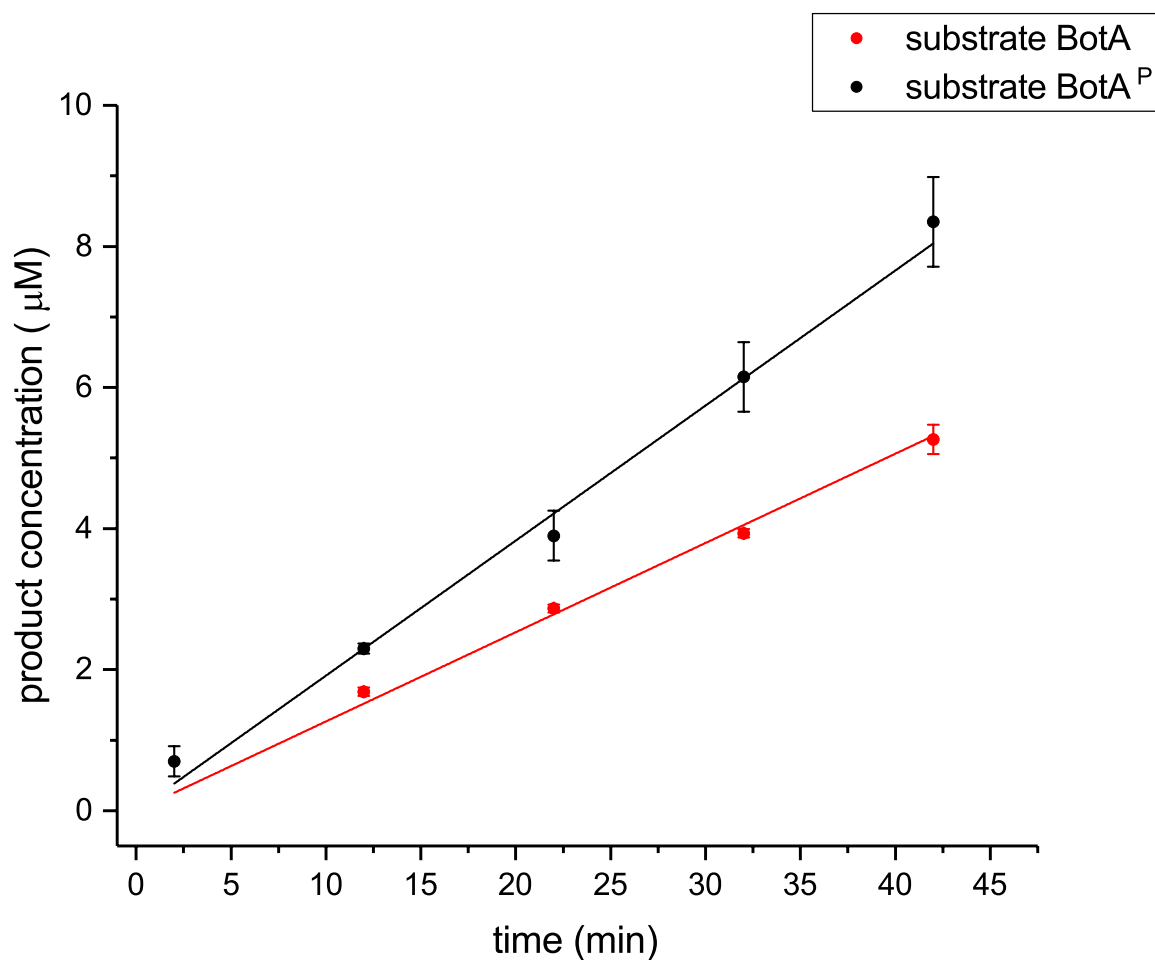




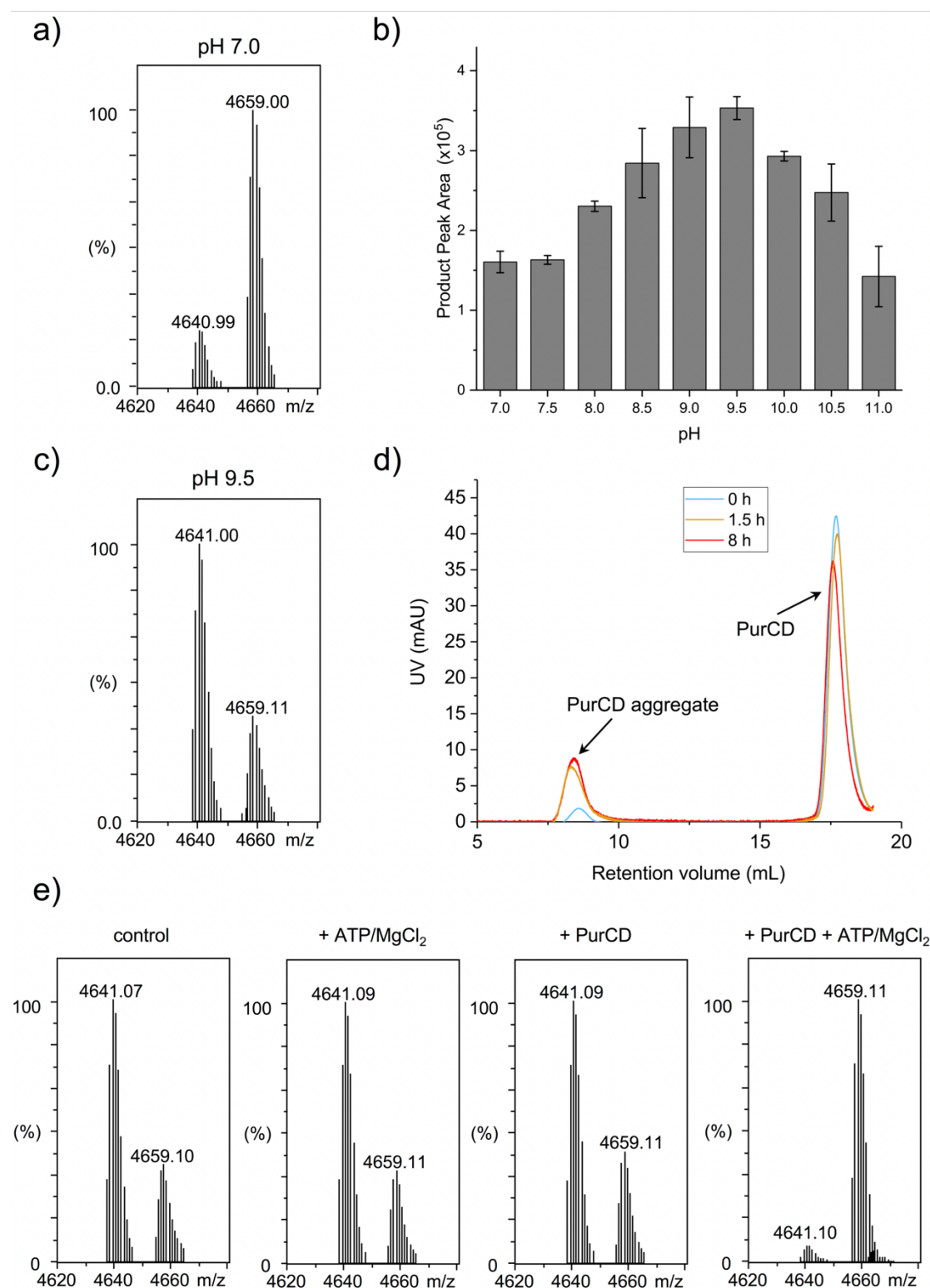


**Figure S6:** a) MS data showing BotA<sup>PC</sup> before and after incubation with PurCD and ATP/MgCl<sub>2</sub>. The product, BotA<sup>PCCD</sup>, shows the characteristic loss of 18 Da, indicating loss of one water molecule b) MS<sup>2</sup> analysis of BotA<sup>PC</sup> treated with YcaO-domain enzyme PurCD (BotA<sup>PCCD</sup>). The fragments visible in the MS<sup>2</sup> are indicated on the peptide sequence and these are detailed in Table S5. The deconvoluted mass spectrum is displayed and is annotated with peptide fragments where possible. It was not possible to fully separate BotA<sup>PC</sup> and BotA<sup>PCCD</sup> for MS<sup>2</sup> analysis, and the mass window for selecting peaks for fragmentation could not be limited to exclude fragmentation of BotA<sup>PC</sup>. Consequently, some fragments relate to BotA<sup>PC</sup> and these are highlighted by stars in the mass spectrum. The presence of sodium salts meant that most fragments contained at least one sodium atom. Fragment b<sub>4</sub>\* relates to a fragment that is characteristically associated with the bottromycin macrocycle, and analogous fragments had previously been observed for all bottromycins and related macrocyclic peptides. Starred peaks indicate fragments that are likely to derive from the BotA<sup>PC</sup> present in this sample.

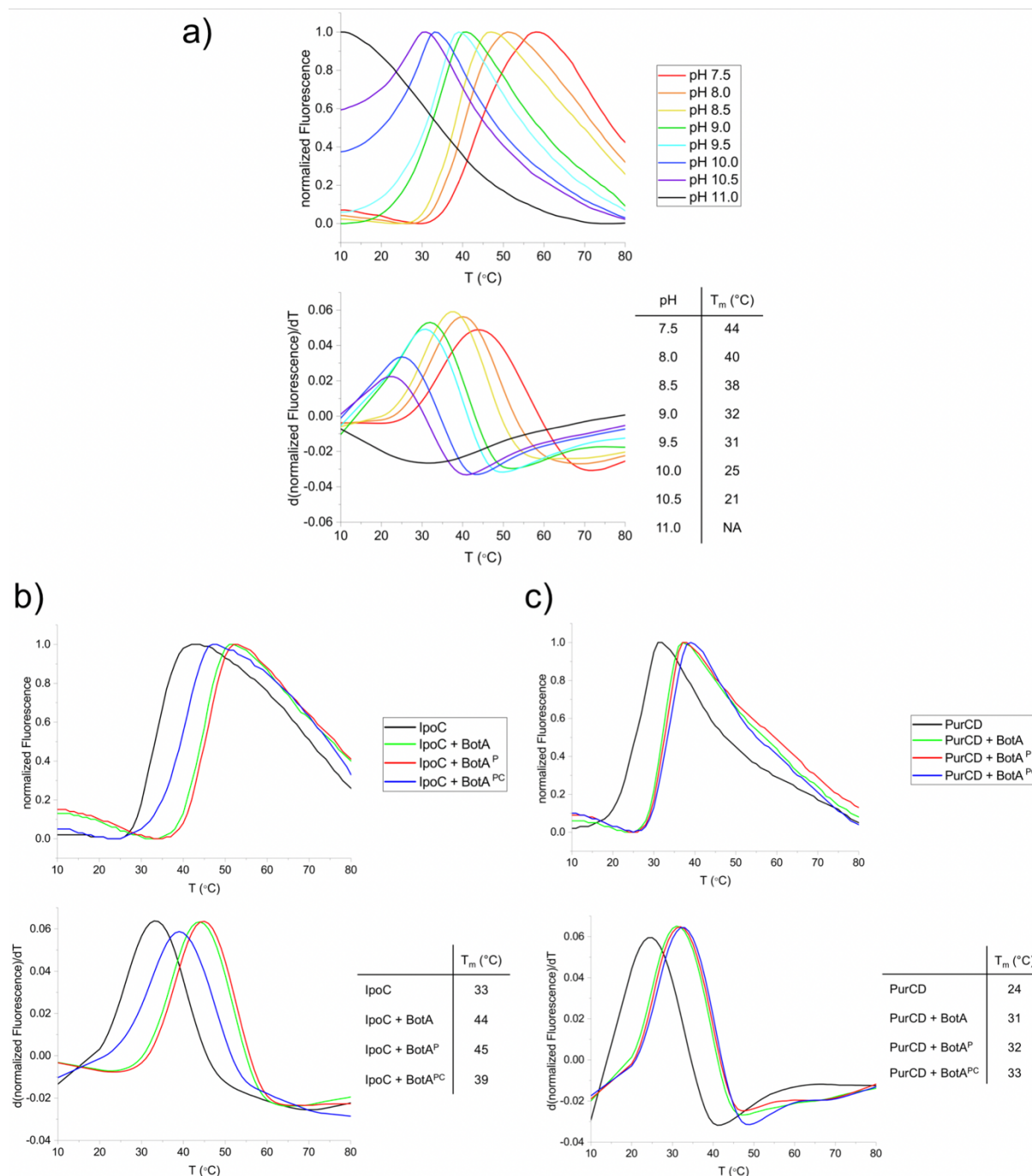




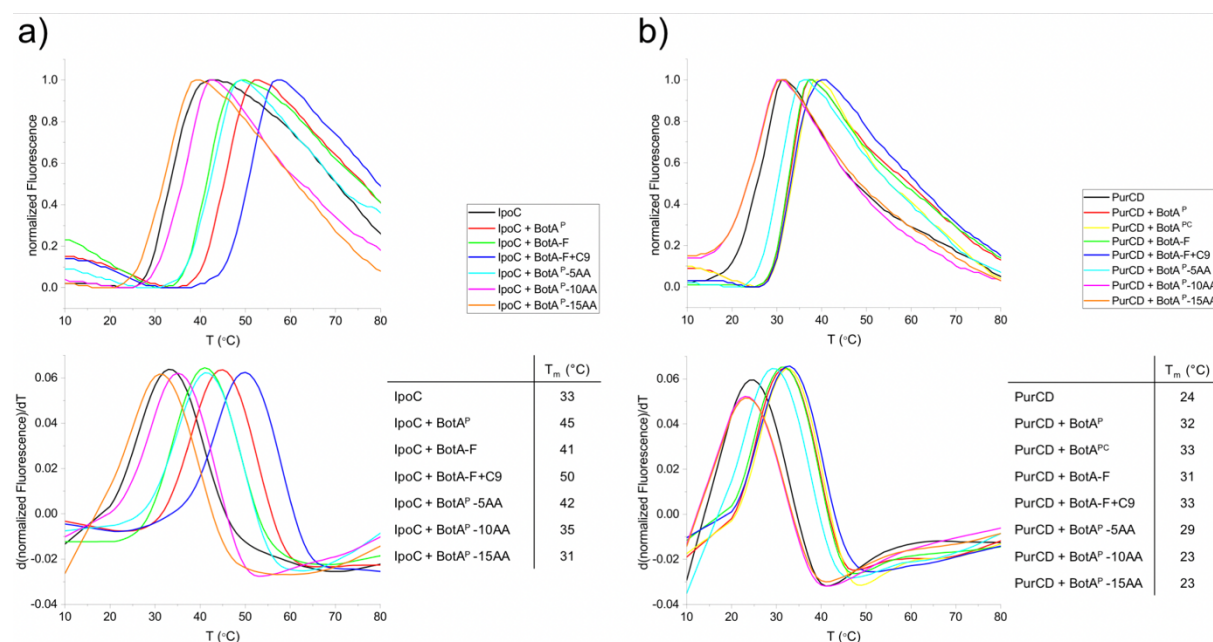
**Figure S7:** Analysis of the turnover rate of IpoC using BotA and BotA<sup>P</sup> as substrates. Using 2 μM enzyme and 100 μM substrate led to a linear increase in product concentration during the first 45 min of the reaction. This linear phase was then used to estimate the turnover rate. Data was fitted using  $y(\text{BotA}^P) = 0.19x$  and  $y(\text{BotA}) = 0.12x$ . The reactions were performed at pH 7.4 and contained 5 mM ATP/Mg<sup>2+</sup>. Data are given as mean ± SD.



**Figure S8:** a) Turnover of PurCD (10  $\mu$ M) when incubated with BotA<sup>PC</sup> (50  $\mu$ M) and 10 mM ATP/MgCl<sub>2</sub> for 90 min at pH 7.0. b) pH dependence of PurCD activity with a maximum at pH 9.5. Data are given as mean  $\pm$  SD. c) Repeat of the experiment displayed in a) at pH 9.5. d) PurCD is stable at RT in macrocyclisation buffer supplemented with 10 mM ATP/MgCl<sub>2</sub>. e) BotA<sup>PCCD</sup> was purified after 90 min from a reaction of BotA<sup>PC</sup> with PurCD (and ATP/MgCl<sub>2</sub>) and then incubated for 12 h at RT in macrocyclisation buffer with 1. no further additive (control) 2. 10 mM ATP/MgCl<sub>2</sub> 3. 10  $\mu$ M PurCD 4. 10  $\mu$ M PurCD and 10 mM ATP/MgCl<sub>2</sub>. The macroamidine is unstable when left in solution with PurCD and 10 mM ATP/MgCl<sub>2</sub> (reaction solution) for an extended period of time.

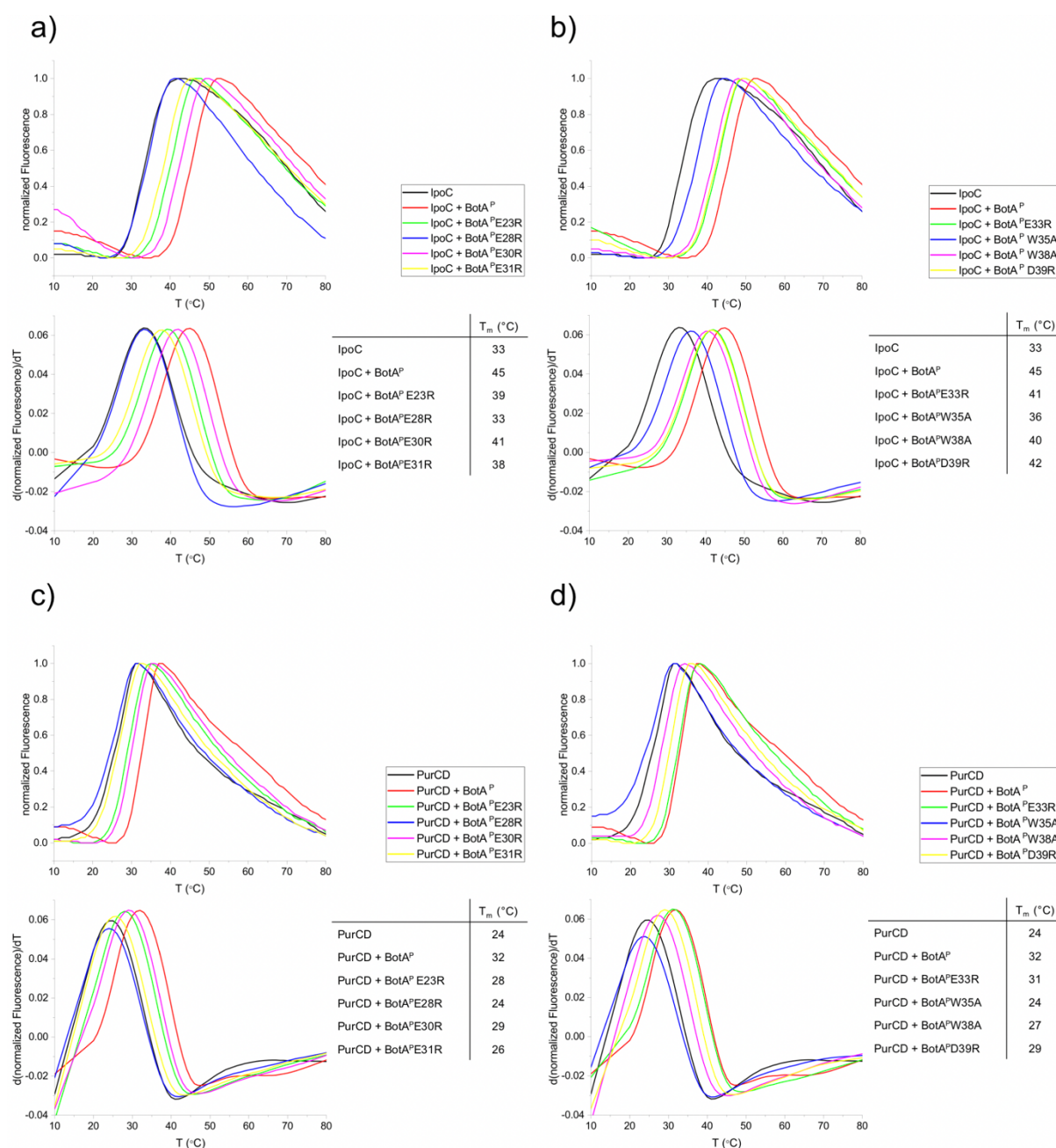


**Figure S9:** TSAs of IpoC and PurCD. a) pH dependence of PurCD stability. PurCD was incubated with ATP/MgCl<sub>2</sub> before the assay to better reflect reaction conditions. The  $T_m$  decreases steadily with increasing pH. b) Effect of substrates and product on the  $T_m$  of IpoC. The product, BotA<sup>PC</sup>, does not stabilize the enzyme to the same extent as both substrates as expected. c) Effect of different modified precursor peptides on the  $T_m$  of PurCD. These melting curves were recorded at pH 8.5 without the addition of ATP. These data indicate that ATP has a marked stabilizing effect on the enzyme. Melting points ( $T_m$ s) were calculated as described in the materials and methods. A complete list of all  $T_m$ s can be found in Table S6.

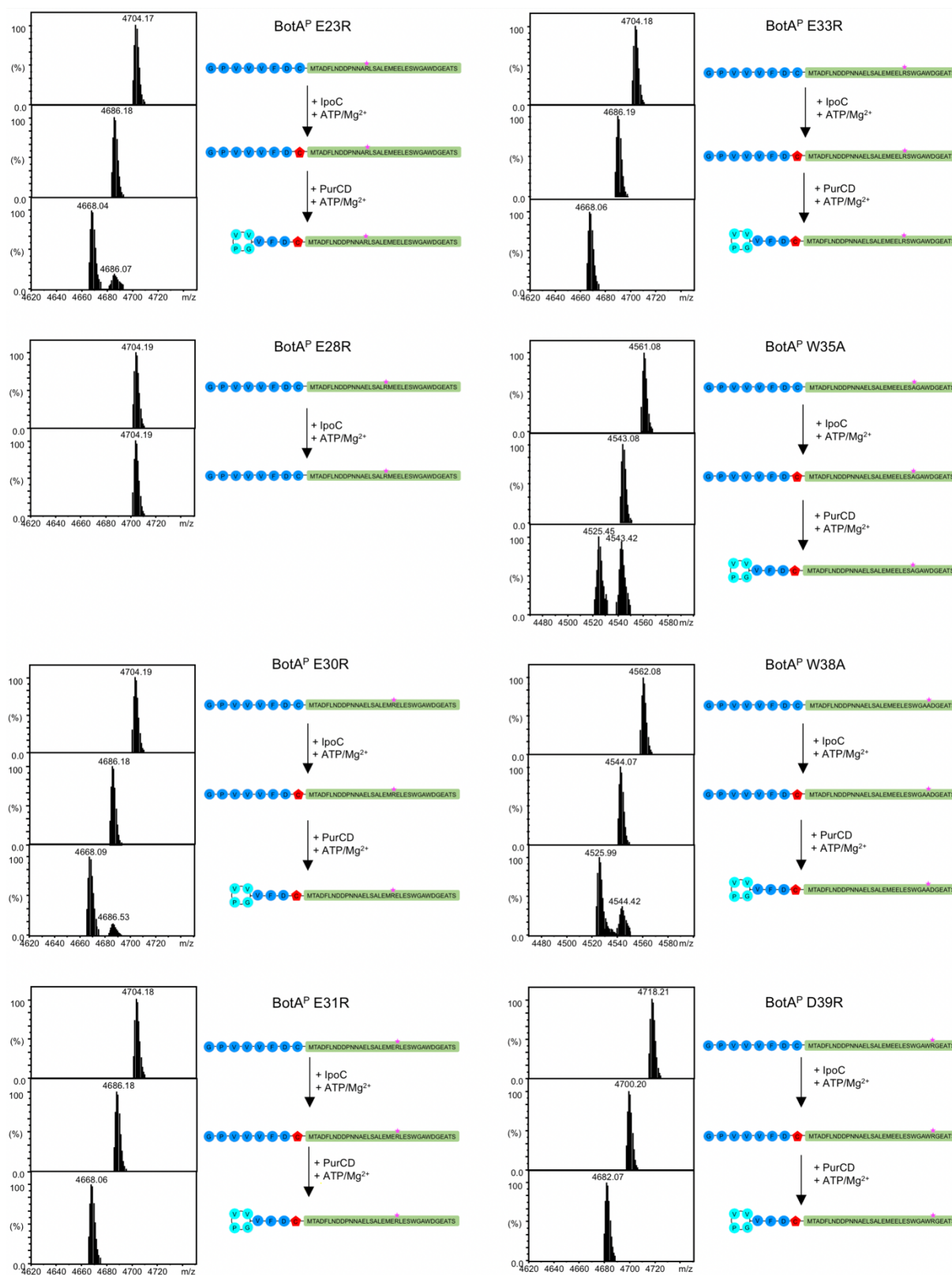


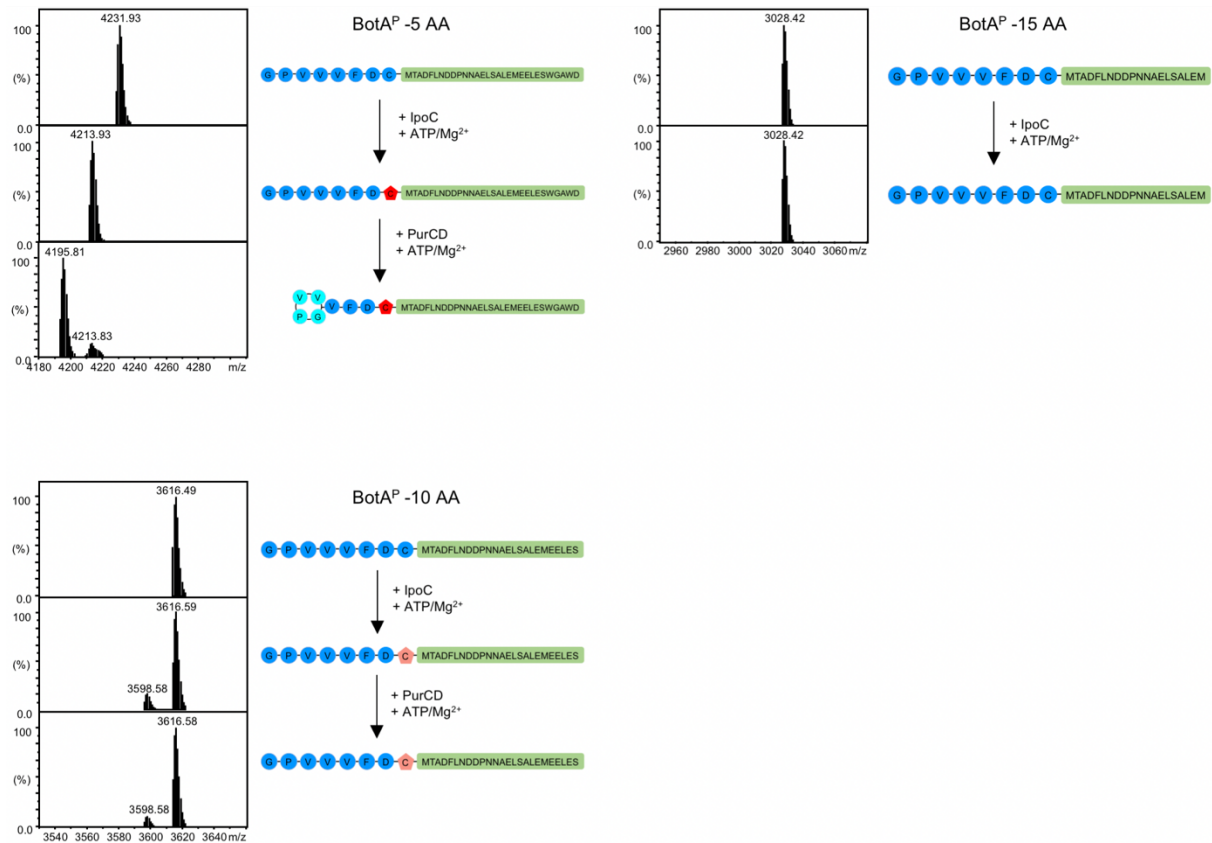
**Figure S10:** TSAs of IpoC (a)) and PurCD (b)) and their complexes with BotA<sup>P</sup>, follower peptide (BotA<sup>F</sup>) and BotA<sup>P</sup> truncations. Interestingly, for IpoC the addition of the final core peptide residue to the follower peptide, the substrate cysteine, had a dramatic stabilizing effect. This would be unexpected for a standard, processive, RiPPs heterocyclase and may indicate specific substrate interactions that processive YcaOs lack. PurCD stability is not affected to the same extent as IpoC by substrate binding. For both enzymes, the stabilizing interactions appear to be mostly contributed by residues between -5 and -10. Melting curves were recorded at pH 8.5 and melting points ( $T_m$ s) were calculated as described in the materials and methods. A complete list of all  $T_m$ s can be found in Table S6.





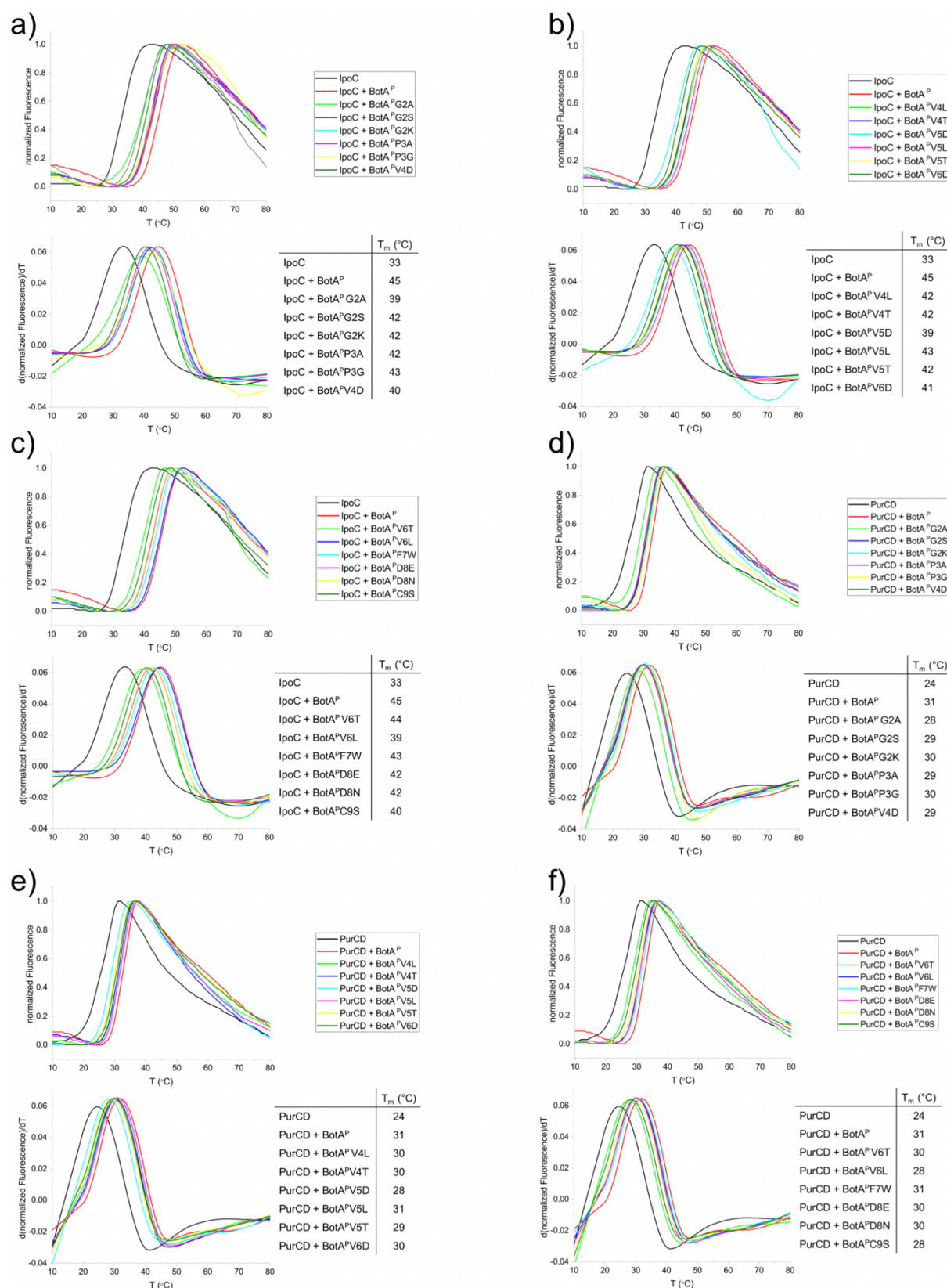
**Figure S11:** TSAs of IpoC (a) and b)) and PurCD (c) and d)) with all follower peptide point mutants reported in this study. While all point mutations lead to a reduced  $T_m$ , only E28R failed to stabilize the enzymes at all. Since this residue lies outside the region identified in Figure S9, it is probable that this residue is normally not involved in substrate binding. Instead, the radical mutation may prevent substrate binding by preventing an off-site steric hinderance. Melting curves were recorded at pH 8.5 and melting points ( $T_m$ s) were calculated as described in the materials and methods. A complete list of all  $T_m$ s can be found in Table S6.





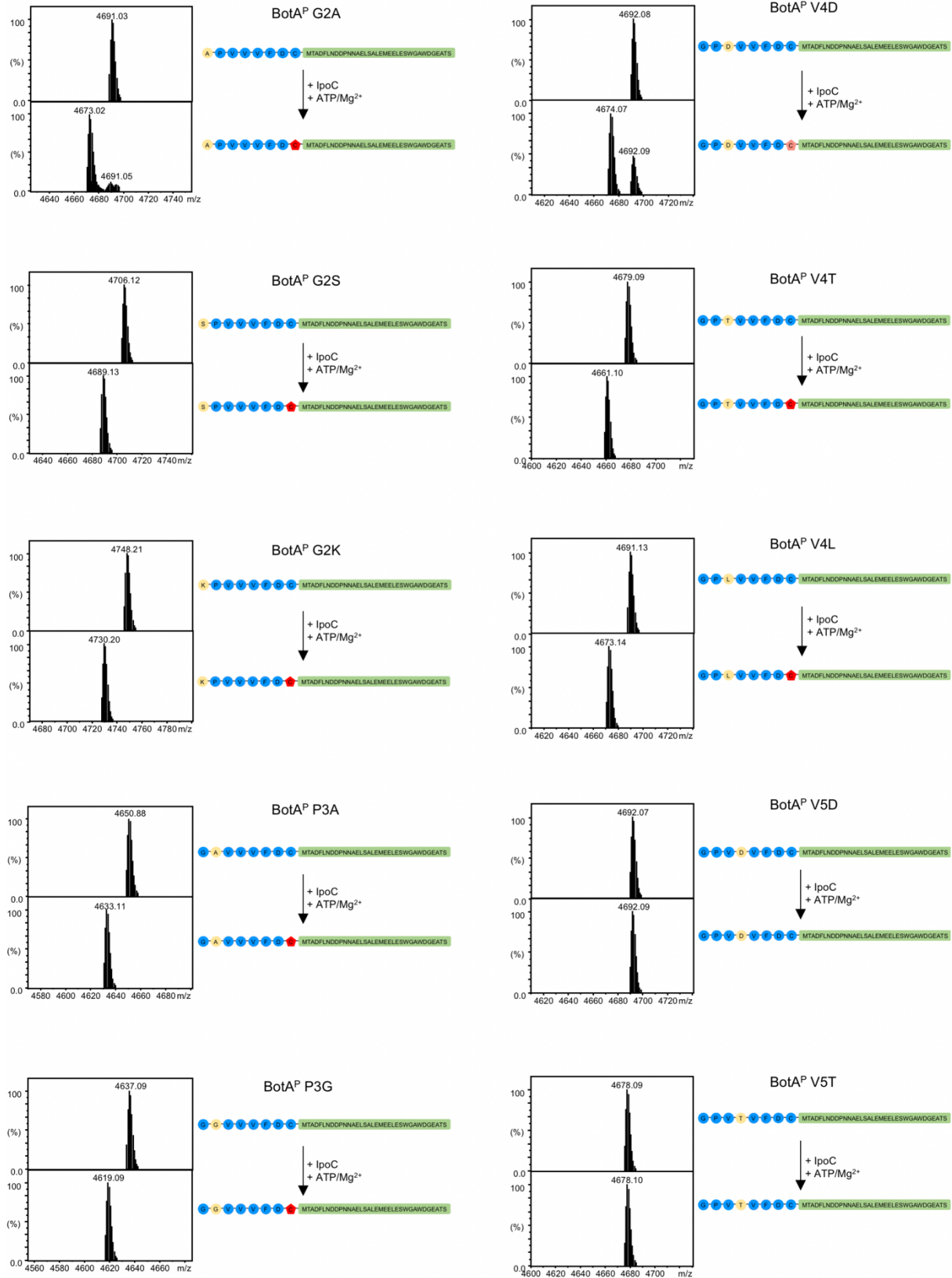
**Figure S12:** MS data showing the reactions of IpoC and PurCD with all BotA<sup>P</sup>s that carry point mutations in their follower peptides. Heterocyclization and macroamide formation result in successive losses of 18 Da each. Overall, IpoC appears to be much more tolerant to point mutations than PurCD, which is most affected by the W35A mutation. These data, when combined with the TSA data, may indicate that both enzymes have different modes of substrate binding. A summary of the results can be found in Tables S7 and S8.

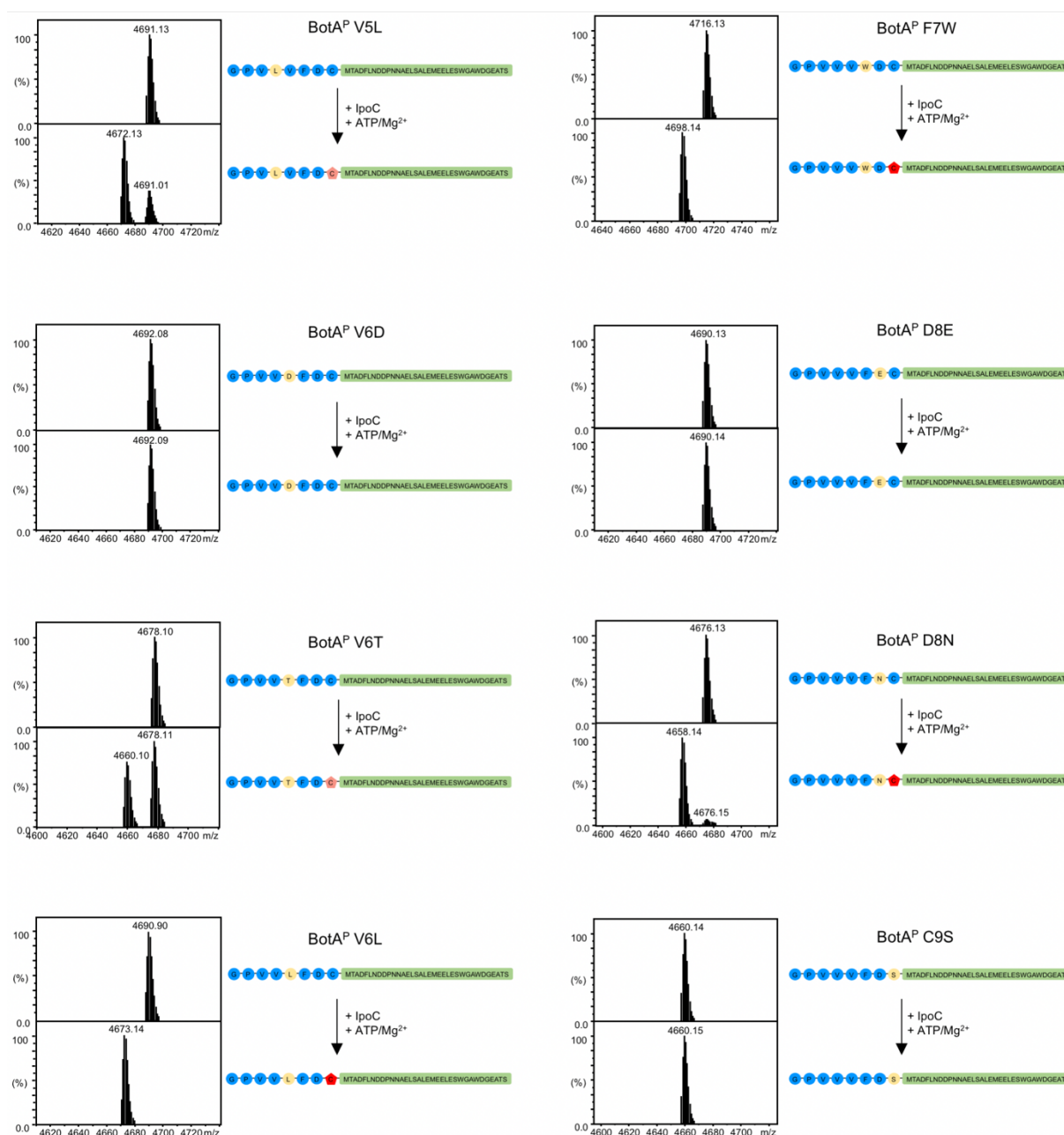




**Figure S13:** TSAs of all core peptide mutants with IpoC and PurCD. Variations of the core peptide caused variations of  $T_m$ s in a very narrow range only when compared to wt BotA. This indicates that while not all mutant BotAs are substrates, they are all still able to bind to the enzymes, echoing the spatial separation between substrate recognition and catalysis paradigm. Melting curves were recorded at pH 8.5 and melting points ( $T_m$ s) were calculated as described in the materials and methods. A complete list of all  $T_m$ s can be found in Table S6.

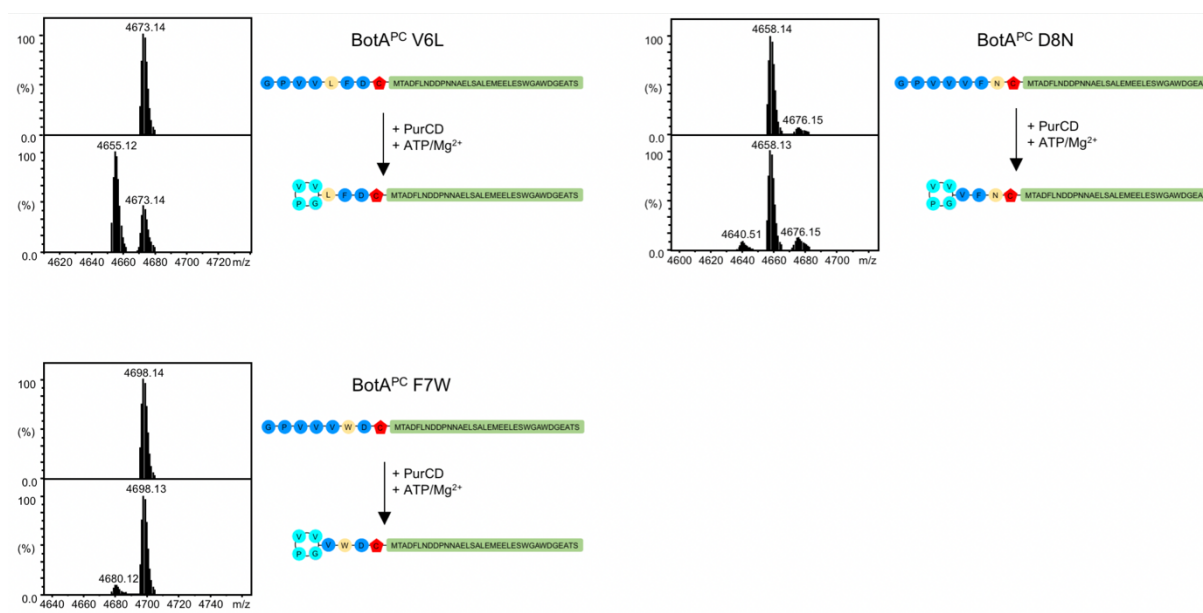






**Figure S14:** MS data showing the reaction of IpoC with all BotA<sup>P</sup>s with mutations in the core peptide used in this study. Thiazoline formation results in a loss of 18 Da. A summary of the results can be found in Table S7.



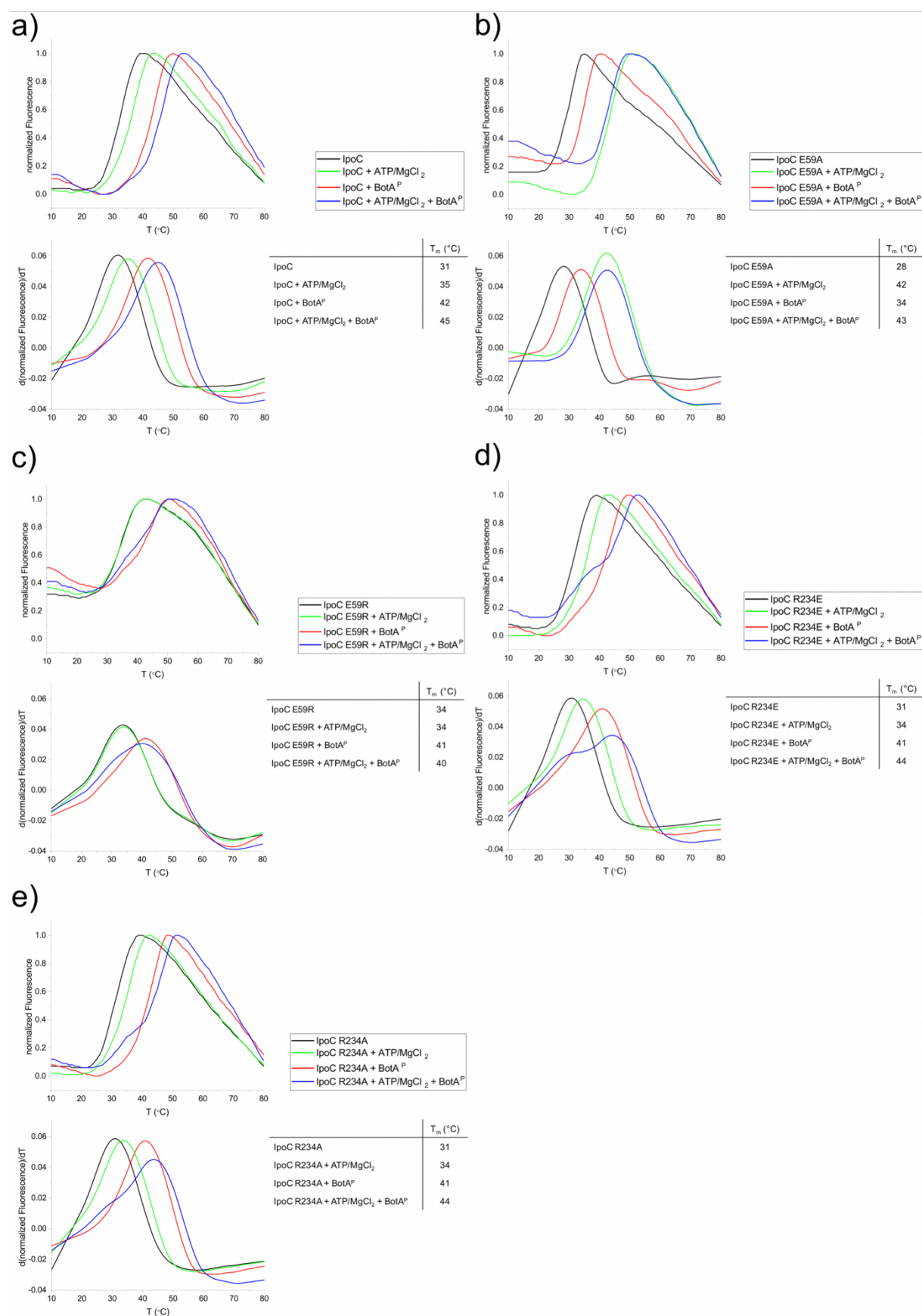


**Figure S15:** MS data showing the reaction of PurCD with all BotA<sup>PC</sup>s with mutations in the core peptide used in this study. Only those mutant BotAs that could be processed with IpoC (contained a heterocycle) could be tested as substrates, since the heterocycle is required for PurCD to catalyze formation of the macroamide (−18 Da). A summary of the results can be found in Table S8.

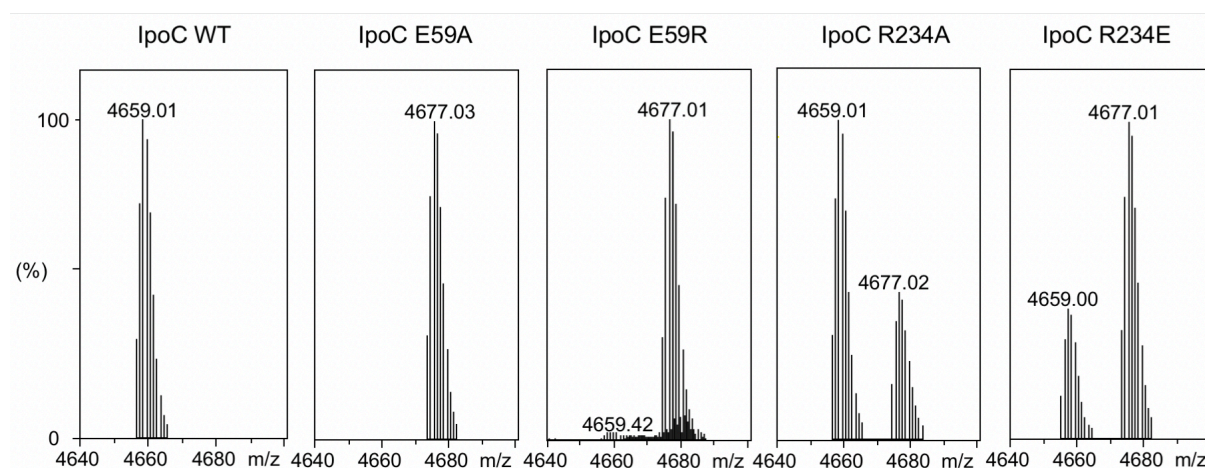
4q85	-----MTQTF-IPG--KDAALEDSIARFQQKLSDLGFQIEEASWLN-	38
4v1t	QHRGFEPLKLESRPKQ-----FTSDGHRGTTPEQTVQKYQHLISPVTVGVV--ELVRI	372
CD	-----MTRPAENDLLTAALLDGTLDKEDVPPERETPLGKALAAVG-----DWLA-	43
C	-----MLEATATECALRE-----	13
	:	
4q85	-----PV-----PNVWSVHIRDKECALCF---TNGKGATKKAALASALGTYFERLST	82
4v1t	TDPANPLVHTYRAGHSFGSATSRLRGLRNTLKHKSSGKKG--TDSQSKASGLCAVERYSG	430
CD	AERLTADIRKYGIDSAPTYEVQLKDADGALCSRSSGKGLGHRSLAS---ALFSAEHYHL	100
C	-----VVYRSYPSERTVTVRCTVVRPAEGSA--RADGYTAATKAVARLKALSFAVERL--	64
	: : : *	
4q85	NYFFADFWLGETIANGPFVHYHYPNEKWFPLTEN-----DDV----	117
4v1t	IFQG-----DEP-----RKRATLAELGDLAIHPEQ-CLCFSDGQYAN	446
CD	DWRR-----DPRASDTVFLSGREIARQPTARRSALLQRLGGIAADAPVLTREYRP---V	151
C	-----VACTPF--ASVVRPPTTSR-----GSGAVPPFPASGVRTVPDGC	101
	.	
4q85	-----PEGLDDRRLRAFYDPENELTGSMLIDLQSGNEDRGICGLPFTQSDNQTVYIP	170
4v1t	RETLNEQAT-VAHDWI PQRFDA SQAIEWT-----PVW-SLTEQTHKYL P	508
CD	AEALAEPADGAAPHRQP VF LRDGGYRNWP-----HPA-DDGSFRT----	190
C	ASRVYRPLTGGGFRRVPLYW-----S-----SPW-TAGEELRAAV-	135
	.	
4q85	-----MNIIGNLYVSNMGMSAGNTRNEARVQGLSEVFFERYVKNRIIAESI----	214
4v1t	TALCYHYPLPPEHRFARGDSNGNAAGNTLEEAILQGFMELVERDGVALWYNNRL----	563
CD	-----LWHYTSSAGYAAGATLHEALVHAVNELVERDAWSYQLARSYFGLAD	236
C	-----LTAPPEARLSTVGWAVAPTPEALHGALLLETELNHGVFLHRSIA----	181
	: * : . . * . * . . * : *	
4q85	SLPEIPADVLARYPAVVEA--IETLEAEGFPIFAYDGS LGGQYPVICVVL FNP--ANGTC	270
4v1t	---RRPAVDLGSFN EYPYFVQLQQFYRENDRLDWL DLTADLGIPAFAGVSNRKTGS SERL	620
CD	GGPALRVVDPATLPAELRQLARRVEAVRDAPVLVVDVTCDDTDPAYVVCDAESRE--DVR	294
C	-GPRRPAA-----GDETLVLPLGGPVRTPTVLAVAYGRGR--RMP	218
	. . *	
4q85	FASFGAHPDFGVALERTVTELLQGRGLKDLDVFTPTFDDEEVAEHTNLETHFIDSSGLI	330
4v1t	ILGFGAHLDPTIAILSAVTEVNQIGLELDKVPDENLKSD-----ATDWLITEKLA	670
CD	LIGSGASPVRTYALQALLEYLQVRTMFEHGPVD--AD-----TEARQIGTALA	341
C	ATGLGCGATRAEATDALLELAQAETMWRSNPTA---VP-----AERFF-LRRFE	264
	. *. * *: * *	
4q85	SWDLFKQDADYPFVDWNFSGT-----TEEEFATLMAIFNKEDKEYVI-----	372
4v1t	DHPYLLPDTTQPLKTAQDYPKRWSDDIYTDVMTVCVNIAQQAGLETLV-----	717
CD	RYPRHLAAARFDIHELPH-----E-----QVAFASDGGLSAAASPEDLLRHLVDRL	387
C	RWPLLGRCATLDFDLSGH-GTPYDD--ERSVASPLEELEAGGISVWADSGA-----	312
	: : : . . . .	
4q85	-----ADYEH LGVYACRIIVPGMSDIYPAEDLWLANN SMGSHLRETILSLPGSEW	422
4v1t	----IDQTR---PDIGLNVVKVTVPGMRHF-----	740
CD	RAVRVDVHYRVLTRPGAVTVVDVVPGL EMLDKA-----	421
C	----VDISGPD-TPRTRL CFAHVVS DPQPLGLV-----	341
	: : :	
4q85	EKEDYLN LIEQLDEEGFDDFTRVRELGLATGSDNGWYTLRIGELKAMLALAGGDLEQAL	482
4v1t	-----WSRFGEGRLYDVVP-----KL	756
CD	-----RAGYPVLPTGRL-----	433
C	-----RAGIPVFDTG EV-----	353
	: * . :	
4q85	VWTEWTMEFNSSVFS PERANYRCLQ TLLLLAQEEDRQPLQYLN AFV-RMYGADAVEAAS	541
4v1t	GWLDEPLTE-----AQMNPTMPFF-----	775
CD	---AERLRS-----RGSER-A-----	445
C	---RRTLDP-----PRDR-PAHHARSADHRAAGRG-----	380
	: :	
4q85	AAMSGEAAFYGLQPVDSDLHAFAAHQSL LKAYEKLQRAKA AFWAK	586
4v1t	-----	775
CD	-----	445
C	--RQGRRTA-----	387

**Figure S16:** Sequence alignment of the only two YcaO enzymes for which structures in complex with nucleotide cofactor have been reported to date (PDB IDs 4v1t and 4q85) with IpoC and PurCD. Residues directly involved ATP or  $Mg^{2+}$  binding in the two protein structures are highlighted in yellow. The two residues chosen for mutation in this study are among those highlighted in yellow and have been highlighted in red for clarity. Non-conserved residues are marked in grey.

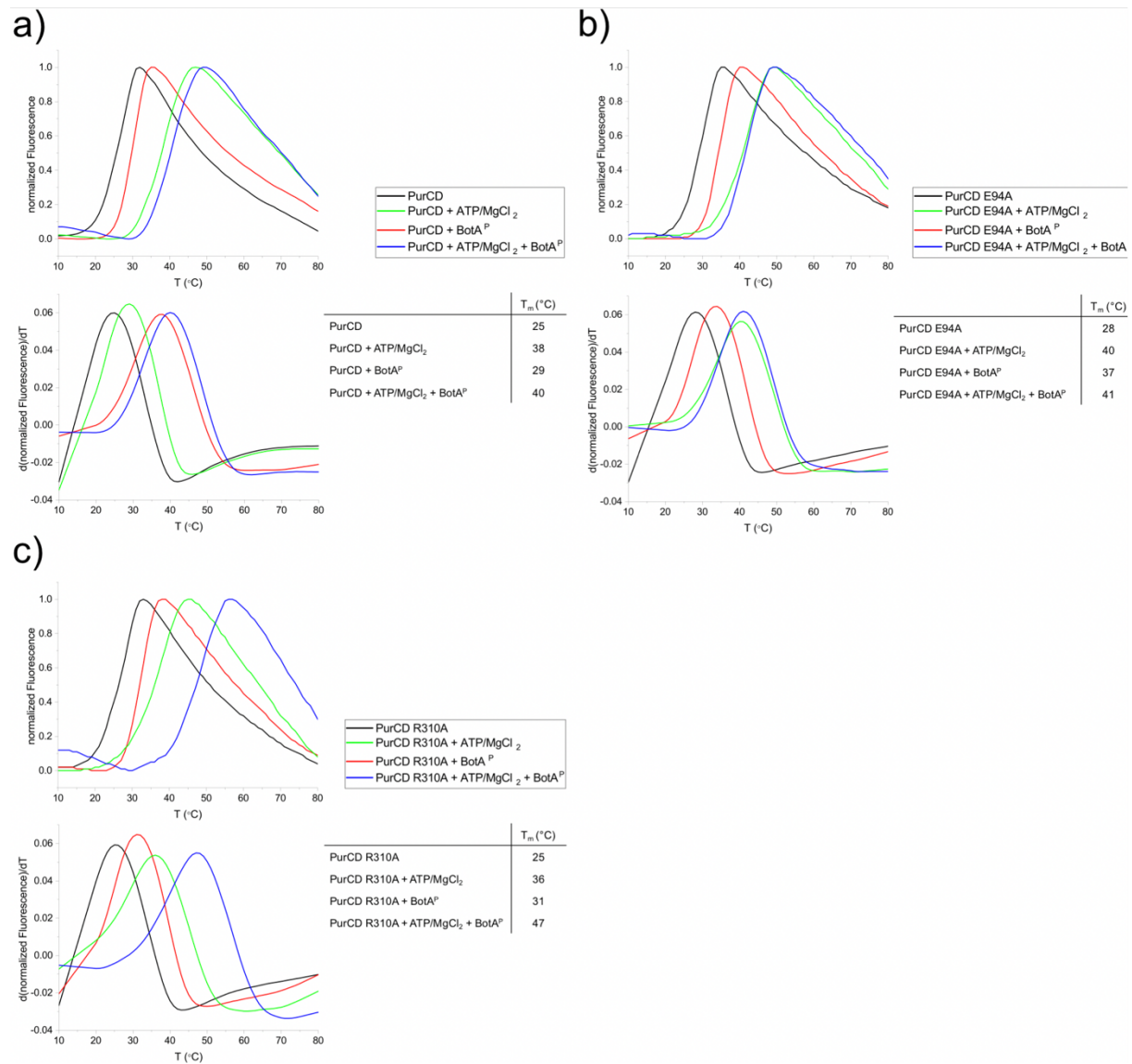




**Figure S17:** TSAs of all IpoC constructs used in this study using 1. apo enzyme, or enzyme in complex with 2. ATP/MgCl<sub>2</sub>, 3. BotA<sup>P</sup>, 4. + ATP/MgCl<sub>2</sub> and BotA<sup>P</sup>. Melting points ( $T_m$ s) were calculated as described in the materials and methods. A complete list of all  $T_m$ s can be found in Table S6.

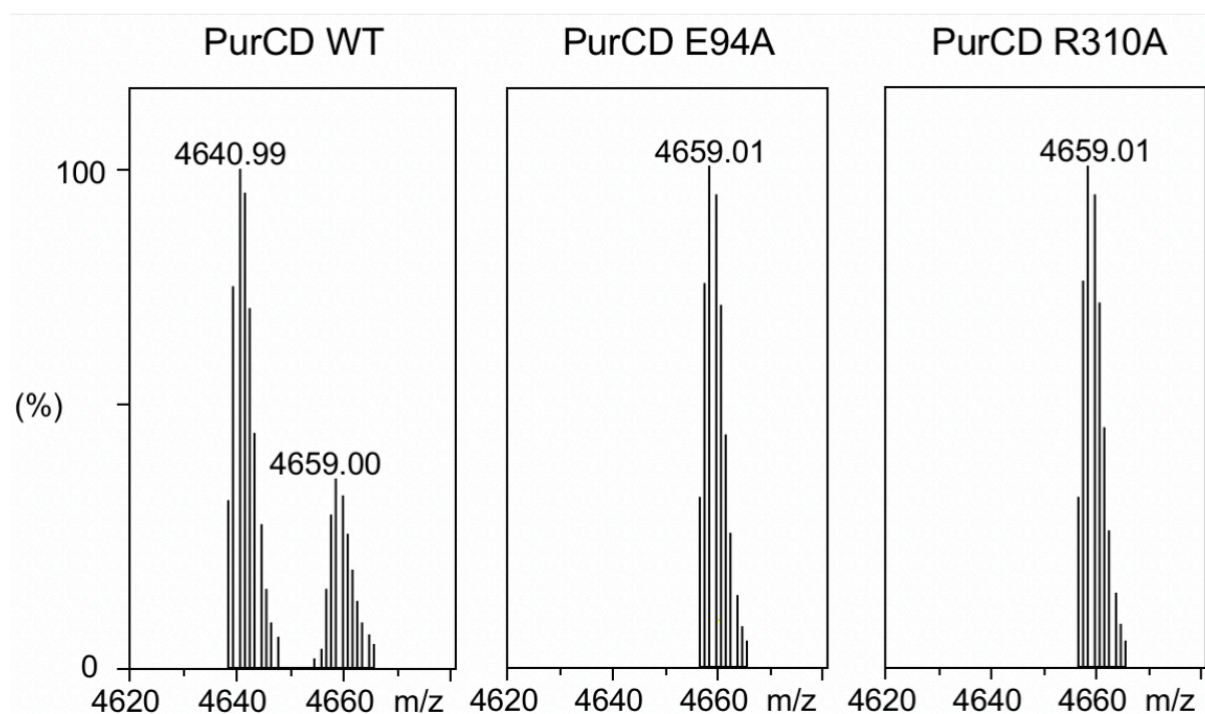


**Figure S18:** MS data comparing the activity of IpoC with that of IpoC mutants. Enzymatic activity of the E59 mutant was almost completely abolished (trace amounts of product could be found, data not shown), while R234 mutants retained significant activity.



**Figure S19:** TSAs of all PurCD constructs used in this study using 1. apo enzyme, or enzyme in complex with 2. ATP/MgCl<sub>2</sub>, 3. BotA<sup>P</sup>, 4. + ATP/MgCl<sub>2</sub> and BotA<sup>P</sup>. Interestingly, the E94A mutation appears to stabilize the enzyme when it is incubated with BotA<sup>P</sup>, but the effect was not additive, as the co-incubation with ATP/MgCl<sub>2</sub> and BotA<sup>P</sup> corresponds to the wt enzyme. PurCD R310A on the other hand is stabilized more than wt when coincubated with nucleotide and BotA<sup>P</sup>. Answering how these effects are exerted will have to await structure determination. Melting points ( $T_m$ s) were calculated as described in the materials and methods. A complete list of all  $T_m$ s can be found in Table S6.





**Figure S20:** MS data comparing the activity of PurCD with that of PurCD mutants. Only two mutants could be expressed and purified. Both were completely inactive.

**Table S1** Peptide fragments observed in MS<sup>2</sup> analysis of BotA

Seq.	b <sub>n</sub>	Obs. b	Calc. b	y <sub>n</sub>	Obs. y	Calc. y
M	1	-	-	44	4894.1	4894.0 (+4Na)
G	2	-	-	43	4762.9	4762.9 (+4Na)
P	3	-	-	42	4706.0	4705.9 (+4Na)
V	4	-	-	41	-	-
V	5	-	-	40	4509.9	4509.8 (+4Na)
V	6	-	-	39	-	-
F	7	-	-	38	4311.8	4311.7 (+4Na)
D	8	867.4105	867.4045 (+Na)	37	4142.6	4142.6 (+3Na)
C	9	-	-	36	4027.6	4027.6 (+3Na)
M	10	1101.4	1101.5 (+Na)	35	3924.6	3924.6 (+3Na)
T	11	-	-	34	3793.6	3793.5 (+3Na)
A	12	-	-	33	-	-
D	13	1388.6	1388.6 (+Na)	32	-	-
F	14	-	-	31	3506.5	3506.4 (+3Na)
L	15	-	-	30	-	-
N	16	-	-	29	-	-
D	17	1877.8, 1899.8	1877.8 (+Na), 1899.8 (+2Na)	28	-	-
D	18	1992.8, 2014.8	1992.8 (+Na), 2014.8 (+2Na)	27	3017.3	3017.2 (+3Na)
P	19	-	-	26	2880.2, 2902.2	2880.2 (+2Na), 2902.2 (+3Na)
N	20	-	-	25	-	-
N	21	-	-	24	2669.1	2669.1 (+2Na)
A	22	-	-	23	2555.1	2555.1 (+2Na)
E	23	2540.0	2540.0 (+2Na)	22	-	-
L	24	-	-	21	-	-
S	25	-	-	20	-	-
A	26	-	-	19	-	-
L	27	-	-	18	-	-
E	28	3075.3	3075.3 (+3Na)	17	-	-
M	29	-	-	16	-	-
E	30	3335.4	3335.4 (+3Na)	15	-	-
E	31	3464.4	3464.4 (+3Na)	14	1559.6, 1581.6	1559.6 (+Na)
L	32	-	-	13	1430.6	1430.6 (+Na)
E	33	3706.6	3706.5 (+3Na)	12	-	1317.5 (+Na)
S	34	3793.6	3793.6 (+3Na)	11	1188.5	1188.5 (+Na)
W	35	-	-	10	1101.4	1101.4 (+Na)
G	36	4035.7	4036.7 (+3Na)	9	-	-
A	37	-	-	8	858.3565	858.3240 (+Na)
W	38	4293.8	4293.8 (+3Na)	7	-	-
D	39	4408.9	4408.8 (+3Na)	6	-	-
G	40	-	-	5	486.1849	486.1807 (+Na)
E	41	4616.9	4616.9 (+4Na)	4	-	-
A	42	4688.0	4687.9 (+4Na)	3	300.1185	300.1166 (+Na)
T	43	-	-	2	-	-
S	44	4876.1	4876.0 (+4Na)	1	-	-

**Table S2** Peptide fragments observed in MS<sup>2</sup> analysis of BotA<sup>C</sup>

Seq.	b <sub>n</sub>	Obs. b	Calc. b	y <sub>n</sub>	Obs. y	Calc. y
M	1	-	-	44	4876.0	4876.0 (+4Na)
G	2	-	-	43	-	-
P	3	286.1208	286.122	42	4687.9	4687.9 (+4Na)
V	4	385.1889	385.1905	41	-	-
V	5	484.2556	484.2589	40	4469.8 <sup>a</sup>	4469.8 (+3Na)
V	6	583.3237	583.3273	39	4370.7	4370.7 (+3Na)
F	7	-	-	38	4271.7	4271.7 (+3Na)
D <sup>b</sup>	8	-	-	37	-	-
C <sup>b</sup>	9	-	-	36	-	-
M	10	1061.5	1061.5	35	-	-
T	11	-	-	34	-	-
A	12	1277.5	1277.5 (+2Na)	33	3692.5	3692.5 (+3Na)
D	13	1370.6	1370.6 (+Na)	32	-	-
F	14	-	-	31	3506.4	3506.4 (+3Na)
L	15	-	-	30	-	-
N	16	-	-	29	-	-
D	17	1859.8, 1881.8	1859.8 (+Na), 1881.8 (+2Na)	28	3110.2	3110.2 (+2Na)
D	18	1974.8, 1996.8	1974.8 (+Na), 1996.8 (+2Na)	27	2995.2	2995.2 (+2Na)
P	19	-	-	26	2880.2, 2902.2	2880.2 (+2Na), 2902.2 (+3Na)
N	20	-	-	25	-	-
N	21	-	-	24	2669.1	2669.1 (+2Na)
A	22	-	-	23	2533.0	2533.1 (+Na)
E	23	2522.0	2522.0 (+2Na)	22	-	-
L	24	-	-	21	2333.0	2333.0 (+Na)
S	25	-	-	20	2219.9	2219.9 (+Na)
A	26	2793.1	2793.2 (+2Na)	19	-	-
L	27	-	-	18	-	-
E	28	3035.3	3035.3 (+2Na)	17	-	-
M	29	-	-	16	1819.7	1819.7 (+Na)
E	30	-	-	15	-	-
E	31	3446.4	3446.4 (+3Na)	14	1559.6	1559.6 (+Na)
L	32	-	-	13	1430.6	1430.6 (+Na)
E	33	-	-	12	-	-
S	34	3775.5	3775.6 (+3Na)	11	1188.5	1188.5 (+Na)
W	35	-	-	10	1101.4	1101.4 (+Na)
G	36	-	-	9	915.3486	915.3455 (+Na)
A	37	-	-	8	858.3315	858.3240 (+Na)
W	38	-	-	7	-	-
D	39	4390.8	4390.8 (+3Na)	6	-	-
G	40	4469.8 <sup>a</sup>	4469.8 (+4Na)	5	486.1788	486.1807 (+Na)
E	41	4598.8	4598.8 (+4Na)	4	-	-
A	42	-	-	3	300.1156	300.1166 (+Na)
T	43	-	-	2	-	-
S	44	-	-	1	-	-

a. 4469.8 fragment could be either b<sub>40</sub>+4Na or y<sub>40</sub>+3Na.

b. Cys9 heterocyclization with Asp8 carbonyl.

**Table S3** Peptide fragments observed in MS<sup>2</sup> analysis of BotA<sup>P</sup>

Seq.	b <sub>n</sub>	Obs. b	Calc. b	y <sub>n</sub>	Obs. y	Calc. y
G	1	-	-	43	4763.0	4762.9 (+4Na)
P	2	155.0824	155.0816	42	-	-
V	3	254.1499	254.1500	41	-	-
V	4	353.2191	353.2184	40	4487.9	4487.8 (+3Na)
V	5	452.2879	452.2868	39	4388.8	4388.7 (+3Na)
F	6	599.3584	599.3552	38	4289.7	4289.7 (+3Na)
D	7	736.3655	736.3641 (+Na)	37	4142.6	4142.6 (+3Na)
C	8	-	-	36	4027.6	4027.6 (+3Na)
M	9	-	-	35	3924.6	3924.6 (+3Na)
T	10	-	-	34	3793.6	3793.5 (+3Na)
A	11	-	-	33	-	-
D	12	1257.5	1257.5 (+Na)	32	-	-
F	13	-	-	31	3506.4	3506.4 (+3Na)
L	14	-	-	30	-	-
N	15	-	-	29	-	-
D	16	1768.7	1768.7 (+2Na)	28	-	-
D	17	1861.8, 1883.8	1861.8 (+Na), 1883.8 (+2Na)	27	2995.2	2995.2 (+2Na)
P	18	-	-	26	2880.2, 2902.2	2880.2 (+2Na), 2902.2 (+3Na)
N	19	-	-	25	-	-
N	20	-	-	24	-	-
A	21	-	-	23	-	-
E	22	-	-	22	-	-
L	23	-	-	21	2355.0	2355.0 (+2Na)
S	24	-	-	20	-	-
A	25	-	-	19	-	-
L	26	2793.2	2793.2 (+2Na)	18	-	-
E	27	-	-	17	-	-
M	28	-	-	16	-	-
E	29	-	-	15	-	-
E	30	3333.4	3333.4 (+3Na)	14	1559.6	1559.6 (+Na)
L	31	-	-	13	-	-
E	32	-	-	12	-	-
S	33	-	-	11	1188.5	1188.5 (+Na)
W	34	-	-	10	1101.4254	1101.4 (+Na)
G	35	3905.6	3905.6 (+3Na)	9	915.3485	915.3455 (+Na)
A	36	-	-	8	858.3399	858.3240 (+Na)
W	37	4184.7	4184.7 (+4Na)	7	-	-
D	38	4277.7, 4299.8	4277.8 (+3Na), 4299.7 (+4Na)	6	601.2091	601.2076 (+Na)
G	39	-	-	5	486.1823	486.1807 (+Na)
E	40	-	-	4	-	-
A	41	-	-	3	300.1170	300.1166 (+Na)
T	42	-	-	2	-	-
S	43	4744.9	4744.9 (+4Na)	1	-	-

**Table S4** Peptide fragments observed in MS<sup>2</sup> analysis of BotA<sup>PC</sup>

Seq.	b <sub>n</sub>	Obs. b	Calc. b	y <sub>n</sub>	Obs. y	Calc. y
G	1	-	-	43	4701.0	4701.0 (+2Na)
P	2	155.0823	155.0816	42	-	-
V	3	254.1498	254.1500	41	4546.9	4546.9 (+2Na)
V	4	353.2189	353.2184	40	4447.9	4447.8 (+2Na)
V	5	452.2878	452.2868	39	4348.8	4348.8 (+2Na)
F	6	599.3575	599.3552	38	4249.7	4249.7 (+2Na)
D <sup>a</sup>	7	-	-	37	4102.6	4102.6 (+2Na)
C <sup>a</sup>	8	-	-	36	-	-
M	9	930.4138	930.4212	35	3902.7	3902.6 (+2Na)
T	10	1031.5	1031.5	34	-	-
A	11	-	-	33	3670.5	3670.5 (+2Na)
D	12	1239.5	1239.5 (+Na)	32	-	-
F	13	-	-	31	3484.5	3484.4 (+2Na)
L	14	-	-	30	3338.4	3337.4 (+2Na)
N	15	-	-	29	3224.3	3224.3 (+2Na)
D	16	1728.7	1728.7 (+Na)	28	3110.3	3110.2 (+2Na)
D	17	1843.8	1843.8 (+Na)	27	2995.2	2995.2 (+2Na)
P	18	-	-	26	2858.2, 2880.2	2858.2 (+Na), 2880.2 (+2Na)
N	19	-	-	25	-	-
N	20	-	-	24	-	-
A	21	-	-	23	2533.1	2533.1 (+Na)
E	22	-	-	22	-	-
L	23	-	-	21	2333.0	2333.0 (+Na)
S	24	-	-	20	2219.9	2219.9 (+Na)
A	25	2662.1	2662.1 (+2Na)	19	2132.9	2132.9 (+Na)
L	26	2775.2	2775.2 (+2Na)	18	2061.9	2061.8 (+Na)
E	27	-	-	17	1948.8	1948.8 (+Na)
M	28	-	-	16	1819.7	1819.7 (+Na)
E	29	3164.3	3164.3 (+2Na)	15	1688.7	1688.7 (+Na)
E	30	3293.4	3293.4 (+2Na)	14	1559.6	1559.6 (+Na)
L	31	-	-	13	1430.6	1430.6 (+Na)
E	32	3535.5	3535.5 (+2Na)	12	1317.5	1317.5 (+Na)
S	33	3622.5	3622.5 (+2Na)	11	1188.5	1188.5 (+Na)
W	34	3786.5	3786.6 (+Na)	10	1101.4	1101.4 (+Na)
G	35	-	-	9	915.3567	915.3455 (+Na)
A	36	3936.7	3936.7 (+2Na)	8	858.3267	858.3240 (+Na)
W	37	-	-	7	765.3071, 787.2891	765.3050, 787.2869 (+Na)
D	38	-	-	6	601.2090	601.2076 (+Na)
G	39	-	-	5	486.1815	486.1807 (+Na)
E	40	-	-	4	-	-
A	41	-	-	3	-	-
T	42	4595.9	4595.9 (+2Na)	2	-	-
S	43	4683.0	4683.0 (+2Na)	1	-	-

a. Cys8 heterocyclization with Asp7 carbonyl.

**Table S5** Peptide fragments observed in MS<sup>2</sup> analysis of BotA<sup>PCCD</sup>

Seq.	b <sub>n</sub>	Obs. b	Calc. b	y <sub>n</sub>	Obs. y	Calc. y
G <sup>a</sup>	1	-	-	43	4704.9	4704.9 (+3Na)
P	2	-	-	42	-	-
V	3	-	-	41	-	-
V <sup>a</sup>	4	335.2086	335.2078	40	-	-
V	5	434.2777	434.2762	39	4370.8	4370.7 (+3Na)
F	6	-	-	38	4271.7	4271.7 (+3Na)
D <sup>b</sup>	7	-	-	37	4124.6	4124.6 (+3Na)
C <sup>b</sup>	8	-	-	36	-	-
M	9	-	-	35	3924.6	3924.6 (+3Na)
T	10	-	-	34	-	-
A	11	-	-	33	3692.5	3692.5 (+3Na)
D	12	-	-	32	3621.5	3621.4 (+3Na)
F	13	-	-	31	3484.5, 3506.4	3484.4 (+2Na), 3506.4 (+3Na)
L	14	-	-	30	-	-
N	15	-	-	29	-	-
D	16	1732.7	1732.7 (+2Na)	28	3131.2	3132.2 (+3Na)
D	17	1825.7, 1847.7	1825.8 (+Na), 1847.7 (+2Na)	27	2995.2	2995.2 (+2Na)
P	18	-	-	26	2880.2, 2902.2	2880.2 (+2Na), 2902.2 (+3Na)
N	19	-	-	25	-	-
N	20	-	-	24	2669.1	2669.1 (+2Na)
A	21	-	-	23	-	-
E	22	2373.0	2373.0 (+2Na)	22	2484.0	2484.0 (+2Na)
L	23	-	-	21	-	-
S	24	-	-	20	-	-
A	25	-	-	19	2154.9	2154.9 (+2Na)
L	26	-	-	18	-	-
E	27	-	-	17	1970.7	1970.7 (+2Na)
M	28	3039.2	3039.3 (+3Na)	16	1819.7, 1841.7	1819.7 (+Na), 1841.7 (+2Na)
E	29	-	-	15	-	-
E	30	-	-	14	1559.6	1559.6 (+Na)
L	31	-	-	13	-	-
E	32	-	-	12	-	-
S	33	-	-	11	-	-
W	34	-	-	10	1101.4	1101.4 (+Na)
G	35	-	-	9	-	-
A	36	-	-	8	-	-
W	37	-	-	7	-	-
D	38	-	-	6	601.2114	601.2076 (+Na)
G	39	-	-	5	486.1826	486.1807 (+Na)
E	40	-	-	4	-	-
A	41	-	-	3	-	-
T	42	-	-	2	-	-
S	43	-	-	1	-	-






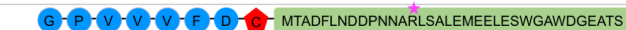
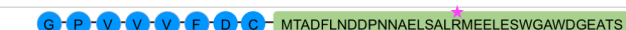


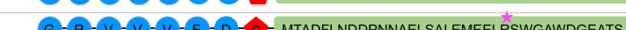
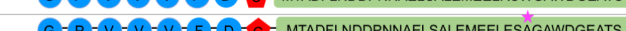




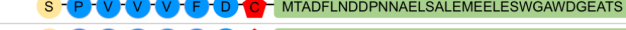








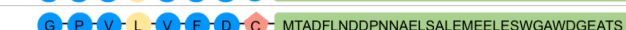

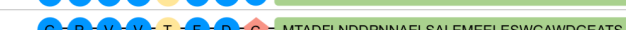

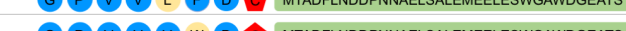
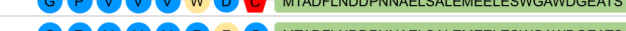

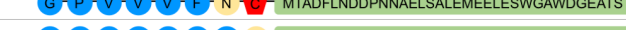



a. Gly1 macrocyclization with Val4 carbonyl.

b. Cys8 heterocyclization with Asp7 carbonyl.

**Table S6:** Summary of all TSA data presented in this manuscript.




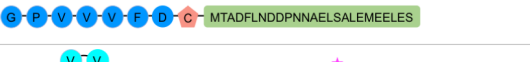

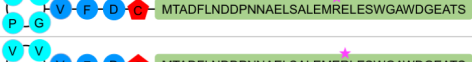
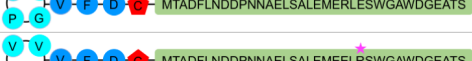
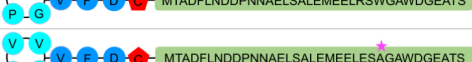
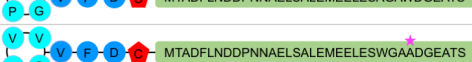
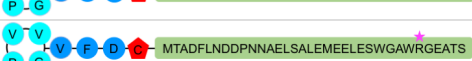

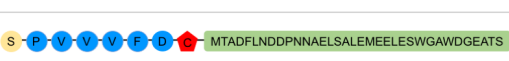
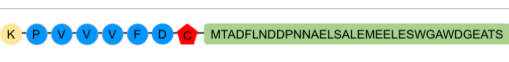



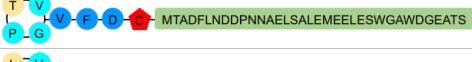


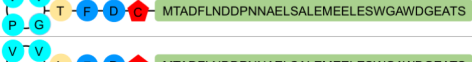
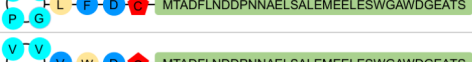
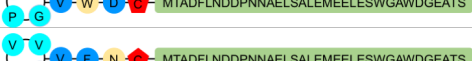
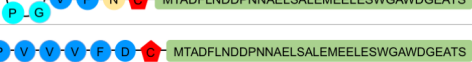
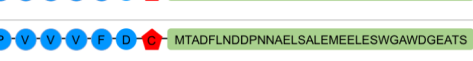


Ligand	IpoC			PurCD		
		T <sub>m</sub> (°C)	ΔT <sub>m</sub> (°C)		T <sub>m</sub> (°C)	ΔT <sub>m</sub> (°C)
-	wt	33	0	wt	24	0
ATP/Mg <sup>2+</sup>	wt	35	2	wt	38	14
BotA	wt	44	11	wt	31	7
BotA <sup>P</sup>	wt	45	12	wt	32	8
BotA <sup>PC</sup>	wt	39	6	wt	33	9
ATP/Mg <sup>2+</sup> + BotA <sup>P</sup>	wt	45	12	wt	40	16
BotA <sup>P</sup> -5 AA	wt	42	9	wt	29	5
BotA <sup>P</sup> -10 AA	wt	35	2	wt	23	-1
BotA <sup>P</sup> -15 AA	wt	31	-2	wt	23	-1
BotA <sup>P</sup> E23R	wt	39	6	wt	28	4
BotA <sup>P</sup> E28R	wt	33	0	wt	24	0
BotA <sup>P</sup> E30R	wt	41	8	wt	29	5
BotA <sup>P</sup> E31R	wt	38	5	wt	26	2
BotA <sup>P</sup> E33R	wt	41	8	wt	31	7
BotA <sup>P</sup> W35A	wt	36	3	wt	24	0
BotA <sup>P</sup> W38A	wt	40	7	wt	27	3
BotA <sup>P</sup> D39R	wt	42	9	wt	29	5
BotA <sup>P</sup> G2A	wt	39	6	wt	28	4
BotA <sup>P</sup> G2S	wt	42	9	wt	29	5
BotA <sup>P</sup> G2K	wt	42	9	wt	30	6
BotA <sup>P</sup> P3A	wt	42	9	wt	29	5
BotA <sup>P</sup> P3G	wt	43	10	wt	30	6
BotA <sup>P</sup> V4D	wt	40	7	wt	29	5
BotA <sup>P</sup> V4T	wt	42	9	wt	30	6
BotA <sup>P</sup> V4L	wt	42	9	wt	30	6
BotA <sup>P</sup> V5D	wt	39	6	wt	28	4
BotA <sup>P</sup> V5T	wt	42	9	wt	29	5
BotA <sup>P</sup> V5L	wt	43	10	wt	31	7
BotA <sup>P</sup> V6D	wt	41	8	wt	30	6
BotA <sup>P</sup> V6T	wt	44	11	wt	30	6
BotA <sup>P</sup> V6L	wt	39	6	wt	28	4
BotA <sup>P</sup> F7W	wt	43	10	wt	31	7
BotA <sup>P</sup> D8E	wt	42	9	wt	30	6
BotA <sup>P</sup> D8N	wt	42	9	wt	30	6
BotA <sup>P</sup> C9S	wt	40	7	wt	28	4
BotA-F (AA 10-44)	wt	41	8	wt	31	7
BotA-F+C9 (AA 9-44)	wt	50	17	wt	33	9
-	E59A	28	0	E94A	28	0
ATP/Mg <sup>2+</sup>	E59A	42	14	E94A	40	12
BotA <sup>P</sup>	E59A	34	6	E94A	37	9
ATP/Mg <sup>2+</sup> + BotA <sup>P</sup>	E59A	43	15	E94A	41	13
-	E59R	34	0	R310A	25	0
ATP/Mg <sup>2+</sup>	E59R	34	0	R310A	36	11
BotA <sup>P</sup>	E59R	41	7	R310A	31	6
ATP/Mg <sup>2+</sup> + BotA <sup>P</sup>	E59R	40	6	R310A	47	22
-	R234A	31	0			
ATP/Mg <sup>2+</sup>	R234A	34	3			
BotA <sup>P</sup>	R234A	41	10			
ATP/Mg <sup>2+</sup> + BotA <sup>P</sup>	R234A	44	13			
-	R234E	31	0			
ATP/Mg <sup>2+</sup>	R234E	34	3			
BotA <sup>P</sup>	R234E	41	10			
ATP/Mg <sup>2+</sup> + BotA <sup>P</sup>	R234E	44	13			

**Table S7:** Summary of all MS data of IpoC reactions presented in this manuscript.

IpoC	BotA	Schematic representation of the detected products	Heterocycle formation
wt	BotA		yes
wt	BotA <sup>P</sup>		yes
wt	BotA <sup>P</sup> -5 AA		yes
wt	BotA <sup>P</sup> -10 AA		partial
wt	BotA <sup>P</sup> -15 AA		no
wt	BotA <sup>P</sup> E23R		yes
wt	BotA <sup>P</sup> E28R		no
wt	BotA <sup>P</sup> E30R		yes
wt	BotA <sup>P</sup> E31R		yes
wt	BotA <sup>P</sup> E33R		yes
wt	BotA <sup>P</sup> W35A		yes
wt	BotA <sup>P</sup> W38A		yes
wt	BotA <sup>P</sup> D39R		yes
wt	BotA <sup>P</sup> G2A		yes
wt	BotA <sup>P</sup> G2S		yes
wt	BotA <sup>P</sup> G2K		yes
wt	BotA <sup>P</sup> P3A		yes
wt	BotA <sup>P</sup> P3G		yes
wt	BotA <sup>P</sup> V4D		partial
wt	BotA <sup>P</sup> V4T		yes
wt	BotA <sup>P</sup> V4L		yes
wt	BotA <sup>P</sup> V5D		no
wt	BotA <sup>P</sup> V5T		no
wt	BotA <sup>P</sup> V5L		partial
wt	BotA <sup>P</sup> V6D		no
wt	BotA <sup>P</sup> V6T		partial
wt	BotA <sup>P</sup> V6L		yes
wt	BotA <sup>P</sup> F7W		yes
wt	BotA <sup>P</sup> D8E		no
wt	BotA <sup>P</sup> D8N		yes
wt	BotA <sup>P</sup> C9S		no
E59A	BotA <sup>P</sup>		no
E59R	BotA <sup>P</sup>		trace
R234A	BotA <sup>P</sup>		partial
R234E	BotA <sup>P</sup>		partial



**Table S8:** Summary of all MS data of PurCD reactions presented in this manuscript.

PurCD	BotA	Schematic representation of the detected products	Macrocycle formation
wt	BotA <sup>P</sup>		partial
wt	BotA <sup>PC</sup>		yes
wt	BotA <sup>PC</sup> -5 AA		yes
wt	BotA <sup>PC</sup> -10 AA		no
wt	BotA <sup>PC</sup> E23R		yes
wt	BotA <sup>PC</sup> E30R		yes
wt	BotA <sup>PC</sup> E31R		yes
wt	BotA <sup>PC</sup> E33R		yes
wt	BotA <sup>PC</sup> W35A		yes
wt	BotA <sup>PC</sup> W38A		yes
wt	BotA <sup>PC</sup> D39R		yes
wt	BotA <sup>PC</sup> G2A		no
wt	BotA <sup>PC</sup> G2S		no
wt	BotA <sup>PC</sup> G2K		no
wt	BotA <sup>PC</sup> P3A		yes
wt	BotA <sup>PC</sup> P3G		partial
wt	BotA <sup>PC</sup> V4D		yes
wt	BotA <sup>PC</sup> V4T		yes
wt	BotA <sup>PC</sup> V4L		partial
wt	BotA <sup>PC</sup> V5L		partial
wt	BotA <sup>PC</sup> V6T		yes
wt	BotA <sup>PC</sup> V6L		yes
wt	BotA <sup>PC</sup> F7W		partial
wt	BotA <sup>PC</sup> D8N		partial
E94A	BotA <sup>PC</sup>		no
R310A	BotA <sup>PC</sup>		no

**Table S9** Bottromycin related metabolites observed in *S. scabiei* *btm* mutants. Compounds containing the macrocycle were only detected in *btmE* R239 mutants.

Strain <sup>a</sup>	Bottromycin-related compounds <sup>b</sup> (m/z)
<i>S. scabiei</i> $\Delta btmF$	406.27, 606.30, 723.32, 737.34, 873.45, 887.47
<i>S. scabiei</i> $\Delta btmF$ pIB139-RBS- <i>btmF</i> E117A	406.27, 606.30, 723.32, 737.34, 873.45, 887.47
<i>S. scabiei</i> $\Delta btmF$ pIB139-RBS- <i>btmF</i> E117R	406.27, 606.30, 723.32, 737.34, 873.45, 887.47
<i>S. scabiei</i> $\Delta btmF$ pIB139-RBS- <i>btmF</i> R341A	406.27, 606.30, 723.32, 737.34, 873.45, 887.47
<i>S. scabiei</i> $\Delta btmF$ pIB139-RBS- <i>btmF</i> R341E	406.27, 606.30, 723.32, 737.34, 873.45, 887.47
<i>S. scabiei</i> $\Delta btmE$	406.27
<i>S. scabiei</i> $\Delta btmE$ pIB139-RBS- <i>btmE</i> E59A	406.27
<i>S. scabiei</i> $\Delta btmE$ pIB139-RBS- <i>btmE</i> E59R	406.27
<i>S. scabiei</i> $\Delta btmE$ pIB139-RBS- <i>btmE</i> R239A	406.27, 795.42, 809.44, 823.45 <sup>c</sup>
<i>S. scabiei</i> $\Delta btmE$ pIB139-RBS- <i>btmE</i> R239E	406.27, 795.42, 809.44, 823.45 <sup>c</sup>

a. The genes *btmF*, *btmE* and *bmtD* from *S. scabiei* encode proteins that are homologs to BotCD/PurCD, BotC/IpoC and BotA, respectively, and the mutated residues are equivalent to the residues mutated for PurCD and IpoC.

b. See Crone et al. 2016<sup>1</sup> for further details on these bottromycin-related compounds. No detected compound contains the macrocycle.

c. 795.42, 809.44 and 823.45 are differentially methylated forms of mature bottromycin. This indicates that the BtmE mutants are active, which is consistent with *in vitro* activity of IpoC mutants.

**Table S10** Primers used to generate *in vivo* mutations in *S. scabiei*.

Primer	Sequence <sup>a</sup>
BtmF start	aggaggac <u>atatg</u> acccggtcccgc
BtmF end	acaggg <u>aattc</u> cagggcggtgccac
BtmF E117A F	gcatggcgagcgcgctgttcGCCagcgccgagcactaccac
BtmF E117A R	gtggtagtgtctcggcgctGGCgaacagcgcgctcgccatgc
BtmF E117R F	gcatggcgagcgcgctgttcCGCagcgccgagcactaccac
BtmF E117R R	gtggtagtgtctcggcgctGCCgaacagcgcgctcgccatgc
BtmF R341A F	cggacgtacgcggtgcagGCCgcgctgctcgagtacctc
BtmF R341A R	gaggtactcgagcagcgcGGCctgcaccgcgtacgtccg
BtmF R341E F	cggacgtacgcggtgcagGAGgcgctgctcgagtacctc
BtmF R341E R	gaggtactcgagcagcgcCTCctgcaccgcgtacgtccg
BtmE start	gggagagggc <u>atatg</u> cgcgaagcg
BtmE end	ctgccgga <u>aattc</u> gtctcgcgctc
BtmE E59A F	cgggcgaaagccctgtcgGCCgccgtggagcggctggtg
BtmE E59A R	caccagccgctccacggcGGCcgacagggctttcgcccg
BtmE E59R F	cgggcgaaagccctgtcgCGGgccgtggagcggctggtg
BtmE E59R R	caccagccgctccacggcCCGcgacagggctttcgcccg
BtmE R239A F	gtcgcggaggcgaccggcGCCgcgctgctggagctggcg
BtmE R239A R	cgccagctccagcagcgcGGCgccggtcgctccgcgac
BtmE R239E F	gtcgcggaggcgaccggcGAGgcgctgctggagctggcg
BtmE R239E R	cgccagctccagcagcgcCTCgccggtcgctccgcgac

a. Underlined nucleotides indicate a restriction site while capital letters indicate mutated codons.

## Materials and Methods

### Cloning of native and mutant PurCD and IpoC

Synthetically produced, codon-optimized genes of the genes with accession numbers WP\_019887079 and WP\_048819674.1 were obtained from Eurofins Genomics and named “PurCD” and “IpoC”, respectively. Using the built-in N-terminal *NcoI* as well as C-terminal *HindIII* restriction sites, the genes were cut out of the provided pEX-K4 cloning vector and ligated with standard T4 DNA ligase protocols using a 1:3 vector to plasmid molar ratio into the expression vectors pHisTEV (PurCD) and pHisSUMOTEV (IpoC), both of which were a gift from Dr. Huanting Liu, St. Andrews University.<sup>[2]</sup> Clones were sequenced using T7 promotor as well as T7 terminator sequencing (LGC genomics) and yielded the expression constructs pHisTEV-PurCD and pHisSUMOTEV-IpoC. Point mutations were designed using the overlap extension method,<sup>[3]</sup> after which the PCR products were cloned into the respective plasmids for each gene.

### Cloning of native BotA and BotA point mutants

Two different vectors were used for BotA precursor peptide expression. In both vectors the restriction sites *NcoI* and *HindIII* were used for cloning:

1. For the production of wt BotA and follower peptide mutants, codon optimized *botA* was cloned into a modified pHisSUMOTEV vector, from which the TEV protease site had been removed. *botA* was amplified using a 5' primer that added a TEV site before the N-terminal methionine of BotA, leading to a TEV site with the amino acid sequence ENLYFQM, where the “M” is the N-terminal methionine of BotA. After TEV cleavage, BotA is released with its native N- terminus.
2. For the production of BotA<sup>P</sup> core peptide mutants, the follower peptide (BotA-F) and follower+C9 (BotA-F+C9), we used unmodified pHisSUMOTEV. Expression of BotA from pHisSUMOTEV when using the *NcoI*-site for cloning of *botA*, would have resulted in peptides with N-terminal cloning artefacts after TEV cleavage (GAMAGPPVVV... (BotA residues are underlined)). To make the native N-terminus of BotA<sup>P</sup> available, we modified the 5' primers to give the N- terminus GAMAGKGPVVV... (BotA residues underlined). This would allow proteolytic processing with Trypsin during BotA<sup>P</sup> purification and liberate the native N-terminus of the peptide.

**Expression of native and mutant PurCD and IpoC**

Expression vectors with verified sequence were transformed into chemically competent *E. coli* Lemo21(DE3) cells (New England Biolabs) using a standard heat shock procedure. Using double selection plates with both Kanamycin (50 µg/mL) as well as Chloramphenicol (34 µg/mL), colonies with both the expression as well as the Lemo21-plasmid were obtained. For expression, a fresh colony was used to inoculate 100 mL of LB Broth starter culture using the Luria-Miller formulation with both antibiotics and was incubated at 37 °C and 200 rpm for 16 h. Large scale expression of pHisTEV-PurCD was carried out in LB Broth whereas large scale expression of pHisSUMOTEV-IpoC was carried out in modified Terrific Broth. Modified Terrific Broth was obtained by adding 1 mM MgSO<sub>4</sub> and NPS buffer<sup>[4]</sup> to a final concentration of 1x after autoclaving instead of the usual potassium phosphate buffer system. Large scale expression for both proteins was carried out by inoculating each medium with both antibiotics at the aforementioned concentrations with the starter culture using a 1/100 dilution. Cultures were grown at 37 °C and 200 rpm until the OD<sub>600</sub> reached 0.8, at which point the agitation was stopped and the culture equilibrated for 1 h while reducing the temperature of the shaker to 16 °C and 18 °C for pHisTEV-PurCD and pHisSUMOTEV-IpoC, respectively. While reaching the desired temperature, protein expression was induced by adding IPTG to a final concentration of 0.1 mM. Cultures were then incubated at 200 rpm for 16 h and 24 h for pHisTEV-PurCD and pHisSUMOTEV-IpoC, respectively.

**Expression of native and mutant BotA**

Expression vectors were transformed into chemically competent *E. coli* Lemo21(DE3) cells (New England Biolabs) using a standard heat shock procedure. Using double selection plates with both Kanamycin (50 µg/mL) as well as Chloramphenicol (34 µg/mL), colonies with both the expression as well as the Lemo21-plasmid could be obtained. For expression, a fresh colony was added into 100 mL of LB Broth starter culture using the Luria-Miller formulation with both antibiotics and was incubated at 37 °C and 200 rpm for 16 h. Large scale expression of both native and mutant BotAs was carried out in LB Broth. Large scale expression was carried out by inoculating LB Broth with both antibiotics at the aforementioned concentrations with the starter culture using a 1:100 dilution. Cultures were grown at 37 °C and 200 rpm until the OD<sub>600</sub> reached 0.8. At this point, protein expression was induced by adding IPTG to a final concentration of 0.5 mM and dropping the temperature to 18 °C. Cultures were incubated at 200 rpm for 16 h.

**Purification of native and mutant PurCD and IpoC**

Cells were harvested by centrifuging the culture at 4,000 rpm and 4 °C for 10 min. The pellet was retrieved and frozen at -80 °C until further use. Two different lysis buffers were used for purification. The pellet of pHisTEV-PurCD was resuspended in PurCD lysis buffer (200 mM NaCl, 50 mM Tris pH 8.0, 20 mM Imidazole pH 8.0, 3 mM  $\beta$ -ME, 10 % Glycerol, final pH 8.7 at RT). Pellets of pHisSUMOTEV-IpoC were resuspended in IpoC lysis buffer (500 mM NaCl, 20 mM Bis-Tris pH 6.8, 20 mM Imidazole pH 8.0, 3 mM  $\beta$ -ME, 10 % Glycerol). In both cases, 100 mL of lysis buffer was used per 25 g of wet cell pellet. Only for IpoC, the lysis buffer was supplemented with cOmplete EDTA-free protease inhibitor tablets (Roche), while DNase (0.4 mg/g wet cells, Sigma) was added to lysis buffers of both proteins. The cell suspension was lysed via passage through a cell disruptor (30 kpsi, Microfluidics Corp.), and the cell debris was removed by centrifugation (40,000 x g, 4 °C, 20 min). The supernatant was filtered through a 0.45  $\mu$ m filter and passed over a 5 mL Ni-NTA superflow cartridge (Qiagen) at a flow rate of 5 mL/min pre-equilibrated in lysis buffer. After application of the lysate and an extensive column wash (30 CV lysis buffer), the bound protein was eluted in one step using lysis buffer supplemented with 250 mM Imidazole. Imidazole was removed by passing the eluate over a desalting column (16/10 GE Healthcare) at a flow rate of 10 mL/min pre-equilibrated in lysis buffer. The protein was then incubated with TEV protease for 14 h at 4 °C at a 1:10 protein:TEV ratio to cleave the expression and purification tags. Digested protein was reapplied to a 5 mL Ni-NTA-Superflow cartridge equilibrated in lysis buffer and the flow-through was collected at a flow rate of 5 mL/min. Finally, proteins were polished by size exclusion chromatography (Superdex 200 16/600, GE Healthcare) pre-equilibrated in the corresponding gel filtration buffers (PurCD: 200 mM NaCl, 50 mM Tris, 10 % Glycerol, 0.5 mM TCEP, pH 8.7 at RT; IpoC: 300 mM NaCl, 10 mM HEPES, 10 % Glycerol, 0.5 mM TCEP, final pH 7.4 at RT) at a flow rate of 1 mL/min. IpoC and PurCD and the respective enzyme point mutants eluted at a retention volume of 88 mL and 85 mL, respectively. Protein purity and integrity was verified by SDS-PAGE as well as intact protein mass spectrometry using a previously described method<sup>[5]</sup>. The enzymes were concentrated using a 30 kDa cutoff filter (Sartorius AG). The protein concentration was determined at a Spectrophotometer (Thermo Fisher Scientific) using the extinction coefficients calculated from the amino acid sequence by the ExPASy ProtParam Server<sup>[6]</sup> (<http://web.expasy.org/protparam/>). In general, all purification steps were performed at 4 °C. Enzymes at a concentration of 120  $\mu$ M were flash frozen in single-use aliquots and stored at -80 °C until further use.

### Purification of native and mutant BotA

All BotA purifications were carried out in the same lysis buffer (500 mM NaCl, 20 mM Tris pH 8.0, 20 mM imidazole pH 8.0, 3 mM  $\beta$ -ME). For every 25 g of wet cell pellet, 100 mL of lysis buffer was added. The lysis buffer was supplemented with cOmplete EDTA-free protease inhibitor tablets (Roche) and DNase (0.4 mg/g wet cell pellet, Sigma). The cell suspension was lysed via passage through a cell disruptor (30 kpsi, Microfluidics Corp.), and the cell debris was removed by centrifugation (40,000  $\times$  g, 4 °C, 20 min). The supernatant was filtered through a 0.45  $\mu$ m filter and applied to a His-Trap HP 5 mL column (GE Healthcare) pre-equilibrated in lysis buffer at a flow rate of 5 mL/min. After application of the lysate and an extensive column wash (30 CV lysis buffer), the bound protein was eluted using a lysis buffer supplemented with 250 mM imidazole. The proteins were passed over a desalting column (16/10 GE Healthcare) at a flow rate of 10 mL/min pre-equilibrated in lysis buffer. The proteins were subsequently incubated with TEV protease for 2 h at room temperature at a 1:10 ratio to remove the His<sub>6</sub>-tagged SUMO. Digested protein was passed over a 5 mL His Trap HP column at a flow rate of 5 mL/min to remove the expression and purification tag and the flow-through was collected. Peptides expressed using the pHisSUMOTEV vector were subsequently incubated with trypsin at a ratio of 1:100 for 1 h at 37 °C to generate the native N-terminus of the peptides. Finally, the proteins were passed over a Superdex 30 16/60 size exclusion chromatography column (GE Healthcare) pre-equilibrated in gel filtration buffer (150 mM NaCl, 10 mM HEPES, 0.5 mM TCEP, pH 7.4) at a flow rate of 1 mL/min. This purification step yielded a pure and homogenous sample that eluted at a retention volume of 64 mL. Following the size exclusion chromatography, BotA follower mutants were incubated with 2  $\mu$ M BotP and 200  $\mu$ M CoCl<sub>2</sub> for 1 h at 37 °C to remove the N-terminal methionine. Afterwards, they were passed over a Superdex Peptide 10/300 GL to remove the BotP. Protein purity and integrity was verified by SDS-PAGE as well as mass spectrometry.

### LC-MS analysis

All measurements to analyse the mass of processed or unprocessed BotA peptides were performed on a Dionex Ultimate 3000 RSLC system using a BEH C18, 100  $\times$  2.1 mm, 3.5  $\mu$ m dp column equipped with a C18 precolumn (Waters). Samples of 1  $\mu$ L were separated by a gradient from (A) H<sub>2</sub>O + 0.1 % formic acid to (B) ACN + 0.1 % formic acid at a flow rate of 800  $\mu$ L/min and 45 °C. The gradient was initiated by a 0.5 min isocratic step at 5 % B, followed by an increase to 35 % B in 2.5 min, to 42.5 % B in 3 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. Full length BotA and BotA<sup>P</sup> peptides showed a distinct peak at a retention time of 4.6 min. Peptides with a thiazoline, introduced by IpoC, had a retention time of 4.3 min, while peptides with a macroamidine cycle had a retention time of 4.2 min. Due to the size, BotA truncations showed different retention times and eluted later from the column. For quantification of processed BotAs in enzyme reactions the peak area of [M+H]<sup>3+</sup> was integrated and concentrations were calculated using a calibration curve obtained from serial dilutions of the respective purified BotA using the TASQ 1.1 software (Bruker Daltonics).

### **Thermal shift assay (TSA)**

Protein melting temperatures  $T_m$  were determined by monitoring protein unfolding using SYPRO orange as fluorescence probe. Enzymes were diluted to 5  $\mu$ M in buffer containing 5x SYPRO orange (Sigma). For the determination of stabilising effects of the enzymes by peptides or other additives, buffer containing 50 mM HEPES, 200 mM NaCl and 10 % glycerol (pH 8.5) was used. Final peptides concentrations of 30 and 50  $\mu$ M for IpoC and PurCD, respectively, were used. To investigate the effects of ATP and MgCl<sub>2</sub>, final concentrations of 5 mM each were used. To determine the stability of PurCD at different pH, buffers containing 50 mM buffering agent, 200 mM NaCl and 10 % glycerol were used. For pH 7.0 to 9.0 Tris and pH 9.5 to 11.0 CHES were used as buffering agents. Samples (40  $\mu$ L/well) were analysed in 96 well plates (qPCR semi-skirted 96well PCR plates (peqlab)) that were sealed (adhesive qPCR film (peqlab)). Measurements were carried out with a realtime PCR machine (peqSTAR 96Q) using the SYBR Green I filter and a temperature gradient from 10 to 80 °C with stepwise increments of 1 °C and 1 min hold. After each temperature step the fluorescence intensity was measured. The melting temperatures ( $T_m$ ) were obtained using the derivative method. All conditions were tested in triplicates and mean values were calculated for the graphic presentation.

### ***o*-Phthalaldehyd peptide concentration assay**

Generally, peptide concentrations were determined by measuring the absorbance at 280 nm using a spectrophotometer (Thermo Scientific) and calculated using extinction coefficients (see above). For truncated BotA variants without tryptophan residues (BotA-10 and BotA-15AA), *o*-phthalaldehyd (OPA)<sup>[7]</sup> was used as fluorogenic reagent, which reacts in presence of 2-mercaptoethanol with primary amino groups (e.g. of peptides) and forms a blue coloured fluorescent product. OPA reagent solution was freshly made by dissolving 10 mg OPA (Fluka) in 250  $\mu$ L ethanol, 9.8 mL PBS



and 20  $\mu$ L 2-mercaptoethanol. The OPA peptide concentration assay was performed using 96-well flat-bottom plates. Full length BotA was used as the standard in the concentration range 0-100  $\mu$ M. 200  $\mu$ L OPA reagent was added to 30  $\mu$ L of sample/well and incubated 5 min at room temperature. Fluorescence intensity was measured at 360 nm excitation and 430 nm emission using an Infinite 200 Pro (Tecan) plate reader. Samples and standards were tested in triplicates.

### Heterocyclization of BotA with IpoC

For heterocyclization reactions, in general, 50  $\mu$ M substrate (BotA, BotA<sup>P</sup>, or BotA<sup>P</sup> core or follower mutants) was incubated with 5  $\mu$ M wt IpoC and 5 mM ATP/MgCl<sub>2</sub> in gel-filtration buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 1 mM TCEP) for 12 h at 37 °C. Turnover of the different BotA variants by IpoC was analyzed by LC-MS. For further use (e.g. enzyme reactions with PurCD or MS<sup>2</sup>-analysis), large scale reactions of heterocyclized BotA (BotA<sup>PC</sup> or BotA<sup>C</sup>) were purified by size-exclusion chromatography on a Superdex Peptide 10/300 GL at 0.5 mL/min using gel-filtration buffer. Peptides eluted between 9.5 and 11.5 mL, were concentrated using 3 kDa cutoff filters (Sartorius AG) and peptide concentrations were determined spectrophotometrically as described above. To determine the time until complete turnover was reached, enzyme reactions using BotA and BotA<sup>P</sup> were set up as described above using wt IpoC. Samples were taken and directly frozen in liquid nitrogen after 0, 0.5, 1, 2, 4, 8 and 16 h. The conversion of BotA and BotA<sup>P</sup> after the different incubation times was subsequently analyzed by LC-MS (see above for details). To test the enzymatic activity of IpoC mutants (IpoC E59A, E59R, R234A, R234E), BotA<sup>P</sup> was used as the substrate and reactions were set up and analyzed as describe above. For the determination of the initial conversion rates of BotA and BotA<sup>P</sup> by IpoC, 100  $\mu$ M substrate was incubated with 2  $\mu$ M wt IpoC and 5 mM ATP/MgCl<sub>2</sub> in gel filtration buffer. Samples were incubated in the LC sample holder at 37 °C and 1  $\mu$ L aliquots were automatically taken and analyzed every 10 min by the LC-MS system in the time range from 2 to 42 min. All reactions (different substrates, reaction conditions, IpoC mutants) were tested in triplicates.

### Macrocyclization reaction by PurCD

For macrocyclization reactions by PurCD, 50  $\mu$ M BotA peptide (BotA<sup>P</sup>, BotA<sup>PC</sup>) was incubated with 10  $\mu$ M enzyme and 10 mM ATP/MgCl<sub>2</sub> in macrocyclization buffer (50 mM Tris pH 9.5, 200 mM NaCl, 10 % glycerol) for 90 min at room temperature. Macroamidine formation was determined by LC-MS analysis as described above. Product (BotA<sup>PCCD</sup>) was purified for MS<sup>2</sup>-analysis and further experiments by size-exclusion chromatography on a Superdex peptide 10/300 GL column using gel filtration buffer at a flow rate of 0.5 mL/min. BotA<sup>PCCD</sup> eluted between 10.0 and 12.0 mL. Peptides were concentrated using 3 kDa cutoff filters (Sartorius AG) and peptide concentrations were determined spectrophotometrically as described above. To determine the effect of different pHs on PurCD

turnover, buffers containing 50 mM buffering agent, 200 mM NaCl and 10 % glycerol were used. For pHs 7.0 to 8.5 Tris and pHs 9.0-11.0 CHES were used as buffering agents. Reactions were set up as described above (50  $\mu$ M BotA<sup>PC</sup>, 10  $\mu$ M PurCD, 10 mM ATP/MgCl<sub>2</sub>) and analyzed by LC-MS. The ability of PurCD to use different BotA<sup>P</sup> follower and core mutants as substrate was tested. As PurCD strongly favored IpoC-cyclized BotA<sup>P</sup> peptides, 50  $\mu$ M substrate was first incubated with 5  $\mu$ M IpoC and 10 mM ATP/MgCl<sub>2</sub> in macrocyclization buffer for 12 h at 37 °C. Samples were cooled down to room temperature, 10  $\mu$ M PurCD was added and samples were incubated for a further 90 min at room temperature. The conversion of the BotA<sup>P</sup> mutants was analyzed by LC-MS. To determine the conversion rate of BotA<sup>PC</sup> by PurCD, 50  $\mu$ M substrate was incubated with 10  $\mu$ M PurCD and 10 mM ATP/MgCl<sub>2</sub>. Samples were incubated in the LC sample holder at 20 °C and 1  $\mu$ L aliquots were automatically taken and analyzed every 10 min by the LC-MS system in the time range from 2 to 122 min. To test the enzymatic activity of PurCD mutants (PurCD E94A, R310A), BotA<sup>PC</sup> was used as substrate (50  $\mu$ M BotA<sup>PC</sup>, 10  $\mu$ M PurCD mutant, 10 mM ATP/MgCl<sub>2</sub> in macrocyclization buffer). The enzymatic activity of the PurCD mutants was analyzed by LC-MS as described above. We observed macroamidine ring opening after extended incubation time of the reactions. To test the enzyme dependence of the ring opening, 50  $\mu$ M purified BotA<sup>PCCD</sup> was incubated for 12 h at RT in macrocyclization buffer with 1. no additive, 2. 10 mM ATP/MgCl<sub>2</sub>, 3. 10  $\mu$ M PurCD, 4. 10  $\mu$ M PurCD + 10 mM ATP/MgCl<sub>2</sub>. Macroamidine ring opening was analyzed by LC-MS-analysis as described above. To determine if PurCD was stable in the macrocyclization buffer for at least 8 h and thus catalyze re-opening of the macroamidine, we performed analytical gel filtration experiments. 10  $\mu$ M PurCD was incubated at room temperature in macrocyclization buffer (200 mM NaCl, 50 mM Tris, 10 % Glycerol, pH 9.5) with addition of 10 mM ATP and 10 mM MgCl<sub>2</sub>. After 0, 1.5 and 8 h, the full reaction mixture was injected onto a Superose 6 Increase 10/300 GL column at a flow rate of 0.5 mL/min. The resulting chromatogram was analyzed by plotting the absorption at 280 nm against the retention volume. All reactions (different substrates, reaction conditions, PurCD mutants) were tested in triplicates.

### **MS<sup>2</sup> analysis of BotA peptides**

Tandem MS data were acquired on a Synapt G2-Si mass spectrometer (Waters) operated in electrospray ionization positive ion mode with a scan time of 1 s (inter scan time of 0.015 s) in a mass range of  $m/z$  50 to 5000. Samples were diluted into 50% methanol/water containing 0.1% formic acid and infused into the mass spectrometer at 5-10  $\mu$ L/min over a period of 2 min using a Harvard Apparatus syringe pump. Data were collected in continuum mode and the following parameters were used: collision gas = argon; capillary voltage = 2.60 kV; cone voltage = 40 V; lockspray capillary voltage = 3.5 kV; source temperature = 100 °C; desolvation temperature = 250 °C; cone gas flow = 10 L/h; desolvation gas flow = 500 L/h. For each peptide, the most intense multiply charged peak in the MS1

was selected for MS<sup>2</sup> fragmentation, where the trap collision energy was ramped stepwise from 10 V until the precursor was fully fragmented, which usually occurred in the range of 40-45 V. Acquisition and analysis of data was carried out using MassLynx 4.1 (Waters). The reported MS<sup>2</sup> spectra are the average of approximately 100 scans, and MS<sup>2</sup> spectra were deconvoluted using the MaxEnt 3 tool in MassLynx 4.1.

### Construction of *btmF* point mutants in *Streptomyces scabies*

To assess the *in vivo* effect of mutations to a BotCD-like protein in a native producer, four different point mutants were generated in *btmF* (E117A, E117R, R341A, R341E), the gene encoding the BotCD homolog in the *S. scabies* bottromycin gene cluster. These mutations are equivalent to the ones tested for *in vitro* activity of PurCD. These were generated by fusion PCR, using internal overlapping primers containing the desired codon change with external primers covering the start and end of *btmF* (see Table S10 for primers). The final fusion PCR products were digested and ligated into the NdeI and EcoRI sites of the pIB139-RBS vector<sup>[1]</sup> to generate pIB139-RBS-*btmF* E117A, pIB139-RBS-*btmF* E117R, pIB139-RBS-*btmF* R341A and pIB139-RBS-*btmF* R341E. These four constructs were introduced by intergeneric conjugation from *E. coli* ET12567/pUZ8002 into *S. scabies*  $\Delta$ *btmF*<sup>[1]</sup> to study the effect of each *btmF* point mutation in the complementation of the chromosomal deletion (note: pIB139-RBS-*btmF* had previously been determined to fully restore bottromycin production in *S. scabies*  $\Delta$ *btmF*).<sup>[1]</sup> Following selection with apramycin (50 µg/mL) and nalidixic acid (30 µg/mL) in SFM medium, three different exconjugants for each point mutation were independently inoculated into 5 mL of TSB medium and grown for 48 h. 500 µL of each culture were employed as seed cultures to inoculate 5 mL of bottromycin production medium (BPM)<sup>[1]</sup> in 50 mL tubes capped with foam bungs. After 96 h of fermentation (250 rpm, 30 °C), 500 µL of each culture was mixed with 500 µL methanol, shaken for 30 min at RT and centrifuged at 13,000 rpm for 30 min before analysis of the resulting supernatant by LC-MS as previously described.<sup>[1]</sup> Bottromycin-related metabolites were manually identified using Browser and Postrun analysis software (Shimadzu). Untargeted metabolomic analysis using Profiling Solution (Shimadzu) indicated that there were no significant differences between the total metabolome of *S. scabies*  $\Delta$ *btmF* and the metabolomes of any of the *btmF* point mutants.

### Construction of *btmE* point mutants in *Streptomyces scabies*

To assess the *in vivo* effect of mutations to a BotC-like protein in a native producer, four different point mutants were generated in *btmE* (E59A, E59R, R239A, R239E), the gene encoding the BotC homolog in the *S. scabies* bottromycin gene cluster. These mutations are equivalent to the ones tested for *in vitro* activity of IpoC. These were generated by fusion PCR, using internal overlapping primers containing the desired codon change with external primers covering the start and end of *btmE* (see

Table S10 for primers). The final fusion PCR products were digested and ligated into the NdeI and EcoRI sites of pIB139-RBS to generate pIB139-RBS-*btmE* E59A, pIB139-RBS- *btmE* E59R, pIB139-RBS-*btmE* R239A and pIB139-RBS-*btmE* R239E. These four constructs were introduced by intergeneric conjugation from *E. coli* ET12567/pUZ8002 into *S. scabiei*  $\Delta btmE$ <sup>[1]</sup> (note: pIB139-RBS-*btmE* had previously been determined to fully restore bottromycin production in *S. scabiei*  $\Delta btmE$ ).<sup>[1]</sup> Following selection with apramycin (50 µg/mL) and nalidixic acid (30 µg/mL) in SFM medium, three different exconjugants for each point mutation were independently inoculated into 5 mL of TSB medium and grown for 48 h. 500 µL of each culture were employed as seed cultures to inoculate 5 mL of BPM in 50 mL tubes capped with foam bungs. After 96 h of fermentation (250 rpm, 30 °C), 500 µL of each culture was mixed with 500 µL methanol, shaken for 30 min at RT and centrifuged at 13,000 rpm for 30 min before analysis of the resulting supernatant by LC-MS as previously described.<sup>[1]</sup> Bottromycin-related metabolites were manually identified using Browser and Postrun analysis software (Shimadzu). Untargeted metabolomic analysis using Profiling Solution (Shimadzu) indicated that there were no significant differences between the total metabolome of *S. scabiei*  $\Delta btmE$  and the metabolomes of any of the *btmE* E59 point mutants, whereas bottromycin was detected in the R239 mutants (Table S9). This is consistent with the *in vitro* activity of IpoC mutants (Figure S18).

## Supplementary References

- [1] Crone, W. J.; Vior, N. M.; Santos-Aberturas, J.; Schmitz, L. G.; Leeper, F. J.; Truman, A. W., *Angew Chem Int Ed Engl* **2016**, 55 (33), 9639-43.
- [2] Liu, H.; Naismith, J. H., *Protein Expr Purif* **2009**, 63 (2), 102-11.
- [3] Higuchi, R.; Krummel, B.; Saiki, R. K., *Nucleic Acids Res* **1988**, 16 (15), 7351-67.
- [4] Studier, F. W., *Protein Expr Purif* **2005**, 41 (1), 207-34.
- [5] von Tesmar, A.; Hoffmann, M.; Pippel, J.; Fayad, A. A.; Dausend-Werner, S.; Bauer, A.; Blankenfeldt, W.; Müller, R., Total Biosynthesis of the Pyrrolo[4,2]benzodiazepine Scaffold Tomaymycin on an In Vitro Reconstituted NRPS System. *Cell Chem Biol* **2017**, 24 (10), 1216-1227 e8.
- [6] Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M.R.; Appel, R.D.; Bairoch, A.; *The Proteomics Protocols Handbook*, Humana Press **2005**, 571-607.
- [7] Roth, M., Fluorescence reaction for amino acids. *Anal Chem* **1971**, 43 (7), 880- 2.

## Chapter 4

# Structural analysis of IpoC, an unusual YcaO cyclodehydratase from the Bottromycin pathway

**Sebastian Adam**<sup>1[a]</sup>, Ramona Duman<sup>1[b]</sup>, Armin Wagner<sup>[b]</sup> and Jesko Koehnke<sup>[a]\*</sup>

*Manuscript in preparation*

### Affiliation

<sup>[a]</sup> Workgroup Structural Biology of Biosynthetic Enzymes, Helmholtz Institute for Pharmaceutical Research, Helmholtz Centre for Infection Research, Saarland University, Universitäts-campus E8 1, 66123, Saarbrücken, Germany.

<sup>[b]</sup> Diamond Light Source, Harwell Science and Innovation Campus, Chilton, Didcot OX11 0DE, UK.

<sup>1</sup> These authors contributed equally to the manuscript

## **Contributions and Acknowledgements**

### **Author's effort:**

The author contributed significantly to the manuscript by establishing the first reliable protein expression and purification protocol for IpoC, allowing him to work on protein crystallization. He applied a combination of several protein engineering techniques, for instance limited proteolysis as well as buffer and construct optimization. He designed and cloned several IpoC extensions and truncations, leading to the first diffraction quality crystals. Furthermore, he collected a first dataset at Swiss Light Source. The author refined the IpoC structure, performed structural analysis and wrote the manuscript.

### **Other's effort:**

Ramona Duman contributed to solving the apo IpoC structure by collecting the Sulphur single anomalous dispersion (S-SAD) datasets at a special beamline at Diamond Light Source, which was needed to obtain the phase information to solve the crystal structure as there were no molecular replacement models available in the PDB.

## 4.1 Abstract

Bottromycins are interesting natural products with promising bioactivity against gram-positive bacteria, which contain two post-translational modifications being catalyzed by YcaO domain enzymes. Recent advances in YcaO protein biochemistry allowed researchers to gain insights into this fascinating superfamily with different biosynthetic scopes found in RiPPs biosynthesis. Unfortunately, they remain poorly understood on a structural basis as there are only a handful of crystal structures available to date. IpoC is a YcaO protein that was shown to catalyze the thiazoline formation in Bottromycins, an essential step for efficient downstream processing by other maturation enzymes. Within the YcaO superfamily, IpoC is a special protein as its functionality was shown to be dependent on binding to the BotA follower peptide rather than binding to an E1-like protein, which was observed for other YcaO domain enzymes. Here, we present the apo crystal structure of IpoC and investigate a possible enzyme-substrate relationship. We demonstrate a potentially biologically relevant dimer interface formation of IpoC and discuss its implications for the biosynthetic machinery. Intriguingly, IpoC shows structural similarity to other YcaO enzymes from different biosynthetic pathways, while the respective sequence homology is very low. Therefore, we can assume that the different biosynthetic scopes of the YcaO superfamily do not necessarily correlate with specific structural features. Even though insights into precursor peptide binding as well as mechanistical differences to other YcaO enzymes require an IpoC complex structure, our data provide the first examination of a YcaO crystal structure involved in Bottromycin biosynthesis.

## 4.2 Introduction

Antimicrobial resistance of pathogens is one of the major health challenges of the 21<sup>st</sup> century, with the number of reported cases for long-approved clinical antibiotics like Carbapenem increasing every year.<sup>[1]</sup> Given the evolutionary-based resistance development of bacteria, it is not surprising that derivatives of known antibiotics share cross-resistance with the latter compound due to similar chemical scaffolds, sometimes to the extent of resistance mechanisms being developed before its introduction to the clinic.<sup>[2]</sup> Therefore, there is an urgent need for anti-infectives with new chemical scaffolds as these molecules can often be associated with a novel mode of action.<sup>[3]</sup> In this context, natural products offer unexplored



molecular frameworks, which can often serve as chemical lead structures for pharmaceutical researchers to further improve the properties of these compounds, sometimes with diverse bioactivities as in the case of Rapamycin.<sup>[4][5]</sup>

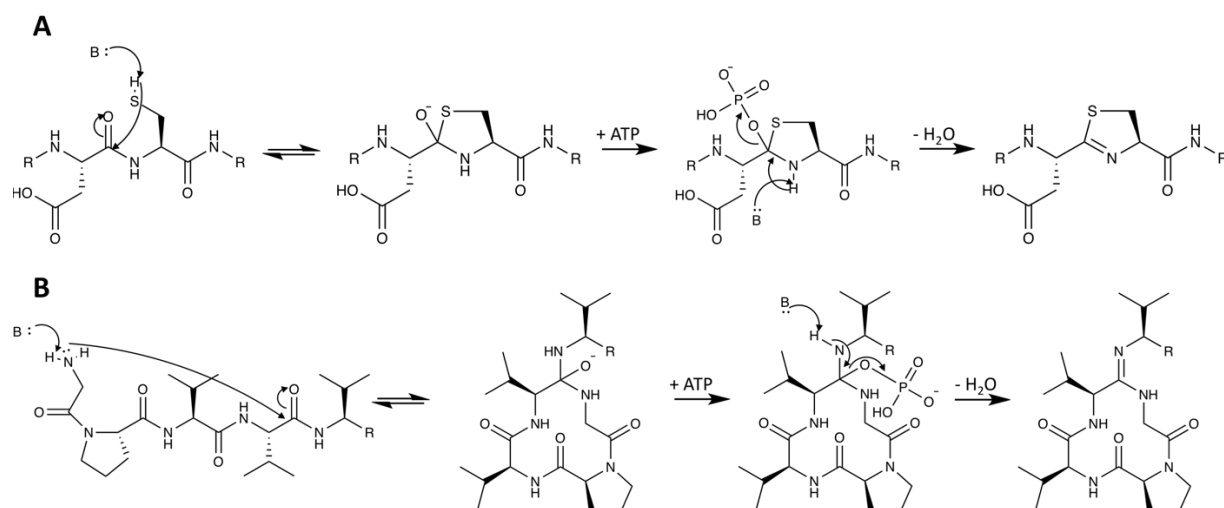
Bottromycins are natural products first isolated from a *Streptomyces* species more than half a century ago, which have gotten into pharmaceutical focus again in the last 10 years due to their as of yet unexploited target, the A-site of the bacterial ribosome, and their activity against clinically relevant pathogens, such as Vancomycin-resistant *Enterococci* (VRE).<sup>[6]</sup> The identification of the biosynthetic gene cluster of Bottromycin led to its classification as a ribosomally synthesized and post-translationally modified peptide (RiPP). This functionally and structurally diverse class of natural products originates from a ribosomally synthesized precursor peptide.<sup>[7][8]</sup> This precursor peptide can be divided into a core peptide, which gets post-translationally modified to become the natural product and a leader, or in the case of Bottromycins follower, peptide, which is associated with substrate recognition of the maturation enzymes.<sup>[9]</sup> As the core peptide in Bottromycins is located at the N-terminus of the precursor peptide, the aminopeptidase BotP is removing the N-terminal methionine as the first biosynthetic step, resulting in the precursor peptide BotA<sup>P</sup>.<sup>[10]</sup> Within the biosynthetic gene cluster of Bottromycin, we find two genes with conserved domains showing sequence homology to the *YcaO* superfamily (formerly DUF181), the biosynthetic genes “C” and “CD” (Figure 1).



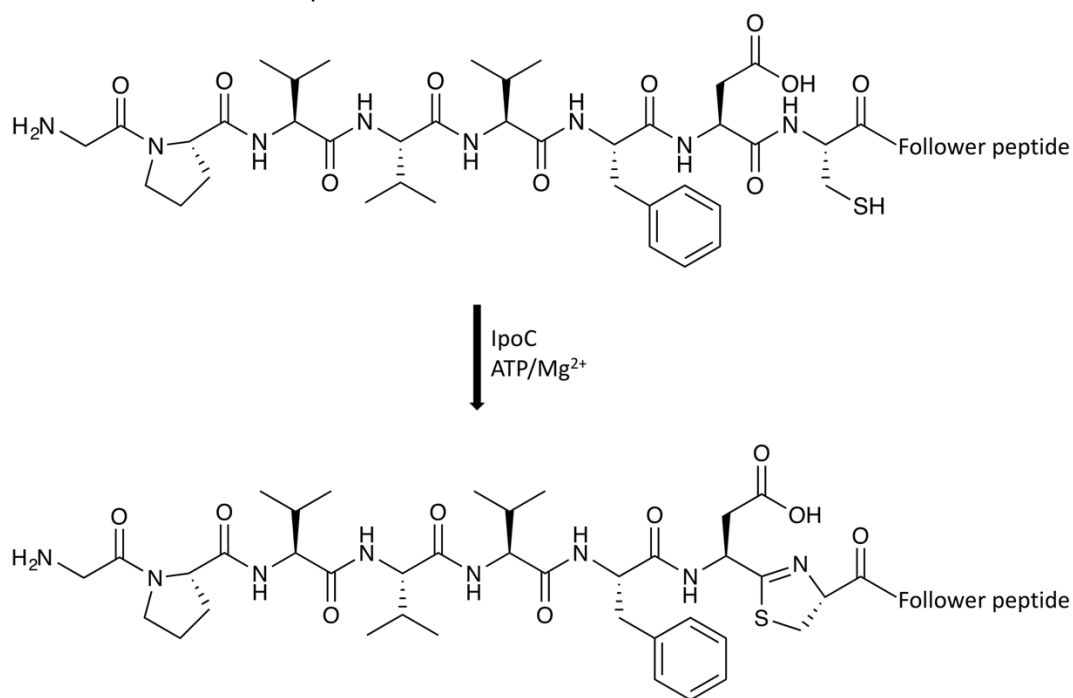
**Figure 1.** Schematic of the annotated Bottromycin biosynthetic gene cluster from *Streptomyces* sp. BC16019.<sup>[7]</sup> The two biosynthetic genes with sequence homology to the *YcaO* superfamily C and CD are shown in mint green, the gene encoding for the precursor peptide BotA is shown in black.

Consequently, the final chemical structure of Bottromycin contains two modifications catalyzed by *YcaO* domain enzymes as was shown by *in vitro* studies, a macroamidine and a thiazoline (which later gets oxidized to become a thiazole).<sup>[11,12]</sup> This observation suggested that *YcaO* functionality proceeds through a common mechanism, which had already been shown before for heterocycle formation,<sup>[13]</sup> the attack of different nucleophiles onto a backbone amide, followed by the ATP-driven phosphorylation of the formed hemiorthoamide intermediate. The elimination of phosphate concludes the reaction, resulting in the loss of a water molecule (Figure 2).<sup>[12]</sup> Furthermore, the *in vitro* reconstitution of both modifications in

recombinantly produced BotA<sup>P</sup> precursor peptide revealed the biosynthetic order of these modifications as efficient catalysis of macroamidine formation was only possible in peptides already harboring the thiazoline.<sup>[12]</sup> Therefore, thiazoline formation catalyzed by the “C” protein (in our case *IpoC*) is a necessary maturation step for the *in vivo* production of Bottromycin. The cysteine-derived thiazoline formation of the core peptide catalyzed by *IpoC* is shown in **Figure 3**.



**Figure 2.** Proposed mechanism of the two *YcaO* catalyzed reactions within the Bottromycin biosynthetic pathway. A) Azoline-forming *YcaO* reaction. Nucleophilic attack of a cysteine side chain onto the preceding backbone amide and formation of a hemiorthoamide. ATP-driven phosphorylation of the amide oxygen, elimination of phosphate and loss of water conclude the reaction. B) Macroamidine-forming *YcaO* reaction. The reaction is suggested to work in a similar fashion using the N-terminal amine as the nucleophile.



**Figure 3.** Thiazoline-formation in the Bottromycin biosynthetic pathway is catalyzed by *IpoC*. Only core peptide residues are shown.

YcaO domain enzymes have recently been mechanistically characterized and functionally divided into 5 different classes.<sup>[14]</sup> By far the best-studied class of YcaO proteins are the azoline-forming YcaO domains, which have been of interest for researchers since the discovery of Microcin B17 (MccB17), which has already been postulated to be post-translationally modified after ribosomal biosynthesis in 1994.<sup>[15]</sup> This azoline-forming YcaO domain enzyme has been found to be engaged in complex formation with a dehydrogenase as well as a protein of the E1-ubiquitin activating (E1-like) superfamily.<sup>[16]</sup> Extensive research in the field established E1-like proteins as essential partners of YcaO domain enzymes, as they were shown to potentiate heterocyclase activity by nearly 1,000-fold through complex formation in the case of a YcaO protein from *Bacillus* sp. Al Hakam.<sup>[13]</sup> E1-like proteins have thus far been found to be comprised of a conserved structural feature, the RiPP recognition element (RRE) involved in substrate peptide binding, which has been found in various RiPP clusters and is linked to guiding the substrate into the catalytic center of the protein.<sup>[17]</sup> In contrast, it was proposed that both YcaO domain enzymes linked to Bottromycin comprise a unique class due to their independence of a partner protein required for substrate recognition (standalone YcaO domain enzymes).<sup>[14]</sup> As such, they might display structural differences concerning substrate binding and catalysis.

Even though a lot of progress has been made regarding YcaO functionality, this enzyme class is still one of the most poorly understood on a structural basis as there are only five different structures available in the Protein Data Bank to date.<sup>[18][19][20][21][22]</sup> This work presents the first apo crystal structure of a potentially unique class of YcaO domain enzymes involved in Bottromycin biosynthesis.

## 4.3 Results and Discussion

### Crystallization of IpoC\_apo

In an attempt to maximize our chances of soluble expression and crystallization of this interesting cyclodehydratase, we set out to clone and express several sequence homologues of this YcaO protein. Interestingly, only four proteins other than homologues from other Bottromycin-producing strains show a sequence homology to IpoC when running a PSI-BLAST with a standard threshold of 0.005. Portrayed as a sequence similarity network, we only found one other distantly related cluster of proteins from various strains showing some sequence homology to IpoC other than the closely related *Streptomyces* strains (**Figure S1**). We randomly selected and cloned 7 different “C” homologues from these strains carrying a Bottromycin biosynthetic gene cluster and attempted protein expression trials in *E. coli*. The only reliable, soluble protein expression with a yield of 3 mg L<sup>-1</sup> could be observed for IpoC (WP\_048819674.1). A table of every cloned and tested homologue can be found in the Supporting Information (**Table S1**). After a large-scale expression and purification protocol for IpoC was developed, this construct was chosen for initial crystal trials.

At first, IpoC crystallization attempts were performed with the construct “WP\_048819674.1”.<sup>[12]</sup> Crystallization trials were initially set up using apo protein and protein pre-incubated with cofactors ATP/MgCl<sub>2</sub>. Additionally, the crystallization properties of IpoC were analyzed with variable pre-incubated substrate peptides as the protein was expected to show a decrease in flexibility upon substrate binding. The amino acid sequence of every pre-incubated substrate peptide is shown in **Table S2**. BotA\_Follower variants were chosen especially as IpoC has shown a specific increase in melting temperature upon incubation with these peptides in thermal shift assays, which is generally correlated with an increase in stability.<sup>[12]</sup> Interestingly, using an engineered follower peptide that contains an N-terminal cysteine residue (BotA\_follower+C9) mimicking the substrate residue of BotA<sup>P</sup> further increases IpoC melting temperature by 9 °C compared to the BotA\_follower peptide, highlighting the importance of this residue.<sup>[12]</sup> These previous observations led to the hypothesis of the core peptide being involved in substrate binding, which has not been observed for other azoline-forming YcaO domain enzymes thus far.

None of these attempts led to initial crystallization hits in sparse matrix screens set up either at 277 or 291 K. A deeper look into the corresponding genome sequence of *Streptomyces*

*ipomoeae* revealed an alternative start codon 12 amino acids upstream of the start codon of “WP\_048819674.1”, which was also found when running a BLAST search in an assembly record by Huguet-Tapia *et al.* (**Figure S2**).<sup>[23]</sup> This construct was cloned and expressed with the premature C-terminal stop codon PAHHA- as this truncated protein was found to be a stable proteolysis construct by mass spectrometry while expressing “WP\_048819674.1” (**Figure S3**). A schematic of every cloned construct that could be expressed and purified is shown in **Figure S4**, the crystallized construct is marked with a star. Lastly, the crystallization buffer was optimized to contain 1 M NaCl, 5% Glycerole, 20 mM HEPES and 1 mM DTT with a final pH of 7.0 as it was suspected to increase the solubility of this protein due to a general correlation between a positive net charge of proteins and low solubility.<sup>[24]</sup> Taken this into account as well as previous observations of low solubility in buffers with low ionic strength, increased the likelihood of higher solubility of this protein in buffers with higher salt concentration. Additionally, a theoretical pI above 9 and the observation of a study determining the optimal pH for initial protein crystallization trials to sometimes be more than 2 units away from the proteins pI made us change the pH to 7.0 while doubling the buffering strength of the solution to a final concentration of 20 mM.<sup>[25]</sup>

### **Structure solution of IpoC\_apo**

Screening of IpoC\_apo with these optimized conditions led to the initial crystallization hit in sparse matrix screens in a well solution of 1 M LiCl, 0.1 M MES pH 6.0 and 20% PEG6000 with a final well solution pH of 6.0 at a protein to well solution ratio of 1 to 2. The crystals were first observed after a five-week incubation time at 277 K and could only be reproduced when using the purification batch that was used for the initial hit. Given these observations, we hypothesized that a potential protease contamination could lead to a stable construct, which could readily form crystals. The crystals belonged to space group  $P 3_2 2$  and diffracted to 2.1 Å at beamline X06DA at the Swiss Light Source with good statistics. The combination of the unit cell parameters of 95, 95, 56 Å with this space group and a theoretical molecular weight of the expressed construct of 40.9 kDa would amount to a solvent content of 32.15% calculated by the *Matthews Probability Calculator*<sup>[26,27]</sup>, which is considered to be statistically unlikely given the resolution diffraction could be observed to. Therefore, a somewhat truncated construct was probable.

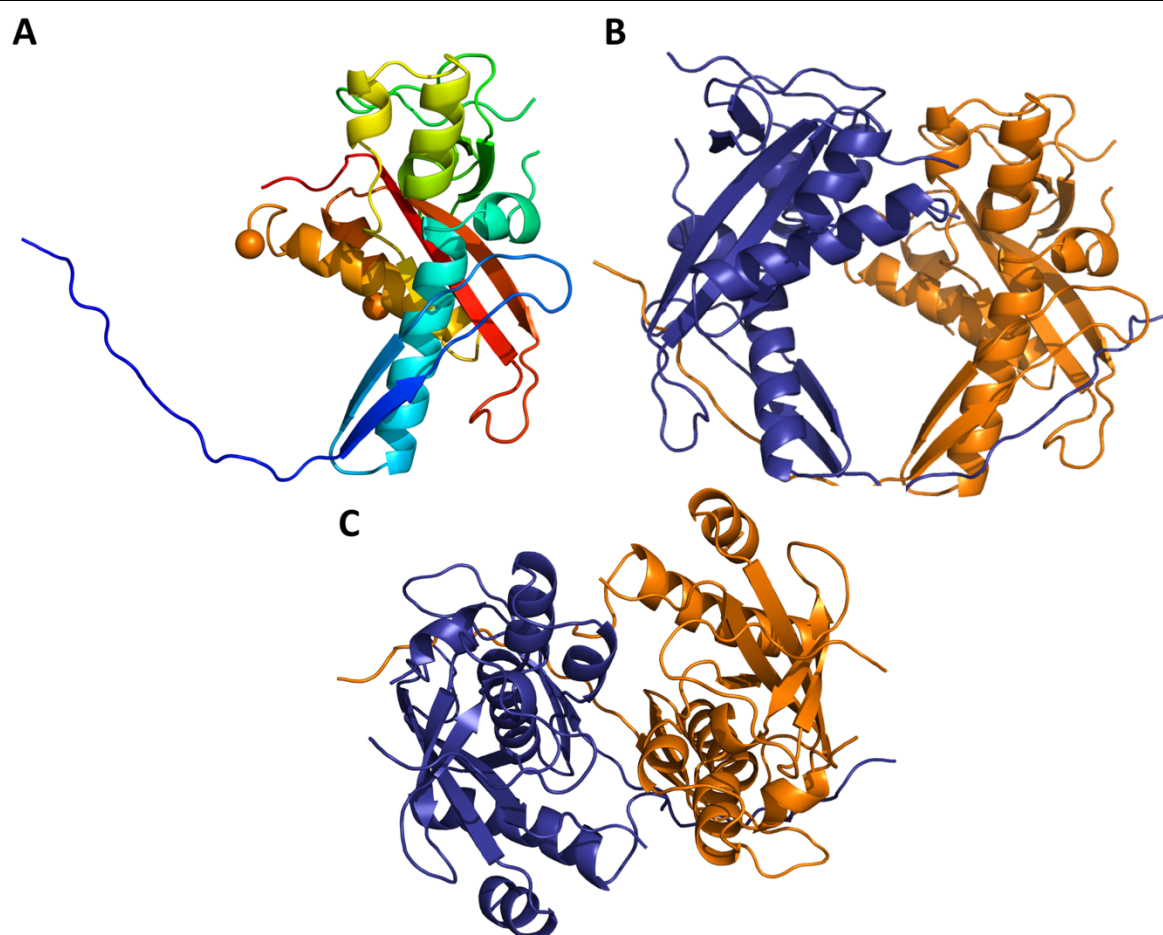
Due to a combination of a missing molecular replacement model in the PDB with sufficient similarity and the problem of reproducing the crystals using fresh protein batches, we decided to try Sulphur Single Anomalous Dispersion (S-SAD) to obtain the required phase information as the naturally high symmetry of the space group allowed us to get the necessary redundancy. In an attempt to maximize our chances of obtaining the required phase information for solving this interesting cyclodehydratase structure, we collected data at the unique I23 beamline located at Diamond Light Source, which allows collecting at long wavelengths with minimal absorption and scattering effects due to the beamline being set up in a vacuum.<sup>[28]</sup> The IpoC\_apo structure could be solved using phase information by amino acid side chain S as well as bound Cl atoms. The position of the respective Chloride as well as Sulphur atoms is visualized in **Figure S5**. The data collection and refinement statistics can be found in **Table 1**.

**Table 1:** Data collection and refinement statistics. Statistics for the highest-resolution shell are shown in parentheses.

	IpoC apo
Resolution range [Å]	46.78-2.10 (2.21-2.10)
Space group	P 3 <sub>2</sub> 2
Unit cell	
<i>a</i> , <i>b</i> , <i>c</i> [Å]	95.18, 95.18, 56.85
<i>α</i> , <i>β</i> , <i>γ</i> [°]	90.00, 90.00, 120.00
Total reflections	470 997 (51 922)
Unique reflections	17 656 (2538)
Multiplicity	26.7 (20.5)
Wavelength [Å]	2.75
Completeness (%)	100.0 (100.0)
Mean <i>I</i> / <i>σ</i> ( <i>I</i> )	21.7 (1.9)
Wilson B-factor	45.68
R <sub>merge</sub>	0.082 (0.790)
R <sub>work</sub>	0.1968 (0.3343)
R <sub>free</sub>	0.2330 (0.3510)
Number of non-hydrogen atoms	1611
Macromolecules	1529
Ligands	10
Water	72
Protein residues	216
RMS (bond lengths)	0.002
RMS (angles)	0.56
Average B-factor	63.20
Macromolecules	63.37
Ligands	78.68
Solvent	57.54
Anomalous completeness	100.0 (100.0)
Anomalous multiplicity	13.6 (10.3)
DelAnom correlation between half-sets	0.536 (0.060)
Mid-slope of Anom Normal Probability	1.295

## **The overall structure of IpoC\_apo**

IpoC is a basic protein with a theoretical pI of 10.02 in its longest form, a net charge of +13 and 12.5% Arginine residues, some of which might be interacting with the 10 negatively charged side chains of the BotA follower peptide (**Table S3**). The asymmetric unit of IpoC\_apo contained one monomer with convincing electron density in the difference map for a large portion of the N-terminus (residues 26-94, 101-192) as well as a portion of the C-terminus (residues 310-354), which is portrayed as a rainbow cartoon in **Figure 4A**. It is possible that we have crystallized a N- and C-terminal truncation as was expected by our predictions beforehand. With the data at hand, the reason for the missing electron density for residues 192-310 (the middle portion of the protein between the two orange spheres in **Figure 4A**) can not be entirely determined. If we consider a protease contamination, it is possible that this portion is cleaved off during the long incubation time before crystals are formed. In contrast, it is not uncommon to observe missing electron density for flexible portions of proteins in crystal structures due to an absence of statistical certainty of their position. The N-terminal part of the structure is mainly built up by 2 bigger  $\alpha$ -helices (cyan and orange), which are oriented perpendicular to each other whereas the C-terminal portion consists of a  $\beta$ -hairpin structure (red). Interestingly, a part of the N-terminus (residues 26-48) is facing away from the rest of the protein and does not form any secondary structure elements. Looking at potential interactions with a symmetry mate, we observed a crystallographic 2-fold rotational axis in between both monomers (**Figure S6**) and close contacts of the N-terminal residues with the dark blue symmetry mate (**Figure 4B**). As such, these residues could be involved in dimer formation in a “strand-swapping” fashion as they mimic an additional beta strand of the respective symmetry mate. Similar events have been observed for various proteins, one example being the cadherins for which it was shown to be a biologically relevant phenomenon leading to dimerization of the cadherin monomers and therefore enabling their mediation of cell-cell contacts.<sup>[29]</sup>



**Figure 4.** Apo structures of IpoC portrayed as a cartoon. A) IpoC monomer found in the asymmetric unit with the N-terminus facing away from the rest of the protein to the left. The Cartoon is colored from blue (N-terminus) to red (C-terminus). The missing part in the middle is visualized by the two orange spheres, which mark the last and first ordered residue. B) IpoC monomer (orange) with graphic visualization of the direct symmetry mate (dark blue), both N-termini are forming close contacts with their respective symmetry mate. C) IpoC monomer (orange) with graphic visualization of the direct symmetry mate (dark blue), top view.

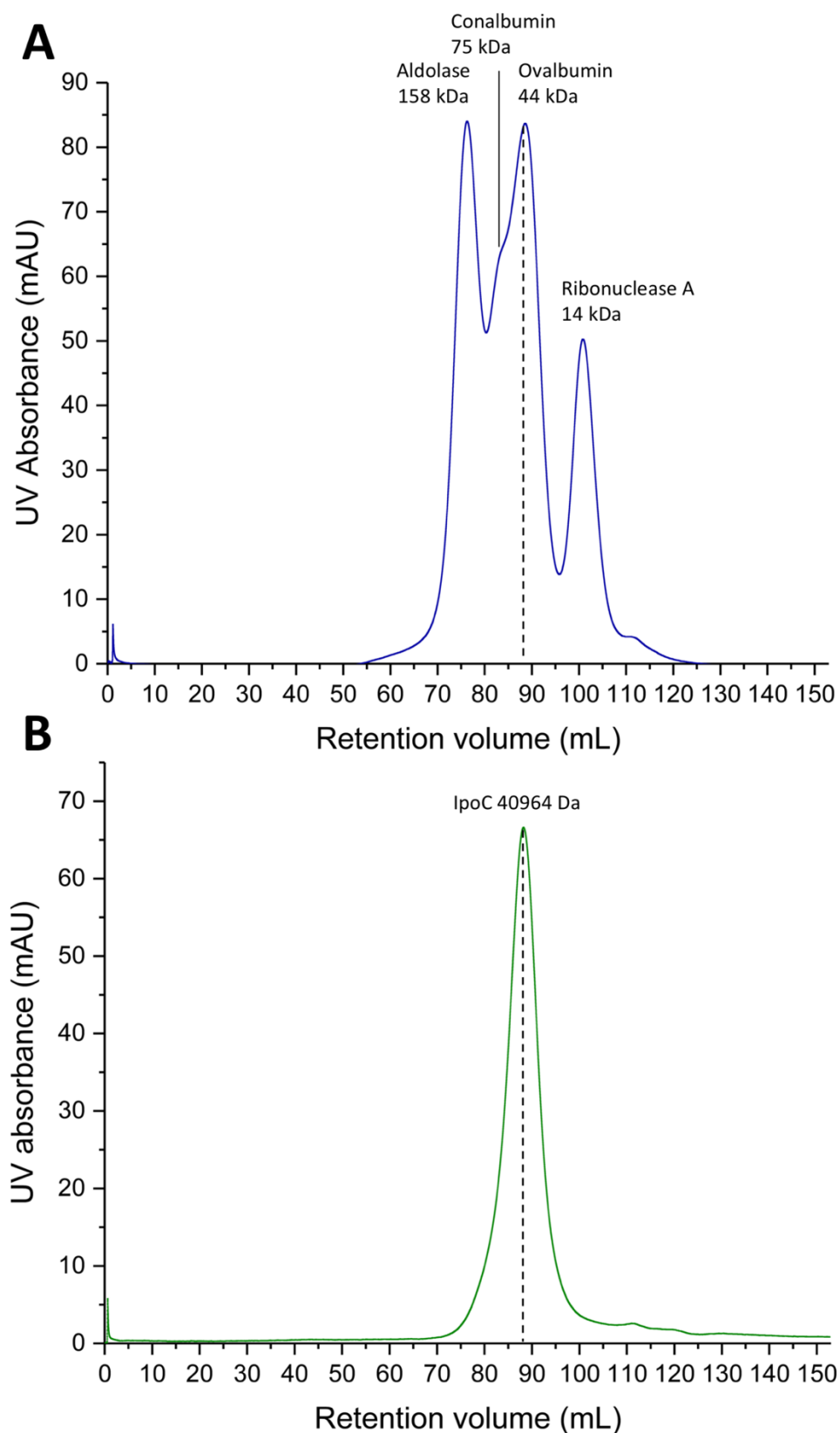
### Structural analysis of IpoC\_apo

If we consider the stacking of protein molecules on top of each other in the same orientation a necessity for crystal formation, a cleaved off C-terminal portion still has to be stabilized by intramolecular interactions to always be oriented in the same way, hence allow for crystallization. Within the missing portion of the protein (residues 193-309) lies a conserved Arginine residue (Arg234), which is predicted to be involved in cofactor binding as a mutation of this residue greatly reduces the activity of IpoC.<sup>[12]</sup> This observation can perhaps be correlated with the hypothesis of this portion being flexible and becoming ordered upon substrate and cofactor binding. In addition, a crystallized construct consisting of the full amino acid chain in the asymmetric unit would have a respective calculated solvent content of 42.3%



with the *Matthews Probability Calculator*<sup>[26,27]</sup>, which is realistic for the given resolution whereas a completely missing portion would result in 64% solvent content, which is unlikely. Even though the unit cell could indeed harbor another 120 amino acids based on its raw size and the fact that we find a big open space close to the last ordered residue of the chain before the missing part, the question remains how these residues would occupy the remaining space as there are close crystal contacts with the symmetry mate right next to the last ordered residue.

Concerning the closely interacting symmetry mates, we wondered whether this was the result of a specific proteolytic cleavage fragment leading to necessary crystal contacts or in fact could be related to a biologically active dimer interface. Analysis of the 206 built residues of each monomer with *PISA*<sup>[30]</sup> revealed that an astonishing 34 % of the residues are involved in interface formation with a total of 21 hydrogen bonds formed between the residues of both symmetry mates. As such, a dimer assembly of both symmetry mates is assumed to be stable in solution with a calculated free dissociation energy of  $\Delta G^{\text{diss}} = 48.3 \text{ kcal mol}^{-1}$ . Contrary to this observation, a comparison of the size of gel filtration standards with the expressed construct would suggest that at the time of purification, the protein is still monomeric in solution (**Figure 5**). In fact, the elution profile of the crystallized IpoC construct (**Figure 5B**) shows a difference of 0.5 mL concerning its peak maximum compared to the model protein Ovalbumin, which has a similar molecular weight. A mass spectrometry analysis of dissolved crystals of IpoC<sub>apo</sub> failed to provide additional insights into the crystallized construct. Unfortunately, further measurements of the apparent size of the construct in solution by Dynamic Light Scattering (DLS) were hindered by the limited stability of the protein at room temperature. In conclusion, an assessment of a biological dimer interface of IpoC awaits a complex structure with ATP/Mg<sup>2+</sup> and the precursor peptide BotA<sup>P</sup>.



**Figure 5.** A) Hiload 16/60 Superdex 200 pg gel filtration calibration run of 4 monomeric model proteins with their respective molecular weight. B) Superdex 200 gel filtration run of the crystallized IpoC construct with its calculated molecular weight.

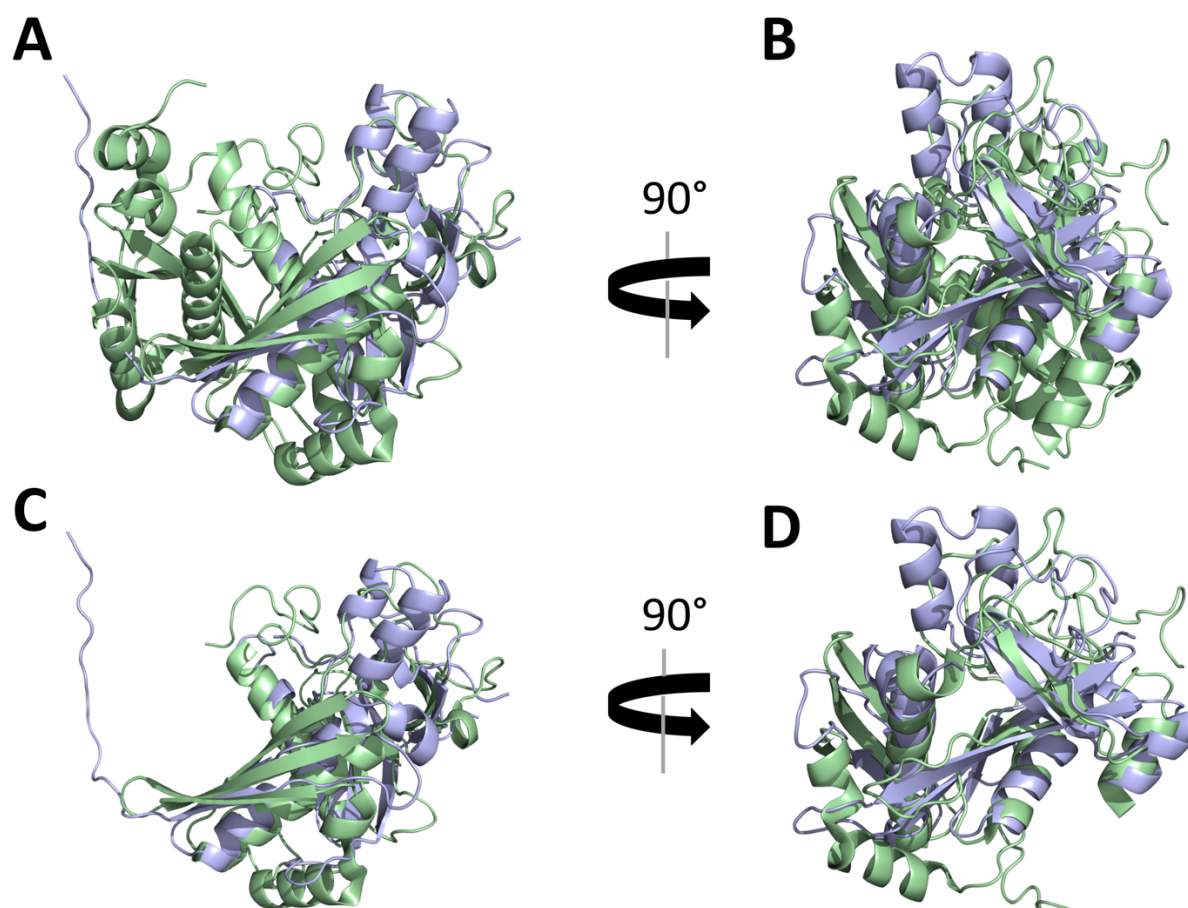
## IpoC\_apo shares structural motifs with other YcaO domains

To date, only a handful of YcaO domain proteins have been characterized structurally. A sequence homology analysis of IpoC displayed in the Sequence Similarity Network (**Figure S1**) revealed no significant sequence homology to proteins other than IpoC homologues from different *Streptomyces* strains harboring a Bottromycin biosynthetic gene cluster. Given these rare findings, we wondered whether these data are correlated to a potential structural homology with other published YcaO domain enzymes. Within the YcaO class of enzymes, 5 functional subfamilies have been suggested and linked to their respective catalytic activity with respective *in vitro* reconstitution data backing up the biosynthetic hypotheses.<sup>[13]</sup> IpoC has been suggested to be a special enzyme as it is one of the few standalone azoline-forming YcaO domains, which does not require an E1-like enzyme with a leader-binding RRE domain. However, it functionally belongs to the best-studied class of YcaO domain enzymes, the backbone azoline-forming YcaOs. These YcaO proteins have originally been found in the biosynthetic gene cluster of Microcin B17, a gyrase inhibitor produced by *E. coli*.<sup>[15]</sup> The corresponding gene McbD has been associated with complex formation with the respective E1-like protein McbB as well as a dehydrogenase McbC and the catalysis of two thiazoles, two oxazoles and two 4,2-fused bisheterocycles, which have been shown to have a crucial effect on substrate binding of the natural product.<sup>[20][31]</sup> In contrast, we observe no similar catalytic promiscuity for IpoC as it did not show catalytic activity for a BotA<sup>P</sup> point mutant with a serine at the cysteine position and was therefore unable to catalyze oxazoline formation.<sup>[11]</sup>

Bioinformatic analysis with the *DALI* server<sup>[32]</sup> revealed significant structural homology with a Z-score above 7 for the 5 already published YcaO domain structures in the PDB. From the 5 available structures with relevant Z-scores for IpoC structural homology, 3 can be assigned to be involved in various backbone azoline-forming mechanisms. LynD (PDB:4V1V)<sup>[22]</sup> and TruD (PDB:4BS9)<sup>[21]</sup> are involved in Cyanobactin biosynthesis of their respective natural products lyngbyabactin and trunkamide, ribosomally synthesized and post-translationally modified cyclic peptides produced by various cyanobacteria species,<sup>[33]</sup> the third protein is McbD (PDB: 6GOS). Intriguingly, the McbD complex with the E1-like protein McbB resembles the overall fold of the domain architecture 1,2 and 3 found in both LynD and TruD.<sup>[20,21,22]</sup>

Additionally, 1 YcaO protein with structural homology to IpoC is suggested to catalyze backbone thioamidation in *Methanopyrus kandleri* (PDB: 6CI7),<sup>[19]</sup> 1 substrate of a YcaO domain enzyme from *E. coli* remains unknown (PDB: 4Q86).<sup>[18]</sup> Interestingly, IpoC shows the

best  $\alpha$ -Carbon structural homology of 3 Å over 97 atoms when compared with the backbone thioamidation YcaO domain from *M. kandleri*, which is also the only one not associated with forming a complex with other biosynthetic enzymes or consisting of more than one domain (**Figure 6A+B**).<sup>[25]</sup> This protein has not been associated with dimer formation, which could be related to its very different substrate scope being the Methyl-coenzyme M reductase (MCR), an essential protein in primary metabolism of anaerobic archaea.<sup>[19]</sup> Additionally, the protein family has been shown to act on 11-residue substrate peptides, whereas IpoC shows no catalytic activity on BotA<sup>P</sup> precursor peptides with a reduced C-terminus of more than 10 amino acids.<sup>[12][19]</sup> As both YcaO proteins have a similar residue count, we wondered if there is a similarity in their overall fold. Interestingly, as we attempted to remove the non-aligned residues in the *M. kandleri* YcaO, we observed a similar orientation of the different protein portions as the missing part in IpoC fits to residues 174-318 in the *M. kandleri* YcaO (**Figures 6C+D**).



**Figure 6.** A) Superposition of IpoC\_apo (blue) with the *M. kandleri* YcaO [(green(PDB:6CI7))]. The  $\alpha$ -Carbon rmsd was calculated to be 3 Å over 97 atoms. B) Superposition of IpoC\_apo (blue) with *M. kandleri* YcaO (green), 90° rotation from A). C) Superposition of IpoC\_apo (blue) with the *M. kandleri* YcaO (green), only aligned residues are shown. D) Superposition of IpoC\_apo (blue) with the *M. kandleri* YcaO (green), only aligned residues are shown, 90 ° rotation from C).

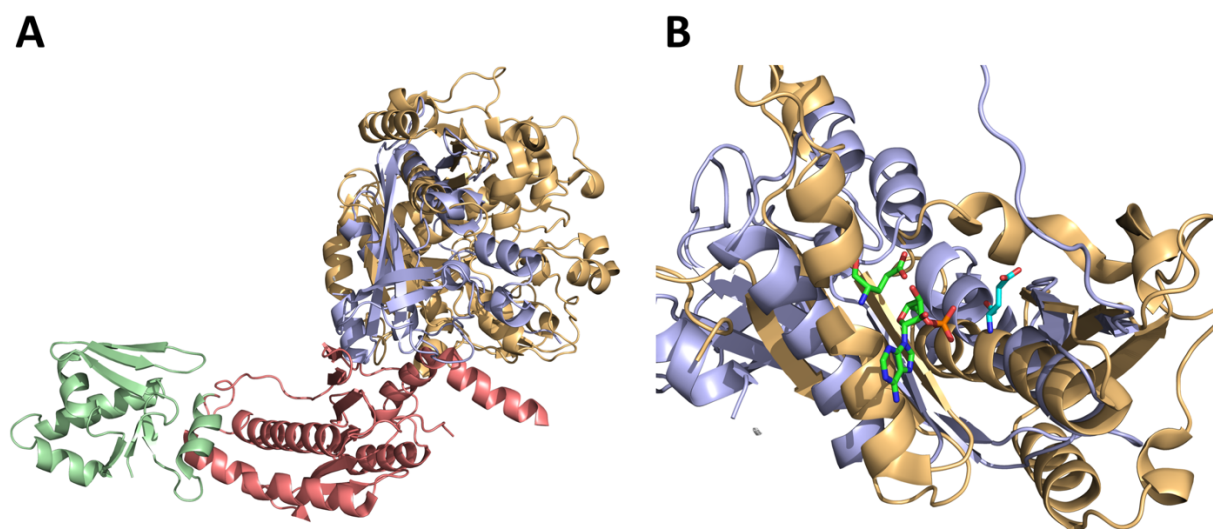
Even though a thorough comparison of structural homology would in theory require a full IpoC chain, we may postulate that the different functions of YcaO domain enzymes are not necessarily correlated to specific structural features.

Concerning the structural homology of IpoC with the Cyanobactin heterocyclase enzymes LynD and TruD, we can directly observe the absence of domains 1 and 2 (green and red) as IpoC (blue) only shows some homology ( $C_{\alpha}$  rmsd = 2.6 over 40 atoms) with the cofactor-binding domain 3 (orange) in LynD, which most likely also processes the core peptide (**Figure 7A**).<sup>[22]</sup> LynD functionality has been shown to be directly related to dimer formation as an interface which is built up between domain 1 of one and domain 3 of the other monomer is required for leader peptide binding. This leader peptide binding induces a conformational change leading to an “active” enzyme, which is necessary for efficient heterocycle formation at two positions of the substrate peptide. In this context, the missing portion of IpoC<sub>apo</sub> could exhibit a specific function by possibly building up a substrate binding tunnel at the interface between both IpoC monomers after a conformational change has been induced. It will be fascinating to see whether IpoC uses Dimer interface formation for its processing and follower binding in the complex structure with BotA<sup>P</sup> and cofactor bound in the same way as it was shown for LynD (**Figure S7**).

In a similar fashion, binding of the leader peptide in TruD enables a processive mechanism with a defined order of reaction starting from the most C-terminal cysteine residue, which has been confirmed by NMR studies and substrate peptide mutagenesis.<sup>[21]</sup> Intriguingly, both LynD and TruD show processing of the most C-terminal cysteine residue in absence of a leader peptide suggesting a specific binding of the immediate environment of this residue. Using binding affinity studies for TruD with point-mutated precursor peptides, a model has been proposed in which the enzyme utilizes varying affinities for both the leader peptide and the respective target residue for substrate binding of each step, accounting for full sequential processing of the precursor peptide and avoiding a combination of different intermediates.<sup>[21]</sup> Given there is only one target residue in the BotA<sup>P</sup> precursor peptide, a similar mechanistic process for IpoC is not required. Therefore, we can imagine a more specific binding mode of the precursor peptide in a fixed position rather than a moving substrate being processed at multiple positions before being released by the enzyme. Surprisingly, IpoC kinetics show rather slow processing of the enzyme needing 8 h for 90% turnover of substrate compared to the 2 h of LynD and TruD, which both catalyze two thiazoline formations per substrate

peptide.<sup>[12]</sup> *In vivo* knockout studies of the bottromycin biosynthetic gene cluster have suggested that incomplete C-methylation reduces the efficiency of various downstream tailoring steps.<sup>[34]</sup> It remains to be seen whether there is a higher binding affinity and a specific binding interaction of (di- or tri-)methylated BotA<sup>P</sup> with IpoC giving it a function as an early regulation mechanism in natural product biosynthesis.

Looking at potential conserved residues for cofactor binding, within domain 3 of LynD lies the cofactor-binding Glu423 residue, which can be functionally correlated to the Glu59 residue in IpoC as was shown by point mutation experiments.<sup>[12]</sup> Comparing both residues respective positions in the structure, we can hypothesize about potential substrate and cofactor binding in IpoC. The corresponding figure reveals a similar orientation of both side chains, IpoC Glu59 in cyan and LynD Glu423 in green, towards the solvent channel where the AMP in the LynD structure is bound (**Figure 7B**). Therefore, we can possibly assume a similar overall conformation and cofactor coordination of an IpoC complex structure even though differences in substrate binding and the mechanistic process are likely.



**Figure 7.** A) Superposition of a cartoon of IpoC (tintblue) with a cartoon of the different domains of LynD. Domain 1 is shown in palegreen, domain 2 is shown in salmon and domain 3 is shown in lightorange. We observe weak structural homology of IpoC with domain 3 of LynD. B) Comparison of cofactor-binding glutamate side chains in IpoC<sub>apo</sub> and LynD crystal structures. IpoC is superposed upon LynD and shown as a cartoon in tintblue with its Glu59 in cyan whereas LynD is shown in lightorange with its Glu423 and AMP bound in green. Overall, we observe a similar orientation of both side chains towards the solvent channel where the AMP in LynD is bound. Only aligned residues are shown.

Our data provide the first structural information of a standalone YcaO domain enzyme, which is not associated with complex formation or an E1-like partner protein and potentially has a distinct mechanism of substrate binding and catalysis within the azoline-forming YcaO

proteins. The obtained information broadens our knowledge of the potential of YcaO domain enzymes for biotechnology and is the first step in engineering Bottromycin derivatives for *in vitro* production of this bioactive natural product.

## 4.4 References

- [1] Burch, D. *Veterinary Record*. **2016** 179, 633-634
- [2] Harbarth S, Balkhy HH, Goossens H, Jarlier V, Kluytmans J, Laxminarayan R, Saam M, Van Belkum A, Pittet D; *Antimicrob Resist Infect Control*. **2015** 4, 49
- [3] Chang J, Kwon HJ *J Ind Microbiol Biotechnol*. **2016** 43 (2-3): 221-31
- [4] Grisoni F, Merk D, Consonni V, Hiss JA, Giani Tagliabue S, Todeschini R, Schneider G *Communications Chemistry* **2018** 1, 44
- [5] Newman DJ, Cragg GM *Future Med Chem*. **2009** 1 (8): 1415-27
- [6] Bologa CG, Ursu O, Oprea TI, Melançon CE 3rd, Tegos GP *Curr Opin Pharmacol*. **2013** 13 (5): 678-87
- [7] Huo L, Rachid S, Stadler M, Wenzel SC, Müller R *Chem Biol*. **2012** 19 (10): 1278-87
- [8] Ortega MA, van der Donk WA *Cell Chem Biol*. **2016** 23 (1):31-44
- [9] Yang X, van der Donk WA *Chemistry* **2013** 19 (24):7662-77
- [10] Mann G, Huo L, Adam S, Nardone B, Vendome J, Westwood NJ, Müller R, Koehnke J *Chembiochem*. **2016** 17 (23):2286-2292
- [11] Schwalen CJ, Hudson GA, Kosol S, Mahanta N, Challis GL, Mitchell DA *J Am Chem Soc*. **2017** 139 (50):18154-18157
- [12] Franz L, Adam S, Santos-Aberturas J, Truman AW, Koehnke J *J Am Chem Soc*. **2017** 139 (50):18158-18161
- [13] Dunbar KL, Melby JO, Mitchell DA *Nat Chem Biol*. **2012** 8 (6):569-75
- [14] Burkhardt BJ, Schwalen CJ, Mann G, Naismith JH, Mitchell DA *Chem Rev*. **2017** 117 (8):5389-5456
- [15] Yorgey P, Lee J, Kördel J, Vivas E, Warner P, Jebaratnam D, Kolter R *Proc Natl Acad Sci U S A*. **1994** 91 (10):4519-23.
- [16] Li YM, Milne JC, Madison LL, Kolter R, Walsh CT *Science* **1996** 274 (5290):1188-93
- [17] Burkhardt BJ, Hudson GA, Dunbar KL, Mitchell DA *Nat Chem Biol*. **2015** 11 (8):564-70
- [18] Dunbar KL, Chekan JR, Cox CL, Burkhardt BJ, Nair SK, Mitchell DA *Nat Chem Biol*. **2014** 10 (10):823-9
- [19] Mahanta N, Liu A, Dong S, Nair SK, Mitchell DA *Proc Natl Acad Sci U S A*. **2018** 115 (12):3030-3035
- [20] Ghilarov D, Stevenson CEM, Travin DY, Piskunova J, Serebryakova M, Maxwell A, Lawson DM, Severinov K *Mol Cell*. **2019** 73 (4):749-762.e5.
- [21] Koehnke J, Bent AF, Zollman D, Smith K, Houssen WE, Zhu X, Mann G, Lebl T, Scharff R, Shirran S, Botting CH, Jaspars M, Schwarz-Linek U, Naismith JH *Angew Chem Int Ed Engl*. **2013** 52 (52):13991-6



- [22] Koehnke J, Mann G, Bent AF, Ludewig H, Shirran S, Botting C, Lebl T, Houssen W, Jaspars M, Naismith JH *Nat Chem Biol.* **2015** 11 (8):558-563
- [23] Huguet-Tapia, J.C., Durkin, A.S., Pettis, G.S. and Badger, J.H. *Direct Submission, John Craig Venter Institute* **2012**
- [24] Kramer RM, Shende VR, Motl N, Pace CN, Scholtz JM *Biophys J.* **2012** 102 (8):1907-15
- [25] Kirkwood J, Hargreaves 1, O'Keefe S, Wilson J *Bioinformatics.* **2015** 31 (9):1444-51
- [26] Matthews BW *J Mol Biol.* **1968** 33, 491-97
- [27] Kantardjieff KA, Rupp B *Protein Science* **2003** 12, 1865-71
- [28] Wagner A, Duman R, Henderson K, Mykhaylyk V *Acta Crystallogr D Struct Biol.* **2016** 72, 430-439
- [29] Patel SD, Ciatto C, Chen CP, Bahna F, Rajebhosale M, Arkus N, Schieren I, Jessell TM, Honig B, Price SR, Shapiro L *Cell.* **2006** 124 (6):1255-68.
- [30] Krissinel E, Henrick K *J Mol Biol.* **2007** 372 (3):774-97
- [31] Zamble DB, Miller DA, Heddle JG, Maxwell A, Walsh CT, Hollfelder F *Proc Natl Acad Sci U S A.* **2001** 98 (14):7712-7
- [32] Holm L, Laakso LM *Nucleic Acids Res.* **2016** 44 (W1):W351-5
- [33] Sivonen K, Leikoski N, Fewer DP, Jokela J *Appl Microbiol Biotechnol.* **2010** 86 (5):1213-25
- [34] Crone WJ, Vior NM, Santos-Aberturas J, Schmitz LG, Leeper FJ, Truman AW *Angew Chem Int Ed Engl.* **2016** 55 (33):9639-43

## Supporting Information

# Structural analysis of IpoC, an unusual YcaO cyclodehydratase from the Bottromycin pathway

**Sebastian Adam**<sup>1[a]</sup>, Ramona Duman<sup>1[b]</sup>, Armin Wagner<sup>[b]</sup> and Jesko Koehnke<sup>[a]\*</sup>

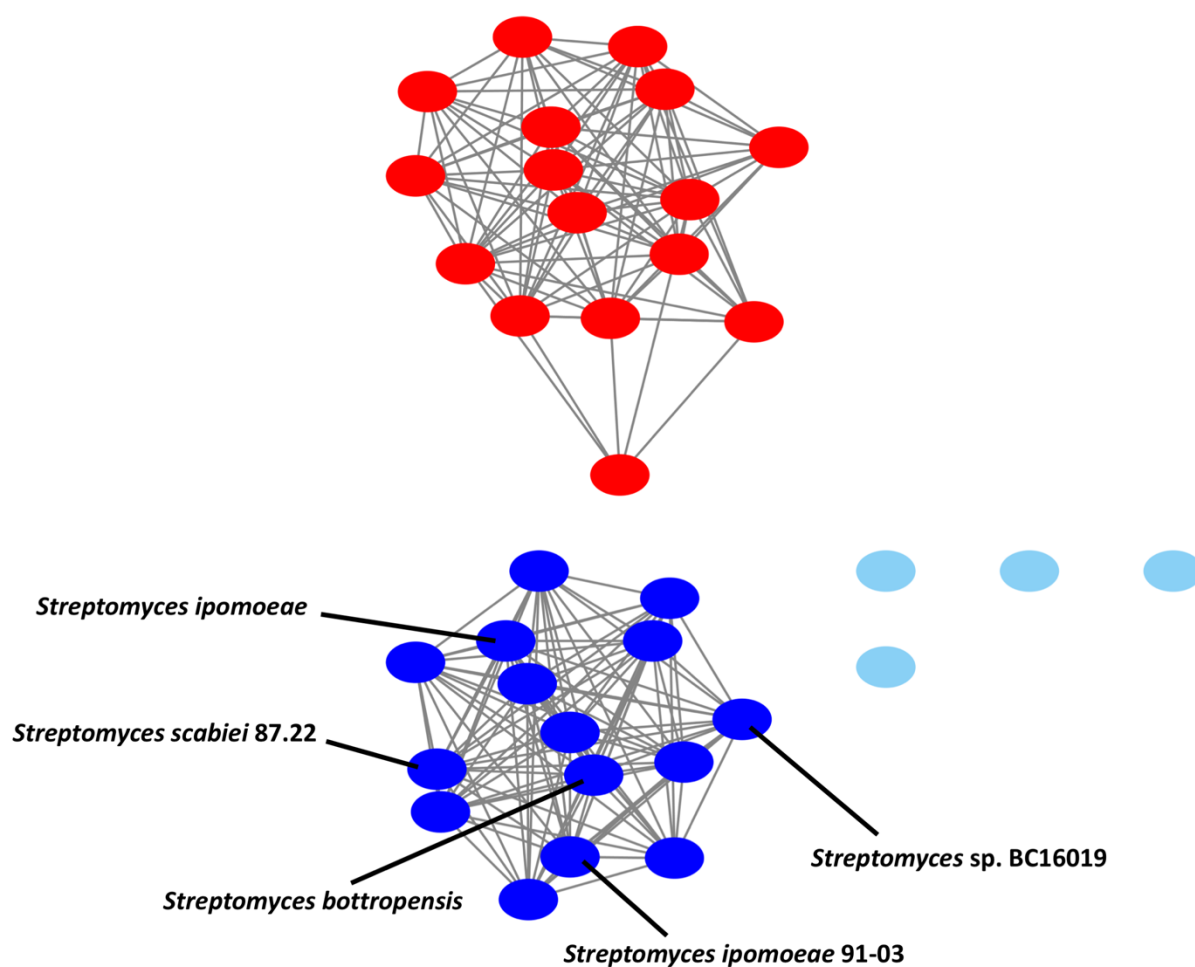
*Manuscript in preparation*

### Affiliation

<sup>[a]</sup> Workgroup Structural Biology of Biosynthetic Enzymes, Helmholtz Institute for Pharmaceutical Research, Helmholtz Centre for Infection Research, Saarland University, Universitäts-campus E8 1, 66123, Saarbrücken, Germany.

<sup>[b]</sup> Diamond Light Source, Harwell Science and Innovation Campus, Chilton, Didcot OX11 0DE, England.

<sup>1</sup> These authors contributed equally to the manuscript



**Figure S1:** Sequence similarity network of IpoC.<sup>[1]</sup> The full length primary amino acid sequence of the crystallized IpoC construct (EKX678557.1) was used for a PSI-BLAST with a threshold of 0.005, all sequences above the threshold were used for the SSN. The name of the producing strain is given if *in vivo* or *in vitro* data of the respective protein is available. The network shows only one other cluster of distantly related proteins from non-related bacterial strains (red) other than the Bottromycin-producing *Streptomyces* strains (blue).

CLUSTAL 2.1 multiple sequence alignment

```

S. ipomoeae 91-03  MKSARRTTGRGSMLEATATECALREVVYRSYPSERTVTVRCTVRPAEGSARADGYGTAAT
S. ipomoeae      -----MLEATATECALREVVYRSYPSERTVTVRCTVRPAEGSARADGYGTAAT
                      *****

S. ipomoeae 91-03  KAVARLKALSEAVERLVACTPFASVVRPPTTSRGSGAVPPFPASGVRTVPDGCASRVYRP
S. ipomoeae      KAVARLKALSEAVERLVACTPFASVVRPPTTSRGSGAVPPFPASGVRTVPDGCASRVYRP
                      *****

S. ipomoeae 91-03  LTGGGPRRVPLYWSSPWTAGEELRAAVLTAPPEARLSSTVGWAVAPTPEAALHGALLELTE
S. ipomoeae      LTGGGPRRVPLYWSSPWTAGEELRAAVLTAPPEARLSSTVGWAVAPTPEAALHGALLELTE
                      *****

S. ipomoeae 91-03  LLNHGVFLHRSLAGPRRPAAGDETLVLPLGGFVRTPTVLAVAYGRGRRMPATGLGCGATR
S. ipomoeae      LLNHGVFLHRSLAGPRRPAAGDETLVLPLGGFVRTPTVLAVAYGRGRRMPATGLGCGATR
                      *****

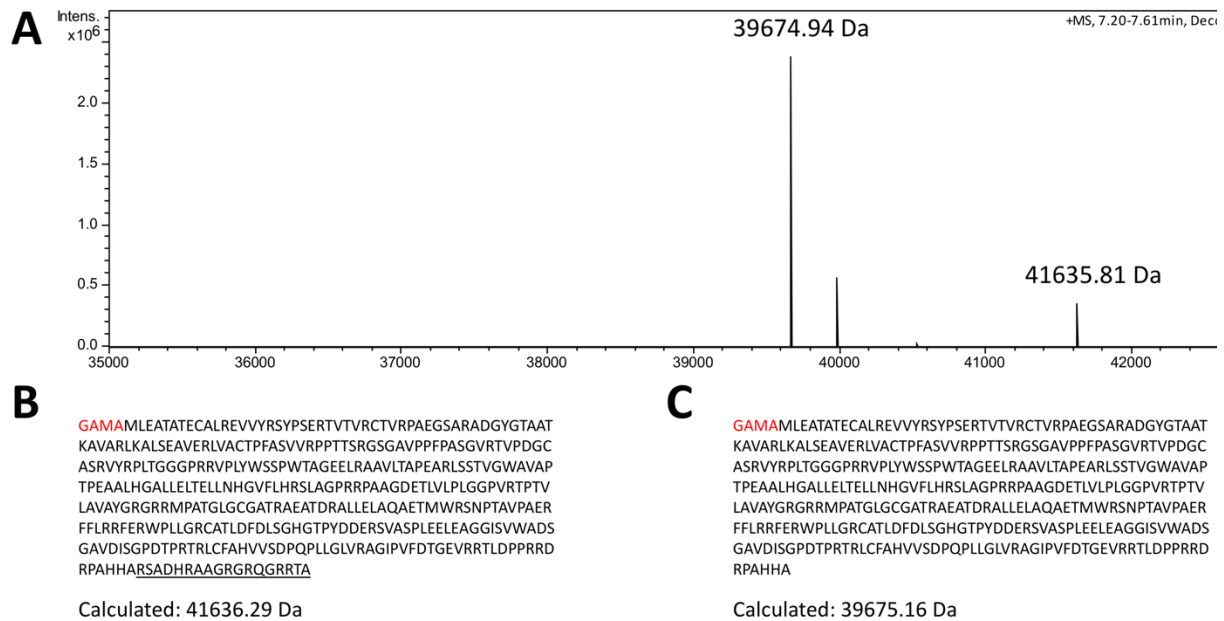
S. ipomoeae 91-03  AEATDRALLELAQAETMWRSNPTAVPAERFFLRFRWPLLGRCATLDFDLSGHGTPYDD
S. ipomoeae      AEATDRALLELAQAETMWRSNPTAVPAERFFLRFRWPLLGRCATLDFDLSGHGTPYDD
                      *****

S. ipomoeae 91-03  ERSVASPLEELEAGGISVWADSGAVDISGPDTPRTRLCFAHVSDPQPLLGLVRAGIPVF
S. ipomoeae      ERSVASPLEELEAGGISVWADSGAVDISGPDTPRTRLCFAHVSDPQPLLGLVRAGIPVF
                      *****

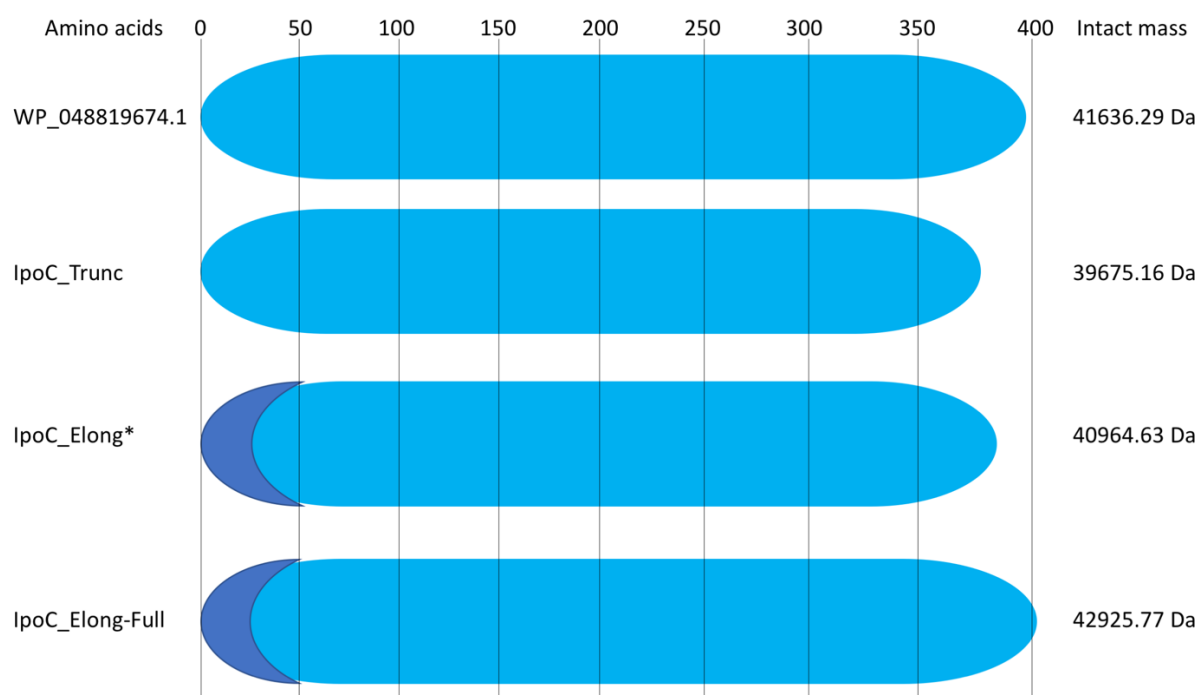
S. ipomoeae 91-03  DTGEVRRRTLDPFRRDRPAHHARSADHRAAGRGRQGRRTA
S. ipomoeae      DTGEVRRRTLDPFRRDRPAHHARSADHRAAGRGRQGRRTA
                      *****

```

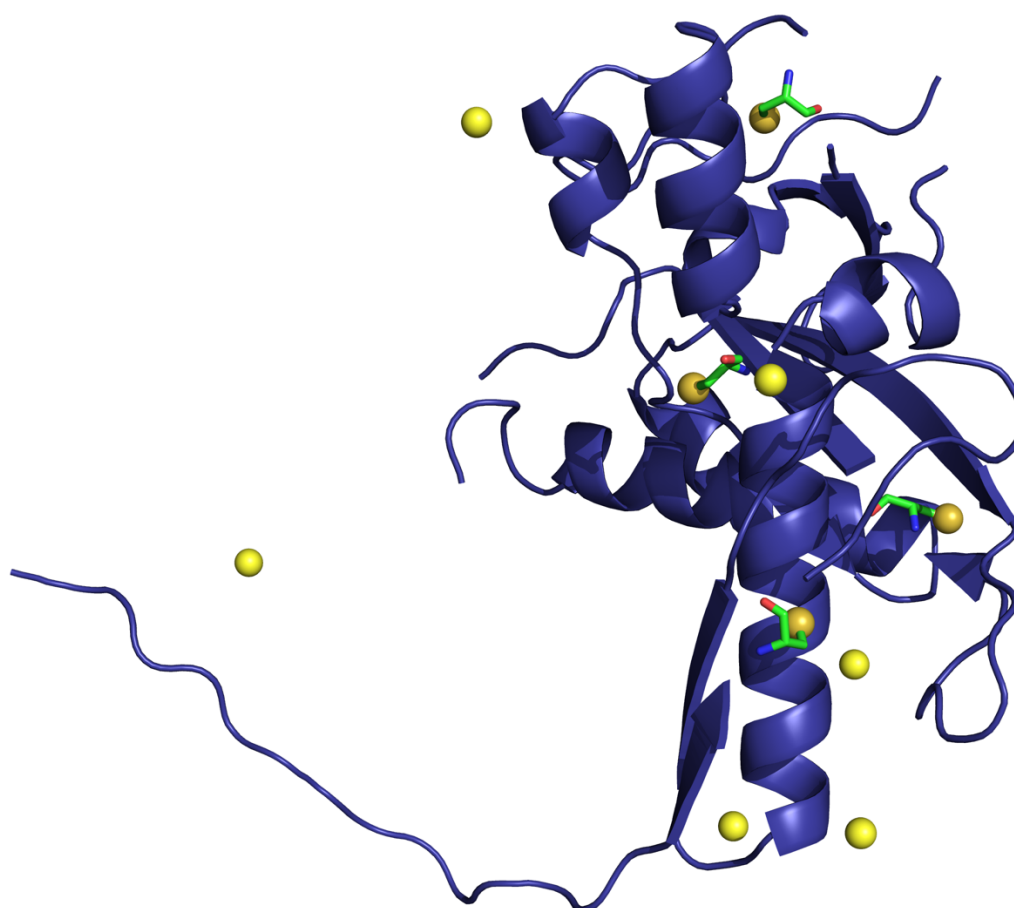
**Figure S2:** Sequence alignment of the primary amino acid sequence of the IpoC construct which was crystallized (*S. ipomoeae* 91-03) with the construct which was originally expressed (*S. ipomoeae*). The strain *S. ipomoeae* 91-03 has been annotated with a 12 amino acid extension at the N-terminus and thus a different start codon. The alignment was generated with *Clustal Omega*.<sup>[2]</sup>



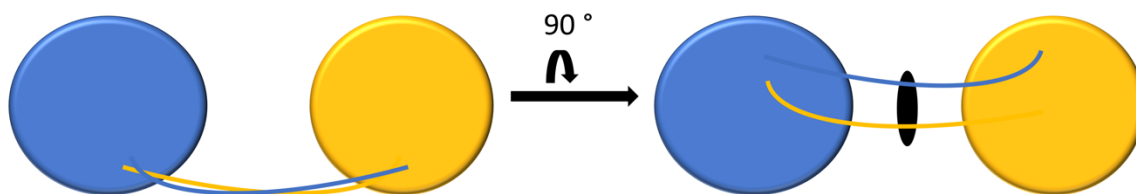
**Figure S3:** A) Intact mass measurement of pure IpoC protein. 10  $\mu$ M protein were injected onto an Aeris 3.6  $\mu$ m widepore XB-C8 column. The resulting protein mass peak on maXis 4G UHR-TOF was integrated and the masses were deconvoluted by maximum entropy deconvolution. B) IpoC construct that was expressed and purified, the red N-terminal residues are the remaining cloning artefact. Underlined residues were found to be cleaved off by a protease. C) IpoC construct which was found to be stable after purification. The resulting intact mass correlates with the given primary sequence.



**Figure S4:** Schematic of different IpoC constructs that could be expressed and purified. The construct which could be crystallized is marked with a star.



**Figure S5:** Visualization of the position of Chloride (yellow) as well as Sulphur (dark yellow) atoms within the IpoC structure (blue) portrayed as a cartoon. The respective Cysteine residues containing the Sulphur atoms are shown as sticks in green, Methionine residues are not found in the final structure.



**Figure S6:** Visualization of the 2-fold rotational symmetry axis found in between the IpoC symmetry mates.



**Figure S7:** Schematic of the proposed conformational change induced by substrate binding in IpoC followed by the thiazoline formation within the active site. The biologically active dimer binds the core peptide and the follower peptide in two distinct positions in the different monomers utilizing the dimer interface in the same way as it was shown for LynD.

**Table S1:** List of “C” YcaO protein homologues from different *Streptomyces* strains.

Strain	Primary amino acid sequence	Accession number
<i>Streptomyces scabiei</i> 87.22	MREATATECELVHRSYPSEGTVTVRCTVRPAEGDARADGYGTATTEAVARAKAL SEAVERLVACTPFTAVARPPATAFADPGTGSGPVPPFPAAGVRTAPDGCADRVRPLT GGGPRRVPLYWSSPWTAGAE LRAGLTAAQARLSSTVGWAVAPTPEAALRGALLEL TELIGYGVFLHRRLAGPPRRPSGDDRTL VVPLEGVGRTPAVLAVAYGRGRMPATGL GCGETVAEATGRALLELAQAETMWRSNPTAEP AERFFLRFRERWPLLTRCATLDFDL KDADAPHGPHDPAPGDDEPRPATPLEELEAGGIRVWADSGALDISGPDIPRTRLCFA HVVSDPQPLLGLVRAGIPVFD TGEVRRTLDP SRPSADPSDPAGRAVPADTRRRERDLR DRSADRGPARRGREGRRSA	WP_013003263
<i>Streptomyces</i> sp. WMMB 272	MADSSATDCVLREVAFRSYRAEHTVTVQCTVRPDSGPGQAEGYGTATTEAVARKKA LSEAVERLLACTPFALIARRPTPTDRSRGTDWPQRLRSAPAGCQVRNYSLSGEQP RQVPLFWSSPWWAGEEMRSANLAPQAARLSSTVGWAVAPSPEALQGALYELTELL NYGAFLYRCLAAPRERPAPGGGPEPREEGRIQTVPISWATRTPTILAVTRSSDRLMPA TGIGSGTTASEAAERAFLELAQAETLWRANSTAVPAERYFMRRFEKWPLLRRCAALD FDLGEHAGPFDDGPHPSPLGALESRAIEVWADPGAVDISGADLRRLRCFAHVVS DP QPLLGLVRAGIPVFD TGEVRKILVCPRRARSADSPSGRVAHSGRPTRA	AFU90403
<i>Streptomyces scabiei</i> 87.22	MREATATECELVHRSYPSEGTVTVRCTVRPAEGDARADGYGTATTKAVARAKAL SEAVERLVACTPFTAVARPPATAFAGPGTGSGPVPPFPAAGVRTAPDGCAGRVRPLT GGGPRRVPLYWSSPWTAGEELRAGLTAAQARLSSTVGWAVAPTPEAALRGALLEL TELIGYGVFLHRRLAGPPRRPAADDRTL VVPLDGVGRTPAVLAVAYGRGRMPATGL GCGATRDEATGRALLELAQAETMWRSNPTAEP AERFFLRFRERWPLLTRCATLDFDL RDANAPHDPAPWDEPRPASPLEELEAGGIRVWADSGALDISGPDIPRTRLCFAHV V SDPQPLLGLVRAGIPVFD TGEVRRTLDP SRRSADPSDPADPADPADRAVPADAR RERRELRDRPADRGPARRGREGRRSA	WP_059081255
<i>Streptomyces ipomoeae</i>	MLEATATECALREVVRYSYPSERTVTVRCTVRPAEGSARADGYGTATKAVARLKALS EAVERLVACTPFASVVRPPTTSRSGSAPVPPFASGVRTVPDGCASRVYRPLTGGGPRR VPLYWSSPWTAGEELRAAVLTAPEARLSSTVGWAVAPTPEAALHGALLELT ELLNHG VFLHRSLAGPPRPAAGDET LVPLGGPVRTPTVLAVAYGRGRMPATGLGCGATRAE ATDRALLELAQAETMWRSNPTAVPAERFFLRFRERWPLLGRCATLDFDLSGHGTPYD DERSVASPLEELEAGGISVWADSGAVDISGPDTPRTRLCFAHVVS DPQPLLGLVRAGI PVFD TGEVRRTLDP PRRDRPAHHARSADHRAAGRGQRRTA	WP_048819674
<i>Streptomyces bottropensis</i>	MREATATECELVHRSYPSERTVTVRCTVRPAEGSAQADGYGTATTEAVARAKAL SEAVERLVACTPFATVARPPATA CADPGTGSGPVPPFPAAGVRTAPEGCAGRVRPLT GGGPRRVPLYWSSPWTAGAE LRAGLTAAEARLSSTVGWAVAPTPEGALRGALLELT ELIDYGVFLHRRLAGPARRPSAGDRTL VVPLGGIGRTPAVLAVAYGRGRMPATGLGC GATRGEATDRALLELAQAETMWRSNPTAEP AERFFLRFRERWPLLARCATLDFDLPD PRDPPARSGPAGGPGPYDDGPCASPLEELEAGGIRVWADSGAVDISGPDIPRTRLCF AHVVS DPQPLLGLVRAGIPVFD TGEVRRTLDP SRRPADRPADRGGHRDRSADRGPA RRGHEGRRSA	WP_005486705
<i>Streptomyces purpureus</i>	MLEASVADCALRDVHRSYRSRDTVTVRCTVHPASGSPPADGYGTATTEAVARKKAL SEAVERLLACTPFASVARPPTHLT TTGDDRNWPAGVRTAPEGCLTRTYRPLTRGAL PRQVPLFWSSPWIAGEELRSALLTAQTARLSSTIGWAVAPSPEALRGALFELTELIN Y GVFLFRALAGPPHDGRDSGPGPLIPIDGATRTPTVLAVARGGGRMPATGLGSGTT RADATDRALLELAQAVTLWRSNPTGIPAERHFMRRFDRWPLLTRCATLDFDLTG YDA PPDGVPRSPLEELEARGIAVWADSGALDLSGPD LARTRLFHAQVADPQPLLGLVR AGIPVFD TGEVRKILDSSRRERPADRRPSGRNAQGGC PARA	WP_019887077
<i>Streptomyces sulphureus</i>	MADSSATDCVLREVAFRSYRAEHTVTVQCTVRPDSGPGQAEGYGTATTEAVARKKA LSEAVERLLACTPFALIARRPTPTDRSRGTDWPQRLRSAPAGCQVRNYSLSGEQP RQVPLFWSSPWWAGEEMRSANLAPQAARLSSTVGWAVAPSPEALQGALYELTELL NYGAFLYRCLAAPRERPAPGGGPEPHEEGRTQTVPISWATRTPTILAVTRSSDRLMPA TGIGSGTTASEAAERAFLELAQAETLWRANSTAVPAERYFMRRFEKWPLLRRCAALD	WP_019549855



	FDLGEHAGPFDDGPHPSPLGALESRAIEVWADPGAVDISGADLRRLRCFAHVVS QPLLGLVRAGIPVFDTGGEVRKILVCPRRARSADLSPSGRVAHSGRPTRA	
<i>Streptomyces</i> sp. BC16019	MLEATATECELVHRSYPSDRTVTVRCTVRPAEGTAQADGYGTATTEAVARAKAL SEAVERLVACTPFATVARPPTTPAGSGSGPVPPFPAAGVRTAPDGCASREYRPLSGD GPRRVPLYWSSPWVAGEELRAGTLSAAEARLSSTVGWAVAPTPEAALRGALLELELI DYGVLHRRLAGPARPRSAGDETLVVPLGGAVRTPTVLAVAYGRGRIMPATGLGCG ASRAEATDRALLELAQAETMWRSNPTAEPAAERFFLRERWPLLTRCATLDFELSDLS GLDGLSGLDGLSRLDGLSRLDHPDLNDPAGRPAPYDDGPHASPLEELEAGGITVW ADSGAVDISGPDIPRTRLCFAHVVSQPLLGLVRAGIPVFDTGGEVRRLDPSRRDAH RDTPRGAPRDAARDSADRGPAGRGRQGRSA	AFV25482

**Table S2:** List of substrate peptides pre-incubated together with IpoC. BotA\_Follower+C9 variants were chosen especially as IpoC has shown a specific increase in melting temperature upon incubation with these peptides.<sup>[3]</sup>

Substrate peptide	Primary amino acid sequence
BotA <sup>P</sup>	GPVVVFDCMTADFLNDDPNNAELSALEMEELESWGAWDGEATS
BotA <sup>P</sup> -5AA	GPVVVFDCMTADFLNDDPNNAELSALEMEELESWGAWD
BotA_Follower	MTADFLNDDPNNAELSALEMEELESWGAWDGEATS
BotA_Follower+C9	CMTADFLNDDPNNAELSALEMEELESWGAWDGEATS
BotA_Follower+C9-5AA	CMTADFLNDDPNNAELSALEMEELESWGAWD

**Table S3:** Colour-coded BotA<sup>P</sup> precursor peptide. Follower peptide residues are highlighted in green, negatively charged residues are highlighted in red. Within 35 follower peptide amino acids, we observe 10 negatively charged side chains and thus a net charge of -10.

BotA <sup>P</sup>	GPVVVFDCMTADFLNDDPNNAELSALEMEELESWGAWDGEATS
-------------------	---

## Materials and Methods

### Cloning and expression of IpoC constructs

Several IpoC expression constructs were designed based on either genomic information on an alternative N-terminus (**Figure S2**), mass spectrometry data (**Figure S3**) or calculation of disordered regions and secondary structure prediction with the *xtalpred* server.<sup>[4]</sup> The corresponding primers were ordered from Sigma-Aldrich and the different gene constructs amplified by PCR with a standard Phusion polymerase protocol with adjusted annealing temperatures. DNA amplicates were subsequently cut with *NcoI* and *HindIII* in a 1 : 0.5 ratio in buffer R at 37 °C for 4 h and ligated into an already cut pSUMO vector (the plasmid was a gift by Dr. David Owen, St. Andrews University<sup>[5]</sup>) treated with alkaline phosphatase with a standard T4 DNA ligase protocol. Ligation samples were transformed into chemically competent *E. coli* HS996 with a standard heat shock protocol, plated on LB-Agar with 50 µg/mL Kanamycin and grown at 37 °C and 200 rpm for 16 h. Single *E. coli* colonies were grown in 10 mL LB medium with 50 µg/mL Kanamycin for 16 h at 37 °C and 200 rpm. 4 mL of grown culture were centrifuged for 10 min at 4000 rpm and 4 °C and the cell pellet was used for a plasmid preparation by alkaline lysis. Extracted plasmids were cut with *NcoI* and *HindIII* at a 1 : 0.5 ratio in buffer R at 37 °C for 4 h and analyzed by agarose gel electrophoresis. Plasmids which were found to carry the correct insert were sent for sequencing by LGC genomics (Berlin) and transformed into chemically competent *E. coli* Lemo21(DE3) cells.

Small and large scale expression and purification of different IpoC constructs was carried out as described earlier with slight changes to the gel filtration buffer.<sup>[3]</sup> As the main goal for crystallization was to increase the solubility of IpoC, different gel filtration buffer adjustments were tried and the resulting characteristics observed in crystallization trials. The final buffer used for crystallization was 1 M NaCl, 20 mM HEPES, 5% glycerole, 1 mM DTT, final pH 7.0.

### Size exclusion properties of IpoC

For the analysis of the assembly of the expressed IpoC construct in solution, 4 mg of 4 monomeric model proteins (Ribonuclease A, Ovalbumin, Conalbumin, Aldolase) of the low and high molecular weight gel filtration calibration kit (GE Healthcare) were applied onto a HiLoad 16/60 Superdex 200 pg gel filtration column pre-equilibrated in IpoC gel filtration buffer and eluted with a flow rate of 1 mL min<sup>-1</sup>. For comparison, 4 mg of the crystallized IpoC construct were eluted with the same conditions. The respective chromatogram values were extracted and the resulting figure was created in Origin (Origin(Pro), Version 2017) OriginLab Corporation, Northampton, MA, USA).

### Crystallization of IpoC\_apo

Several sparse matrix and grid screens (Core, PEGS, pHClear Suites, Qiagen) were set up for initial crystallization trials of the different IpoC constructs using a Crystal Gryphon LCP (Art Robbins Instruments). Furthermore, different protein concentrations, crystallization temperatures and additives were applied to observe differences in the crystallization properties of IpoC. Crystals of IpoC\_apo could be detected after 5 weeks of incubation at 4 °C at a 1 : 2 protein : well solution ratio in 1 M LiCl, 0.1 M MES pH 6.0 and 20% PEG6000 with a final well solution pH of 6.0 and a protein concentration of 200 µM. Within this batch of purified protein, crystals could be reproduced and optimized using a Dragonfly liquid handler (TTP Labtech) with varying concentrations of precipitant or salt and changing the final pH of the well solution. For data collection, crystals were cryoprotected by adding 32% glycerol to 68% well solution and quick soaking the crystals in this cryo solution. Subsequently, crystals were mounted using mounted cryoloops (Hampton Research) and flash frozen in liquid nitrogen until further use.

### Data collection and structural analysis of IpoC\_apo

Data for IpoC\_apo crystals was recorded on beamline I23 located at Diamond Light Source (Didcot, UK). High redundancy datasets were taken in-vacuum at 4.5 keV which were used for a combination of Sulphur and Chloride phasing to solve the structure. The presence of Chloride atoms within the structure was confirmed by taking datasets above (3.0 keV) and below (2.8 keV) the absorption edge of Cl. Several rounds of manual model building and refinement were carried out in *Coot*<sup>[6]</sup> and *phenix.refine*<sup>[7]</sup> within the PHENIX suite<sup>[8]</sup>, respectively, to obtain a model which was used for molecular replacement with *phenix.phaser*<sup>[9]</sup> with a high resolution dataset taken at 2.1 Å. The final PDB coordinates were analyzed with *MolProbity*<sup>[10]</sup> by the *PISA* server<sup>[11]</sup> for the detection of a macromolecular assembly as well as the *DALI* server<sup>[12]</sup> for the detection of structural homology. All structural images portrayed were rendered in *PyMOL* (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC).

## Supplementary References

- [1] Gerlt JA, Bouvier JT, Davidson DB, Imker HJ, Sadkhin B, Slater DR, Whalen KL *Biochim Biophys Acta*. **2015** 1854 (8):1019-37.
- [2] Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R *Nucleic Acids Res*. **2019** [Epub ahead of print]
- [3] Franz L, Adam S, Santos-Aberturas J, Truman AW, Koehnke J *J Am Chem Soc*. **2017** 139 (50):18158-18161
- [4] Slabinski L, Jaroszewski L, Rychlewski L, Wilson IA, Lesley SA, Godzik A *Bioinformatics*. **2007** 23 (24):3403-5
- [5] Liu, H.; Naismith, J. H., *Protein Expr Purif* **2009** 63 (2):102-11
- [6] Emsley P, Lohkamp B, Scott WG, Cowtan K *Acta Crystallogr D Biol Crystallogr*. **2010** 66 (Pt 4):486-501
- [7] Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD *Acta Crystallogr D Biol Crystallogr*. **2012** 68 (Pt 4):352-67
- [8] Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH *Acta Crystallogr D Biol Crystallogr*. **2010** 66 (Pt 2):213-21
- [9] McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ *J Appl Crystallogr*. **2007** 40 (Pt 4):658-674
- [10] Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC *Acta Crystallogr D Biol Crystallogr*. **2010** 66 (Pt 1):12-21
- [11] Krissinel E, Henrick K *J Mol Biol*. **2007** 372 (3):774-97
- [12] Holm L, Laakso LM *Nucleic Acids Res*. 2016 44 (W1):W351-5

---

## Chapter 5

# Characterization of the stereoselective P450 enzyme BotCYP enables the *in vitro* biosynthesis of the Bottromycin core scaffold

**Sebastian Adam**<sup>[a]1</sup>, Laura Franz<sup>[a]1</sup>, Mohammed Milhim<sup>[b]</sup>, Rita Bernhardt<sup>[b]</sup> and Jesko Köhnke<sup>[a]</sup>

*Manuscript close to submission*

Note: The manuscript was already put into the appropriate template for submission, but some figures and tables were not ready for publication before the submission of this thesis due to unexpected experimental delay. Therefore, a fraction of the presented figures (Figure 2A, S3 and S13) and tables (Table S6) feature preliminary data.

### Affiliation

<sup>[a]</sup> Workgroup Structural Biology of Biosynthetic Enzymes, Helmholtz Institute for Pharmaceutical Research, Helmholtz Centre for Infection Research, Saarland University, Universitätscampus E8 1, 66123, Saarbrücken, Germany.

<sup>[b]</sup> Institute of Biochemistry, Saarland University, 66123, Saarbrücken, Germany.

<sup>1</sup> These authors contributed equally to the manuscript

## **Contributions and Acknowledgements**

### **Author's effort:**

The author contributed significantly to this work by establishing the first heterologous protein expression and purification of the Cytochrome P450 enzyme from the bottromycin pathway in our lab. He crystallized SalCYP and its N-terminal truncation, solved and refined the x-ray crystal structures. The author established a protocol for the purification of BotA<sup>PCCDAH</sup> (the substrate of BotCYP) and purified the first batch, most of which was used for the results shown in this work. He performed the binding affinity experiments, designed figures and edited the manuscript.

### **Other's effort:**

Laura Franz contributed significantly by performing and evaluating the presented biochemical experiments qualitatively as well as quantitatively. She performed the cloning for the SalCYP point mutants, purified a batch of substrate and optimized the enzymatic processes for a potential large-scale application. Furthermore, she designed figures and edited the manuscript. Mohammed Milhim contributed by performing the heme reduction experiments, providing plasmids and protocols for the purification of redox partners and suggested an optimized construct for BotCYP.

## 5.1 Abstract

Bottromycins are ribosomally synthesized and post-translationally modified peptide natural product antibiotics that display activity against problematic gram-positive human pathogens such as methicillin-resistant *Staphylococcus aureus*. Here we report the biochemical characterization of the cytochrome P450 enzyme BotCYP from a bottromycin biosynthetic gene cluster. We demonstrate that BotCYP is responsible for the oxidative decarboxylation of a C-terminal thiazoline to yield a thiazole. We determined the structure of the close BotCYP homolog SalCYP, which possesses a deep, highly conserved active site cleft suitable for binding large peptidic substrates. We demonstrate that BotCYP cooperates with the epimerase BotH, which enables the facile in vitro production of the bottromycin core scaffold with quantitative yields. Our work will enable the rapid generation of bottromycin analogues for compound development.

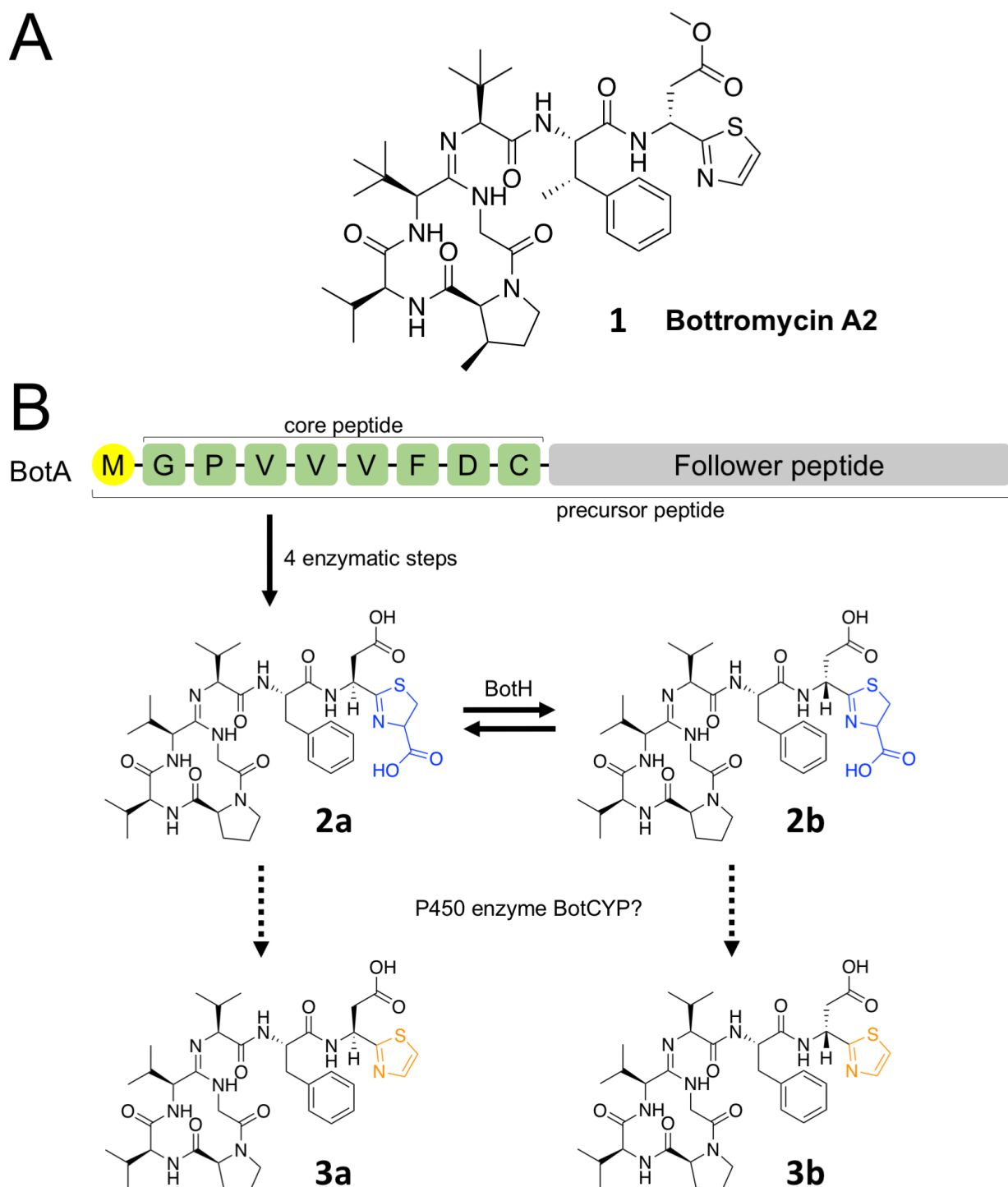
## 5.2 Main text

Ribosomally synthesized and post-translationally modified peptides (RiPPs) are a rapidly growing natural product superfamily with interesting and diverse bioactivities. Bottromycins (**1**, **Scheme 1A**) are RiPPs with potent activity against gram positive bacteria, including the problematic human pathogens MRSA and VRE.<sup>[1,2]</sup> Uniquely amongst bacterial RiPPs, the bottromycin precursor peptide contains an N-terminal core peptide (the eventual natural product) and a follower peptide, which is important for substrate recognition by several of the biosynthetic enzymes.<sup>[3,4]</sup> *In vitro* analyses of the biochemical transformations leading to bottromycins have largely validated the proposed order of steps: Removal of the N-terminal methionine, thiazoline and macroamidine formation, followed by removal of the follower peptide.<sup>[3,5,6]</sup> The resulting intermediate **2a** is then rapidly interconverted enzymatically with **2b** by BotH to yield a mixture of the two epimers (**Scheme 1B**). The *D*-Asp configuration of **2b** is also found in the final natural product bottromycins (**Scheme 1A**).

Five-membered heterocycles that are enzymatically derived from serine, threonine or cysteine residues are frequently found in RiPPs.<sup>[7]</sup> Their oxidation status, azoline or azole, can have a profound effect on RiPPs bioactivity, and all enzymes linked with azoline oxidation thus far are FMN-dependent enzymes. In contrast, it has been suggested that the oxidative

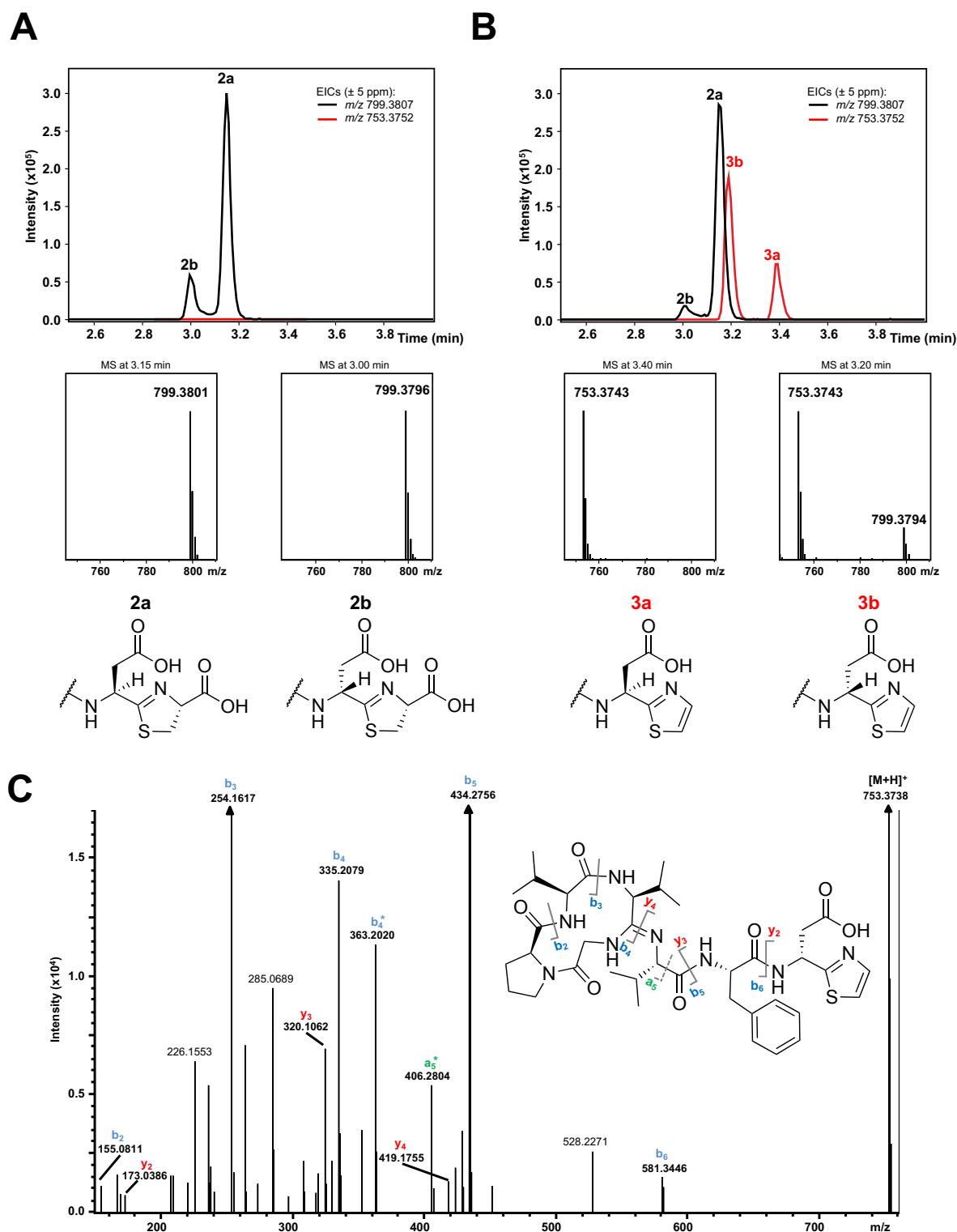
decarboxylation reaction of a thiazoline in bottromycin biosynthesis is catalyzed by a P450 enzyme (BotCYP) found in the bottromycin biosynthetic gene cluster (**Figure S1**).

**Scheme 1.** **A** Structure of bottromycin A2. **B** *In vitro* biosynthesis of **2a/b**, with a proposed role for the P450 enzyme BotCYP to yield the bottromycin core scaffolds **3a/b**.





To investigate the role of this enzyme *in vitro*, we used heterologously expressed and purified BotCYP in spectral analyses to confirm an intact heme-containing protein and incubation with sodium dithionite led to the observation of the typical Soret band at 448 nm in the carbon monoxide spectrum, indicating the chemical reduction of the heme iron (**Figure S2A**).<sup>[8]</sup> Next, we applied different ratios of several P450 reductase-ferredoxin pairs to the enzyme to achieve an enzymatic reduction of the heme iron and observed the highest peak at 448 nm using a P450 reductase-ferredoxin pair from *Bacillus megaterium*<sup>[9]</sup> (BmCPR-Fdx2) in a 1:3:10 (BotCYP : BmCPR : Fdx2) ratio (**Figure S2A**). We then produced the putative substrate mixture **2a/b** biochemically *in vitro* using three enzymes (Details can be found in the supplementary material).<sup>[3,5,6]</sup> We incubated **2a/b** with a further optimized BotCYP : BmCPR : Fdx2 ratio of 1:1:10 and a molar excess of NADPH in the absence and presence of BotCYP at 30 °C overnight. The reaction mixture was subsequently analyzed by high-resolution liquid chromatography-mass spectrometry (HR-LCMS, **Figures 1A** and **1B**). In the absence of BotCYP, we observed no effect on **2a/b** ([M+H]<sup>+</sup>calc.<sub>mono.</sub>: 799.3807 Da; [M+H]<sup>+</sup>obs<sub>mono.</sub>: r.t. 3.15 min 799.3801 Da, error –0.75 ppm, r.t. 3.10 min 799.3796 Da, error –1.38 ppm). In the presence of BotCYP however, two new peaks appeared, which had identical masses that corresponded to the decarboxylated, oxidized reaction product **3** ([M+H]<sup>+</sup> calc.<sub>mono.</sub>: 753.3752 Da, [M+H]<sup>+</sup>obs<sub>mono.</sub>: r.t. 3.40 min 753.3743 Da, error –1.19 ppm, r.t. 3.20 min 753.3743 Da, error –1.19 ppm). Due to their different retention times we termed them **3a** and **3b** (**Figure 1B**). Analysis of both peaks by tandem mass spectrometry (MS<sup>2</sup>) strongly supports their assignment as **3** (**Figures 1C** and **S2B**, **Tables S1** and **S2**).



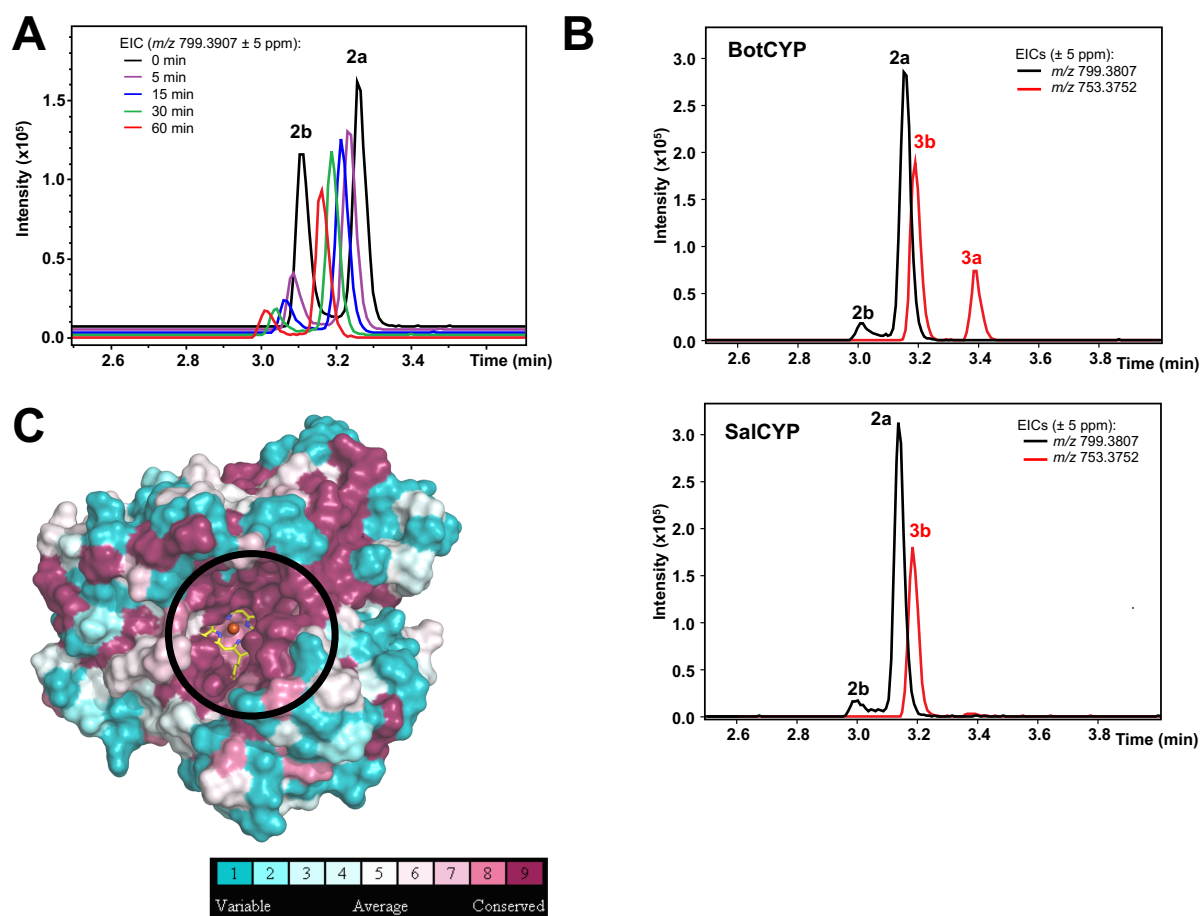
**Figure 1.** **A** Incubation of putative BotCYP substrates **2a/b** with all co-factors results in no detectable change in the HR-LCMS spectrum. **B** Same as A, but with BotCYP added. The new species **3a/b** show a loss of 46 Da, which is in agreement with the expected product. **C** MS<sup>2</sup> spectrum of **3b**. a, b and y ions are indicated.

These data establish the P450 enzyme found in the bottromycin biosynthetic gene cluster as sufficient for the oxidative decarboxylation of the C-terminal thiazoline to a thiazole. We

estimate the yield of this reaction to be poor (< 25%). The presence of two product peaks led us to speculate that the enzyme may not be stereoselective, thus converting both substrate epimers to the respective products. When using an enzymatically produced equimolar mixture of **2a** and **2b**, time-course experiments allowed us to demonstrate that the *D*-Asp containing **2b** is preferentially consumed by BotCYP, but that **2a** is also a substrate (**Figure 2A**).

To better understand BotCYP, we attempted to determine the crystal structure of the full-length protein. Extensive screening did not yield crystals. We thus selected 5 close homologs of the protein and protein expression trials resulted in soluble expression of the homologue SalCYP from *Salinispora tropica*, which shares 68% sequence identity with BotCYP (**Figure S1**). We expressed and purified SalCYP as described for BotCYP. SalCYP was able to bind a **2a/b** mixture with low  $\mu\text{M}$  affinity (**Figure S3**) and was also able to perform the oxidative decarboxylation reaction (**Figures 2B** and **S4**). Intriguingly, SalCYP reactions resulted in only one product peak, **3b** (**Figure 2B**). SalCYP crystals belonged to space group  $P2_1$  and a high-resolution dataset was collected to 1.8 Å. All data collection and refinement statistics can be found in **Table S3**. Molecular replacement was unsuccessful, and the structure was determined using single-wavelength anomalous dispersion. A high-redundancy dataset was collected at the iron edge, using a wavelength of 1.738 Å. The overall structure of SalCYP (**Figure S5**) reveals a heme b bound at the reactive center that is coordinated by Cys354, forms two salt bridges between the two heme carboxylic acid moieties and Arg104 and Arg298 of SalCYP as well as extensive hydrophobic interactions (**Figure S6**). Surprisingly, we found the N-terminus (residues 1 – 20) of the protein to be ordered and extend away from the protein, with the amino group of the N-terminus coordinating the heme iron of a symmetry mate (**Figure S7**). To allow unbiased analysis of the binding pocket, we deleted the N-terminal 20 amino acids, but this proved detrimental to crystallization. After a careful analysis of the contacts between the N-terminal residues and the symmetry mate, we deleted residues 1-13 and introduced mutation Thr14Asp, to create a new crystal contact. This mutant protein, SalCYP<sup>T</sup>, readily crystallized in space group  $P4_1$  and a dataset was collected to 1.5 Å. A search for structural homologs using the DALI server<sup>[10]</sup> revealed the camphor hydroxylase CYP101D2 as the closest structural homolog (PDB ID 4dxy) with a  $C_\alpha$  rmsd of 1.5 Å over 276 residues (**Figure S8**). The structural homology of SalCYP to the well-characterized OleTJE<sup>[11]</sup>, that catalyzes the oxidative decarboxylation of fatty acids, was calculated as a  $C_\alpha$  rmsd of 2.9 Å over 310 residues (PDB ID 5m0n)(**Figure S8**). The most striking difference between SalCYP and

virtually all close structural homologs is the wide and deep active site cleft found in SalCYP (**Figures 2C** and **S9**). The only exception was the structure of TbtJ1 (PDB ID 5vws), which is the only other P450 enzyme structure from a RiPPs pathway.<sup>[12]</sup> This protein catalyzes the hydroxylation of thiomuracin<sup>[13-14]</sup> and also possesses a large, albeit smaller wide, open active site topology. Taken together these structures thus indicate that RiPPs P450 enzymes have evolved to perform selective reactions within complex linear and cyclic RiPP core peptides (**Figure S9**).

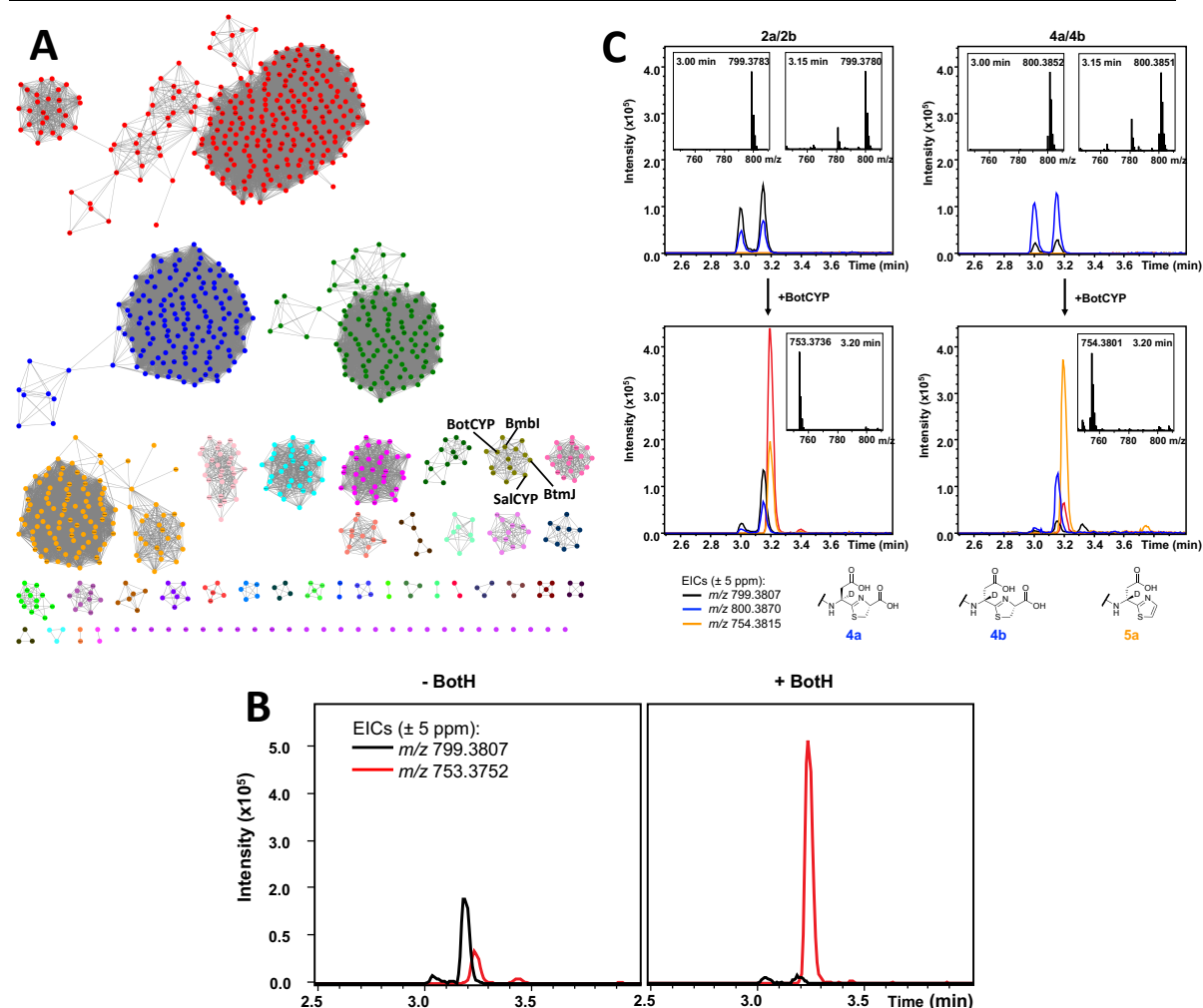


**Figure 2:** **A** LC-MS data show that BotCYP preferentially consumes **2b**. **B** Comparison of BotCYP and SalCYP reactions with **2a/b**. Only one product peak is observed when SalCYP is used. **C** ConSurf map showing the conservation of residues around the SalCYP active site (circle) for SalCYP homologues from other bottromycin biosynthetic gene clusters (**Figure 3A**).

We generated a sequence similarity network for SalCYP and found it to be part of a small node that exclusively contained P450 enzymes from (putative) bottromycin biosynthetic gene clusters (**Figure 3A**), which echoes the findings for the amidohydrolase PurAH. It thus appears that the enzymes involved in bottromycin biosynthesis have evolved to fulfil a very specialized function. To understand the difference in stereoselectivity between SalCYP and BotCYP, we aligned all sequences belonging to the SalCYP node (**Figure S10**). Only two residues in the

vicinity of the heme iron at the active site are variant, Thr242 and Val243 (SalCYP numbering). In all homologues, a single point mutation occurred (Thr242Cys), with 69% of all homologues including BotCYP also carrying another mutation (Val243Thr). To analyze the effect of these mutations on the stereoselectivity of the enzyme, we produced two SalCYP mutant enzymes: SalCYPV243T, and SalCYPT242C/V243T. Both mutant enzymes retained their stereoselectivity (**Figure S11**), which implies that additional mutations, distant from the active site, cause slight changes in the protein architecture to facilitate stereoselectivity. Alternatively, BotCYP, part of an intact biosynthetic gene cluster, may simply be more reactive and thus turn over the wrong epimer *in vitro* and *in vivo*.

The exchange between **2a** and **2b** is slow, and **2b** (*D*-Asp) was the preferred substrate for BotCYP. We thus speculated that the poor yield of the BotCYP reaction may be the result of depleting **2b** in the reaction mixture. We thus added the recently characterized bottromycin epimerase BotH to the reaction mixture. This enzyme leads to the rapid interconversion of **2a** and **2b**, even at enzyme concentrations that have no impact on the relative abundance of these two species. Analysis of reactions containing BotH and BotCYP by HR-LCMS demonstrates that the oxidative decarboxylation reaction now goes almost to completion (**Figure 3B**). Critically, BotCYP now produces a single peak, as observed for SalCYP. We know that **2a** contains *L*-Asp and **2b** *D*-Asp, and determined that **2b** is preferentially consumed by BotCYP. It may thus be intuitive to assume that the reaction product **3b** also harbors the *D*-Asp found in the natural product. To probe the stereochemistry at the Asp C $\alpha$ -position of **3b**, we enzymatically produced substrate selectively labeled with a deuteron at this position (**4a/b**) [M+H]<sup>+</sup>calc.<sub>mono.</sub>: 800.3870 Da; r.t. 3.72 min (**4a**) [M+H]<sup>+</sup>obs<sub>mono.</sub>: 800.3866 Da, error -0.50 ppm, and r.t. 3.57 min (**4b**) [M+H]<sup>+</sup>obs<sub>mono.</sub>: 800.3868 Da, error -0.25 ppm) (**Figure S12**). We then used **4a/b** as a substrate in a BotCYP reaction, the resulting peak had the same fragmentation pattern as **3b** (**Figures 3C** and **S13**). Crucially, the product peak (**5a**) retained the mass shift of + 1 Da ([M+H]<sup>+</sup>calc<sub>mono.</sub>: 754.3815 Da, observed: 754.3806 Da, error -1.19 ppm), indicating that the configuration of the Asp C $\alpha$ -position is retained during the oxidative decarboxylation reaction. Intriguingly, upon incubation of **5a** in H<sub>2</sub>O in combination with BotH, an exchange of the deuteron at the Asp C $\alpha$ -position is no longer observed. Presence of the deuteron at the Asp-position was confirmed by MS<sup>2</sup> (**Figure S13**). These data strongly imply that **3b** represents the bottromycin core scaffold harboring the *D*-Asp moiety found in bottromycins.



**Figure 3:** **A** Sequence similarity network generated using the 1000 closest SalCYP homologs identified by a PSI-BLAST search. The network was generated with a minimum/maximum sequence length of 350/500 and an alignment score of 100. **B** Comparison of BotCYP reaction with (right) and without (left) BotH. Addition of BotH enables almost complete turnover. **C** BotCYP reaction using **4a/b** as a substrate. The product **5a** retains its mass shift of + 1 Da.

We have demonstrated that the P450 enzyme found in the bottromycin biosynthetic gene cluster is sufficient for the oxidative decarboxylation of the C-terminal thiazoline found in the bottromycin biosynthetic intermediate. To function efficiently, it needs to cooperate with the epimerase BotH. Curiously, we were unable to detect complex formation between the two proteins (data not shown). This mirrors the results for the bottromycin biosynthetic enzymes BotCD and BotAH, which are both required for efficient macroamidine formation. Our data also suggest that the use of SalCYP for the heterologous production of bottromycins may be advantageous, since it should quash the production of the wrong epimer and thus improve yields. While P450 enzymes play pivotal roles in the biosynthesis of many bacterial natural products,<sup>[15]</sup> they are not very common in RiPPs pathways. Only a handful of RiPPs P450 enzymes have an assigned function to date: MibO from microbisporicin biosynthesis, TbtJ1

and TbtJ2 from thiomuracin biosynthesis and TsrR from thiostrepton biosynthesis all catalyze the hydroxylation of amino acids.<sup>[13-14, 16]</sup> GetJ from GE37468 biosynthesis catalyzes the conversion of isoleucine to  $\delta$ -hydroxyproline and TsrP, also involved in thiostrepton biosynthesis, which catalyzes an epoxidation that triggers additional changes.<sup>[16-17]</sup> None of these proteins catalyze an oxidative decarboxylation reaction, nor are they involved in heterocycle biochemistry. BotCYP thus expands the catalytic scope of P450 enzymes involved in RiPPs biosynthesis. With the reconstitution of BotCYP, we also present a facile *in vitro* route to the bottromycin core scaffold with quantitative yields, which will enable the rapid generation of bottromycin analogues for antibiotic activity testing.

### 5.3 Acknowledgements

We acknowledge the use of ESRF beamlines ID29 and ID30a3. Jesko Koehnke is the recipient of an Emmy Noether Fellowship (KO4116/3-1). We thank Asfandyar Sikandar for providing BotH.

## 5.4 References

- [1] Hou Y, Tianero MD, Kwan JC, Wyche TP, Michel CR, Ellis GA, Vazquez-Rivera E, Braun DR, Rose WE, Schmidt EW, Bugni TS *Org Lett*. **2012** 14 (19):5050-3
- [2] Shimamura H, Gouda H, Nagai K, Hirose T, Ichioka M, Furuya Y, Kobayashi Y, Hirono S, Sunazuka T, Omura S *Angew Chem Int Ed Engl*. **2009** 48 (5):914-7
- [3] Franz L, Adam S, Santos-Aberturas J, Truman AW, Koehnke J *J Am Chem Soc*. **2017** 139 (50):18158-18161
- [4] Schwalen CJ, Hudson GA, Kosol S, Mahanta N, Challis GL, Mitchell DA *J Am Chem Soc*. **2017** 139 (50):18154-18157
- [5] Mann G, Huo L, Adam S, Nardone B, Vendome J, Westwood NJ, Müller R, Koehnke J *Chembiochem*. **2016** 17 (23):2286-2292
- [6] Sikandar A, Franz L, Melse O, Antes I, Koehnke J *J Am Chem Soc*. **2019** 141 (25):9748-9752
- [7] Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, Camarero JA, Campopiano DJ, Challis GL, Clardy J, Cotter PD, Craik DJ, Dawson M, Dittmann E, Donadio S, Dorrestein PC, Entian KD, Fischbach MA, Garavelli JS, Göransson U, Gruber CW, Haft DH, Hemscheidt TK, Hertweck C, Hill C, Horswill AR, Jaspars M, Kelly WL, Klinman JP, Kuipers OP, Link AJ, Liu W, Marahiel MA, Mitchell DA, Moll GN, Moore BS, Müller R, Nair SK, Nes IF, Norris GE, Olivera BM, Onaka H, Patchett ML, Piel J, Reaney MJ, Rebuffat S, Ross RP, Sahl HG, Schmidt EW, Selsted ME, Severinov K, Shen B, Sivonen K, Smith L, Stein T, Süßmuth RD, Tagg JR, Tang GL, Truman AW, Vederas JC, Walsh CT, Walton JD, Wenzel SC, Willey JM, van der Donk WA *Nat Prod Rep*. **2013** 30 (1):108-60
- [8] Guengerich FP, Martin MV, Sohl CD, Cheng Q *Nat Protoc*. **2009** 4 (9):1245-51
- [9] Milhim M, Gerber A, Neunzig J, Hannemann F, Bernhardt R *J Biotechnol*. **2016** 231:83-94
- [10] Holm L, Laakso LM *Nucleic Acids Res*. **2016** 44 (W1):W351-5
- [11] Wang JB, Lonsdale R, Reetz MT *Chem Commun (Camb)*. **2016** 52 (52):8131-3
- [12] Gober JG, Ghodge SV, Bogart JW, Wever WJ, Watkins RR, Brustad EM, Bowers AA *ACS Chem Biol*. **2017** 12 (7):1726-1731
- [13] Hudson GA, Zhang Z, Tietz JI, Mitchell DA, van der Donk WA *J Am Chem Soc*. **2015** 137 (51):16012-5
- [14] Zhang Z, Hudson GA, Mahanta N, Tietz JI, van der Donk WA, Mitchell DA *J Am Chem Soc*. **2016** 138 (48):15511-15514
- [15] Greule A, Stok JE, De Voss JJ, Cryle MJ *Nat Prod Rep*. **2018** 35 (8):757-791
- [16] Zheng Q, Wang S, Liao R, Liu W *ACS Chem Biol*. **2016** 11 (10):2673-2678
- [17] Young TS, Walsh CT *Proc Natl Acad Sci U S A*. **2011** 108 (32):13053-8



## Supporting Information

# Characterization of the stereoselective P450 enzyme BotCYP enables the *in vitro* biosynthesis of the Bottromycin core scaffold

**Sebastian Adam**<sup>[a]1</sup>, Laura Franz<sup>[a]1</sup>, Mohammed Milhim<sup>[b]</sup>, Rita Bernhardt<sup>[b]</sup> and Jesko Köhnke<sup>[a]</sup>

*Manuscript close to submission*

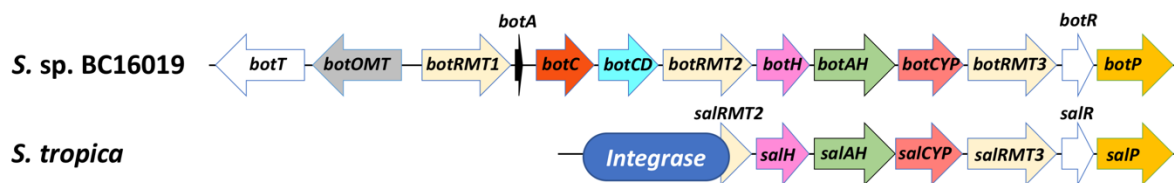
Note: The manuscript was already put into the appropriate template for submission, but some figures and tables were not ready for publication before the submission of this thesis due to unexpected experimental delay. Therefore, a fraction of the presented figures (Figure 2A, S3 and S13) and tables (Table S6) feature preliminary data.

### Affiliation

<sup>[a]</sup> Workgroup Structural Biology of Biosynthetic Enzymes, Helmholtz Institute for Pharmaceutical Research, Helmholtz Centre for Infection Research, Saarland University, Universitätscampus E8 1, 66123, Saarbrücken, Germany.

<sup>[b]</sup> Institute of Biochemistry, Saarland University, 66123, Saarbrücken, Germany.

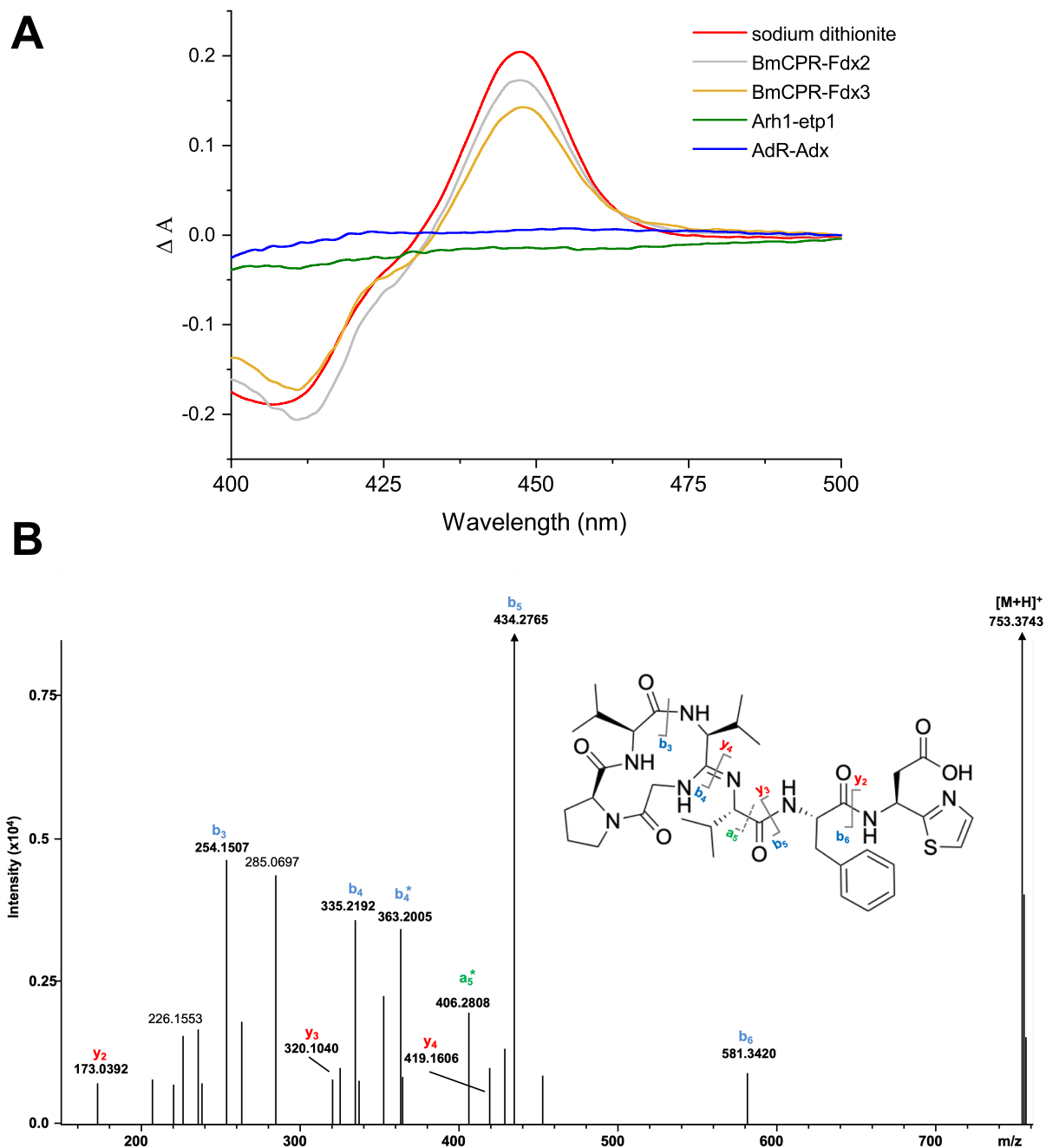
<sup>1</sup> These authors contributed equally to the manuscript

**A****B**

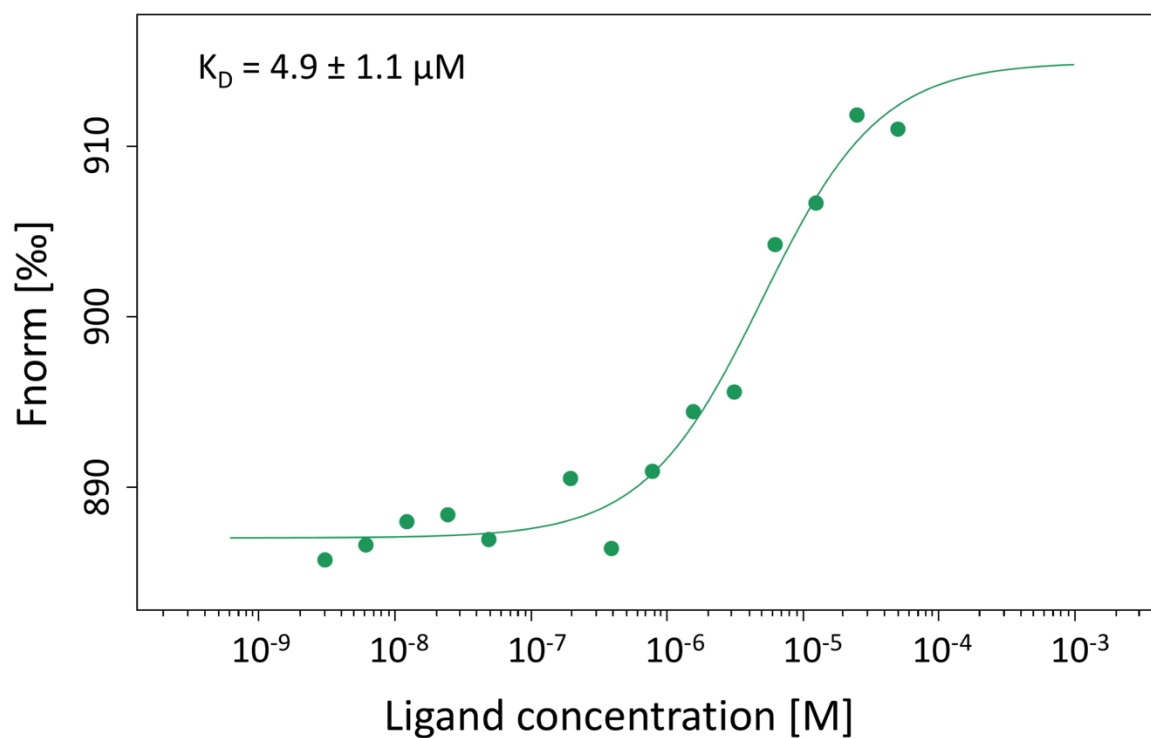
CLUSTAL O(1.2.4) multiple sequence alignment

BotCYP	MQADREPDRGTGREAKRPAEPDHPQATCPVDFDFFAAPQTYRDAAAEHAGEHGAFYS-	59
SalCYP	MQADKCPV-----TGATATTARGDSVEFDFFAAPQAYRRVAAEHARD-GAFHSSR	49
	***: *	:*****:** .***** : **:*
BotCYP	-RGFWVLTTYDGIVDAFKDEGTFTVGRVSAAEGAEERWIPLTVEGREHTAWRQRLGAWF	118
SalCYP	GDSFWVLSTYEGICAAFRDEDTFSVSRVSAADGAEDERWIPLTIQGRTHNEWRRRLAAWF	109
	.*****:**:* **:*.***.*.*****:*****:*****:*** * .**:*.***	
BotCYP	TPQVRRELTPSMRAGARRRIEGFLEKGEVSFNEDFARPYVLENLMTAVGWPPDGFLLIA	178
SalCYP	TPQARDLTPAIRANARRRISAFVDRGEVSFSDEFARPYVLENLMLAVGWPLADLDHLLA	169
	***.*:**:*.*.*****.*.***:*****.:*****:***** ***** .:* **:	
BotCYP	INRAMIDSRAPDPRAAAYGELGLPALERFAREHIARRRAEPAD-DLTASFGWEIDGAE	237
SalCYP	INVAMIRSREAPDPRQAFNAETAFFALQEYVRRHVARRRAEPVEGDLTSATFDWEIDGTP	229
	** ** * .***** * .* .:***:..*.*:*****.: ***:*.*.*****:	
BotCYP	VTDDDRASLLCTLFLAGIDSTVNHLANAVQHLAHHEEDRRRFLAGPEVRPPAVEEFLRAN	297
SalCYP	VSDADRESLLTVLFLAGVDSTVNHMANGIQHLAHHPGDRHRFLRDPEVRPAAVEEFLRVN	289
	*:* ** *** .*****:*****:*.:***** **:**** .***** *****.*	
BotCYP	SCMPGRQAATGGAGGVADRGDTVLLPLALANHDPEVFPEPGRIDFDRTRNPHIAFGTGP	357
SalCYP	SCMPGRLATREGAGGVASQGDVLLPLALANYDPAVFPEPERVDFDREQNPHIAFGTGH	349
	***** *: *****.:*****:*****:* ***** *:***** :*****	
BotCYP	HQCLGAAFARAQIILVALEEWHALVPDYAPSPEQRTAEPFPLRNSYDLRLVW	408
SalCYP	HQCLGAAYARAQILTAWEEWHELIPDYRLP--DPTVEPPFLRNVDLRIVW	398
	*****:*****.* ***** *.*** :*.***** *****:**	

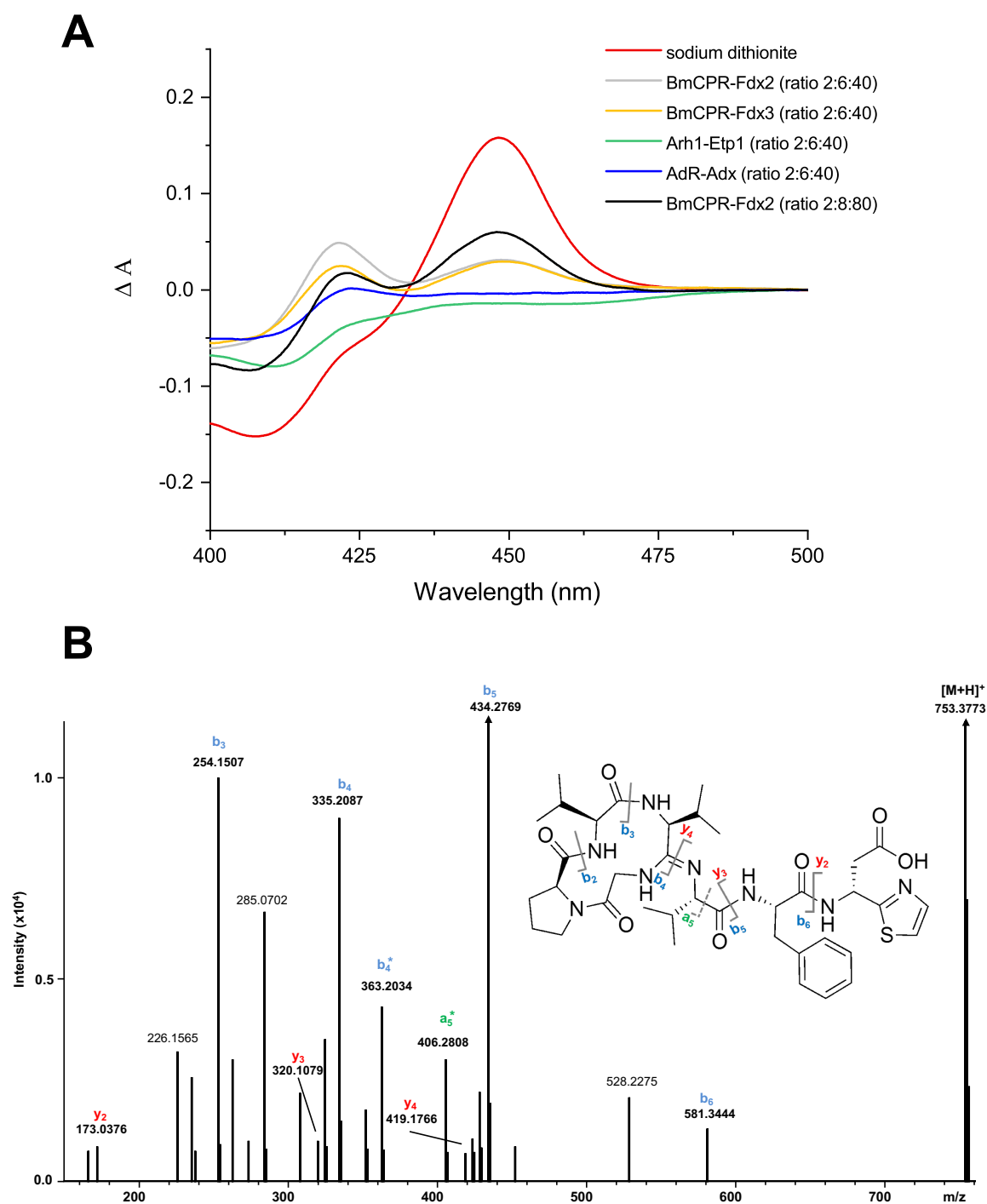
**Figure S1: A** Comparison of the bottromycin biosynthetic gene cluster in *Streptomyces* sp. BC16019 and the truncated gene cluster found in *Salinispora tropica*. As we observe an integrase upstream of *salRMT2*, most likely the first part of the cluster including the precursor peptide A is missing. **B** Pairwise sequence alignment of SalCYP (RefSeq WP\_051425703) with BotCYP (UniProt ID K4MJU2). The sequence identity was determined to be 68%. The alignment was generated with *Clustal Omega*.<sup>[1]</sup>



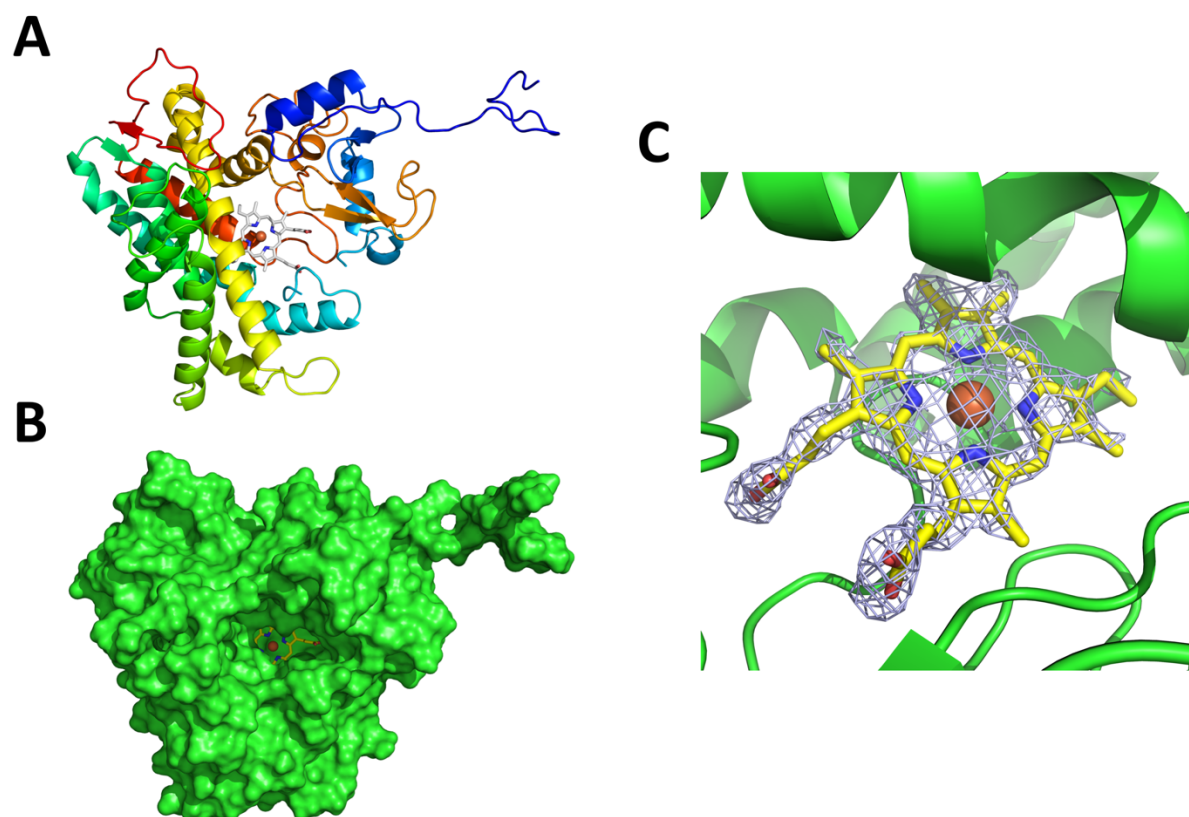
**Figure S2: A** The dithionite reduced CO-difference spectrum (red line) shows the typical peak maximum at 448 nm for the Fe<sup>II</sup>-CO complex. Searching for suitable redox partners, BotCYP reduction using autologous and heterologous electron transfer partners was determined. The dithionite reduced CO-difference spectrum (red line) of BotCYP was compared with the NADPH reduced BmCPR-Fdx2 (grey line), BmCPR-Fdx3 (yellow line), Arh1-Etp1<sup>fd</sup> (green line) and AdR-Adx<sub>4-108</sub> (blue line) CO-complexed spectrum. The NADPH (1 mM) reduced CO-difference spectra were recorded in a 200  $\mu$ l mixture of CYP/ferredoxin/reductase with a 2:40:6 molar ratio in KPP buffer (10 mM K<sub>2</sub>HPO<sub>4</sub> and KH<sub>2</sub>PO<sub>4</sub>). These results showed that the redox partners, BmCPR and Fdx2/3, were able to transfer, at least, the first electron to BotCYP. The BmCPR-Fdx2 redox system was selected for further investigations with BotCYP. **B** MS<sup>2</sup> spectrum of **3a**. a, b and y ions are indicated.



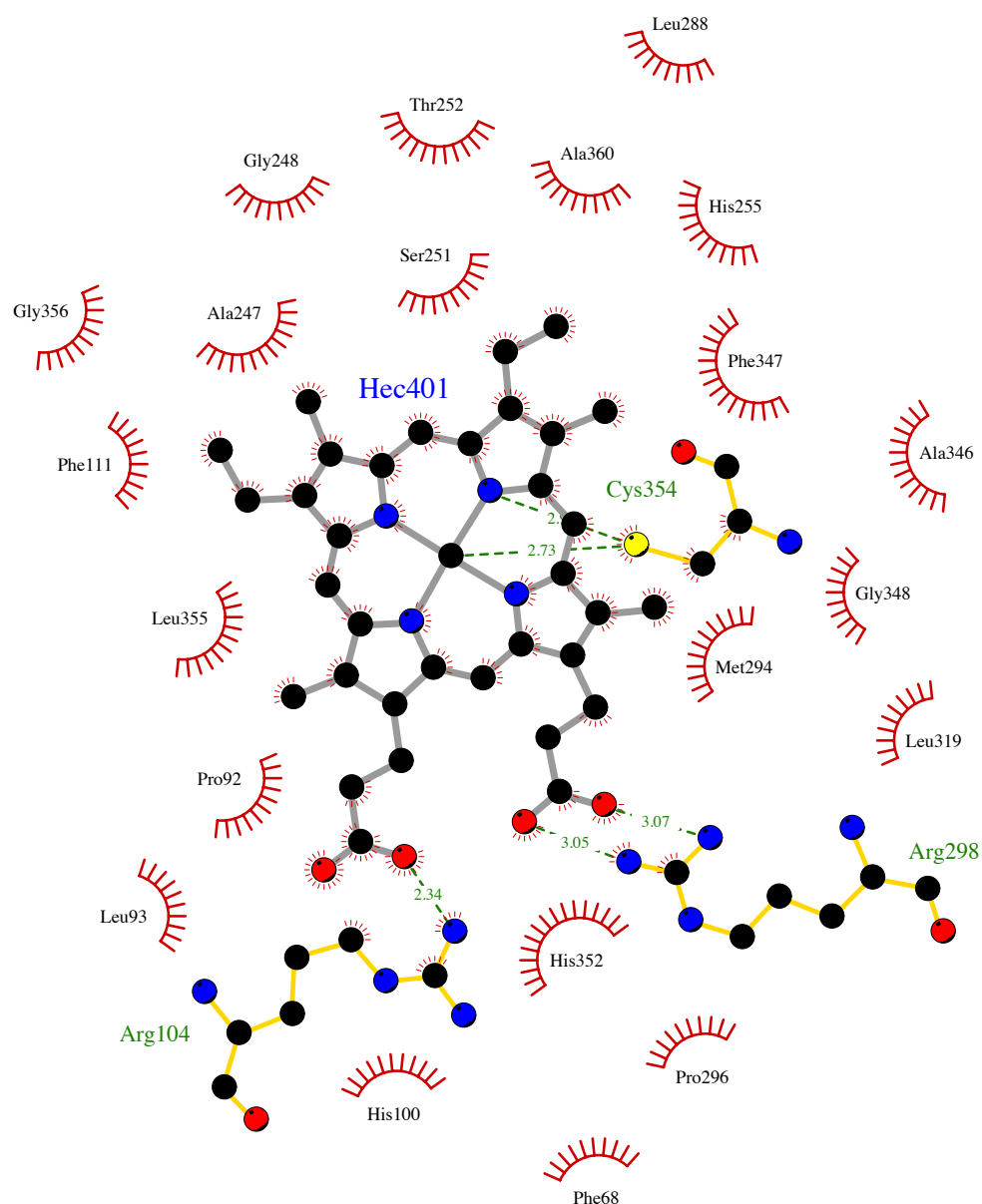
**Figure S3:** Microscale Thermophoresis (MST) measurement of the binding of SalCYP with a mixture of **2a/2b** at the equilibrium state. The dissociation constant ( $K_D$ ) was determined to be  $4.9 \pm 1.1 \mu\text{M}$ .



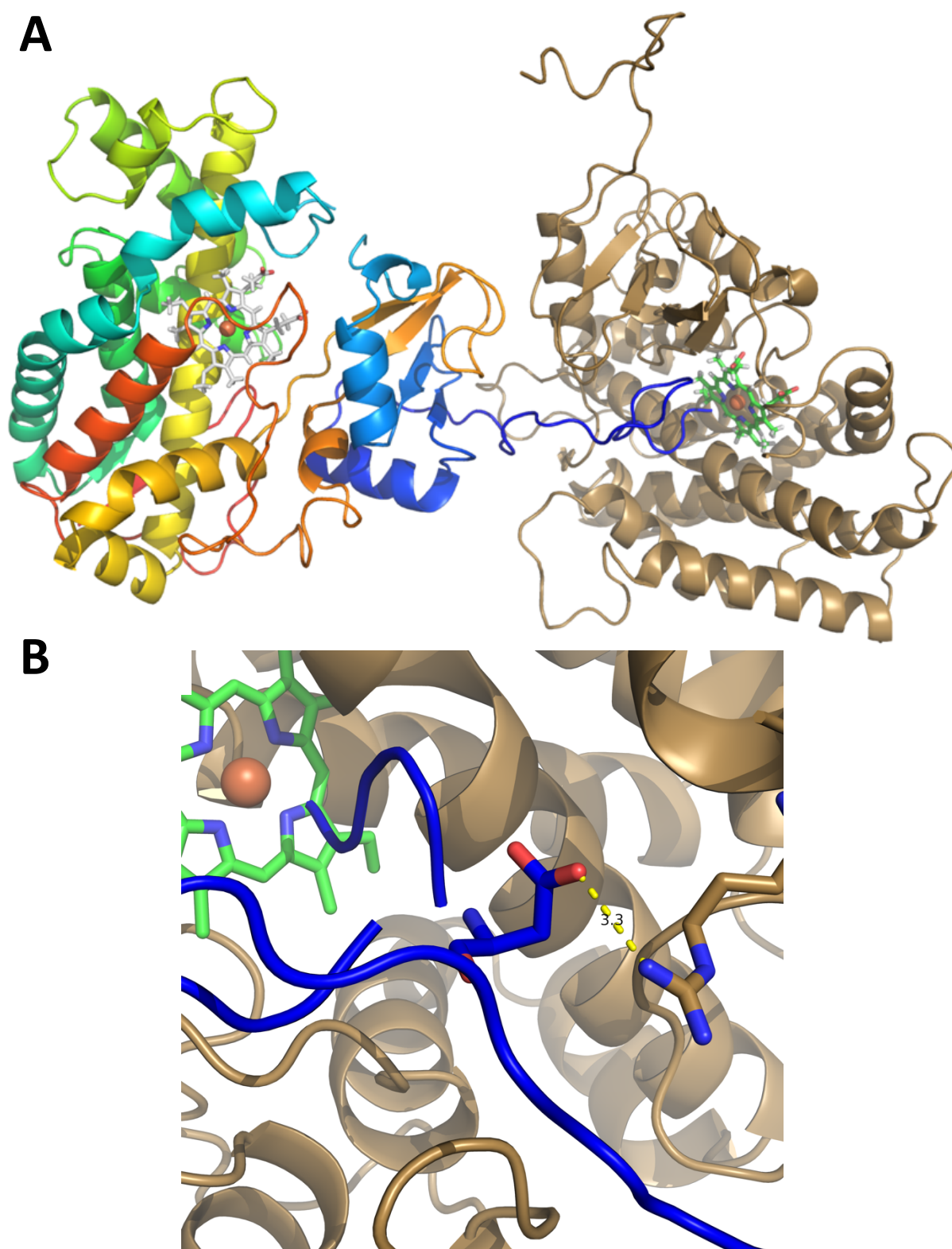
**Figure S4:** **A** The dithionite reduced CO-difference spectrum (red line) shows the typical peak maximum at 448 nm for the Fe<sup>II</sup>-CO complex. Searching for suitable redox partners, SalCYP reduction using autologous and heterologous electron transfer partners was determined. The dithionite reduced CO-difference spectrum (red line) of SalCYP was compared with the NADPH reduced BmCPR-Fdx2 (grey and black line), BmCPR-Fdx3 (yellow line), Arh1-Etp1<sup>fd</sup> (green line) and AdR-Adx<sub>4-108</sub> (blue line) CO-complexed spectrum. The NADPH (1 mM) reduced CO-difference spectra were recorded in a 200  $\mu$ l mixture of CYP/ferredoxin/reductase with a 2:40:6 (and 2:80:8 for BmCPR-Fdx2) molar ratio in KPP buffer (10 mM K<sub>2</sub>HPO<sub>4</sub> and KH<sub>2</sub>PO<sub>4</sub>). These results showed that the redox partners, BmCPR and Fdx2/3, were able to transfer, at least, the first electron to SalCYP. The BmCPR-Fdx2 redox system was selected for further investigations with SalCYP. **B** MS<sup>2</sup> spectrum of the SalCYP product. a, b and y ions are indicated.



**Figure S5:** **A** Rainbow cartoon representation of SalCYP, blue is representing the N- whereas red is representing the C-terminus. We observe the N-terminus extending away from the rest of the protein. **B** Surface representation of SalCYP, highlighting accessibility of the heme b in the active site for potential substrates. **C** Enlarged visualization of the active site heme b in combination with the respective electron difference map. The electron difference map ( $2F_o-F_c$ ) was contoured at  $2.6\sigma$  with a carve of 1.6. Heme molecules are shown as sticks, iron ions as orange spheres.

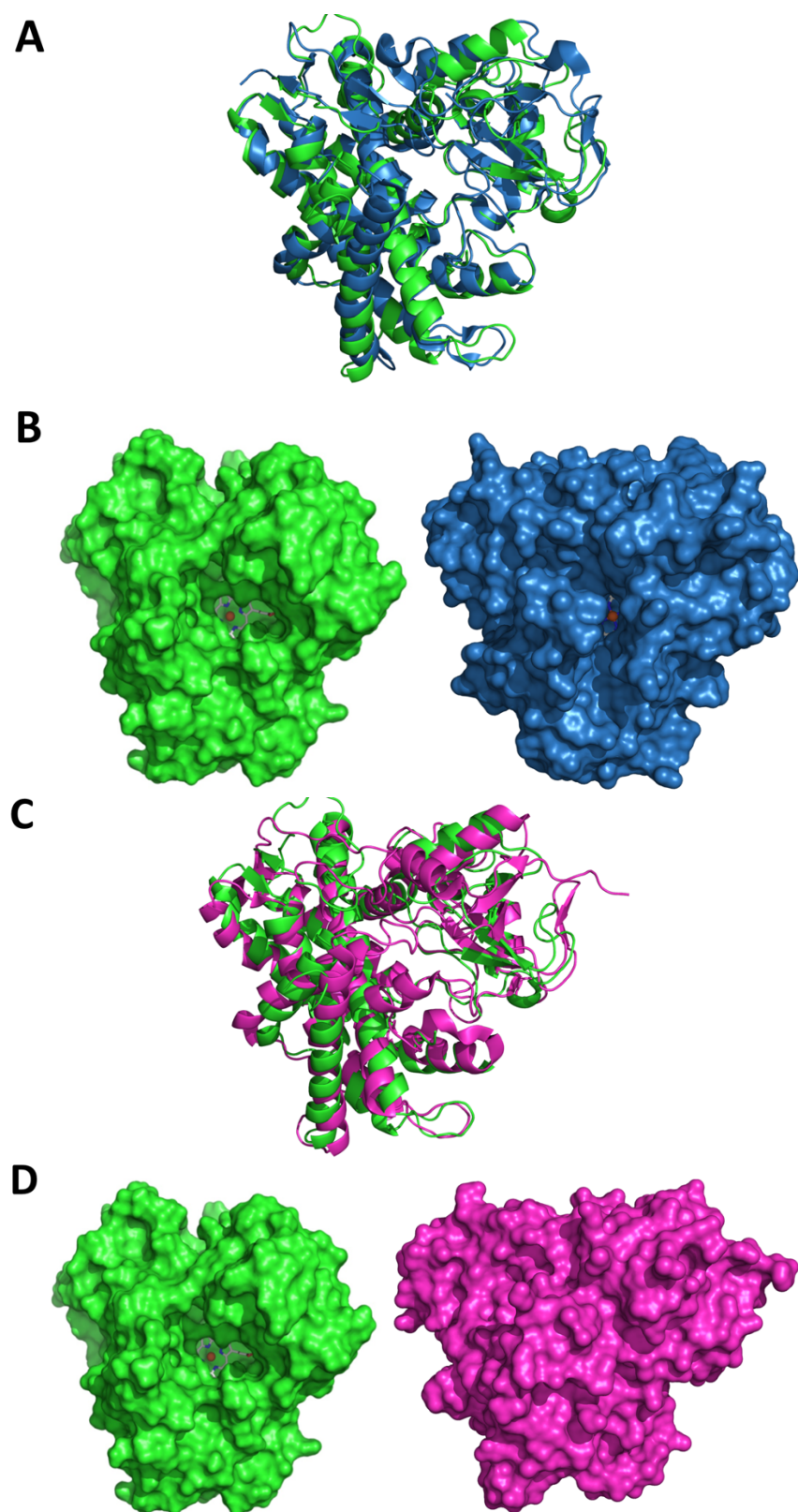


**Figure S6:** LigPlus diagram of the interactions between Heme b and SalCYP. Ligand bonds are grey, protein bonds yellow. Hydrogen bonds and salt bridges are indicated by green, dashed lines and distances given. Hydrophobic interactions are represented as red spoked arcs.

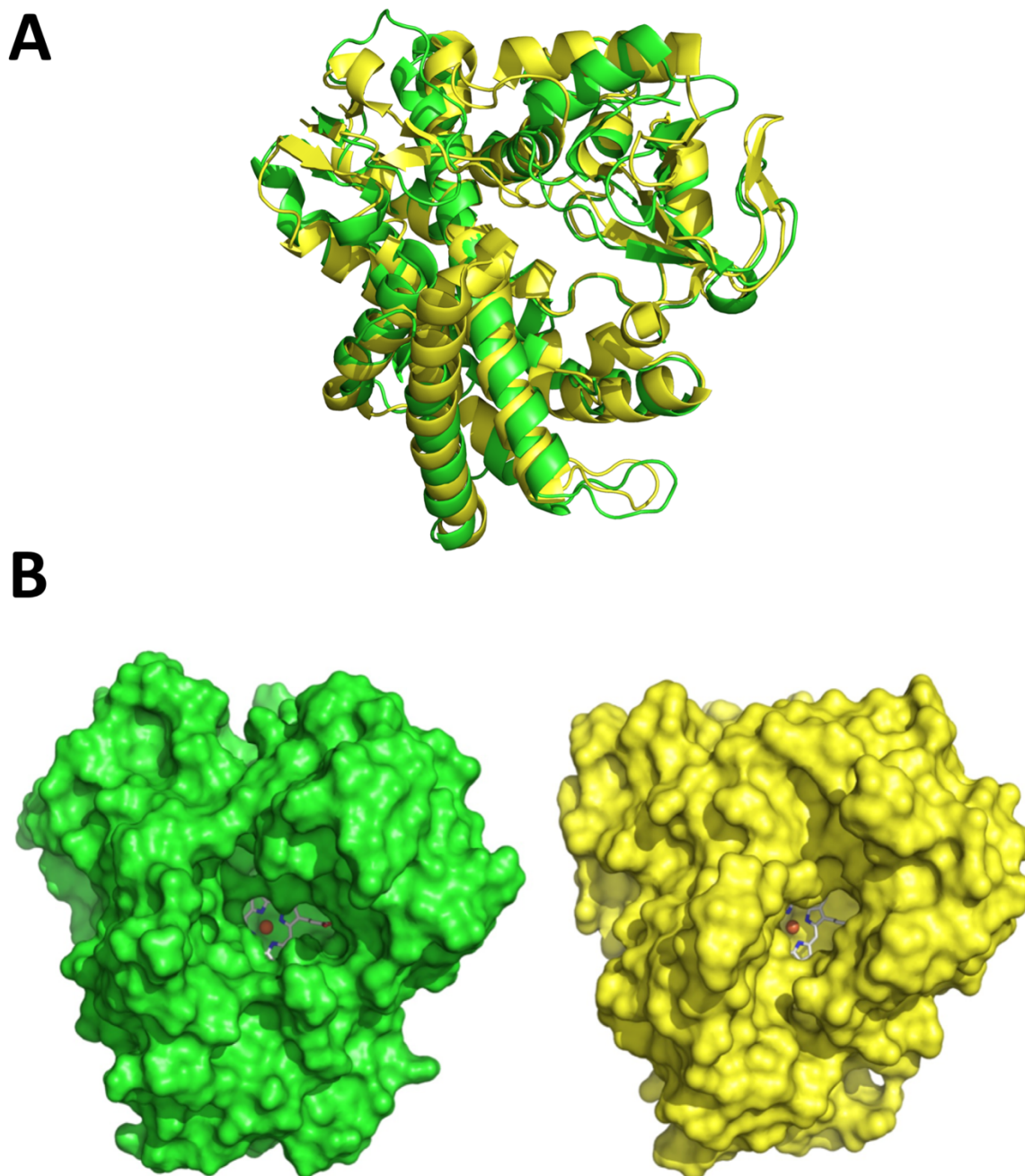


**Figure S7:** **A** Interaction of the SalCYP N-terminus with the heme iron of a symmetry mate. Both structures are shown as cartoon representations, the left using the rainbow color scheme (N-terminus blue), the right in brown. Heme molecules are shown as sticks, iron ions as orange spheres. **B** Formation of a salt bridge between an aspartate and an arginine residue of each monomer. It is a crucial interaction for the coordination of the N-terminus into the active site of the symmetry mate and possibly required for crystallization. Heme molecules and interacting residues are shown as sticks, iron ions as orange spheres. A distance measurement between the interacting side chains is shown.





**Figure S8:** Comparison of the active sites of the closest structural homologue CYP101D2 (blue, PDB ID 4dxy) and a well characterized homologue catalyzing an oxidative decarboxylation OleTJE (magenta, PDB ID 5m0n) with SalCYP (green). **A** Superposition of cartoon representations of SalCYP<sup>T</sup> and CYP101D2 (Structural homology: C<sub>α</sub> rmsd 1.5 Å over 275 atoms). **B** Surface comparison of SalCYP<sup>T</sup> and CYP101D2. **C** Superposition of cartoon representations of SalCYP<sup>T</sup> and OleTJE (Structural homology: C<sub>α</sub> rmsd 2.9 Å over 310 atoms). **D** Surface comparison of SalCYP<sup>T</sup> and OleTJE. Compared to SalCYP, the active site heme of OleTJE is inaccessible for larger substrates.



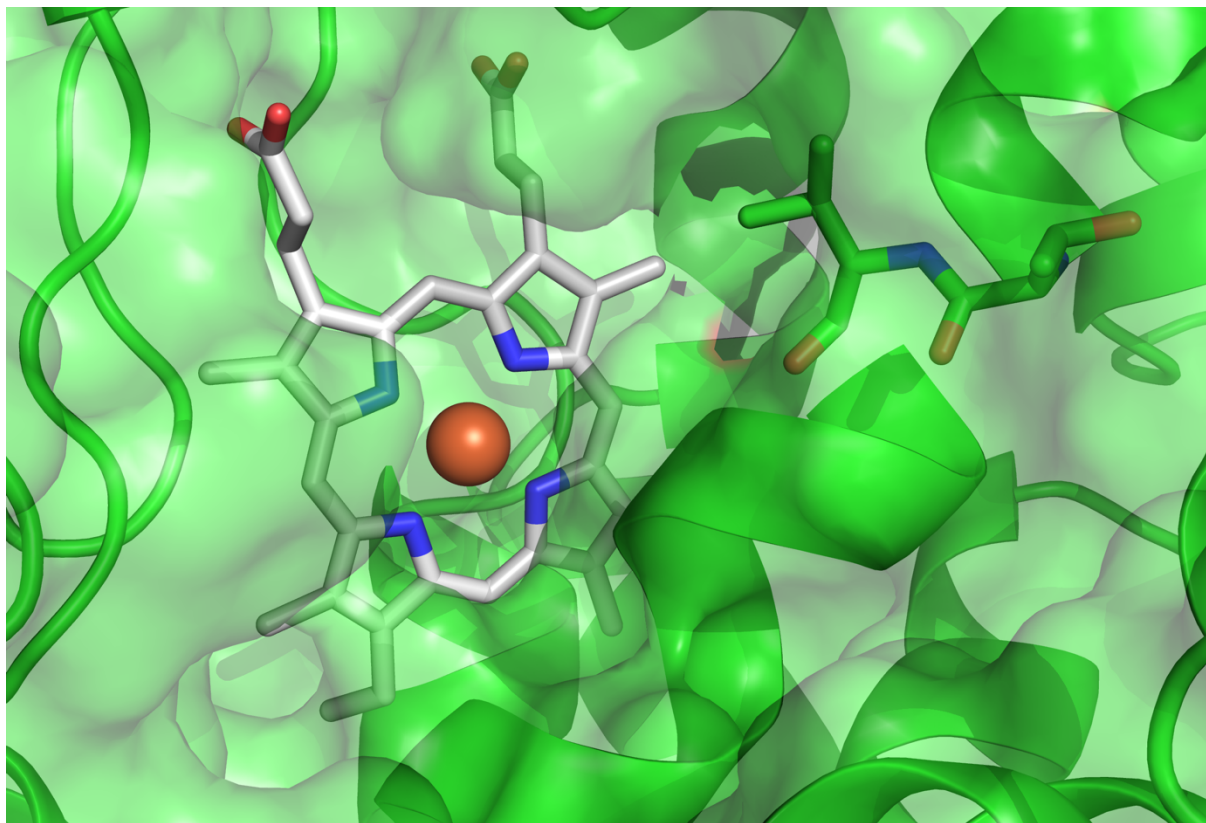
**Figure S9:** Comparison of the accessibility of the active sites of SalCYP (green) and thiomuracin hydroxylase TbtJ1 (PDB ID 5vws, yellow). **A** Superposition of the cartoon representations of both proteins. The structural homology was calculated as a  $C_{\alpha}$  rmsd of 1.6 Å over 285 atoms. **B** Surface representation of both proteins. Heme molecules are shown as white sticks, while iron ions are shown as orange spheres. The active site of TbtJ1 is also open, although less so than SalCYP. Both proteins accept much larger substrates than most small molecule-modifying Cytochrome P450 enzymes.

CLUSTAL O(1.2.4) multiple sequence alignment

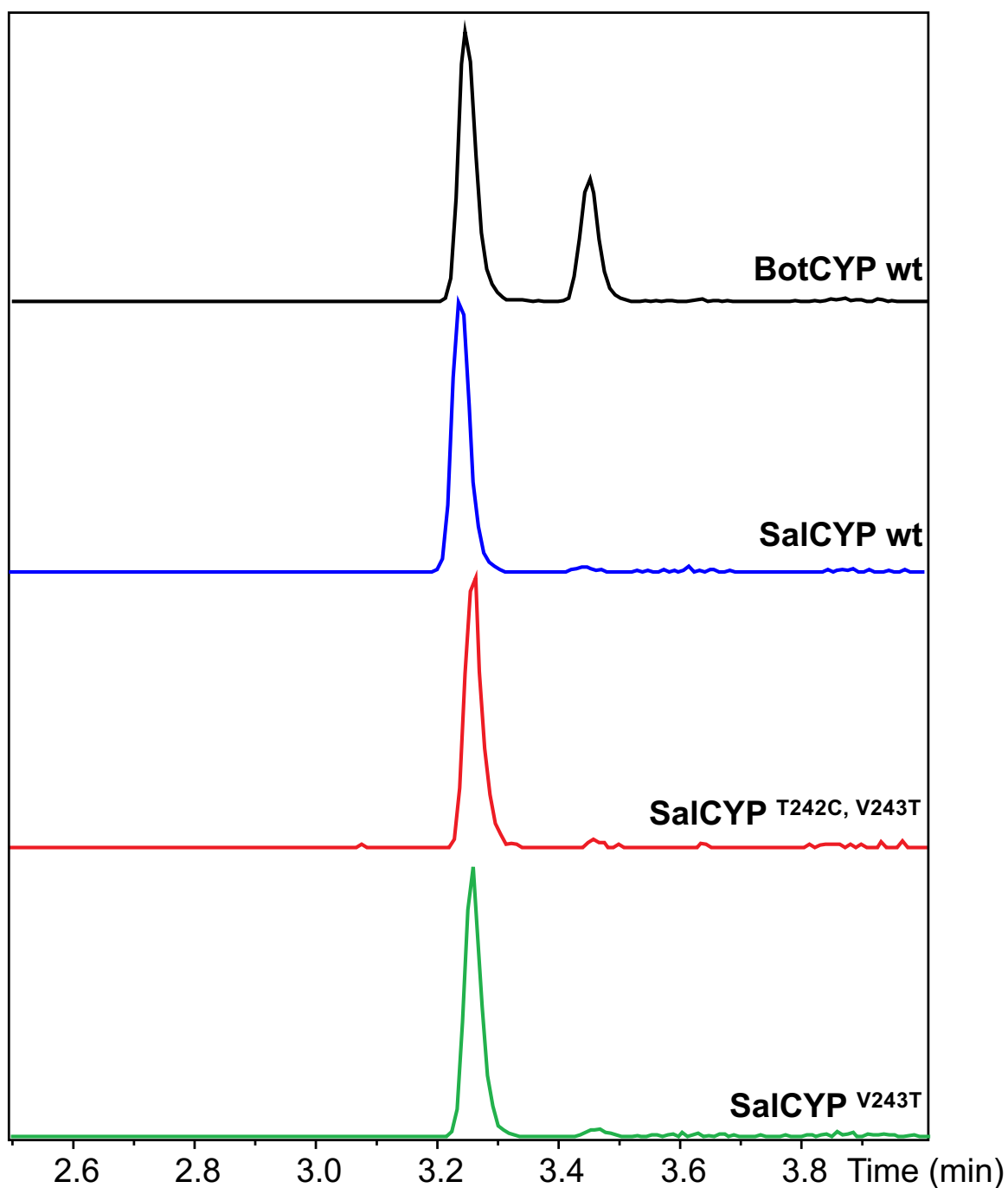
tropica	-----MQADKCPVTGATATTARGDSVEFDFFAAPQAYRR	34
orinoci	-----MDFDFFNAPAEYRR	14
WMMB272	-----MDFEFFNGPSAYRE	14
sulphureus	-----MDFEFFNGPSAYRE	14
purpureus	-----MDFDFFNAPAEYRR	14
ipomoeae	MQADRPQMADQGPDDAGGTGGTAARTEQAAQCPVAPDRAARGAECPVDFDFFAAPQQYRD	60
ipomoeae2	-----MDFDFFFAAPQQYRD	14
BotCYP	-----MQADREPDRGTGREAKRPA-----E--PDHPPQQATCPVDFDFFAAPQTYRD	44
bottropensis	-----MDFDFFFAAPQQYRD	14
bmbI	-----MQADREPDRGTGHEAKRPA-----GTGPAPEGRGAMCPVDFDFFAAPQQYRD	46
scabiei1	-----MDFDFFFAAPQQYRD	14
scabiei2	-----MDFDFFFAAPQQYRD	14
scabiei3	-----MDFDFFFAAPQQYRD	14
scabiei4	-----MQADRETDRPGPGAHEHPA-----GPGPAGPGARCVPVDFDFFAAPQQYRD	44
	:*: **: *	
tropica	VAAEHARD-GAFHSSRGDSFWVLSTYEGICAAFRDEDTFSVSRVSAADGAEDERWIPLTI	93
orinoci	AAAEHARD-GVFRSEHGGGFWVLTTYTGICRAFRNEADFTVGRVSAAGGADAERWIPLTV	73
WMMB272	AAATHADE-GAFYASGDGFWVLTTYEGICEAFRDEEQFTVGRVSAADGAEDRWIPLTI	73
sulphureus	AAATHADE-GAFYASGDGFWVLTTYEGICEAFRDEEQFTVGRVSAADGAEDRWIPLTI	73
purpureus	VAAAHARE-GAFYSGRGSFWVLTTYDGICEAFKDEDTFTVGRVSAAGGAEERWIPLTV	73
ipomoeae	AAAHAAGDHGAFYSD--RGFWVLTTFDGISAFAKDEGTFTVGRVSAAGGAEERWIPLTV	118
ipomoeae2	AAAHAAGDHGAFYSD--RGFWVLTTFDGISAFAKDEGTFTVGRVSAAGGAEERWIPLTV	72
BotCYP	AAAHAAGDHGAFYSD--RGFWVLTTYDGIVDAFKDEGTFTVGRVSAAGGAEERWIPLTV	102
bottropensis	AAAHAAGEQGAFHSD--RGFWVLTTYDGIVDAFRDEDTFTVGRVSAAGGAEERWIPLTV	72
bmbI	AAAHAAGEQGAFHSD--RGFWVLTTYDGIVDAFRDEDTFTVGRVSAAGGAEERWIPLTV	104
scabiei1	AAAHAAGEHGAFYSD--RGFWVLTTYDGIVDAFKDEGTFTVGRVSAAGGAEERWIPLTV	72
scabiei2	AAAHAAGEHGAFYSD--RGFWVLTTYDGIVDAFKDEGTFTVGRVSAAGGAEERWIPLTV	72
scabiei3	AAAHAAGEHGAFYSD--RGFWVLTTYDGIVDAFKDEGTFTVGRVSAAGGAEERWIPLTV	72
scabiei4	AAAHAAGEHGAFYSD--RGFWVLTTYDGIVDAFKDEGTFTVGRVSAAGGAEERWIPLTV	102
	.* * *: *. * .*****: *: * *: *: * *: *.*****: *: *:*****:	
tropica	QGRTHNEWRRRLAAWFTTPQRRDLTPAIRANARRRISAFVDRGEVSFSDEFARPYVLENL	153
orinoci	DGPEHTEWRRRLASWFTTPQVRREMPWIRQONARRRIAALRKDGASFSDEFARPYVLENL	133
WMMB272	EGRQHTEWRRRLGAWFTTPQVRRELTSPSIEANARRRIEGFVDDGGEVSFNEDFARPVLENL	133
sulphureus	EGRQHTEWRRRLGAWFTTPQVRRELTSPSIEANARRRIEGFVDDGGEVSFNEDFARPVLENL	133
purpureus	EGDQHTQWRRRLASWFTTPQVRRELTPEIRRNARRRIEAFRDKGEVSFNEDFARPYVLENL	133
ipomoeae	EGREHTAWRQRLGAWFTTPQVRRELTSGIRENARRRIEGFLEKGEVSFNEDFARPYVLENL	178
ipomoeae2	EGREHTAWRQRLGAWFTTPQVRRELTSGIRENARRRIEGFLEKGEVSFNEDFARPYVLENL	132
BotCYP	EGREHTAWRQRLGAWFTTPQVRRELTSPVRAGARRRIEGFLEKGEVSFNEDFARPVLENL	162
bottropensis	EGREHTAWRQRLGAWFTTPQVRRELTSPVRAGARRRIEGFLEKGEVSFNEDFARPYVLENL	132
bmbI	EGREHTAWRQRLGAWFTTPQVRRELTSPVRAGARRRIEGFLEKGEVSFNEDFARPYVLENL	164
scabiei1	EGREHTAWRQRLGAWFTTPQVRRELTSPVRAGARRRIEEFLDKGEVSFNEDFARPYVDNL	132
scabiei2	EGREHTAWRQRLGAWFTTPQVRRELTSPVRAGARRRIEEFLDKGEVSFNEDFARPYVDNL	132
scabiei3	EGREHTAWRQRLGAWFTTPQVRRELTSPVRAGARRRIEEFLDKGEVSFNEDFARPYVLENL	132
scabiei4	EGREHTAWRQRLGAWFTTPQVRRELTSPVRAGARRRIEEFLDKGEVSFNEDFARPYVLENL	162
	: * *. **: *: *****: *: * *: . ***** : . **.*.: :*****: *: *	
tropica	MLAVGWPLADLDHLLAINVAMIRSREAPDPRQAFNAETAFFPALQEYVRRHVARRRAEPVE	213
orinoci	MTAVGWPLSGDLRLLAIDRAMIAARAEPDPREAFLAEPGLPALEYAREQIARRRDKPA-	192
WMMB272	MTAVGWEFDRFDQLLRIDRAMIDSRYAPDPREAAFEVEVGIPALEKLAREQVERRRAEPA-	192
sulphureus	MTAVGWEFDRFDQLLRIDRAMIDSRYAPDPREAAFEVEVGIPALEKLAREQVERRRAEPA-	192
purpureus	MTAVGWPLDGLGLLLAIDRAMIDSRSAPDPREAAYGELGLPALEKFARRHIELRRRAEPA-	192
ipomoeae	MTAVGWPPDGFDDLAINRAMIDSRSAPDPRAAAYGELGLPALERFARELIARRRAEPA-	237
ipomoeae2	MTAVGWPPDGFDDLAINRAMIDSRSAPDPRAAAYGELGLPALERFARELIARRRAEPA-	191
BotCYP	MTAVGWPPDGFDDLIAINRAMIDSRSAPDPRAAAYGELGLPALERFARELIARRRAEPA-	221
bottropensis	MTAVGWPGDGFDDLAIIDRAMIDSRSAPDPRAAAYGELGLPALERFAREHIARRRAEPA-	191
bmbI	MTAVGWPGDGFDDLAIIDRAMIDSRSAPDPRAAAYGELGLPALERFAREHIARRRAEPA-	223
scabiei1	MTAVGWPPDAFDLLAIIDRAMIDSRSAPDPRAAAYGELGLPALERFAREHVARRRAEPA-	191
scabiei2	MTAVGWPPDAFDLLAIIDRAMIDSRSAPDPRAAAYGELGLPALERFAREHVARRRAEPA-	191
scabiei3	MTAVGWPRDAFDLLAIIDRAMIDSRSAPDPRAAAYGELGLPALERFAREHVARRRAEPA-	191
scabiei4	MTAVGWPRDAFDLLAIIDRAMIDSRSAPDPRAAAYGELGLPALERFAREHVARRRAEPA-	221
	* **** :. *: *: * * * * * * *: * *: . *: *: * * *	
tropica	GDLTASFDFWEIDGTPVSDADRESLLTVLFLAGVDSTVNHMANGIQHLAHPGDRHRFLR	273
orinoci	QDLVTATFDWKVGRPVTDERTSLLCVLFLAGIDSTVNHLANAVQHLACHPEDRRRFS	252
WMMB272	DDLTTATFGWEVEGRPVTDERRASLLCVLFLAGVDSTVNHLANAVQHLVCHDAKRRFLA	252
sulphureus	DDLTTATFGWEVEGRPVTDERRASLLCVLFLAGVDSTVNHLANAVQHLVCHDAKRRFLA	252
purpureus	DDLTTASFDFWEIDGTPVSDDDRASLLCVLFLAGIDSTVNHMNAIQHLARDEEDRARFLA	252
ipomoeae	DDLATASFGEIDGAPVGDDDRASLLCTLFLAGIDSTVNHMNAVQHLAHPEDRGRFMS	291
ipomoeae2	DDLATASFGEIDGAPVGDDDRASLLCTLFLAGIDSTVNHMNAVQHLAHPEDRGRFMS	257
BotCYP	DDLTTASFGEIDGAETVDDDRASLLCTLFLAGIDSTVNHLANAVQHLAHPEDRGRFMS	281

bottropensis	DDLTTASFGWEIDGVPVTD DDDRASLLCTLFLAGVDSTVNHLANAVQH LAHHKEDRRRFLA	251
bmbI	DDLTTASFGWEIDGVPVTD DDDRASLLCTLFLAGVDSTVNHLANAVQH LAHHKEDRRRFLA	283
scabiei1	DDLTTASFGWEIDGTPVS DDDRASLLCTLFLAGVDSTVNHLANAVQH LAHHEEDRRRFLA	251
scabiei2	DDLTTASFGWEIDGTPVS DDDRASLLCTLFLAGVDSTVNHLANAVQH LAHHEEDRRRFLA	251
scabiei3	DDLTTASFGWEIDGTPVS DDDRASLLCTLFLAGVDSTVNHLANAVQH LAHHEEDRRRFLA	251
scabiei4	DDLTTASFGWEIDGTPVS DDDRASLLCTLFLAGVDSTVNHLANAVQH LAHHEEDRRRFLA **.:*.:*.: * * * * * .*****:*****:*.:.***. . *: **	281
tropica	DPEVRPAAVEEFLRVNSCMYPGRLATREGAGGVASQGD TVLLPLALANYDPAVFPEPERV	333
orinoci	DEAVRPAAVEEFLRAFSCMYPGRMTAREGACFVTGRGETVLLPLALANHDPDVFDPGRV	312
WMMB272	SREVRPRAVEEFLRVNSCMYPGRLAARDGAGGTARRGETVLLPLAVANYDPEVFHDPEQV	312
sulphureus	SREVRPRAVEEFLRVNSCMYPGRLAARDGAGGTARRGETVLLPLAVANYDPEVFHDPEQV	312
purpureus	SREIRPAAVEEFLRTNSCMYPGRMAVHDGAGGHAGRGQTVLLPLALANHDPPEVFPEPERV	312
ipomoeae	SPEVRPPAVEEFLRVNSCMYPGRMAVTADAGGVADRGDTVLLPLALANHDPPEVFPEPARI	357
ipomoeae2	SPEVRPPAVEEFLRVNSCMYPGRMAVTADAGGVADRGDTVLLPLALANHDPPEVFPEPARI	311
BotCYP	GPEVRPPAVEEFLRVNSCMYPGRQAATGGAGGVADRGDTVLLPLALANHDPPEVFPEPGRI	341
bottropensis	APEARPPAVEEFLRVNSCMYPGRQAATGGAGGVADRGDTVLLPLALANHDPPEVFPEPARV	311
bmbI	APEARPPAVEEFLRVNSCMYPGRQAATGGAGGVADRGDTVLLPLALANHDPPEVFPEPARV	343
scabiei1	APEARPPAVEEFLRVNSCMYPGRQAATGGAGGVADRGDTVLLPLALANHDPPEVFPEPGRI	311
scabiei2	APEARPPAVEEFLRVNSCMYPGRQAATGGAGGVADRGDTVLLPLALANHDPPEVFPEPGRI	311
scabiei3	APEARPSAVEEFLRVNSCMYPGRQAATGGAGGVADRGDTVLLPLALANHDPPEVFPEPGRI	311
scabiei4	APEARPSAVEEFLRVNSCMYPGRQAATGGAGGVADRGDTVLLPLALANHDPPEVFPEPGRI ** *****. ***** :. . * : :*:*****:***:*** ** :* ::	341
tropica	DFDREQNPHIAFGTGHHQCLGAAYARAQILTAWEEWHELIPDYRLPDP--TVEPPFLRVN	391
orinoci	DFEREHNPHIAFGTGPHQCLGAALARAQILTALEEWHALIPRYSLPPEQDSTAPPFLRNS	372
WMMB272	DFDRERNPHIAFGTGAHQCLGAALARAQIRALTADVWHELIPHYGFPPQPEPSEPRFLRNA	372
sulphureus	DFDRERNPHIAFGTGAHQCLGAALARAQIRALTADVWHELIPHYGFPPQPEPSEPRFLRNA	372
purpureus	DFDRERNPHIAFGTGPHQCLGAAFARAQILTALEEWHDLLPGYGVPEQAGSPAPFLRND	372
ipomoeae	DFDRERNPHIAFGTGPHQCLGAAFARAQILVALEEWHALVPDYVPPEQRTAEPFLRNS	417
ipomoeae2	DFDRERNPHIAFGTGPHQCLGAAFARAQILVALEEWHALVPDYVPPEQRTAEPFLRNS	371
BotCYP	DFDRTRNPHIAFGTGPHQCLGAAFARAQILVALEEWHALVPDYAPSPPEQRTAEPFLRNS	401
bottropensis	DFDRPRNPHIAFGTGPHQCLGAAFARAQILVALEEWHTLVPDYGPPPEQRGTEPAFLRNS	371
bmbI	DFDRPRNPHIAFGTGPHQCLGAAFARAQILVALEEWHTLVPDYGPPPEQRGTEPAFLRNS	403
scabiei1	DFGRPRNPHIAFGTGPHQCLGAAFARAQILVALEEWHALVPDYAPHPGQRATEPRFLRNS	371
scabiei2	DFGRPRNPHIAFGTGPHQCLGAAFARAQILVALEEWHALVPDYAPHPGQRATEPRFLRNS	371
scabiei3	DFGRPRNPHIAFGTGPHQCLGAAFARAQILVALEEWHALVPDYAPHPGQRADEPRFLRNS	371
scabiei4	DFGRPRNPHIAFGTGPHQCLGAAFARAQILVALEEWHALVPDYAPHPGQRADEPRFLRNS ** * :***** ***** ***** . * : * * * : * * ****	401
tropica	YDLRIWV	398
orinoci	YDLRLIW	379
WMMB272	YDLRLVW	379
sulphureus	YDLRLVW	379
purpureus	YDLRLAW	379
ipomoeae	YDLRLIW	424
ipomoeae2	YDLRLIW	378
BotCYP	YDLRLVW	408
bottropensis	YDLRLVW	378
bmbI	YDLRLVW	410
scabiei1	YDLRLVW	378
scabiei2	YDLRLVW	378
scabiei3	YDLRLVW	378
scabiei4	YDLRLVW ****.: *	408

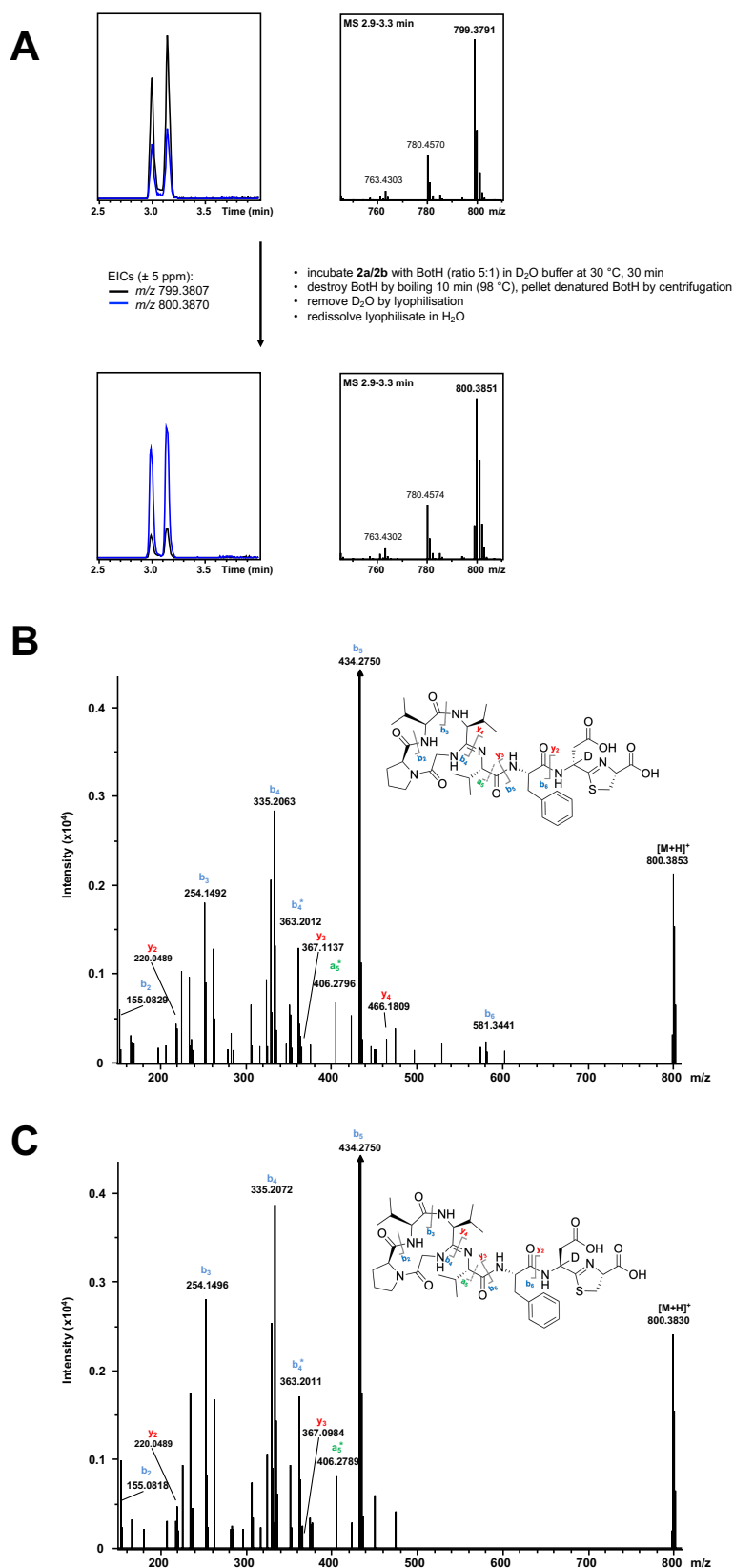


**B**

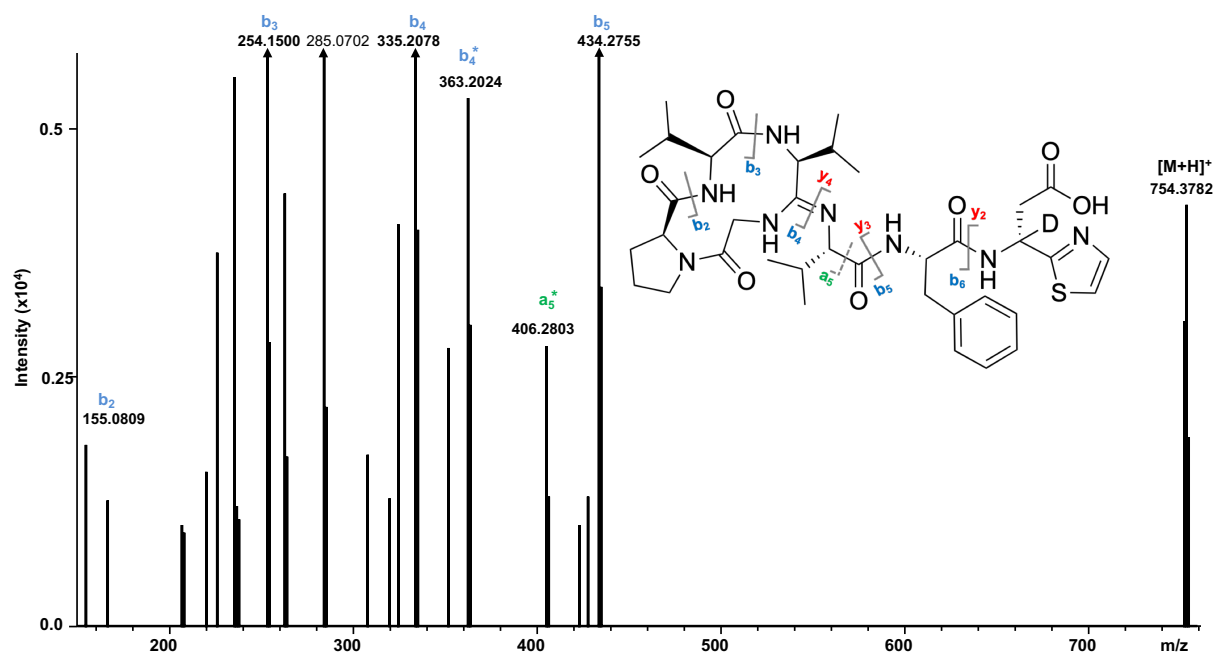
**Figure S10:** **A** Sequence alignment of every SalCYP homologue with a pairwise identity above 65% (every sequence found within the bottromycin-associated cluster in the Sequence Similarity Network (**Figure 3A**)). A yellow box is highlighting the only residues that are mutated in the proximity of the active site of SalCYP compared to BotCYP. The sequence alignment was generated with *Clustal Omega*<sup>[1]</sup> **B** Enlarged view of the SalCYP active site showing the mutated amino acids in SalCYP compared to other bottromycin producers as sticks. The side chain of the valine residue is pointing towards the solvent channel close to the active site heme.



**Figure S11:** EICs of **3a/b** ( $m/z$  753.3752  $\pm$  5 ppm) generated by BotCYP wt (black), SalCYP wt (blue), and the two generated SalCYP point/double mutants SalCYP<sup>T242C, V243T</sup> (red) and SalCYP<sup>V243T</sup> (green). The appearance of only one peak in all SalCYP proteins indicates that the introduced point mutations do not have an effect on SalCYPs stereospecificity.



**Figure S12: A** Production of substrates **4a/4b**, which contain a deuterium at the  $\alpha$ -Carbon of the aspartate residue. Substrates **4a/4b** show the same retention times that were observed for **2a/2b** and a mass shift of +1 Da. **B** MS<sup>2</sup> spectrum of **4b**. a, b and y ions are indicated. **C** MS<sup>2</sup> spectrum of **4a**. a, b and y ions are indicated.



**Figure S13:** MS<sup>2</sup> spectrum of **5a**. a, b and y ions are indicated. The observed mass shift of +1 Da during the production of **4a/4b** is retained after the oxidative decarboxylation has occurred.

**Table S1:** MS<sup>2</sup> fragmentation data for **3a** from BotCYP reaction.

Seq.	b <sub>n</sub>	Obs. b	Calc. b	y <sub>n</sub>	Obs. y	Calc. y
G <sup>a</sup>	1			8	753.3743	753.3752
P	2	-	155.0821	7	-	696.3543
V	3	254.1507	254.1605	6	-	599.3016
V <sup>a</sup>	4	335.2192	335.2083	5	-	500.2332
V	5	434.2765	434.2767	4	419.1606	419.1753
F	6	581.3420	581.3451	3	320.1040	320.1069
D <sup>b</sup>	7			2	173.0392	173.0385
C <sup>b</sup>	8			1		

a. Gly1 macrocyclisation with Val4 carbonyl

b. Cys8 heterocyclisation with Asp7 carbonyl and decarboxylation

Additionally, fragment a<sub>5</sub> (obs. 406.2808; calc. 406.2818) and b<sub>4</sub><sup>\*</sup> (obs. 363.2005; calc. 363.2032), which is characteristically associated with the bottromycin macrocycle and has been previously observed for bottromycins and macrocyclic peptides, was detected.



**Table S2: MS<sup>2</sup> fragmentation data for 3b from BotCYP reaction, corresponding to Figure 1C.**

Seq.	b <sub>n</sub>	Obs. b	Calc. b	y <sub>n</sub>	Obs. y	Calc. y
G <sup>a</sup>	1			8	753.3746	753.3752
P	2	155.0811	155.0821	7	-	696.3543
V	3	254.1617	254.1605	6	-	599.3016
V <sup>a</sup>	4	335.2079	335.2083	5	-	500.2332
V	5	434.2756	434.2767	4	419.1755	419.1753
F	6	581.3446	581.3451	3	320.1062	320.1069
D <sup>b</sup>	7			2	173.0386	173.0385
C <sup>b</sup>	8			1		

a. Gly1 macrocyclisation with Val4 carbonyl

b. Cys8 heterocyclisation with Asp7 carbonyl and decarboxylation

Additionally, fragments a<sub>5</sub> (obs. 406.2804; 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 previously observed for bottromycins and macrocyclic peptides, were detected.

**Table S3: Data collection and refinement statistics**

	SalCYP	SalCYP <sup>T</sup>
<b>Data collection</b>		
Space group	P 21	P 41
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	50.44 97.40 93.28	96.53 96.53 49.99
α, β, γ (°)	90.00 104.57 90.00	90.00 90.00 90.00
Resolution (Å)	48.82 - 1.85	43.17 - 1.5
	(1.92 - 1.85)	(1.55 - 1.5)
<i>R</i> <sub>merge</sub>	0.158 (0.699)	0.072 (0.889)
<i>I</i> / σ <i>I</i>	17.9 (4.7)	15.1 (2.1)
Completeness (%)	99.61 (96.61)	99.89 (99.90)
Redundancy	28.7 (23.2)	9.8 (9.9)
<b>Refinement</b>		
Resolution (Å)	48.82 - 1.85	48.26 - 1.5
No. reflections	74120 (7163)	73830 (7347)
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.1500/0.1900	0.1730/0.1862
No. atoms	7403	3242
Protein	6214	2894
Ligand/ion	98	49
Water	1091	299
<i>B</i> -factors	18.93	29.23
Protein	17.27	28.36
Ligand/ion	11.20	22.13
Water	29.07	38.76
R.m.s. deviations		
Bond lengths (Å)	0.010	0.019
Bond angles (°)	1.00	1.43

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

**Table S4: MS<sup>2</sup> fragmentation data for 4a.**

Seq.	b <sub>n</sub>	Obs. b	Calc. b	y <sub>n</sub>	Obs. y	Calc. y
G <sup>a</sup>	1	-	-	8	800.3830	800.3870
P	2	155.0818	155.0821	7	-	-
V	3	254.1496	254.1605	6	-	646.3133
V <sup>a</sup>	4	335.2072	335.2083	5	-	547.2449
V	5	434.2750	434.2767	4	-	466.1871
F	6	-	581.3451	3	367.0984	367.1186
D <sup>b,c</sup>	7	-	-	2	220.0489	220.0502
C <sup>b</sup>	8	-	782.3770	1	-	-

- a. Gly1 macrocyclisation with Val4 carbonyl
- b. Cys8 heterocyclisation with Asp7 carbonyl
- c. Asp7 C<sub>α</sub>-positon labeled with deuterium

Additionally, fragment a<sub>5</sub> (obs. 406.2789; calc. 406.2818) and b<sub>4</sub><sup>\*</sup> (obs. 363.2011; calc. 363.2032), which is characteristically associated with the bottromycin macrocycle and has been previously observed for bottromycins and macrocyclic peptides, was detected.

**Table S5: MS<sup>2</sup> fragmentation data for 4b.**

Seq.	b <sub>n</sub>	Obs. b	Calc. b	y <sub>n</sub>	Obs. y	Calc. y
G <sup>a</sup>	1	-	-	8	800.3853	800.3870
P	2	155.0829	155.0821	7	-	-
V	3	254.1492	254.1605	6	-	646.3133
V <sup>a</sup>	4	335.2063	335.2083	5	-	547.2449
V	5	434.2750	434.2767	4	466.1809	466.1871
F	6	581.3341	581.3451	3	367.1137	367.1186
D <sup>b,c</sup>	7	-	-	2	220.0489	220.0502
C <sup>b</sup>	8	-	782.3770	1	-	-

- a. Gly1 macrocyclisation with Val4 carbonyl
- b. Cys8 heterocyclisation with Asp7 carbonyl
- c. Asp7 C<sub>α</sub>-positon labeled with deuterium

Additionally, fragment a<sub>5</sub> (obs. 406.2796; calc. 406.2818) and b<sub>4</sub><sup>\*</sup> (obs. 363.2012; calc. 363.2032), which is characteristically associated with the bottromycin macrocycle and has been previously observed for bottromycins and macrocyclic peptides, was detected.

**Table S6: MS<sup>2</sup> fragmentation data for 5a.**

Seq.	b <sub>n</sub>	Obs. b	Calc. b	y <sub>n</sub>	Obs. y	Calc. y
G <sup>a</sup>	1	-	-	8	-	753.3752
P	2	155.0809	155.0821	7	-	697.3606
V	3	254.1500	254.1605	6	-	600.3078
V <sup>a</sup>	4	335.2078	335.2083	5	-	501.2394
V	5	434.2755	434.2767	4	-	420.1816
F	6	-	581.3451	3	-	321.1132
D <sup>b,c</sup>	7	-	-	2	-	174.0448
C <sup>b</sup>	8	-	-	1	-	-

- a. Gly1 macrocyclisation with Val4 carbonyl

- b. Cys8 heterocyclisation with Asp7 carbonyl and decarboxylation
- c. Asp7 C $\alpha$ -position labeled with deuterium

Additionally, fragment a<sub>5</sub> (obs. 406.2803; calc. 406.2818) and b<sub>4</sub>\* (obs. 363.2024; calc. 363.2032), which is characteristically associated with the bottromycin macrocycle and has been previously observed for bottromycins and macrocyclic peptides, was detected.

## Methods

### **Cloning of the BotCYP, SalCYP, SalCYP<sup>T242C, V243T</sup>, SalCYP<sup>V243T</sup> and SalCYP<sup>T</sup> expression constructs**

The codon-optimized DNA of SalCYP for *E. coli* expression with the artificially added TEV cleavage site “ENLYFQGG” added in front of the gene was obtained from Eurofins Genomics. The gene was cut out of the vector pEX-K4 by a double restriction with *NdeI* and *HindIII* and ligated into a linearized pET28b+-vector treated with the same restriction enzymes beforehand using T4 DNA Ligase. The Thr14Asp mutation for SalCYP<sup>T</sup> was introduced by PCR using a primer carrying the same N-terminal mutation within its DNA sequence. The BotCYP construct was amplified by PCR using a cosmid containing the bottromycin biosynthetic gene cluster from *Streptomyces* sp. BC16019. The amplification products of both BotCYP and SalCYP<sup>T</sup> were excised from the gel, treated with *NdeI* and *HindIII* and cloned into a pET28b+-vector in the same way as the full-length construct. The SalCYP<sup>T242C, V243T</sup>, SalCYP<sup>V243T</sup> mutants were generated by SDM-PCR using complementary mutagenic primers and reagents from the QuikChange Kit (Agilent). Template DNA was destroyed with DnaI before transformation into *E. coli* HS996.

Subsequently, the resulting clones were confirmed to have the fragment by DNA restriction analysis as well as sequencing. Correct plasmids were transformed into *E. coli* Lemo21(DE3) cells using a standard heat shock procedure and plated onto a Luria-Bertani(LB) agar plate using a double selection of Kanamycin and Chloramphenicol at 50 and 34 µg/mL, respectively.

### **Expression of BotCYP, SalCYP, SalCYP<sup>T242C, V243T</sup>, SalCYP<sup>V243T</sup> and SalCYP<sup>T</sup>**

Single colonies of *E. coli* Lemo21(DE3) cells were used to inoculate starter cultures of Luria-Bertani(LB) medium with Kanamycin and Chloramphenicol at 50 and 34 µg/mL, respectively. Large scale expression of all constructs was carried out in modified Terrific Broth (TB) medium with the appropriate antibiotic concentrations added. Modified Terrific Broth was obtained by adding 1 mM MgSO<sub>4</sub> and NPS buffer<sup>[2]</sup> to a final concentration of 1x after autoclaving instead of the usual potassium phosphate buffer system. The cultures were inoculated 1:100 with starter culture and grown at 37 °C and 180 rpm until and OD<sub>600</sub> of 0,8 was reached at which point the temperature was reduced to 25 °C and final concentrations of 0.5 mM 5-aminolevulinic acid as well as 0.25 mM FeCl<sub>3</sub> were added. After growing the cultures for another 30 min at 25 °C, a final concentration of 0.25 mM IPTG was added and the flasks were incubated for another 40 h. After 40 h, the cells were pelleted by centrifugation for 10 min at 5000 rpm and 4 °C and cell pellets were stored at -80 °C until further use.

### **Purification of BotCYP, SalCYP, SalCYP<sup>T242C, V243T</sup>, SalCYP<sup>V243T</sup> and SalCYP<sup>T</sup>**

The protein purification of all constructs were carried out in the same lysis buffer (500 mM NaCl, 20 mM Bis-Tris pH 6.8, 20 mM Imidazole pH 8.0, 3 mM β-mercaptoethanol). For every 25 g of wet

cell pellet, 100 mL of lysis buffer was added. The lysis buffer was supplemented with cComplete EDTA-free protease inhibitor tablets and DNase (0.4 mg/g wet cell pellet). The cell suspension was lysed via passage through a cell disruptor (30 kPSI) and the cell debris was removed by centrifugation (19000 rpm, 4 °C, 15 min). The supernatant was filtered through a 0.45 µm filter and applied to a His-Trap HP 5 mL column pre-equilibrated in lysis buffer at a flow rate of 5 mL/min. After application of the lysate and an extensive column wash (30 CV lysis buffer), 3 mg of purified His-MBP-TEV-protease was dissolved in 5 mL of lysis buffer and loaded onto the column. The TEV cleavage was carried out on the column for 2 h at room temperature. Cleaved CYP proteins were eluted off the column by washing the column with another 20 mL of lysis buffer while gathering 2 mL elution fractions. Eluted fractions with the most color were pooled and loaded onto a Superdex 200 16/600 pre-equilibrated in the respective gel filtration buffers (SalCYP, BotCYP and SalCYP point mutants: 150 mM NaCl, 10 mM HEPES pH 7.4, 0.5 mM TCEP; SalCYP<sup>T</sup>: 10 mM HEPES pH 7.4). Size exclusion chromatography was carried out at 1 mL/min and the fractions with the most color were pooled and concentrated using a 30 kDa cutoff filter. The protein concentration was determined by spectrophotometry at A<sub>280</sub> using the extinction coefficients calculated from the primary amino acid sequence. SalCYP was diluted to 130 µM whereas SalCYP<sup>T</sup> was diluted to 450 µM in their respective gel filtration buffers for crystallography. Protein purity and integrity was confirmed by SDS-PAGE as well as intact protein mass spectrometry.

### X-ray crystallography

Initial screening experiments for both proteins were carried out at their aforementioned concentrations in sitting-drop SwissCI plates with commercially available Qiagen screens using a Gryphon crystallization robot. A needle cluster for SalCYP was observed after 3 days at 277 K in a well solution of 20 % PEG 3350 and 200 mM calcium acetate. Initial crystals were crushed and seeded in different ratios in the protein solution prior to setting up optimization plates with a Dragonfly liquid dispenser using the microseeding technique. Single crystals were cryoprotected in 30 % glycerol and mounted onto a cryo loop. A high redundancy dataset for Fe-SAD was collected at the ESRF Beamline ID-29 at 100 K and a wavelength of 1.738 Å. Data were processed using XDS<sup>[3]</sup> and POINTLESS<sup>[4]</sup>, AIMLESS<sup>[5]</sup> and ctruncate<sup>[6]</sup> implemented in ccp4<sup>[7]</sup>. The structure was solved using Phenix.AutoSol,<sup>[8]</sup> followed by several rounds of manual rebuilding in COOT<sup>[9]</sup> and refinement in Phenix.refine<sup>[8]</sup>. Crystals of SalCYP<sup>T</sup> appeared after 5 days at 277 K in a well solution containing 20 % PEG 8000, 16.6 % glycerol, 200 mM calcium acetate and 80 mM sodium cacodylate pH 6.5. As for SalCYP, microseeding improved the quality as well as the nucleation probability of SalCYP<sup>T</sup> and was used in the same fashion to obtain single crystals which were cryoprotected by incubation with 30 % glycerol and mounted onto a cryo loop. A full dataset of SalCYP<sup>T</sup> was collected at ESRF Beamline MASSIF-3 at 100 K and a wavelength of 0.961 Å. Data were processed as for SalCYP and molecular replacement in Phaser<sup>[10]</sup> was carried out

using SALCYP as a search model, followed by several rounds of manual rebuilding in COOT<sup>[9]</sup> and refinement in Phenix.refine<sup>[8]</sup>. The structures were validated using MolProbity<sup>[11]</sup>, and all images presented were created using PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.). Interaction diagrams were created using Ligplus<sup>[12]</sup>.

### ***In vitro* production of BotA<sup>PCCDAH</sup> (2a/b)**

BotA<sup>P</sup>, IpoC, PurCD and PurAH were expressed and purified as described previously.<sup>[13]</sup> Large scale IpoC reactions were carried out in 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 applied onto a Superdex 30 16/60 size exclusion chromatography column pre-equilibrated in GF buffer. Heterocyclized-peptide-containing fractions were pooled, analyzed by mass spectrometry and concentrated using a 5 kDa cutoff filter. The large scale reactions of PurCD and PurAH could be carried out in combination on a 9 mL scale by incubating the reaction setup 50 μM BotA<sup>PC</sup>, 5 μM PurCD, 0.25 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 PurCD rxn buffer (200 mM NaCl, 50 mM CHES, 10 % glycerole, pH 9.5). The reaction mixture was filtered through a 0.22 μm filter and applied onto a Superdex 30 16/60 size exclusion chromatography column pre-equilibrated in ultrapure water. Every chromatography fraction was analyzed by mass spectrometry and the pure **2a/b**-containing fractions were pooled and lyophilized. Powder of **2a/b** was dissolved in ultrapure water and pipetted into a pre-weighed HPLC tube. All solvent was removed by vacuum centrifugation and remaining powder of **2a/b** was weighed using a precision scale. A concentration of 10 mM for **2a/b** was adjusted in ultrapure water, subsequently analyzed by mass spectrometry and kept on ice until further use.

### **MST**

MST experiments were performed using a Monolith NT.115 instrument in GF buffer without TCEP (150 mM NaCl, 10 mM HEPES pH 7.4) at 22 °C. Proteins were dialysed into freshly prepared GF buffer without TCEP, while a 100 mM stock of **2a/b** in ultrapure water was diluted in GF buffer without TCEP to the starting concentration of 1 mM. The protein was fluorescently labelled using the Monolith His-tag labelling kit RED-tris-NTA according to the manufacturer's instructions. For MST measurements, the manufacturer's instructions were followed using a starting concentration of 0.5 mM **2a/b**, 20% excitation power and 40% MST power. K<sub>D</sub> values were estimated using the *MO.Affinity* software suite (Nanotemper).

### Redox partner purification

BmCPR and Fdx2 were purified as reported previously.<sup>[14,15]</sup> The preparation of the fission yeast proteins Arh1 and Etp1<sup>fd</sup> was carried out as described before.<sup>[16,17]</sup> Recombinant bovine AdR and the Adx4-108 (truncated form of Adx comprising amino acids 4 – 108) were purified as mentioned elsewhere.<sup>[18,19]</sup>

The concentration of recombinant P450 was estimated using the CO-difference spectral assay as described previously with  $\epsilon_{450-490} = 91 \text{ mM}^{-1} \text{ cm}^{-1}$ .<sup>[20]</sup> The concentration of BmCPR was quantified by measuring the flavin absorbance at 456 nm with  $\epsilon_{456} = 21 \text{ mM}^{-1} \text{ cm}^{-1}$  for the oxidized enzyme.<sup>[15]</sup> The concentrations of the AdR and Arh1 were measured using the extinction coefficient  $\epsilon_{450} = 11.3 \text{ mM}^{-1} \text{ cm}^{-1}$ .<sup>[17,21]</sup> The concentrations of Fdx2 and Etp1<sup>fd</sup> were measured using the extinction coefficient  $\epsilon_{390} = 6.671 \text{ mM}^{-1} \text{ cm}^{-1}$  and  $\epsilon_{414} = 9.8 \text{ mM}^{-1} \text{ cm}^{-1}$ , respectively.<sup>[14,22]</sup>

### Investigation of electron transfer partners

The functional interaction of the electron transfer partners for a particular P450 was determined by observing the NADPH reduced CO-complex peak at 450 nm when P450 was coupled with the different redox partners in the absence of substrate. For this, SalCYP or BotCYP was mixed with ferredoxins and ferredoxin reductases with ratios of SalCYP : Fdx2/Etp1<sup>fd</sup>/Adx : BmCPR/Arh1/AdR of 1:20:3  $\mu\text{M}$  (and 1:40:4 for BmCPR-Fdx2) in 50 mM HEPES buffer pH 7.4 and NADPH was added to a final concentration of 1 mM. The spectrum of NADPH-reduced samples was recorded after bubbling the sample with carbon monoxide (CO) gas.

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

Measurements were performed on a Dionex Ultimate 3000 RSLC system (Thermo Fisher Scientific) 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) H<sub>2</sub>O + 0.1% formic acid at a flow rate of 600  $\mu\text{L min}^{-1}$  and 45 °C. 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% ACN before re-equilibration under the initial conditions. UV spectra were recorded by a DAD in the range from 200 to 600 nm.

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 14.5 psi. 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 the mass range from 150 to 2500 m/z. 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  $C_{12}H_{19}F_{12}N_3O_6P_3$ ,  $C_{18}H_{19}O_6N_3P_3F_2$  and  $C_{24}H_{19}F_{36}N_3O_6P_3$  which generate the  $[M+H]^+$  ions of 622.028960, 922.009798 and 1221.990638. LC-MS<sup>2</sup> fragmentation was performed using an automatic precursor selection mode.

Signals in the MS-spectra are labelled with the observed monoisotopic mass. Extracted-ion chromatograms were generated with the Bruker Compass DataAnalysis software (Version 4.2) using the calculated monoisotopic mass with a range of 5 ppm. EICs shown in one figure always have the same time and intensity range, except for **Figure S12**. MS-spectra in one figure are scaled to a relative intensity of 1.

#### ***In vitro* reactions of BotCYP and SalCYP**

To investigate the role of the P450 enzymes *in vitro*, 20  $\mu$ M CYP substrate analogue **2a/b** were incubated with 5  $\mu$ M BotCYP, SalCYP or SalCYP mutants, respectively, 5  $\mu$ M BmCPR, 50  $\mu$ M Fdx2 and 2.5 mM NADPH at 30 °C overnight in reaction buffer (150 mM NaCl, 10 mM HEPES, pH 7.4). Reactions without the P450 enzyme were used as controls. Reactions were stopped and enzymes were precipitated by the addition of two volumes of ACN. The samples were frozen at -80 °C and precipitated enzymes were pelleted by centrifugation at 15.000 x g for 20 min. The turnover was analyzed by LC-MS measurements. Reactions were set up and analyzed in triplicates.

To investigate the effect of the addition of BotH to a CYP reaction, reactions were set up as described above with and without the addition of 1  $\mu$ M BotH. Reactions were stopped after 120 min incubation at 30 °C.

To test if the configuration at the Asp C $\alpha$ -position is retained during the oxidative decarboxylation (**Figure 3C**), **4a/b** was produced by incubating 30  $\mu$ M **2a/b** with 6  $\mu$ M BotH in D<sub>2</sub>O reaction buffer at 30°C for 30 min. Afterwards, BotH was destroyed by incubating the solution at 98 °C for 10 min. Denatured BotH was removed by centrifugation (15.000xg, 20 min, 4 °C). D<sub>2</sub>O was removed by lyophilization. The lyophilizate was redissolved in H<sub>2</sub>O and reactions were performed as described above. In parallel, the same reaction was performed in H<sub>2</sub>O.

To investigate the substrate preference of BotCYP (**Figure 2A**), reactions were set up as described above, but were stopped after 5, 15, 30 and 60 min by the addition of ACN. For this experiment, BotH treated **2a/b** solution was used. Upon treatment with BotH, an approximate ratio of **2a/b** of 1:1 was achieved. BotH was destroyed and removed from the solution before the experiments, as described above.



## Supplementary References

- [1] Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R *Nucleic Acids Res.* **2019** [Epub ahead of print]
- [2] Studier FW *Protein Expr Purif.* **2005** 41 (1):207-34
- [3] Kabsch W *Acta Crystallogr D Biol Crystallogr.* **2010** 66 (Pt 2):125-32
- [4] Evans PR *Acta Crystallogr D Biol Crystallogr.* **2011** 67 (Pt 4):282-92
- [5] Evans PR, Murshudov GN *Acta Crystallogr D Biol Crystallogr.* **2013** 69 (Pt 7):1204-14
- [6] French S, Wilson K *Acta Crystallographica Section A* **1978** 34:517-525
- [7] Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS *Acta Crystallogr D Biol Crystallogr.* **2011** 67 (Pt 4):235-42
- [8] Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH *Acta Crystallogr D Biol Crystallogr.* **2010** 66 (Pt 2):213-21.
- [9] Emsley P, Lohkamp B, Scott WG, Cowtan K *Acta Crystallogr D Biol Crystallogr.* **2010** 66 (Pt 4):486-501
- [10] McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ *J Appl Crystallogr.* **2007** 40 (Pt 4):658-674
- [11] Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC *Acta Crystallogr D Biol Crystallogr.* **2010** 66 (Pt 1):12-21
- [12] Laskowski RA, Swindells MB *J Chem Inf Model.* **2011** 51 (10):2778-86
- [13] Sikandar A, Franz L, Melse O, Antes I, Koehnke J *J Am Chem Soc.* **2019** 141 (25):9748-9752
- [14] Brill E, Hannemann F, Zapp J, Brüning G, Jauch J, Bernhardt R *Appl Microbiol Biotechnol.* **2014** 98 (4):1701-17.
- [15] Milhim M, Gerber A, Neunzig J, Hannemann F, Bernhardt R *J Biotechnol.* **2016** 231:83-94
- [16] Bureik M, Schiffler B, Hiraoka Y, Vogel F, Bernhardt R *Biochemistry.* **2002** 41 (7):2311-21
- [17] Ewen KM, Schiffler B, Uhlmann-Schiffler H, Bernhardt R, Hannemann F *FEMS Yeast Res.* **2008** 8 (3):432-41
- [18] Uhlmann H, Beckert V, Schwarz D, Bernhardt R *Biochem Biophys Res Commun.* **1992** 188 (3):1131-8
- [19] Sagara Y, Wada A, Takata Y, Waterman MR, Sekimizu K, Horiuchi T. *Biol Pharm Bull.* **1993** 16 (7):627-30
- [20] Omur T, Sato R *J Biol Chem.* **1964** 239:2379-85
- [21] Hiwatashi A, Ichikawa Y, Maruya N, Yamano T, Aki K *Biochemistry.* **1976** 15(14):3082-90

[22] Schiffler B, Bureik M, Reinle W, Müller EC, Hannemann F, Bernhardt R *J Inorg Biochem.* **2004** 98(7):1229-37

---

## Chapter 6

# Structure and Biosynthesis of Crocagins: Polycyclic Postranslationally Modified Ribosomal Peptides from *Chondromyces crocatus*

Previously published in:

Konrad Viehrig<sup>[a]</sup>, Frank Surup<sup>[b][c]</sup>, Carsten Volz<sup>[a]</sup>, Jennifer Herrmann<sup>[a]</sup>, Antoine Abou  
Fayad<sup>[a]</sup>, **Sebastian Adam**<sup>[a]</sup>, Jesko Koehnke<sup>[a]</sup>, Dirk Trauner<sup>[d]</sup>, and Rolf Müller<sup>[a][b][c]\*</sup>

**Angew Chem Int Ed Engl.** 2017 Jun 19;**56**(26):7407-7410.

**DOI: 10.1002/anie.201612640**

### Affiliation

<sup>[a]</sup> Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Center for Infection Research and Pharmaceutical Biotechnology, Saarland University, Campus, Building E 8.1, 66123, Saarbrücken, Germany

<sup>[b]</sup> Helmholtz Center for Infection Research (HZI), Department Microbial Drugs, Inhoffenstraße 7, 38124, Braunschweig, Germany

<sup>[c]</sup> German Centre for Infection Research (DZIF), Partner Site Hannover-Braunschweig, Germany

<sup>[d]</sup> Department of Chemistry, Ludwig-Maximilians-Universität München, Butenandtstrasse 5-13, 81377, Munich, Germany

## **Contributions and Acknowledgements**

### **Author's effort:**

The chapter is mainly featured in this thesis to introduce the novel compound class of crocagins, which will be the topic of further research that the author conducted in the next two chapters. Still, the author recombinantly expressed and purified CgnB, which was used in this study to prove that the precursor peptide CgnA is the origin of crocagins.

### **Other's effort:**

Konrad Viehrig contributed by identifying and annotating the crocagin biosynthetic gene cluster. He designed and cloned the inactivation constructs, performed Cm c5 mutagenesis and cultivation as well as MS analysis. Frank Surup contributed by initially isolating crocagins, carrying out structure elucidation as well as labeling experiments. Carsten Volz contributed by performing the binding affinity studies of CgnA with CgnB. Jennifer Herrmann contributed by performing the microscopy as well as evaluating the biological activity testing. Antoine Abou Fayad contributed by synthesizing the fluorescein derivative of RNA B.

## 6.1 Abstract

Secondary metabolome mining efforts in the myxobacterial multiproducer of natural products, *Chondromyces crocatus* Cm c5, resulted in the isolation and structure elucidation of crocagins, which are novel polycyclic peptides containing a tetrahydropyrrolo[2,3-b]indole core. The gene cluster was identified through an approach combining genome analysis, targeted gene inactivation in the producer, and *in vitro* experiments. Based on our findings, we developed a biosynthetic scheme for Crocagin biosynthesis. These natural products are formed from the three C-terminal amino acids of a precursor peptide and thus belong to a novel class of ribosomally synthesized and post-translationally modified peptides (RiPPs). We demonstrate that Crocagin A binds to the carbon storage regulator protein CsrA, thereby inhibiting the ability of CsrA to bind to its cognate RNA target.

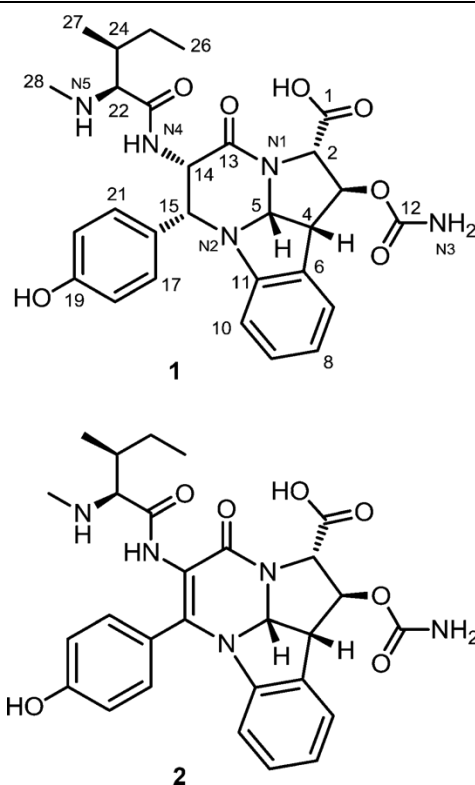
## 6.2 Main text

Myxobacteria are Gram-negative soil bacteria that exhibit unusual features, such as coordinated swarming behavior and the formation of multicellular fruiting bodies.<sup>[1]</sup> Dozens of novel compound classes have been isolated from myxobacteria, including many that appear to be exclusively produced by them.<sup>[2,3]</sup> The full extent of their potential to form secondary metabolites has only emerged in recent years, since fully sequenced genomes and computational genome analysis tools have become available. Predictions based on genome sequence analysis indicate a secondary metabolite potential in myxobacteria that far exceeds the currently known compounds.<sup>[4]</sup> *Chondromyces crocatus* Cm c5 is an exceptionally prolific myxobacterial producer of multiple bioactive natural product families, such as the chondramides,<sup>[5]</sup> ajudazols,<sup>[6]</sup> crocacins,<sup>[7]</sup> crocapeptins,<sup>[8]</sup> and thuggacins.<sup>[9]</sup> A computational estimation of the encoded secondary metabolite potential using antiSMASH<sup>[10]</sup> revealed an outstanding 19 gene clusters with nonribosomal peptide synthetase (NRPS) or polyketide synthase (PKS) genes,<sup>[11]</sup> less than a third of which have been correlated to secondary metabolites to date. In addition, several gene clusters encoding ribosomally synthesized and post-translationally modified peptides (RiPPs) belonging to the bacteriocin and lantipeptide families have been predicted. RiPPs are a highly diverse group of natural products with a broad spectrum of biological functions.<sup>[12]</sup> RiPPs originate from ribosomally assembled precursor peptides, which undergo post-translational modification to gain their biological function. Typically, the precursor peptide contains a leader, which is essential for recognition by parts

of the biosynthetic machinery, and a core peptide that forms the final product. Thus far, no RiPPs have been reported from myxobacteria. In an effort to tap into the secondary metabolite potential of Cm c5, the strain was used in a compound mining approach that combined mutagenesis and comparative HR-MS analysis, as described previously.<sup>[13]</sup> We discovered that a group of mass peaks were not abolished in any of our NRPS/PKS knockout mutants and finally assigned them to the Crocagins, the discovery of which we report here.

Crocagin A (**1**) was obtained from cultures of Cm c5 as a white, amorphous powder. HR-MS data suggested a molecular formula of  $C_{28}H_{33}N_5O_7$  and 15 degrees of unsaturation. Combined interpretation of  $^1H$  and  $^1H,^1H$  HSQC NMR data enabled the assignment of eight aromatic as well as eight aliphatic methines, one methylene, three methyl groups, and two H/D-exchangeable protons. The  $^{13}C$  spectrum revealed the presence of four aromatic quaternary and four carboxyl carbons. From  $^1H,^1H$  COSY and TOCSY spectra, five proton spin systems were identified, which together with inner residue  $^1H,^{13}C$  HSQC and HMBC correlations provided the partial structures of N-methylisoleucine (NMelle), a 3-substituted tyrosine and a tryptophan-derived moiety. A carbamate group in **1** was suggested by the loss of a carbamic acid fragment ( $CH_3NO_2$ ) in the MS/MS (**Table S1** in the Supporting Information) and by the hydrolytic removal of a carbamoyl residue (**Figure S7** in the Supporting Information). Combined interpretation of the NMR data, that is, the  $^1H$  and  $^1H,^1H$  HSQC correlations,  $^1H,^1H$  COSY and TOCSY correlations,  $^1H,^{13}C$  HSQC and HMBC correlations, and  $^1H,^{15}N$ -HMBC correlations, led to the elucidation of the planar Crocagin structure, which contains a highly substituted central tetrahydropyrimidinone ring moiety (**Figures S1** and **S2**). The relative configuration of **1** was assigned from the  $^1H,^1H$  ROESY correlations and  $^1H,^1H$  as well as  $^1H,^{13}C$  coupling constants. Advanced Marfey's method and the phenylglycine methyl ester (PGME) method determined the absolute configuration of **1** as 2*S*,3*S*,4*R*,5*R*,14*S*,15*R*,23*S*,24*S*.

Crocagin B (**2**) was detected in small amounts in extracts of Cm c5. HR-ESI-MS data suggested the molecular formula as  $C_{27}H_{33}N_5O_7$ , which has two fewer protons than **1**. The UV spectrum of **2** with absorption maxima of 290 and 346 nm indicated an enlarged  $\pi$  system. The NMR data for **2** were highly similar to those for **1**, with key difference being the replacement of methines C14 and C15 by quaternary olefinic carbons, providing the  $\Delta^{14,15}$  double bond (**Figure 1**). An additional minor derivative of **1** was observed by HPLC-UV HR-ESI-MS in crude extracts of Cm c5. Here, MS data indicated the loss of the N-methyl group at the Ile residue.

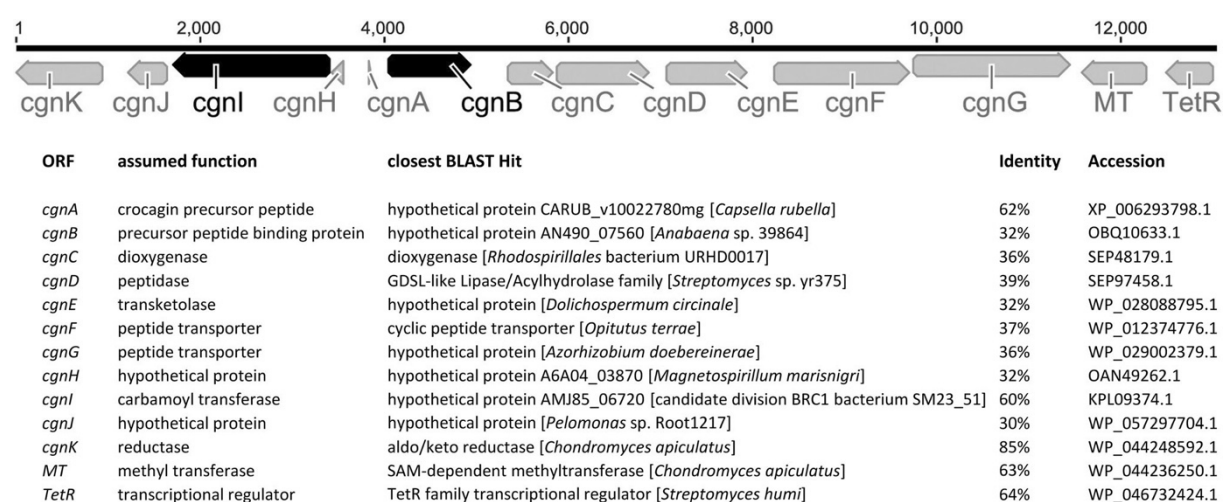


**Figure 1.** Chemical structures of Crocagin A (1) and B (2), isolated from *Chondromyces crocatus* Cm c5.

Crocagin A (**1**) showed no antimicrobial or cytotoxic activity in our standard screening<sup>[14]</sup> but was later found to be active in a screen for inhibitors of carbon storage regulator protein A (CsrA).<sup>[15]</sup> CsrA binds to the 5' untranslated region of mRNA molecules, thereby affecting their translation and stability. This conserved protein was found in many bacteria and represents a promising target for anti-infective drug development since it is involved in the regulation of several virulence factors in Gram-negative pathogens such as *Helicobacter pylori*<sup>[16]</sup> and *Salmonella typhimurium*.<sup>[17,18]</sup> Compound **1** competitively inhibits the interaction of CsrA with its RNA ligand in a dose-dependent manner, and direct binding of **1** to CsrA was confirmed in a fluorescence polarization (FP)-based competition assay (**Figure S8 A**) and by surface plasmon resonance (SPR) affinity experiments (**Figure S8 B**) with a calculated dissociation constant ( $K_D$ ) of  $12.9 \pm 5.8 \mu\text{M}$ . Because of its *in vitro* activity, **1** was initially considered a potential small-molecule inhibitor of virulence factor production in opportunistic pathogens. However, when cellular uptake of **1** was investigated by MS and fluorescence microscopy, it was found not to be taken up by bacterial cells (**Figures S10–S13**).

The chemical structure of Crocagin A (**1**) consists of the tripeptide sequence Ile-Tyr-Trp (IYW), of which the tyrosine and tryptophan residues are covalently crosslinked through two C-N bonds. NRPS genes with suitable predicted substrate specificity for this peptide are not found

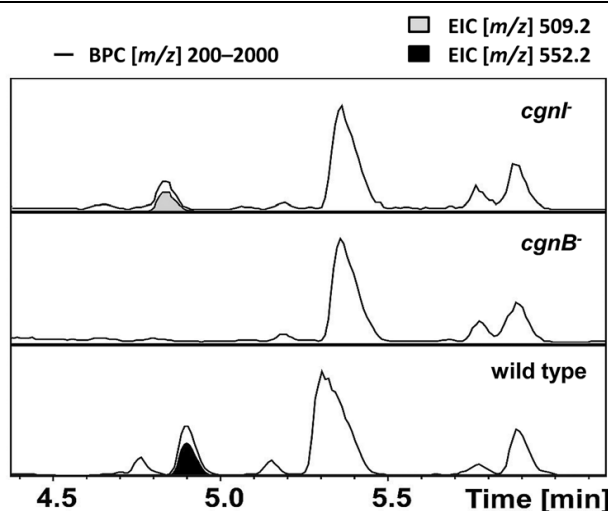
in the Cm c5 genome sequence and none of our NRPS inactivation mutants led to the loss of Crocagin production. As an alternative biosynthetic route, a ribosomal origin was considered, in which **1** would be derived from a hypothetical precursor peptide. Screening the Cm c5 genome sequence for the IYW motif yielded one IYW-containing 21aa hypothetical protein, which was labeled CgnA. The gene cluster around *cgnA* was annotated as shown in **Figure 2**, and targeted inactivation of the annotated open reading frames (ORFs) was pursued with the integrative plasmid pSUP\_*hyg*<sup>[19]</sup> to confirm their involvement in the biosynthesis of **1**. Mutant cultures with insertions in *cgnB* and *cgnI* were obtained, genetically verified (**Figures S15** and **S16**), and analyzed by LC-MS. Representative chromatograms of mutant and wild-type extracts are shown in **Figure 3**.



**Figure 2.** Proposed Crocagin biosynthetic gene cluster. Black arrows show genes that were inactivated in Cm c5 through insertional mutagenesis and found to be involved in Crocagin biosynthesis. Further annotated ORFs expected to be involved in the biosynthesis of **1** are shown as grey arrows. Predicted gene functions as obtained from BLAST analysis and the closest homologue of each annotated ORF are shown in the table below the gene cluster.

In mutants with inactivated *cgnB*, production of **1** was abolished, while mutants with inactivated *cgnI* produced a compound with a *m/z* of 509.2 instead, which is not present in other mutants or the wild-type producer. This mass signal corresponds to the  $[M+H]^+$  ions of **1** without the carbamate group, which was later corroborated by chemical hydrolysis of the carbamate group from **1** (**Figure S7**) and comparison to this compound. In extracts of *cgnB* mutants, no potential precursor peaks were observed. Masses corresponding to the putative precursor peptide (2450 Da) were not detected in wild-type or mutant extracts.



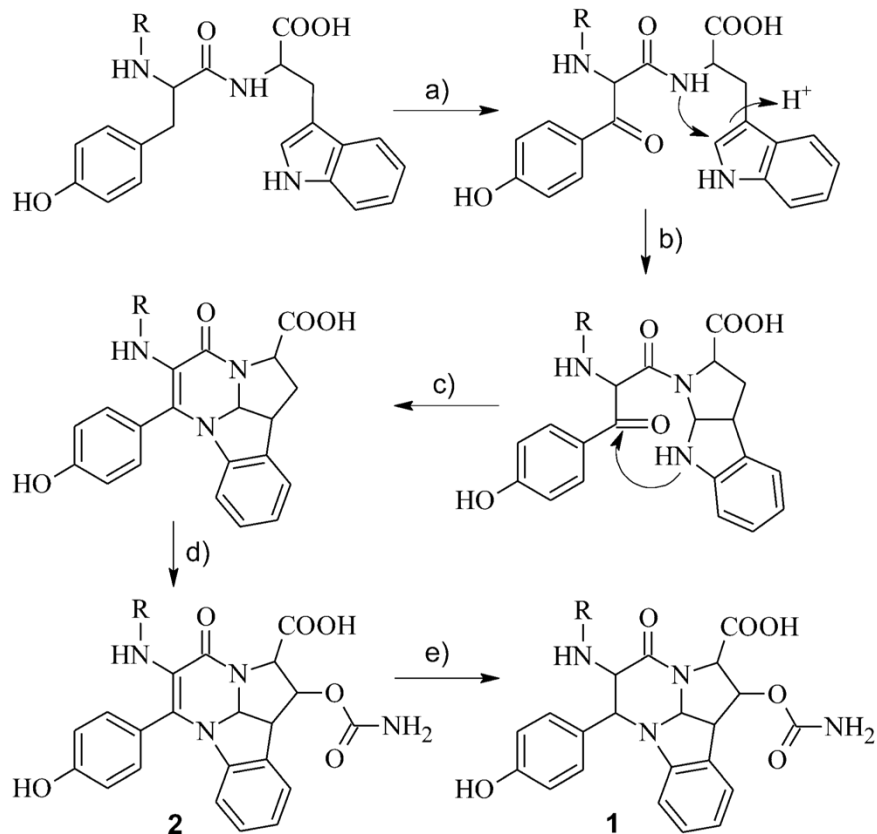


**Figure 3.** MS analysis of Cm c5 mutants, showing the effect of gene inactivation on the production of Crocagin A(**1**, black peak). In *cgkB* mutants, the production of **1** is abolished, while *cgnt* mutants exclusively produce a precursor (grey peak) with  $m/z$  of 509.2.

Inactivation of *cgkB* and *cgnt*, both encoded in the vicinity of *cgmA*, confirmed their involvement in the biosynthesis of **1**. The existence of CgnA as the precursor peptide of **1** remained, however, hypothetical. From the observed loss of Crocagin production in *cgkB* mutants, we assumed that CgnB is a key enzyme involved in the biosynthesis of **1** and should exhibit specific binding to CgnA. To investigate the direct interaction of CgnB and CgnA *in vitro*, recombinant CgnB was used in SPR experiments, in which concentration-dependent specific binding of CgnB to the precursor peptide of CgnA was observed (**Figure S17**). Assuming a 1:1 interaction, CgnB binds to CgnA with a binding constant ( $K_D$ ) of  $0.716 \pm 0.19 \mu\text{M}$ . No conversion or modification of CgnA was observed upon incubation of CgnA and CgnB *in vitro*.

The combined results from mutagenesis and *in vitro* experiments confirmed the genetic locus for Crocagin biosynthesis and their ribosomal origin. Crocagins therefore belong to the class of ribosomally synthesized and post-translationally modified peptides (RiPPs). The sequence of CgnB does not contain conserved motifs that hint towards a catalytic function and we therefore expect that CgnB either mediates the binding of the precursor peptide CgnA to other modifying enzymes, or is part of a protein complex with the substrate CgnA and other tailoring enzymes, in a manner akin to that found in lassopeptide biosynthesis.<sup>[20]</sup> The heavily cross-linked structure of **1** suggests the involvement of redox enzymes, which are expected to carry out oxidation, reduction, and crosslinking reactions. Obvious candidates for these functionalities are the putative dioxygenase CgnC and the reductase CgnK. The tetracyclic system in **1** is hypothesized to be formed in a manner similar to that found in ardeemin biosynthesis,<sup>[21]</sup> which would require hydroxylase, dehydratase, and reductase functionalities. A possible

biosynthesis scheme is shown in **Figure 4**. Briefly, the ring-closing reactions are expected to take place after or in parallel with the hydroxylation of C15 and oxidation to the respective ketone, followed by cyclization and water elimination steps (**Figure 4**) to yield Crocagin B. Subsequent reduction of the double bond by CgnK would lead to the formation of Crocagin A.



**Figure 4.** Proposed biosynthesis of the tetrahydropyrimidinone system in **1**. R = rest of the precursor peptide. After or in parallel to the oxidation of Tyr at C15, which is assumed to be catalyzed by the dioxygenase CgnC (a), ring closing reactions and water elimination take place (b, c). Further modifications of the resulting scaffold, that is, the second hydroxylation at C3 and carbamoylation occur (d) and result in formation of Crocagin B. Eventually, reduction of the double bond, for example, by the reductase CgnK, leads to the formation of Crocagin A. The exact order of the biosynthetic steps is unknown. N-methylation and peptide cleavage are not shown.

Candidate genes for the further modifications observed in **1** are also found in the gene cluster. The function of CgnI as a carbamoyl transferase was confirmed by mutagenesis and the predicted methyl transferase encoded in the cluster is considered to be responsible for methylation of the N terminus after cleavage from the leader peptide. The latter is expected to occur after partial or complete modification of the core peptide, as is generally observed in RiPPs biosynthesis. The putative hydrolase CgnD is considered a likely candidate for cleavage of the processed precursor peptide and release of the tripeptide, but shows no sequence homology to peptidases of known RiPPs gene clusters. Genes *cgnF* and *cgnG* are predicted to

encode peptide transporter proteins but may also contribute to precursor peptide cleavage as known from bacteriocin biosynthesis.<sup>[22]</sup> The minor metabolite **2** is considered a precursor (or possibly a side product) of **1** and may also stem from autoxidation of **1**. It cannot be excluded that further enzymes are involved in the biosynthesis of **1**, which are encoded elsewhere in the genome.

Crocagin A contains a tetrahydropyrrolo[2,3-b] indoline core, a motif known from alkaloids that has been pursued in synthetic studies<sup>[23]</sup> but is unprecedented for bacterial natural products. A few bacterial metabolites exhibiting some similarity to this compound have been described but they differ significantly in their core structure.<sup>[24]</sup> In addition, Crocagins do not group into any of the RiPPs compound classes provided by Arnison *et al.*,<sup>[12]</sup> because of both their unprecedented tetracyclic structure and biosynthesis. We therefore suggest a new class of RiPPs for this scaffold. To further corroborate this assumption, studies to elucidate the Crocagin biosynthesis *in vitro* with recombinant proteins are ongoing.

### 6.3 Acknowledgements

We thank Christel Kakoschke and Alberto Plaza for recording NMR spectra, Manfred Nimtz, Aileen Gollasch, and Thomas Hoffmann for ESI-MS measurements, Christine Maurer for fluorescence polarization measurements, Diana Telkemeyer for strain maintenance, the fermentation team for large-scale cultivation, Victor Wray for helpful discussions and Rolf Jansen for MM+ calculations in addition to proofreading the manuscript. Research in R.M.'s laboratory was funded by the Deutsche Forschungsgemeinschaft (DFG) and the Bundesministerium für Bildung und Forschung (BMBF).

## 6.4 References

- [1] M. Dworkin *Microbiol. Rev.* **1996**, *60*, 70 – 102.
- [2] a) K. J. Weissman, R. Müller *Nat. Prod. Rep.* **2010**, *27*, 1276 – 1295; b) J. A. Herrmann, A. Abou Fayad, R. Müller *Nat. Prod. Rep.* **2017**, *34*, 135 – 160.
- [3] T. F. Schäberle, F. Lohr, A. Schmitz, G. M. König *Nat. Prod. Rep.* **2014**, *31*, 953 – 972.
- [4] a) S. C. Wenzel, R. Müller *Curr. Opin. Drug Discov. Devel.* **2009**, *12*, 220 – 230 ; b) S. C. Wenzel, R. Müller *Nat. Prod. Rep.* **2009**, *26*, 1385 – 1407.
- [5] B. Kunze, R. Jansen, F. Sasse, G. Höfle, H. Reichenbach *J. Antibiot.* **1995**, *48*, 1262 – 1266.
- [6] B. Kunze, R. Jansen, G. Höfle, H. Reichenbach *J. Antibiot.* **2004**, *57*, 151 – 155.
- [7] B. Kunze, R. Jansen, G. Höfle, H. Reichenbach *J. Antibiot.* **1994**, *47*, 881 – 886.
- [8] K. Viehrig, F. Surup, K. Harmrolfs, R. Jansen, B. Kunze, R. Müller *J. Am. Chem. Soc.* **2013**, *135*, 16885 – 16894.
- [9] K. Buntin, H. Irschik, K. J. Weissman, E. Luxemburger, H. Blocker, R. Müller *Chem. Biol.* **2010**, *17*, 342 – 356.
- [10] T. Weber, K. Blin, S. Duddela, D. Krug, H. U. Kim, R. Brucoleri, S. Y. Lee, M. A. Fischbach, R. Müller, W. Wohlleben *Nucleic Acids Res.* **2015**, *43*, W237 – W243.
- [11] N. Zaburannyi, B. Bunk, J. Maier, J. Overmann, R. Müller, *Appl. Environ. Microbiol.* **2016**, *82*, 1945 – 1957.
- [12] 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, W. A. van der Donk *Nat. Prod. Rep.* **2013**, *30*, 108 – 160.
- [13] a) D. Pistorius, R. Müller *ChemBioChem* **2012**, *13*, 416–426; b) L. Keller, A. Plaza, C. Dubiella, M. Groll, M. Kaiser, R. Müller *J. Am. Chem. Soc.* **2015**, *137*, 8121 – 8130.
- [14] F. Surup, B. Thongbai, E. Kuhnert, E. Sudarman, K. D. Hyde, M. Stadler *J. Nat. Prod.* **2015**, *78*, 934 – 938.
- [15] C. K. Maurer, M. Fruth, M. Empting, O. Avrutina, J. Hossmann, S. Nadmid, J. Gorges, J. Herrmann, U. Kzmaier, P. Dersch, R. Müller, R. W. Hartmann *Future Med. Chem.* **2016**, *8*, 931 – 947.
- [16] F. M. Barnard, M. F. Loughlin, H. P. Fainberg, M. P. Messenger, D. W. Ussery, P. Williams, P. J. Jenks *Mol. Microbiol.* **2004**, *51*, 15–32.
- [17] C. Altier, M. Suyemoto, S. D. Lawhon *Infect. Immun.* **2000**, *68*, 6790 – 6797.
- [18] A. K. Heroven, K. Bohme, P. Dersch *Adv. Exp. Med. Biol.* **2012**, *954*, 315 – 323.
- [19] S. Rachid, D. Krug, B. Kunze, I. Kochems, M. Scharfe, T. M. Zabriskie, H. Blocker, R. Müller *Chem. Biol.* **2006**, *13*, 667 – 681.
- [20] S. Zhu, C. D. Fage, J. D. Hegemann, A. Mielcarek, D. Yan, U. Linne, M. A. Marahiel *Sci. Rep.* **2016**, *6*, 35604.
- [21] S. W. Haynes, X. Gao, Y. Tang, C. T. Walsh *ACS Chem. Biol.* **2013**, *8*, 741 – 748.
- [22] J. Michiels, G. Dirix, J. Vanderleyden, C. Xi *Trends Microbiol.* **2001**, *9*, 164 – 168.
- [23] A. H. Jackson, P. V. R. Shannon, D. J. Wilkins *J. Chem. Soc. Chem. Commun.* **1987**, 653 – 654.

[24] M. Okada, I. Sato, S. J. Cho, H. Iwata, T. Nishio, D. Dubnau, Y. Sakagami *Nat. Chem. Biol.* **2005**, 1, 23 – 24.

## Supporting Information

# Structure and Biosynthesis of Crocagins: Polycyclic Postranslationally Modified Ribosomal Peptides from *Chondromyces crocatus*

Previously published in:

Konrad Viehrig<sup>[a]</sup>, Frank Surup<sup>[b][c]</sup>, Carsten Volz<sup>[a]</sup>, Jennifer Herrmann<sup>[a]</sup>, Antoine Abou  
Fayad<sup>[a]</sup>, **Sebastian Adam**<sup>[a]</sup>, Jesko Koehnke<sup>[a]</sup>, Dirk Trauner<sup>[d]</sup>, and Rolf Müller<sup>[a][b][c]\*</sup>

**Angew Chem Int Ed Engl.** 2017 Jun 19;**56**(26):7407-7410.

**DOI: 10.1002/anie.201612640**

### Affiliation

<sup>[a]</sup> Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Center for Infection Research and Pharmaceutical Biotechnology, Saarland University, Campus, Building E 8.1, 66123, Saarbrücken, Germany

<sup>[b]</sup> Helmholtz Center for Infection Research (HZI), Department Microbial Drugs, Inhoffenstraße 7, 38124, Braunschweig, Germany

<sup>[c]</sup> German Centre for Infection Research (DZIF), Partner Site Hannover-Braunschweig, Germany

<sup>[d]</sup> Department of Chemistry, Ludwig-Maximilians-Universität München, Butenandtstrasse 5-13, 81377, Munich, Germany

## General Information

**Instrumentation:** Optical rotations were determined with a Perkin-Elmer 241 instrument; UV spectra were recorded with a Shimadzu UV-Vis spectrophotometer UV2450. NMR spectra were recorded with Bruker ARX 600 ( $^1\text{H}$  600 MHz,  $^{13}\text{C}$  150 MHz) and Bruker Avance III 700 ( $^1\text{H}$  700 MHz,  $^{13}\text{C}$  175 MHz) spectrometers. HRESIMS mass spectra were obtained with an Agilent 1200 series HPLC-UV system combined with an ESI-TOF-MS (Maxis, Bruker) [column 2.1×50 mm, 1.7 mm,  $\text{C}_{18}$  Acquity UPLC BEH (Waters), solvent A:  $\text{H}_2\text{O}$  + 0.1% formic acid; solvent B:  $\text{AcCN}$  + 0.1% formic acid, gradient: 5% B for 0.5 min increasing to 100% B in 19.5 min, maintaining 100% B for 5 min, FR = 0.6 mL min $^{-1}$ , UV detection 200-500 nm].

**Cultivation:** *Chondromyces crocatus*, strain Cmc5 was cultivated in 70 L of a medium containing probion 3 g/L, soluble starch 3 g/L,  $\text{MgSO}_4 \times 7 \text{H}_2\text{O}$  2 g/L,  $\text{CaCl}_2 \times 2 \text{H}_2\text{O}$  0.5 g/L in a biofermenter Biostat UE100 (Braun Melsungen, Melsungen, D) at 30 °C for 11 days. 1% adsorber resin XAD-16 was added to the fermentation broth 3 days after inoculation. Mutagenesis of Cmc5: Gene disruption in Cm c5 via single crossing-over was performed as described previously,<sup>[16]</sup> as were cultivation, extraction and MS analysis of the resulting mutants. Genetic verification of Cm c5 mutants was carried out by PCR, see SI-6.

**Extraction and Isolation:** For isolation of **1** and **2**, XAD-16 and wet cell mass were harvested by centrifugation. Cells were separated from XAD-16 by flotation and discarded. The XAD was washed with water and 50% methanol, and subsequently eluted by methanol (5 L) and acetone (2 L). All fractions were analyzed by HPLC-UVMS. The concentrated methanol extract, which contained **1** and **2**, was subjected to solvent partition. The extract was dissolved in 85 % aqueous methanol (1 L) and extracted twice with heptane (0.5 L); subsequently the combined methanol phase was adjusted to 70% methanol and extracted twice with dichloromethane. HPLC-MS analysis revealed that the methanol layer contained the main amount of **1** and **2**, which was fractionated by RP MPLC [column 480 x 30mm (Kronlab) with ODS/AQ C18(YMC), solvent 33% methanol for 15 min, gradient 33% to 66% methanol in 45 min, gradient 66% to 100% methanol in 30 min, flow 30 mL/min, UV peak detection at 210 nm]. Fractions containing **1** and **2** were combined and further purified by preparative RP HPLC [column 250 x 21 mm, VP Nucleodur 100-10 C18 EC (MachereyNagel), solvents containing 50 mM sodium acetate, gradient 15% to 40% methanol in 25 min, flow 20 mL/min]. A final step of preparative RP HPLC [column 250 x 21 mm, VP Nucleodur C18 Gravity 5  $\mu\text{m}$  (Macherey Nagel), gradient 10% to 20% acetonitril in 25 min, 0.2% acetic acid, flow 20 mL/min] provided 6.5 mg of **1** and 1.2 mg of **2**.

**Crocagin A (1):** colorless, amorphous powder,  $[\alpha]_D + 71.6^\circ$  (c 0.5, MeOH); UV (MeOH) $I_{\max}$  (log e) 230 (sh), 277 nm (3.49); IR (KBr) 3374, 2966, 2932, 1672, 1607, 1516, 1481, 1434, 1385, 1272, 1210, 1142, 1025  $\text{cm}^{-1}$ ;  $^1\text{H}$  and  $^{13}\text{C}$  NMR (DMSO- $d_6$ ) see table S2; HRESIMS  $m/z$  552.2463 ( $[\text{M}+\text{H}]^+$ , calcd. for  $\text{C}_{28}\text{H}_{34}\text{N}_5\text{O}_7$ , 552.2453), 550.2299 ( $[\text{MH}]^-$ , calcd. for  $\text{C}_{28}\text{H}_{32}\text{N}_5\text{O}_7$ , 550.2307).

**Crocagin B (2):** colorless, amorphous powder,  $^1\text{H}$  and  $^{13}\text{C}$  NMR (MeOH- $d_4$ ): see table S3; HRESIMS  $m/z$  550.2307 ( $[\text{M}+\text{H}]^+$ , calcd. for  $\text{C}_{28}\text{H}_{32}\text{N}_5\text{O}_7$ , 550.2296),

**Preparation of the (S)- and (R)-PGME amides of 1:** Two portions of **1** (1.6 mg each; 2.9  $\mu\text{mol}$ ) were dissolved in 2 mL of dry pyridine and 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (4.5 mg, 29  $\mu\text{mol}$ ), 4-dimethylaminopyridine (50  $\mu\text{g}$ , 0.4  $\mu\text{mol}$ ) and (S)-(+)- respectively (R)-(-)-2-Phenylglycine methyl ester hydrochloride (5.8 mg, 29  $\mu\text{mol}$ ) were added. After stirring at room temperature for 16 h, the solvent was removed in vacuo. The residue was dissolved in 500  $\mu\text{L}$  of methanol and separated by preparative HPLC (column 250 x 21 mm, VP Nucleodur C18 Gravity 5  $\mu\text{m}$ , gradient 28% to 100% methanol in 30 min, 0.5% acetic acid, flow 20 mL/min) to give 0.46 mg of the R- and 0.53 mg of the S-PGME derivatives.

**Crocagin A-(S)-PGME derivative (1a):** colorless, amorphous powder,  $^1\text{H}$  (600 MHz, DMSO- $d_6$ ):  $\delta$  7.24 (N4-H), 7.21 (H7), 6.64 (H9), 6.51 (H8), 5.99 (H5), 5.37 (H3), 5.32 (H10), 5.20 (H14), 5.11 (H15), 4.61 (H2), 3.90 (H4); HRESIMS  $m/z$  699.3133 ( $[\text{M}+\text{H}]^+$ , calcd. for  $\text{C}_{37}\text{H}_{43}\text{N}_6\text{O}_8$ , 699.3137).

**Crocagin A-(R)-PGME derivative (1b):** colorless, amorphous powder,  $^1\text{H}$  (600 MHz, DMSO- $d_6$ ):  $\delta$  7.28 (N4-H), 7.08 (H7), 6.62 (H9), 6.45 (H8), 5.98 (H5), 5.25 (H3), 5.33 (H10), 5.21 (H14), 5.13 (H15), 4.63 (H2), 3.88 (H4); HRESIMS  $m/z$  699.3133 ( $[\text{M}+\text{H}]^+$ , calcd. for  $\text{C}_{37}\text{H}_{43}\text{N}_6\text{O}_8$ , 699.3137).

**Advanced Marfey's Analysis of Amino Acids:** **1** was hydrolyzed in 6N HCl at 90  $^\circ\text{C}$  for 17 h. The hydrolysate was evaporated to dryness and dissolved in  $\text{H}_2\text{O}$  (100  $\mu\text{L}$ ). 1 N  $\text{NaHCO}_3$  (20  $\mu\text{L}$ ) and 1% 1-fluoro-2,4-dinitrophenyl-5-L-leucinamide (100  $\mu\text{L}$  in acetone) was added and the mixture was heated at 40  $^\circ\text{C}$  for 40 min. After cooling to room temperature, the solution was neutralized with 2 N HCl (20  $\mu\text{L}$ ) and evaporated to dryness. The residue was dissolved in MeOH and analyzed by LC-MS using a linear gradient from 15-60% aqueous  $\text{CH}_3\text{CN}$  as mobile phase. As a standard, L-NMelle was converted with L-FDLA and D-FDLA as described above. Retention times of L-FDLA derivatized hydrolysate was NMelle 17.2 min; retention times of the authentic standards were L-NMelle-L-FDLA 17.3, L-NMelle-D-FDLA 20.2  $m/z$  438.199  $[\text{M}-\text{H}]^-$ .



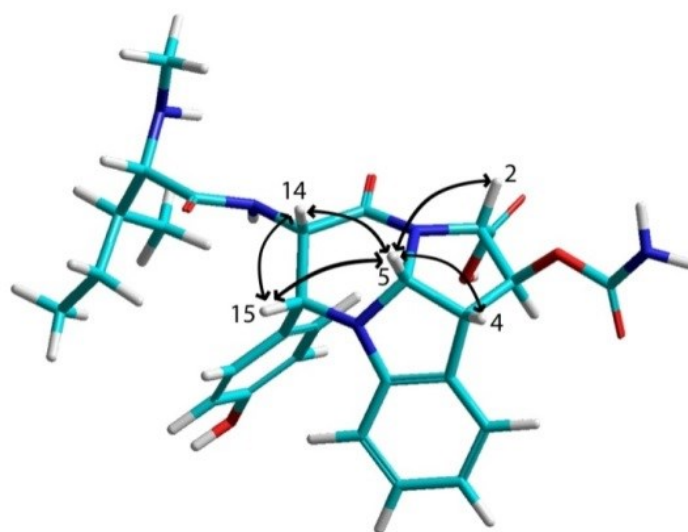
**SI-1: MS and NMR data of Crocagins**

**Table S1.** MS-MS fragmentation of **1**. <sup>a</sup> For convenience the following abbreviations are used for the fragments: A = [M+H]<sup>+</sup>; Mlle = C<sub>7</sub>H<sub>13</sub>NO, N-methylisoleucine; Carb = CH<sub>3</sub>NO<sub>2</sub>, Carbamate; Phe = C<sub>6</sub>H<sub>5</sub>O, Phenyl.

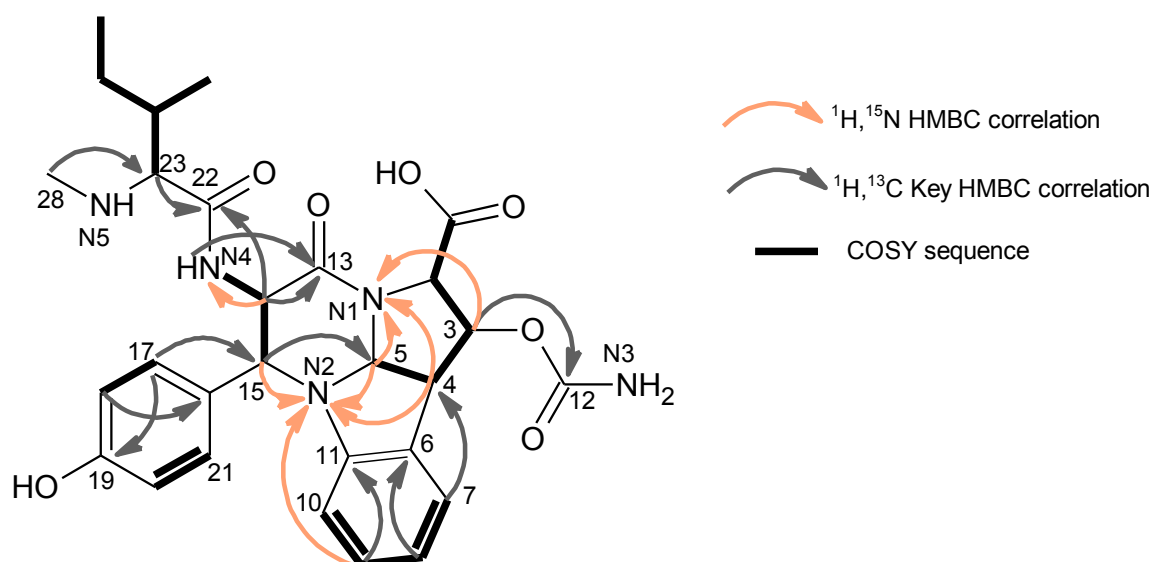
<i>m/z</i>	fragment <sup>a</sup>
552.245	[M+H] <sup>+</sup> = A
491.229	A - Carb
447.239	A - Carb - CO <sub>2</sub>
425.146	A - Mlle
364.129	A - Mlle - Carb
331.104	A - Mlle - Phe
320.139	A - Carb - CO <sub>2</sub> - Mlle
307.108	A - Mlle - C <sub>3</sub> H <sub>6</sub> O <sub>3</sub> N <sub>2</sub>
270.087	A - Mlle - Phe - Carb

**Crocagin isolation and structure elucidation:** Crocagin A (**1**) was detected in extracts of Cm c5 by HPLC-UV-HRESIMS as a mass and UV signal peak with an UV absorption maximum at 280 nm. From the corresponding [M+H]<sup>+</sup> ion *m/z* 552.2470 the molecular formula C<sub>28</sub>H<sub>33</sub>N<sub>5</sub>O<sub>7</sub> and thereby 15 degrees of unsaturation were calculated. Isolation from culture extracts of strain Cm c5 by preparative reversed phase HPLC provided **1** as white, amorphous powder. In the NMR spectra some very broad aromatic <sup>1</sup>H and <sup>13</sup>C signals were barely detectable at 298 K, but measurement at 333 K provided sharper signals that allowed detecting their correlations and concurrently suggested a steric hindrance of a bond rotation within the molecule. Combined interpretation of <sup>1</sup>D <sup>1</sup>H and <sup>13</sup>C NMR spectra and a <sup>1</sup>H, <sup>13</sup>C HSQC NMR spectrum enabled the identification of eight aromatic as well as eight aliphatic methines, one methylene, three methyls, four quaternary aromatic and four carboxylcarbons, and two exchangeable protons (**Table S1**). From <sup>1</sup>H, <sup>1</sup>H COSY and TOCSY spectra five proton spin systems were identified, which together with inner residue HMBC correlations provided the partial structures of N-methylisoleucine (NMlle), a 3-substituted tyrosine and a tryptophan derived moiety. A carbamate group in **1** was insinuated by the loss of carbamic acid in the MS/MS fragmentation (**Table S1**) and by the hydrolytic removal of a carbamoyl residue upon treatment with NaOH (see SI-4). The <sup>1</sup>H, <sup>13</sup>C HMBC correlation between H3 and the carbamate carbonyl C12 located the carbamate at oxymethine carbon C3, thus explaining the acyl shift of H3 ( $\delta_H$  5.33). Key <sup>1</sup>H, <sup>13</sup>C HMBC correlations connecting the structural units were observed for NH-4 and H14 to carbonyl C22, and between H15 and C5. For the identification of further bonds between the substructures a <sup>1</sup>H, <sup>15</sup>N-HMBC experiment proved to be conclusive. Both H15 and H5 strongly correlate to nitrogen atom N2 (**Figure S1, S2**), thus proving the C15/N2 bond as connection between these residues. Conveniently, the chemical shift of N2 at  $d_N$  87.6 characterized an aryl amine, which was further indicated by a weak correlation with H9. Another strong correlation of H5 was observed with the amide nitrogen atom N1 ( $d_N$  142.2) indicating a direct C5/N1 bond. The lack of further residues-connecting HMBC correlations seriously hampered assigning the last missing ring

closure required from the number of un saturations. Derivatization of **1** provided the essential information: Firstly, reaction with diazomethane provided the mono methyl ester **3** as main product (**Figure S5**). The  $^1\text{H}$ ,  $^{13}\text{C}$  HMBC spectrum of **3** indicated the additional methyl group ( $d_{\text{H,C}}$  3.59, 52.4) to be connected to C1 ( $d_{\text{C}}$  168.5). Thus, carboxyl group C1 was a free carboxylic acid in **1**. Secondly, acetylation with acetanhydride in pyridine provided the diacetylated derivative **4**. The  $^1\text{H}$  and  $^{13}\text{C}$  shifts revealed the acylation of N5 and of the phenol at C19 in **4**. Consequently, the remaining open bonds at carbonyl C13 and the tertiary amide N1 were connected to give another six-membered ring thus finally accounting for the last remaining calculated unsaturation. The proximity of the bulky substituents at the tetrahydropyrimidinone ring in **1** explains the hindered rotation of the phenol ring, causing the signal broadening in the NMR spectra at room temperature as well as the unusual low chemical shift of the aromatic proton H10 ( $d_{\text{H}}$  5.35) by the shielding effect of the neighboring phenol ring. The relative configuration of the stereo centers C23/C24 was assigned by  $^1\text{H}$ ,  $^1\text{H}$ ROESY correlations and *J*-based configuration analysis (**Figure S3**) whereas the advanced Marfey's and the phenylglycine methyl ester (PGME, **Figure S4**) methods demonstrated the absolute configuration of C23 and C2, respectively, thus finalizing the stereo chemical assignment as 2*S*,3*S*,4*R*,5*R*,14*S*,15*S*,23*S*,24*S*. Crocagin B (**2**) was detected in minor amounts in extracts of Cm c5. HR-ESI-MS data provided the molecular formula as  $\text{C}_{27}\text{H}_{33}\text{N}_5\text{O}_7$ , two protons less compared to **1**. The UV spectrum of **2** with absorption maxima of 290 and 346 nm indicated an enlarged  $\pi$  system. The NMR data of **2** were highly similar to **1**. The key difference was the replacement of methines C14 and C15 by quaternary olefinic carbons, providing the  $\Delta_{14,15}$  double bond. A further minor metabolite of **1** was observed by HPLC-UV-HRESIMS in crude extracts of Cm c5. Here, the MS data indicated the loss of the N-methyl group at the Ile residue.



**Figure S1.** Model of **1** calculated by PM3 with HyperChem. Key ROESY correlations used to establish the relative stereochemistry of the core of **1** are shown as black arrows.



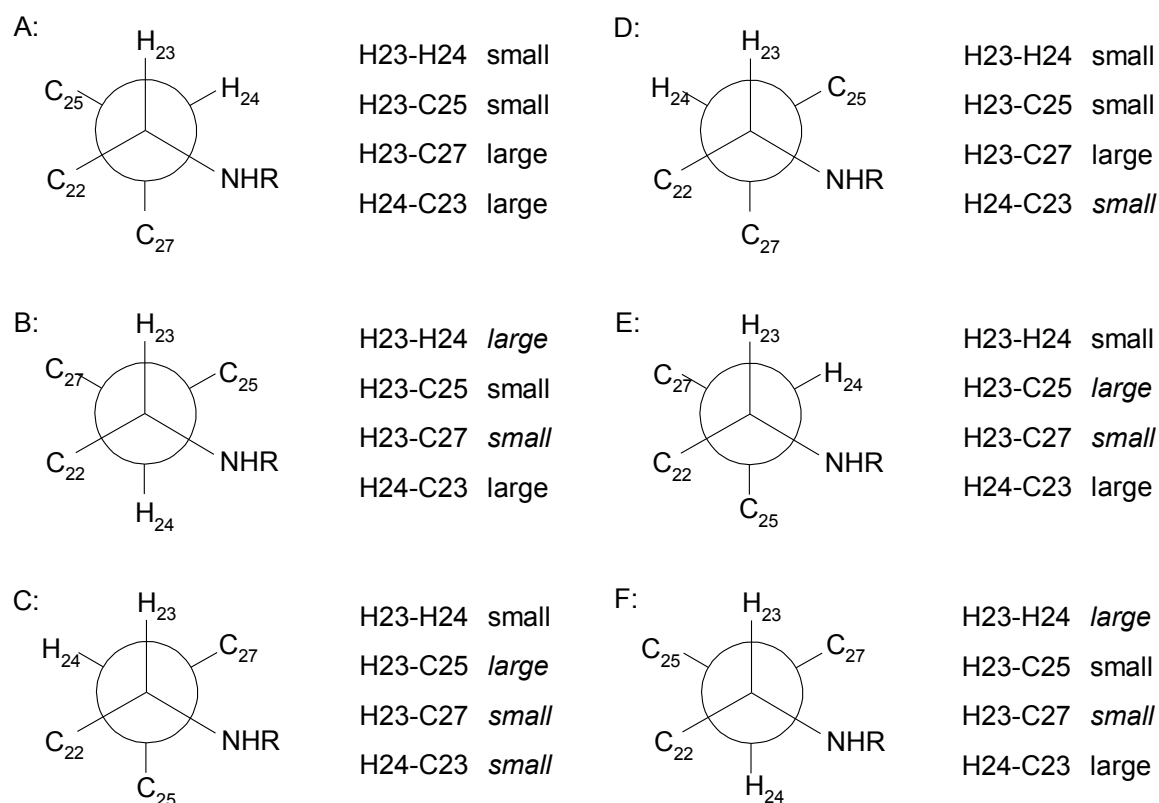
**Figure S2:** Key  $^1\text{H}, ^1\text{H}$  COSY,  $^1\text{H}, ^{13}\text{C}$  HMBC and  $^1\text{H}, ^{15}\text{N}$  HMBC correlations for Crocagin A (1).

**Table S2:** NMR data ( $^1\text{H}$  700 MHz,  $^{13}\text{C}$  175 MHz) of crocagin A (1) in  $\text{DMSO}-d_6$ .

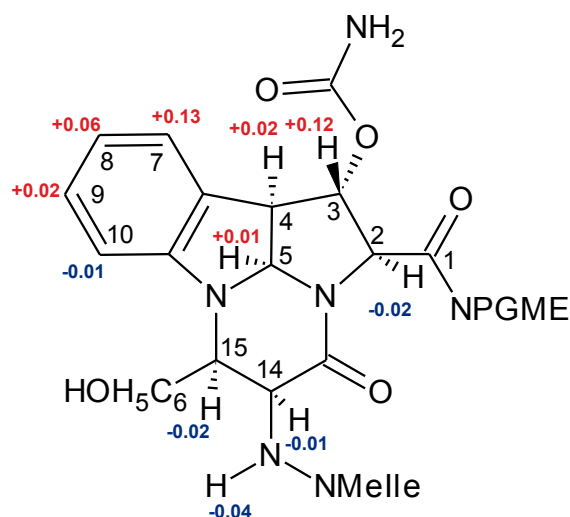
Atom #	$\delta_{\text{C}}$ , mult.	$\delta_{\text{H}}$ , mult.	COSY	NOESY	HMBC
1	168.9, C				
2	64.0, CH	4.25, d (3.9)	3	3	1, 3
3	81.7, CH	5.38, m	2, 4	2, 4, 7	1, 5, 12
4	51.7, CH	3.96, br d (8.3)	3, 5	3, 5, 7	11, 6, 3, 2
5	75.8, CH	5.98, d (8.3)	4	4, 13	11, 6, 3, 2
6	126.0, C				
7	124.5, CH	7.24, br d (7.7)	8	4, 3, 8	11, 9
8	117.4, CH	6.54, t, (7.7)	9, 7	7	10, 6
9	127.2, CH	6.68, t, (7.7)	10, 8	10	11, 7
10	110.6, CH	5.39, d (7.7)	9	9	11, 6, 8
11	147.7, C				
12	155.2, C				
13	164.7, C				
14	54.1, CH	5.32, dd (7.2, 6.1)	N4	5	14, 13, 22, 16
15	59.3, CH	5.27, br d (7.2)		17/21	17/21, 5, 13, 14
16	124.2, C				
19	157.0, C				
18/ 20	114.7, CH	6.57, m	17/21	17/21	
17/21	130.9, CH	7.11, br d (8.3)	18/20	15, 18/20, 28	15, 19
22	166.6, C				
23	64.6, CH	3.71, m	24		22, 24, 27
24	35.7, CH	1.78, m	23, 25, 27	26, 27	22, 25
25	24.2, $\text{CH}_2$	1.51, ddq (14.0, 11.0, 7.5) 1.11, dqd (14.5, 7.5, 6.5)	25, 26 24, 25, 26	25, 26, 27 25	24, 26, 27 23, 24, 26, 27
26	10.7, $\text{CH}_3$	0.83, t (7.5)	25	24, 25	24, 25
27	14.0, $\text{CH}_3$	0.89, d (7.0)	24	24, 25	23, 24, 25
28	32.2, $\text{CH}_3$	2.38, s		17/21	23
19OH		9.26, br s			18/20
N4-H		7.91, br d (6.1)	14		13, 22

**Table S3:** NMR data ( $^1\text{H}$  700 MHz,  $^{13}\text{C}$  175 MHz) of crocagin B (**2**) in  $\text{MeOH-}d_4$ .

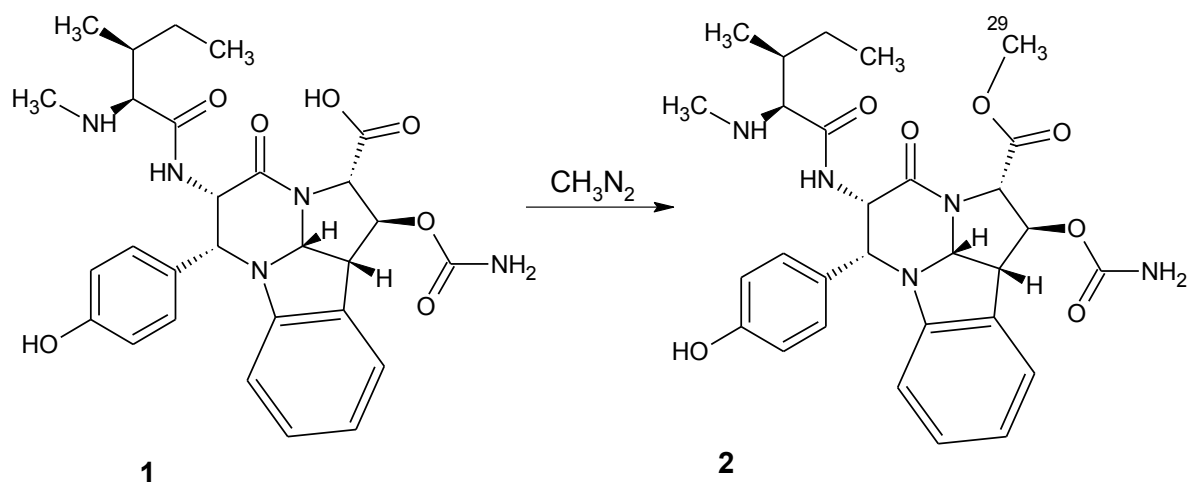
Atom #	$\delta_{\text{C}}$ , mult.	$\delta_{\text{H}}$ , mult.	COSY	HMBC
1	170.3, C			
2	67.1, CH	4.68, s		1, 3, 4, 5
3	82.7, CH	5.58, s		5, 12, 1, 2, 6
4	53.0, CH	4.31, d (7.2)	5	6, 11, 3, 7, 2
5	80.4, CH	6.09, d (7.2)	4	2, 3, 11, 6
6	131.1, C			
7	126.8, CH	7.39, m	8	11, 4, 9
8	123.9, CH	6.87, m	7	6
9	130.2, CH	6.86, m	10	7, 10
10	114.1, CH	5.72, m	9	6, 8
11	147.3, C			
12	158.4, C			
13	164.1, C			
14	111.6, C			
15	152.0, C			
16	124.4, C			
19	162.1, C			
20, 18	117.1, CH	6.92, d (8.6)	21, 17	16, 19
21, 17	133.1, CH	7.56, d (8.6)	20, 18	20, 18, 19, 15
22	167.3, C			
23	67.9, CH	3.66, d (5.16)	24	22, 24, 27, 25, 28
24	38.2, CH	1.93, m	27, 23, 25	
25	26.3, $\text{CH}_2$	1.67, m	26, 25	27, 23, 26, 24
25	26.3, $\text{CH}_2$	1.30, m	26, 25, 24	23, 24, 27, 26
26	11.9, $\text{CH}_3$	1.00, t (7.3)	25, 25	24, 25
27	14.9, $\text{CH}_3$	1.05, d (6.9)	24	25, 23, 24
28	33.0, $\text{CH}_3$	2.21, s		23



**Figure S3.** *J*-based analysis of six hypothetical rotamers with 23*S*\*,24*S*\* (A – C) and 23*S*\*,24*R*\* (D – F) configuration to determine the relative stereochemistry of the *N*-methylisoleucin moiety in crocagin A (**1**). Expected couplings contrary (shown in *italic*) to the observed ones ( $^3J_{\text{H}23,\text{H}24} = 4.0$  Hz, small;  $^2J_{\text{H}23,\text{C}25} = 2.5$  Hz, small;  $^3J_{\text{H}23,\text{C}27} = 4.8$  Hz, large;  $^2J_{\text{H}24,\text{C}23} = 5.9$  Hz, large) exclude all conformations except A.

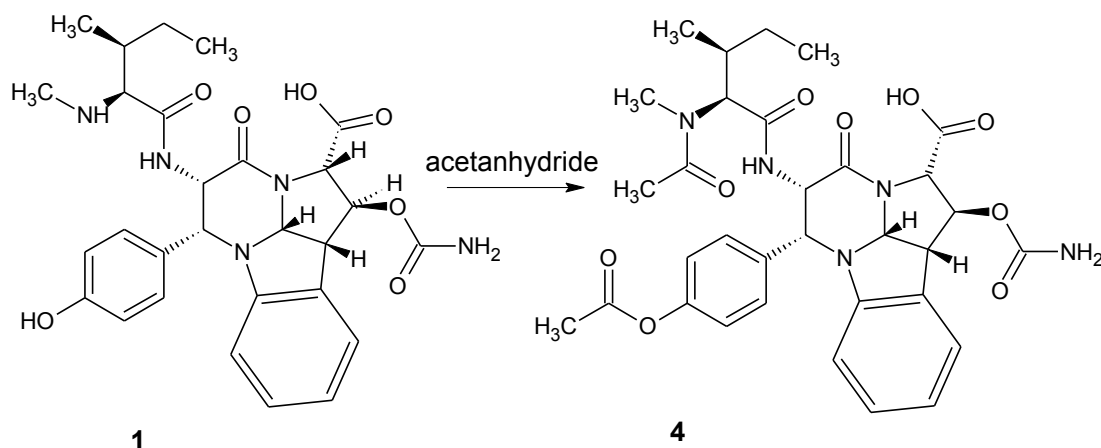


**Figure S4:** Determination of the absolute configuration of **1** using the PGME method.  $\Delta\delta = \delta(\text{S-PGME amide}) - \delta(\text{R-PGME amide})$ .

SI-2: -Preparation of the methyl ester of crocagin (**3**)**Figure S5:** Preparation of the methyl ester of crocagin (**3**)

Diazomethane (1.7 mmol) in ether (6 mL), which was freshly produced in a Sigma-Aldrich Diazomethane Generator according to the manufacturer's instructions, was added stepwise to **1** (6.0 mg, 10.9  $\mu\text{mol}$ ) in aqueous methanol (50 %, 1 mL). The reaction was monitored by analytical HPLC and stopped by removing the solvent in vacuo after all starting material had been converted. Preparative RP HPLC (column 250 x 21 mm, VP Nucleodur C18 Gravity 5  $\mu\text{m}$ , gradient 28 % to 100 % methanol in 25 min, 0.5 % acetic acid, flow 15 mL/min, UV peak detection at 220 nm) provided 2.1 mg (34 % yield) of the methyl ester of crocagin (**2**).

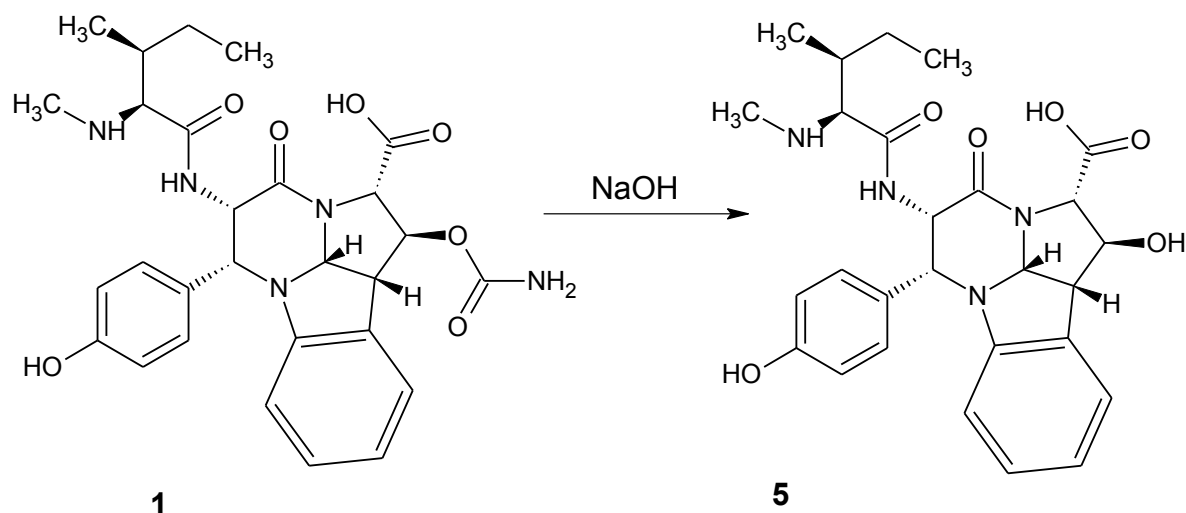
HR-ESI-MS  $m/z$  564.2461  $[\text{M} - \text{H}]^-$  (calcd for  $\text{C}_{29}\text{H}_{34}\text{N}_5\text{O}_7$ , 564.2464);  $^1\text{H}$  NMR ( $\text{DMSO}-d_6$ ), 7.36 (NH, d, 7.3), 7.21 (d, 7.3 Hz, H7), 6.99 (m, H17, H21), 6.71 (t,  $J = 7.3$ , H9), 6.57 (br d,  $J = 7.3$ , H18, H20), 6.54 (t,  $J = 7.3$ , H8), 5.97 (m, H5), 5.44 (m, H10), 5.26 (m, H3), 5.22 (m, H14), 5.13 (m, H15), 4.28 (d,  $J = 5.1$ , H2), 3.97 (dd,  $J = 8.4, 3.3$ , H4), 3.59 (s,  $\text{H}_{329}$ ), 2.65 (d,  $J = 5.1$ , H23), 2.13 (s,  $\text{H}_{328}$ ), 1.46 (m, H24), 1.26 (ddd,  $J = 13.8, 7.5, 4.4$ , H25a), 0.97 (ddd,  $J = 13.8, 9.2, 7.5$ , H25b), 0.78 (t,  $J = 7.5$ ,  $\text{H}_{326}$ ), 0.75 (d,  $J = 7.3$ ,  $\text{H}_{327}$ );  $^{13}\text{C}$  ( $\text{DMSO}-d_6$ ), 173.0 (C-22, qC), 168.5 (C-1, qC), 165.8 (C-13, qC), 157.0 (C-19, qC), 155.5 (C-12, qC), 147.8 (C-11, qC), 136.6 (C-17, C-21, CH), 128.9 (C-16, qC), 127.7 (C-9, CH), 126.5 (C-6, qC), 124.6 (C-7, CH), 117.7 (C-8, CH), 114.8 (C-18, C-20, CH), 110.7 (C-10, CH), 81.9 (C-3, CH), 75.7 (C-5, CH), 69.3 (C-23, CH), 63.6 (C-2, CH), 59.7 (C-15, CH), 53.2 (C-14, CH), 52.4 (C-29,  $\text{CH}_3$ ), 51.3 (C-4, CH), 37.5 (C-24, CH), 35.4 (C-28,  $\text{CH}_3$ ), 24.4 (C-25,  $\text{CH}_2$ ), 15.6 (C-27,  $\text{CH}_3$ ), 11.6 (C-26,  $\text{CH}_3$ ).

SI-3: Preparation of diacetyl derivative of crocagin (**4**)**Figure S6:** Preparation of the diacetyl derivative of crocagin (**4**)

**1** (11.2 mg; 20.3  $\mu\text{mol}$ ) was dissolved in 4 mL of dry pyridine. After the addition of acetic acid anhydride (2 mL) the mixture was stirred for 16 h at 25  $^{\circ}\text{C}$ . Water (2 mL) was added and the solvent was concentrated in vacuo to about 2 mL. By preparative HPLC (column 250 x 21 mm, VP Nucleodur C18 Gravity 5  $\mu\text{m}$ , gradient 28 % to 100 % methanol in 30 min, 0.5 % acetic acid, flow 20 mL/min) 0.8 mg of diacetyl derivative **3** were isolated.

$^1\text{H}$  NMR ( $\text{DMSO}-d_6$ )  $\delta$  = 7.99 (d,  $J$  = 5.9, NH), 7.15 (m, H7), 7.00 (d,  $J$  = 8.1, H17, H21), 6.77 (m, H9), 6.65 (m, H18, H20), 6.56 (m, H8), 6.16 (m, H5), 5.58 (d,  $J$  = 8.1, H10), 5.38 (d,  $J$  = 7.3, H15), 5.33 (m, H14), 5.31 (m, H3), 4.63 (d, 11.0, H23), 4.01 (m, H4), 2.91 (s, H<sub>3</sub>28), 2.23 (s, H<sub>3</sub>Ac), 2.00-1.90 (m, H24), 1.90 (s, H<sub>3</sub>Ac'), 1.25 (m, H25a), 0.90 (m, H25b), 0.85 (m, H<sub>3</sub>26), 0.84 (m, H<sub>3</sub>27); HR-ESI-MS  $m/z$  636.2657 [ $\text{M} + \text{H}$ ] $^{+}$  (calcd for  $\text{C}_{32}\text{H}_{38}\text{N}_5\text{O}_9$ , 636.2664).

## SI-4: Preparation of decarbamoylderivative of crocagin (5)

**Figure S7:** Preparation of decarbamoylderivative of crocagin (5)

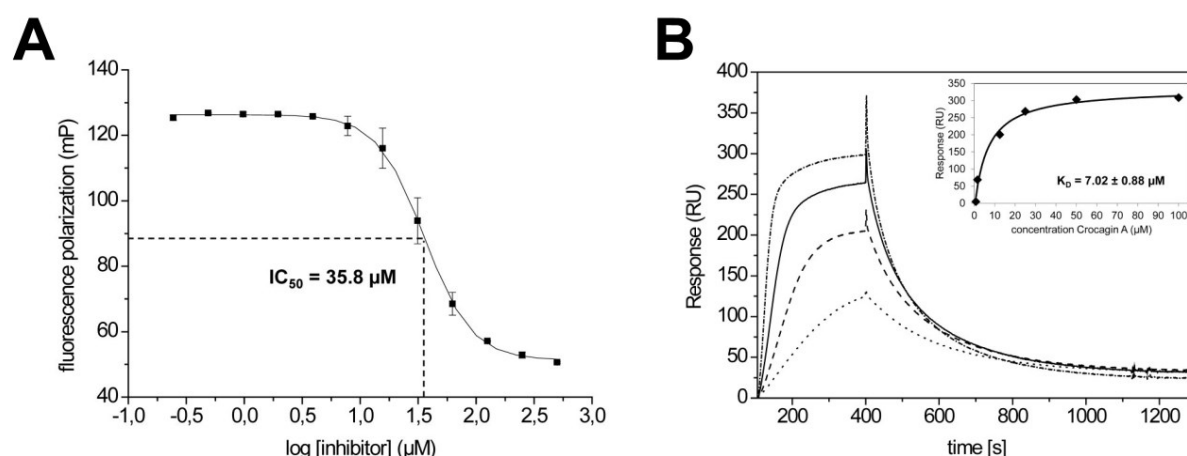
**1** (12.8 mg, 23  $\mu$ mol) was incubated in 1 mL 0.01 M NaOH for 16 h at 25 °C. Analytical HPLC revealed only a slow reaction progress. Therefore the mixture was adjusted to pH 12.5 with 2 M NaOH and incubated for 5 h. Because the conversion was still incomplete, the mixture was adjusted to pH 13 and incubated for 16 h, ensuing partial decomposition. 7/10 of the crude material was then treated with diazomethane in ether as described above. 0.8 mg of unmethylated decarbamoyl derivative **4** were recovered by preparative HPLC [column 250 x 21 mm, VP Nucleodur C18 Gravity 5  $\mu$ m (Macherey Nagel), gradient 28 % to 100 % methanol in 30 min, 0.5 % acetic acid, flow 20 mL/min].

$^1\text{H}$  NMR (DMSO- $d_6$ ):  $\delta$  = 7.13 (d,  $J$  = 6.6 Hz, H7), 7.10 (m, H17, H21), 6.66 (m, H9), 6.52 (m, H18, H20), 6.49 (m, H8), 6.23 (d,  $J$  = 5.1 Hz, 3OH), 5.93 (d,  $J$  = 8.1 Hz, H5), 5.37 (d,  $J$  = 8.1 Hz, H10), 5.23 (m, H14), 5.13 (m, H15), 4.48 (m, H3), 4.06 (d,  $J$  = 5.1 Hz, H2), 3.73 (dd,  $J$  = 8.1, 3.7 Hz, H4), 3.65 (d,  $J$  = 5.1 Hz, H23), 2.29 (s, H<sub>3</sub>28), 1.64 (m, H24), 1.40 (m, H25a), 1.03 (m, H25b), 0.80 (m, H<sub>3</sub>26), 0.79 (m, H<sub>3</sub>27). HR-ESI-MS  $m/z$  509.2401 [ $M + H$ ] $^+$  (calcd. for C<sub>27</sub>H<sub>33</sub>N<sub>4</sub>O<sub>6</sub>, 509.2395).



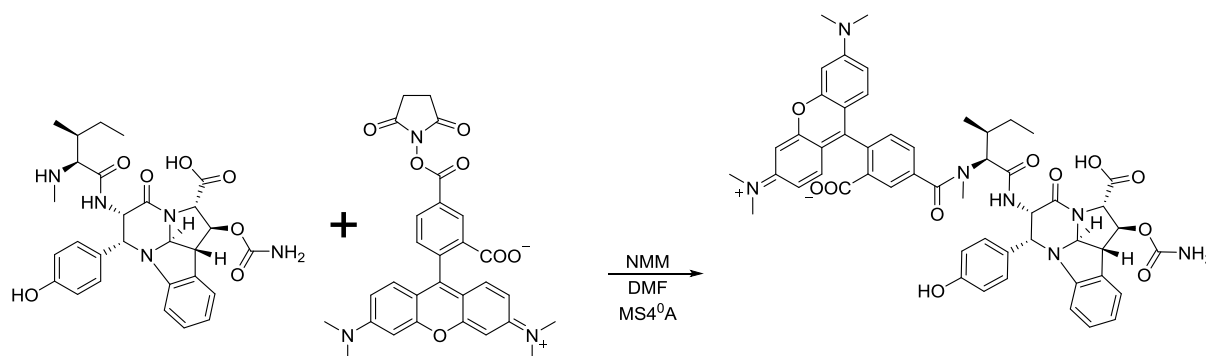
## SI-5: CsrA inhibition assay

Crocagin A showed no antimicrobial or cytotoxic activities and was included into the DZIF (German Centre for Infection Research) myxobacterial natural product library, which was screened for inhibitors of the interaction between carbon storage regulator protein A (CsrA) and its cognate RNA target.<sup>[9]</sup> CsrA binds to the 5' untranslated region of mRNA molecules and by this affects their translation and stability. This conserved protein is found in many bacteria and is a promising target for anti-infective drug development as it is involved in the regulation of several virulence factors in Gram-negative pathogens such as *Helicobacter pylori* and *Salmonella thyphimurium*. Micromolar concentrations of **1** competitively inhibit the interaction of CsrA with an RNA ligand containing the conserved ANGGA(N) core binding motif in a dose-dependent manner (**Figure S8A**). The direct binding of **1** to CsrA was then confirmed by surface plasmon resonance (SPR) affinity experiments (**Figure S8B**). The responses at equilibrium were fitted to a steady-state affinity model yielding a dissociation constant ( $K_D$ ) of  $12.9 \pm 5.8 \mu\text{M}$  determined in three independent experiments.



**Figure S8A:** Dose-response curve of crocagin A (**1**) in a fluorescence polarization (FP)-based competition assay with  $0.4 \mu\text{M}$  of CsrA and  $15 \text{ nM}$  of fluorescein-tagged RNA<sub>B</sub>\* as probe. The representative curve of one independent experiment is depicted with data points representing averaged FP values of duplicates  $\pm$  standard deviation. Curves were fit to a four-parameter dose-response model to calculate an  $\text{IC}_{50}$  value of  $35.8 \mu\text{M}$ . **Figure S8B:** Surface plasmon resonance (SPR) analyses revealed that crocagin A (**1**) binds to CsrA in a dose-dependent manner and equilibrium responses were fitted to a steady-state affinity model yielding a  $K_D$  value of  $12.9 \pm 5.8 \mu\text{M}$ . Representative referenced sensorgrams of one independent experiment and its corresponding affinity plot are depicted.

## SI-6: Preparation and purification of Rhodamine-tagged crocagin A and testing bacterial cell entry

**Figure S9:** Preparation and purification of Rhodamine-tagged Crocagin A

To a solution of Crocagin A (**1**) (2 mg, 0.0036 mmol) in anhydrous DMF (0.5 mL) with molecular sieves 4 Å, a solution of 5-carboxy-tetramethylrhodamine N-succinimidyl ester (3 mg, 0.0057 mmol) was added in anhydrous DMF (0.5 mL) followed by N-methylmorpholine (2.5 µL, 0.144 mmol). The mixture was stirred for 6 days at room temperature and monitored by LC-MS. The mixture was then centrifuged and the 'pinkish' supernatant was directly subjected to preparative RP-HPLC without further workup. Lyophilisation of the fraction containing the desired product yielded 0.5 mg of a dark pink powder. HRMS (ESI, +ve) calculated for  $C_{53}H_{53}N_7O_{11}$   $m/z$   $[M+H]^+$ : 964.3876; found: 964.3828 and  $[M+2H]^+$ : 482.6974; found: 482.6963.

HPLC conditions: the reaction mixture ( $\cong$  1 mL) was purified in 50 µL portions.

Column: Phenomenex Luna® 5 µM, C18(2), 100 Å, 250 mm x 10 mm;

Mobile phase (flow 7 mL/min): A: water with 0.1% formic acid, B: acetonitrile with 0.1% formic acid.

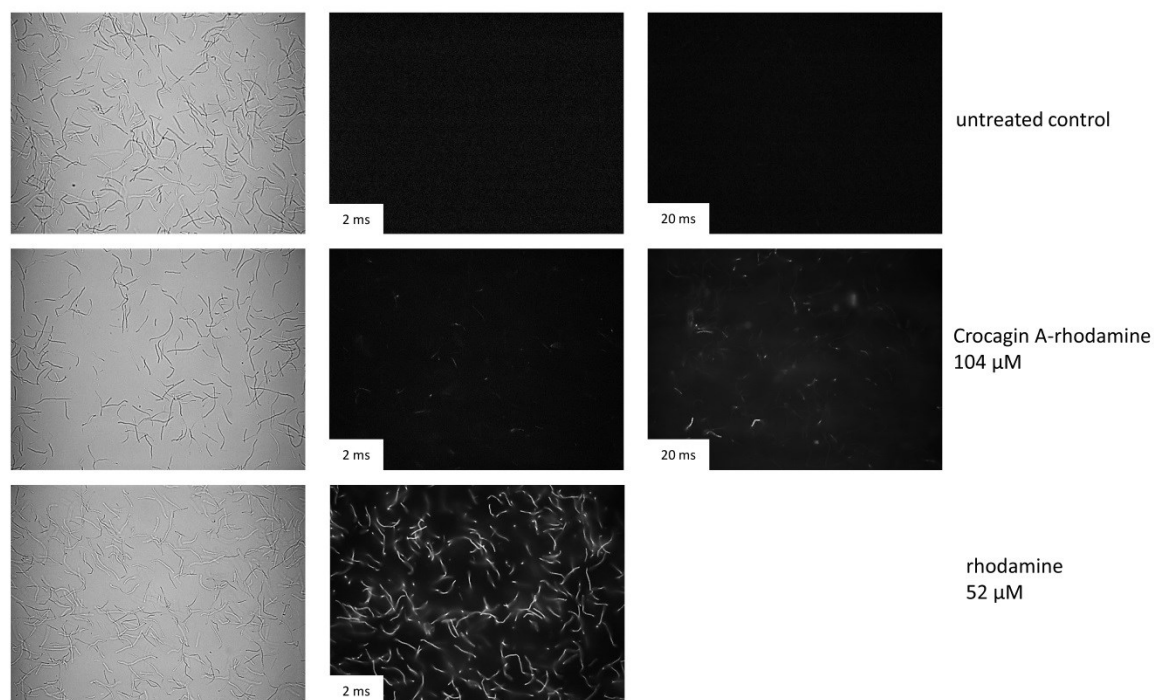
Gradient: 1% B to 60% B in 40 min, followed by 60% B to 95% B in 1 min then 95% B for 3 min, then 95% B to 1% B in 1 min, followed by 1% B for 3 min.

Temperature: 40 °C (column oven)

Detection: UV (557 nm), PDA

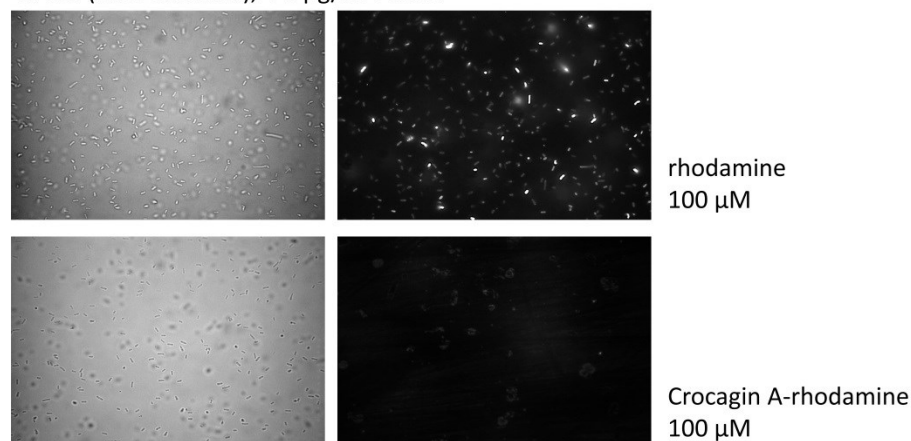
Cell entry of rhodamine-tagged crocagin A was tested in gram-positive and gram negative bacterial cells by fluorescence microscopy of treated cultures. As shown in **Figures S10** and **S11**, only cells incubated with free rhodamine show fluorescence while cultures treated with rhodamine-tagged Crocagin A show very little to no fluorescent cells. Additionally, cells and supernatant of *E. coli* TolC- cultures treated with either Crocagin A (**1**) or free rhodamine were extracted separately and subjected to LCMS to compare incorporation into the cells, as shown in **Figure S12** and **S13**. Only free rhodamine is taken up by the cells.

***B. megaterium* DSM-32**  
 30 min, 37 °C  
 (40x objective; BF/F-Ph)

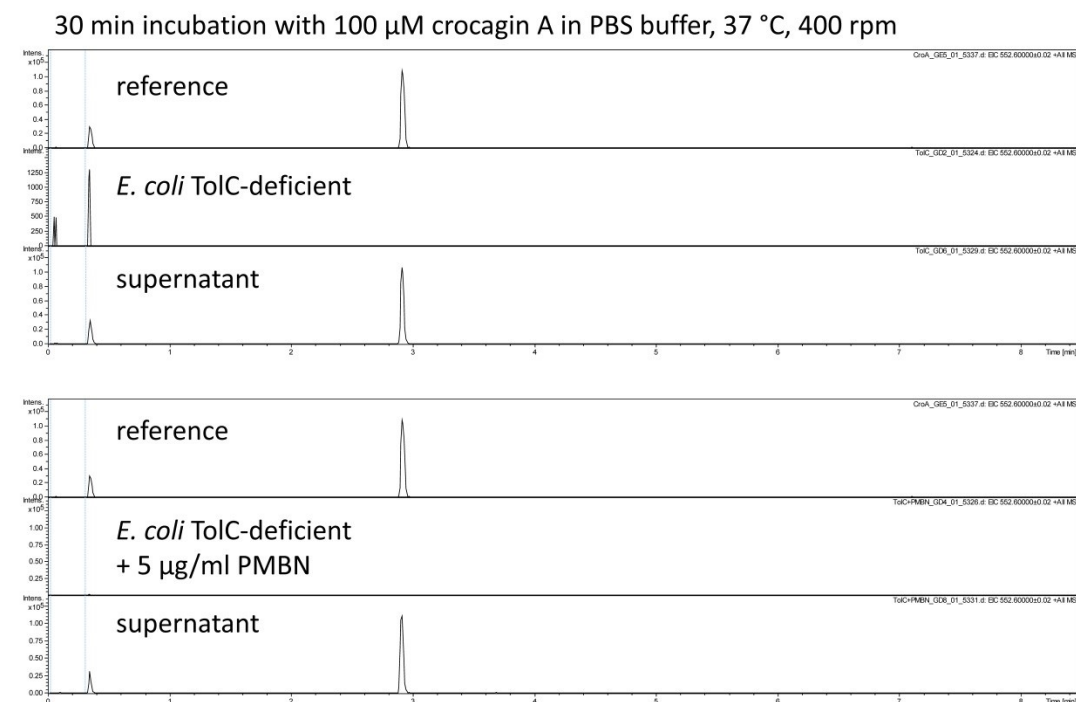


**Figure S10:** Tests for bacterial cell entry of rhodamine-tagged Crocagin A into the gram-positive bacterium *Bacillus megaterium* DSM-32.

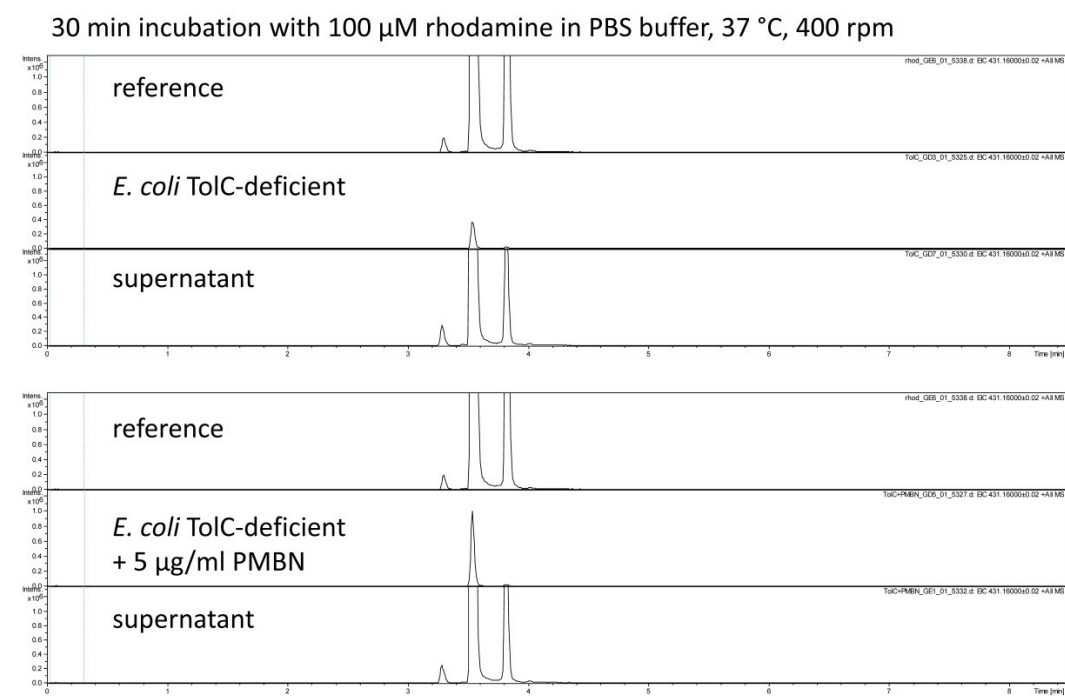
*E. coli* (TolC-deficient), + 3  $\mu\text{g/ml}$  PMBN



**Figure S11:** Tests for bacterial cell entry of rhodamine-tagged Crocagin A the gram-negative bacterium *Escherichia coli* (TolC-deficient mutant).



**Figure S12:** Tests for bacterial uptake of Crocagin A (**1**) by TolC-deficient *E. coli*. Under the conditions tested, **1** is not detected in cell extracts.

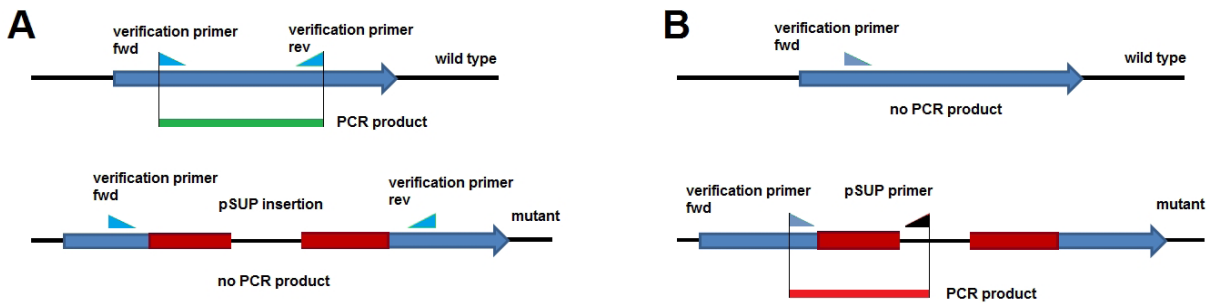


**Figure S13:** Tests for bacterial uptake of free rhodamine by TolC-deficient *E. coli*. In contrast to **1**, Rhodamine is detected in cell extracts.

SI-6: Cm c5 Mutagenesis and Genetic Verification

Derivatives of the plasmid pSUP\_hyg were used for mutagenesis of Cm c5, each containing a homologous fragment for integration into the gene of interest in the Cm c5 wild type. Plasmid integration at the desired locus results via single crossing-over results in target gene disruption by the plasmid backbone, which confers resistance to hygromycin.

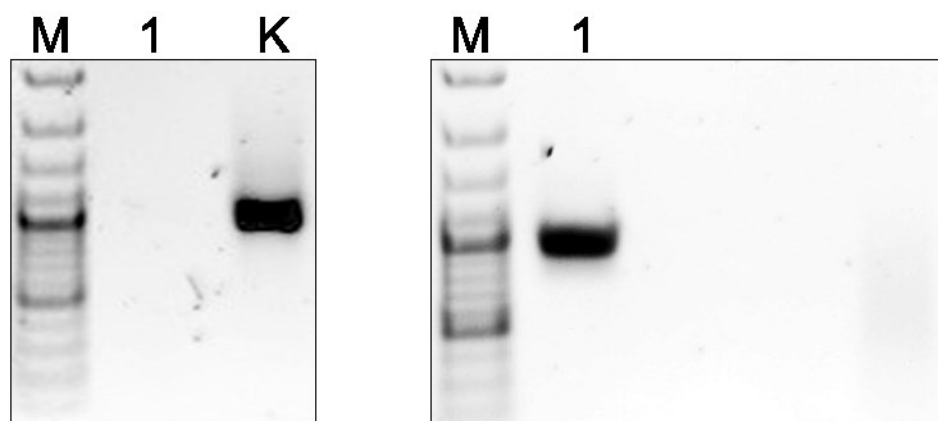
To verify the genotype of an isolated mutant, gDNA of mutant cells grown in liquid culture was extracted according to standard protocols and plasmid integration was tested with PCR reactions, testing for presence of the intact gene as shown in **Fig. S14A** and the integration of the plasmid backbone at the desired gene locus, as shown in **Fig S14B**. While the first PCR must only yield a product in the wild type cells or other mutants, the second PCR must yield a product only in the respective mutant. Oligonucleotide Primer sequences for PCR of insert fragments for target gene inactivation and mutant verification are shown in **Table S3**. The agarose gels showing the results of mutant verification PCR are shown in **Figures S15 and S16**.



**Figure S14:** Genetic verification of Cm c5 mutants by PCR. A: PCR with primers that bind outside the plasmid integration site on the gene of interest. The mutants must not yield a product in this reaction. B: The PCR is performed with one primer binding on the plasmid sequence. Only the mutant DNA template must yield a product in this reaction.

**Table S3:** Oligonucleotide Primers for PCR of homologous fragments of *cgn* genes

	Primer name	Sequence 5'-3'
Inactivation of <i>cgnI</i>	552carb_for	TGACCTTATGGCGAGCCCACCTTC
	552carb_rev	TTCCAGTCGTTACACGTCCGC
Inactivation of <i>cgnB</i>	552hypo_for	TGATTGATGCAGGCCTGGGAGAC
	552hypo_rev	TCACTCCTTGCCGACGATCAGCTC
Verification of <i>cgnI</i> mutants	552carb_veri_for	AGTGCCTTCTGGGCGCTCTT
	552carb_veri_rev	AAGCTGTGGACCGAGCGGAA
Verification of <i>cgnB</i> mutants	552hypo_veri_for	TGCGTACTGCCCCCTTCGTT
	552hypo_veri_rev	TCTTTCCCCATCCCCACGCC
Primers binding pSUP_hyg	psup_hyg_fwd	ATGTAGCACCTGAAGTCAGC
	psup_hyg_rev	GCATATAGCGCTAGCAGC



**Figure S15:** Genetic verification of Cm c5 *cgnB*- mutants. **Left:** PCR to test for *cgnB* disruption. DNA template of the wild type yields the PCR product, indicating intact *cgnB* (lane K). DNA from mutant *cgnB*- yields no product (lane 1). **Right:** PCR to verify pSUP\_*hyg* integration into *cgnB*. The mutant DNA template yields the expected product (lane 1).



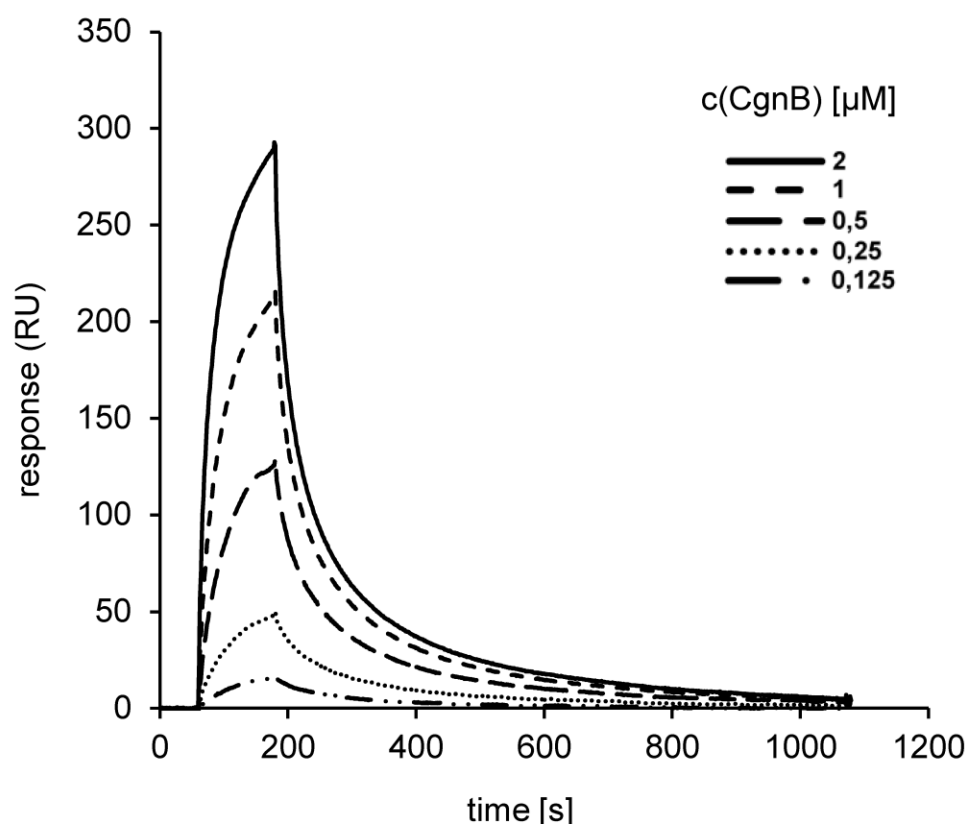
**Figure S16:** Genetic verification of Cm c5 *cgnI*- mutants. **Left:** PCR to verify gene *cgnI* disruption. DNA template of the wild type yields the PCR product (lane K), indicating intact *cgnI*. DNA from *cgnI*- mutants yield no product (lane 1-5). **Right:** PCR to verify pSUP\_*hyg* integration into the *cgnI*. Mutant DNA templates yields the expected product (lanes 1-5).

All genetically verified clones were cultivated in liquid Pol03 medium supplemented with 100 mg/L Hygromycin B and 1% w/v Amberlite XAD-16 adsorber resin. Cells and XAD were extracted twice with 50% v/v MeOH after 5 days of cultivation. 1  $\mu$ L of 10x concentrated extracts (corresponding to culture volume) was injected for each measurement. HRESIMS mass spectra were obtained with an Agilent 1200 series HPLC-UV system combined with an ESI-TOF-MS (Maxis, Bruker) [column 2.1 $\times$ 50 mm, 1.7  $\mu$ m, C18 Acquity UPLC BEH (Waters), solvent A: H<sub>2</sub>O + 0.1% formic acid; solvent B: AcCN + 0.1% formic acid, gradient: 5% B for 0.5 min increasing to 100% B in 19.5 min, maintaining 100% B for 5 min, FR = 0.6 mL min<sup>-1</sup>, UV detection 200-600 nm].

**SI-7: Purification of recombinant CgnB and surface plasmon resonance (SPR) analysis**

Full-length *cgnB* was cloned into pEHISSUMOTEV (A gift from David Owen, University of St Andrews) and expressed in *E. coli* BL21 DE3 cells grown in auto-induction medium at 18 °C for 48 h. Cells were harvested by centrifugation (4000 g, 4 °C, 15 min) and resuspended in lysis buffer (500 mM NaCl, 20 mM imidazole, 20 mM Tris pH 8.0, 3 mM  $\beta$ -mercaptoethanol (BME)) supplemented with cOmplete EDTA-free protease inhibitor tablets (Roche) and DNase (0.4 mg g<sup>-1</sup> wet cells, Sigma). The cell suspension was lysed via passage through a cell disruptor (20 kpsi, Microfluidics), and the cell debris was removed by centrifugation (40 000 g, 4 °C, 20 min). The supernatant was filtered through a 0.45  $\mu$ m filter and applied to a FF-Ni-NTA (GE Healthcare) pre-equilibrated in lysis buffer. The protein was eluted in lysis buffer supplemented with extra imidazole (250 mM). The protein eluent was passed over a desalting column (16/10 GE Healthcare) into “low imidazole, low salt” buffer (150 mM NaCl, 20 mM imidazole, 20 mM Tris pH 8.0, 3 mM BME) and subsequently incubated with TEV protease at room temperature for 2 h to cleave off His<sub>6</sub>-tagged-SUMO. Digested protein was loaded onto a FF-Ni-NTA, and the flow through was collected. A homogenous and pure sample as obtained via passage down a Superdex 75 gel filtration column (GE Healthcare) pre-equilibrated in gel filtration buffer (150 mM NaCl, 10 mM HEPES pH 7.4, 1 mM TCEP). Protein purity was assessed by SDS-PAGE and integrity confirmed by MS.

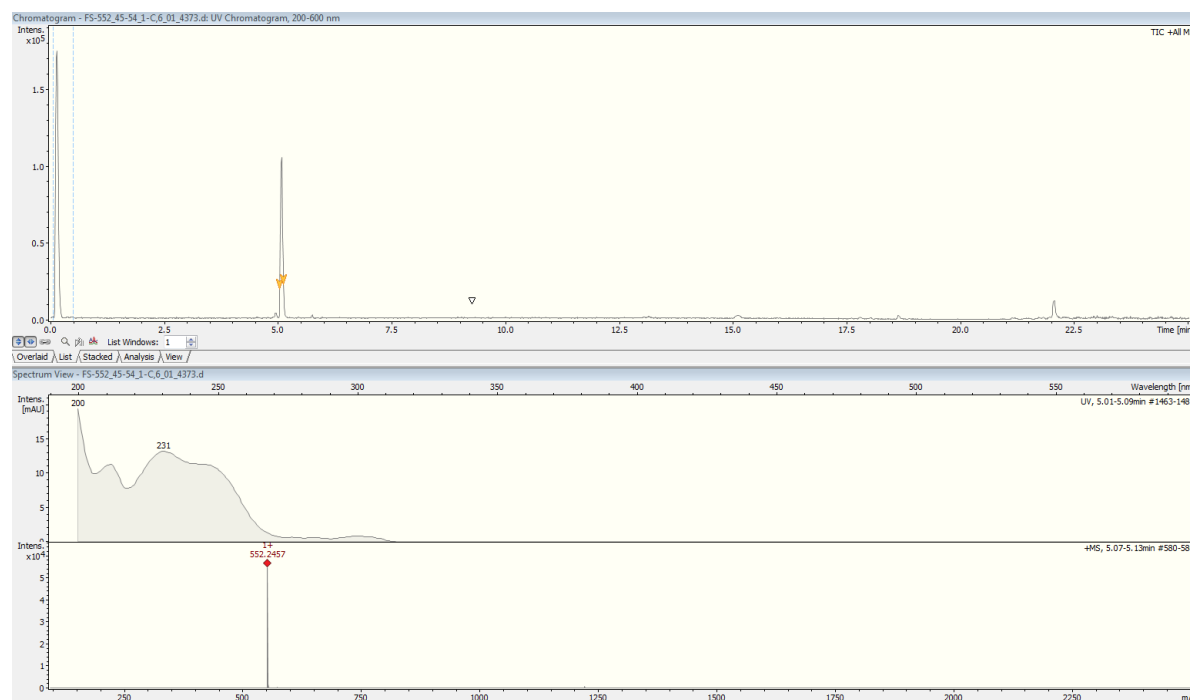
Interaction studies of purified CgnB and the CgnA precursor peptide were carried out using an X100 Biacore Surface Plasmon Resonance System (GE Healthcare) and a CM5 sensor chip. The CgnA precursor peptide obtained from GeneCyst (1 mg/mL in H<sub>2</sub>O) was used as the ligand immobilized via standard amine coupling in flow cell 2. In parallel, the tripeptide IYW (GeneCyst) was immobilized in the reference cell (flow cell 1). CgnB in SPR buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 1 mM TCEP) was used as the analyte in a serial 1:1 dilution ranging from 2  $\mu$ M to 0.125  $\mu$ M and perfused over both channels followed by SPR buffer. Binding and dissociation was measured in real time at 25°C. Data analysis was performed using BIAevaluation software (GE Healthcare). The experiments were performed in triplicates.



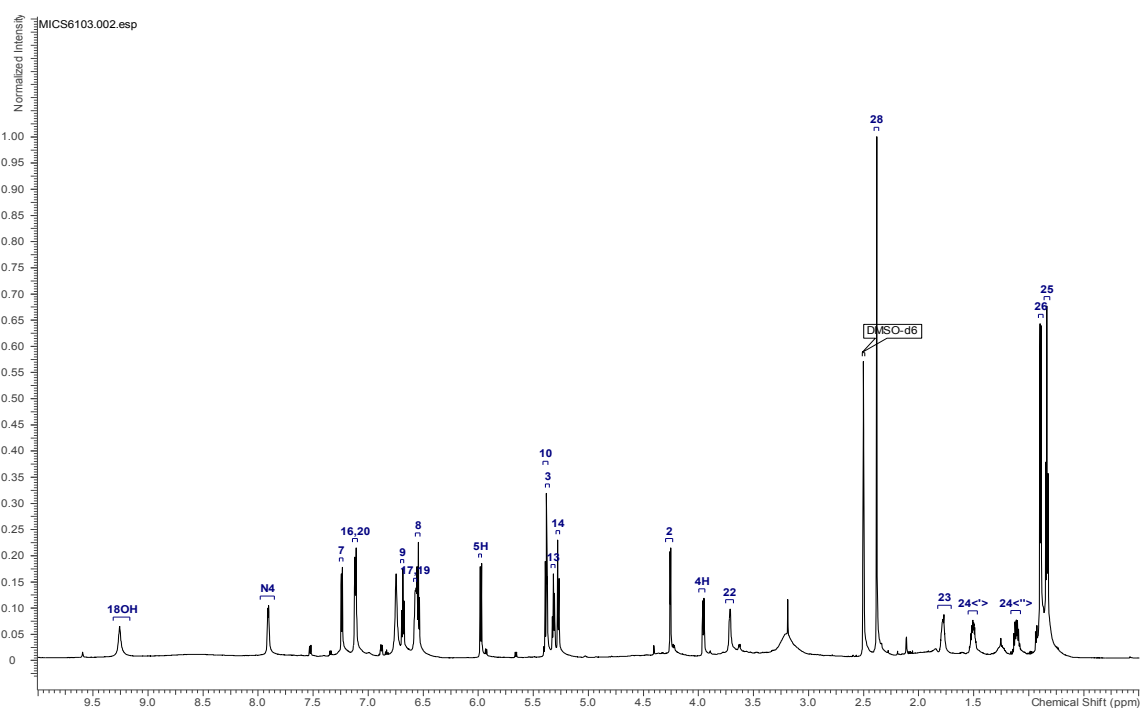
**Figure S17:** Surface plasmon resonance (SPR) analysis reveals a dose-dependent specific binding of CgnB to the 21 aa precursor peptide CgnA. Representative referenced sensorgrams of one independent experiment are shown. The response is given in response units (RU) in dependency of time [s].

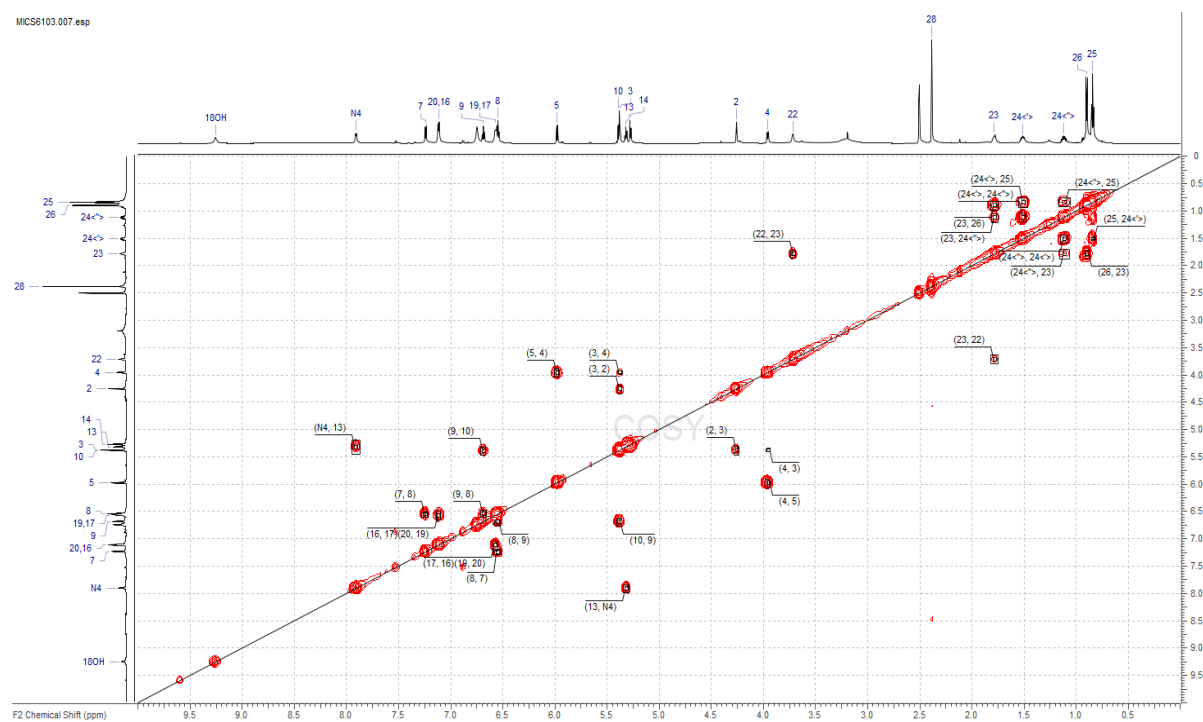
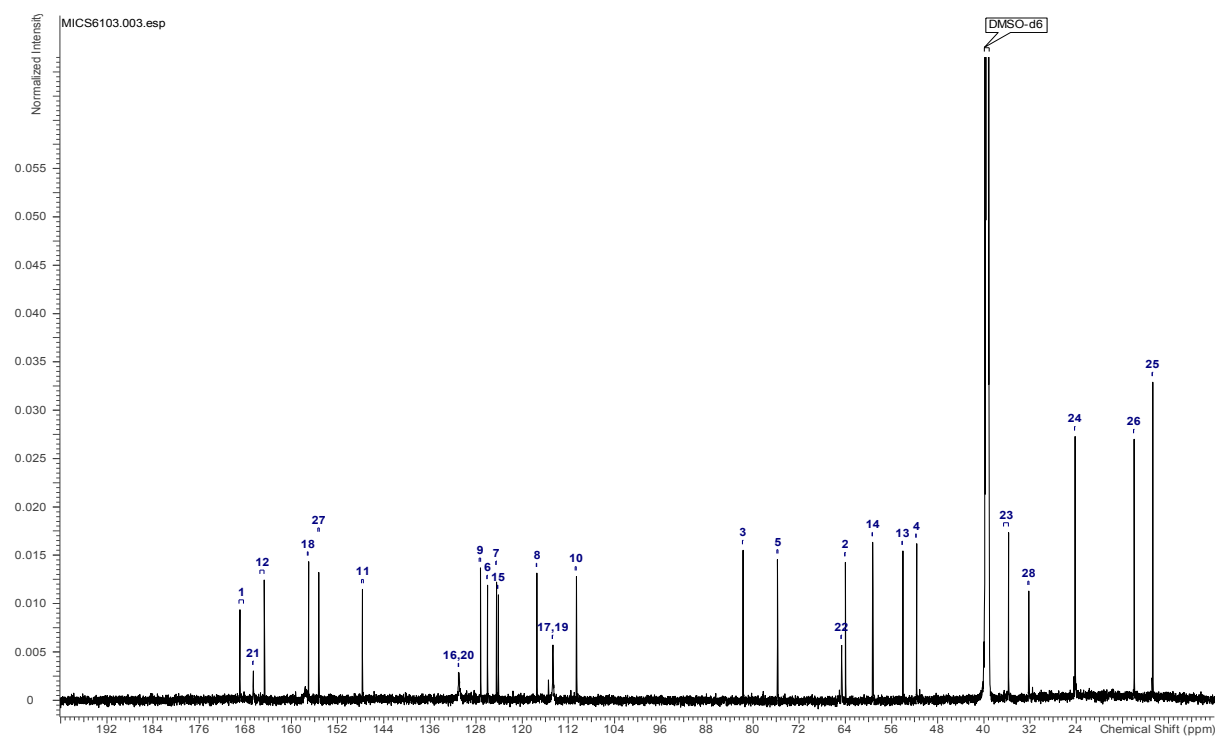


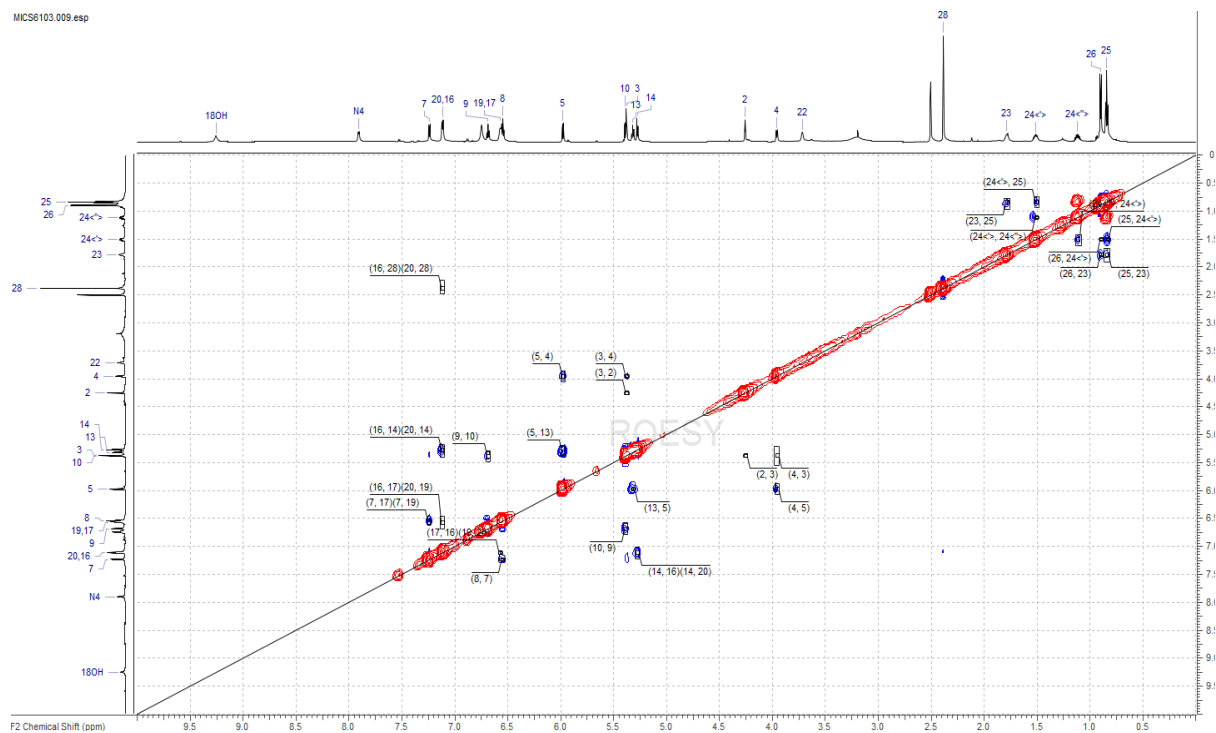
## SI-8: Crocagin LC-MS chromatograms and NMR spectra



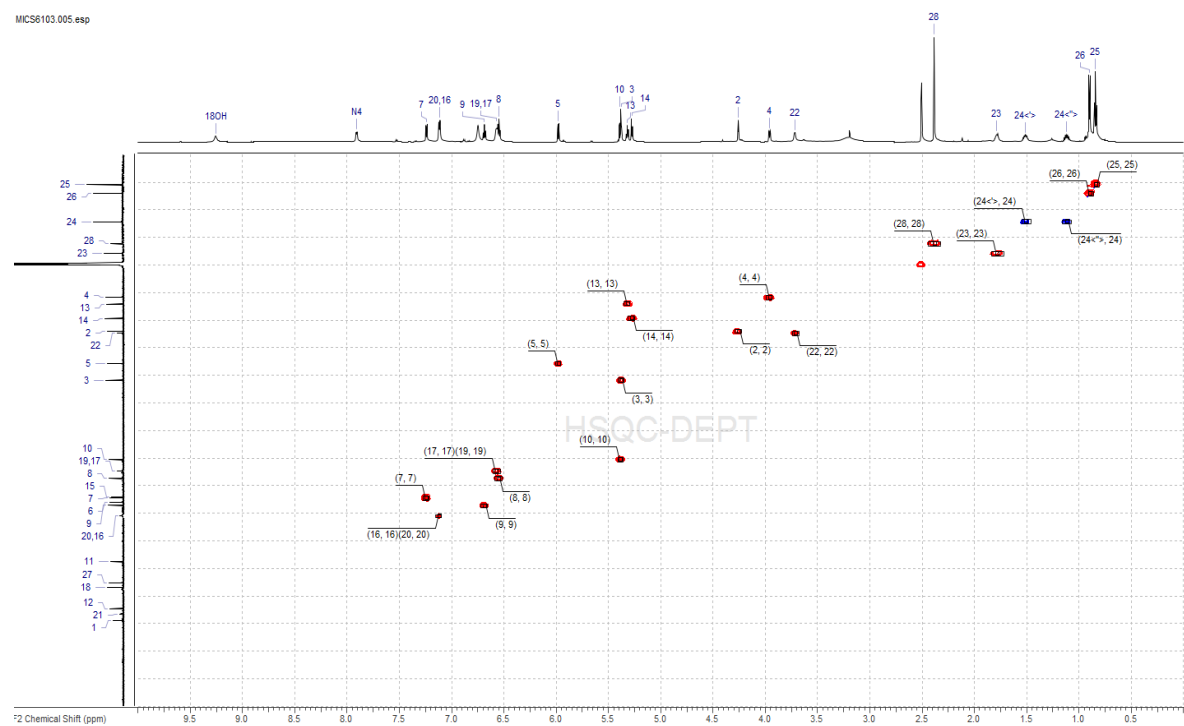
HPLC-HRESIMS spectrum of crocagin A (1).

Proton NMR spectrum (700 MHz, DMSO-*d*<sub>6</sub>, 333 K) of crocagin A (1).

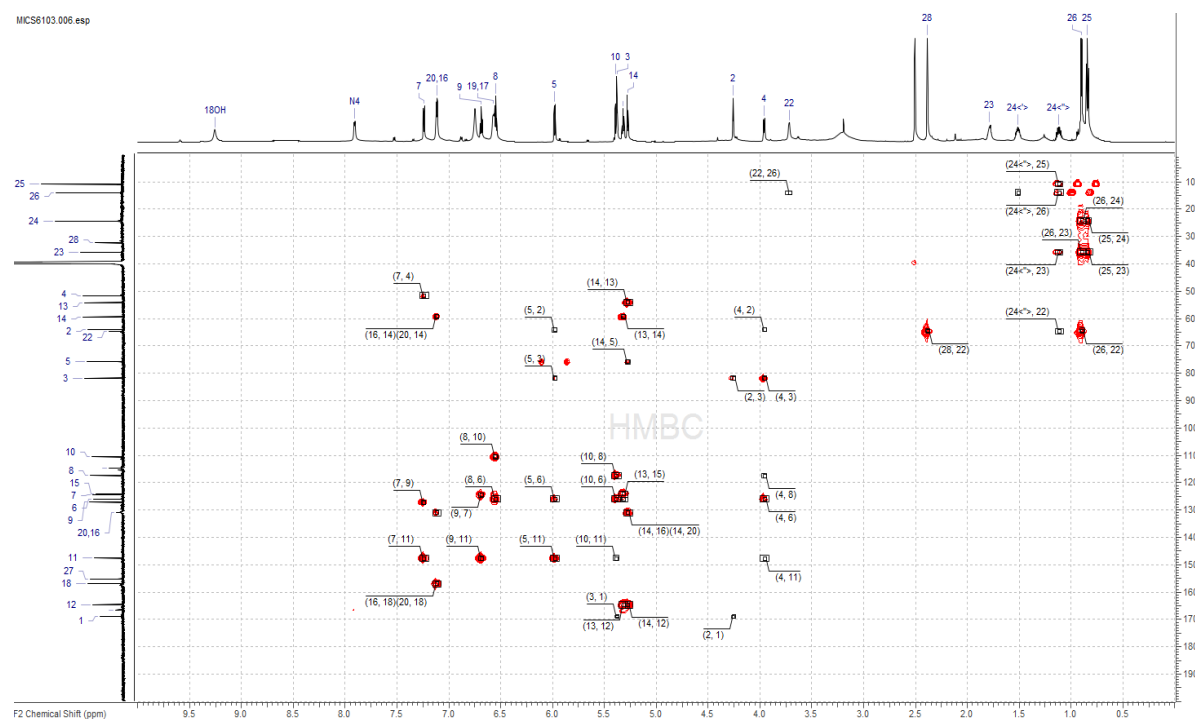
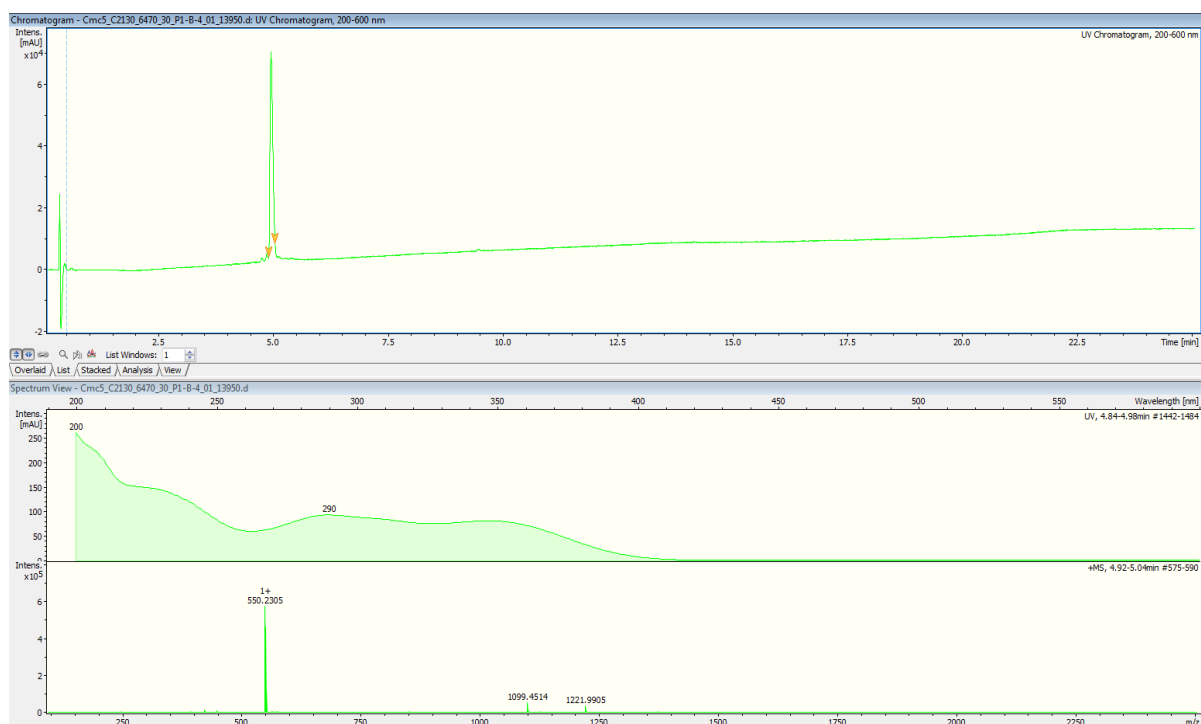




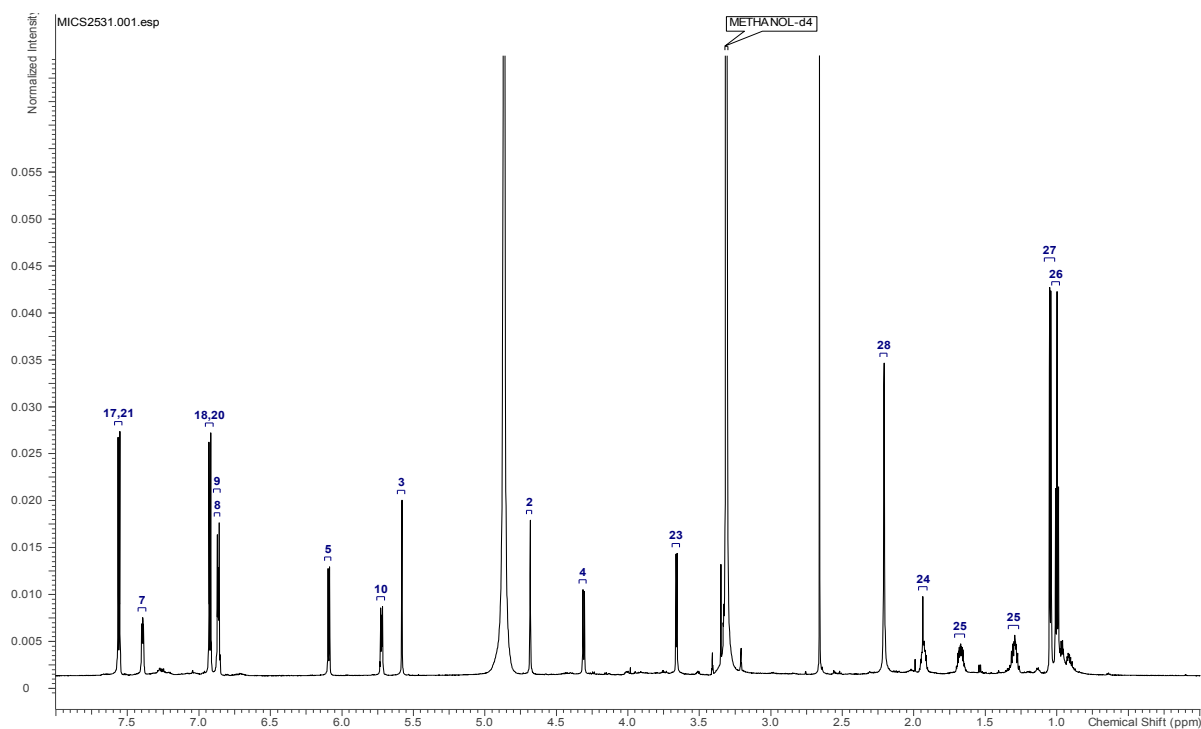
ROESY NMR spectrum (700 MHz, DMSO- $d_6$ , 333 K) of crocagin A (**1**).



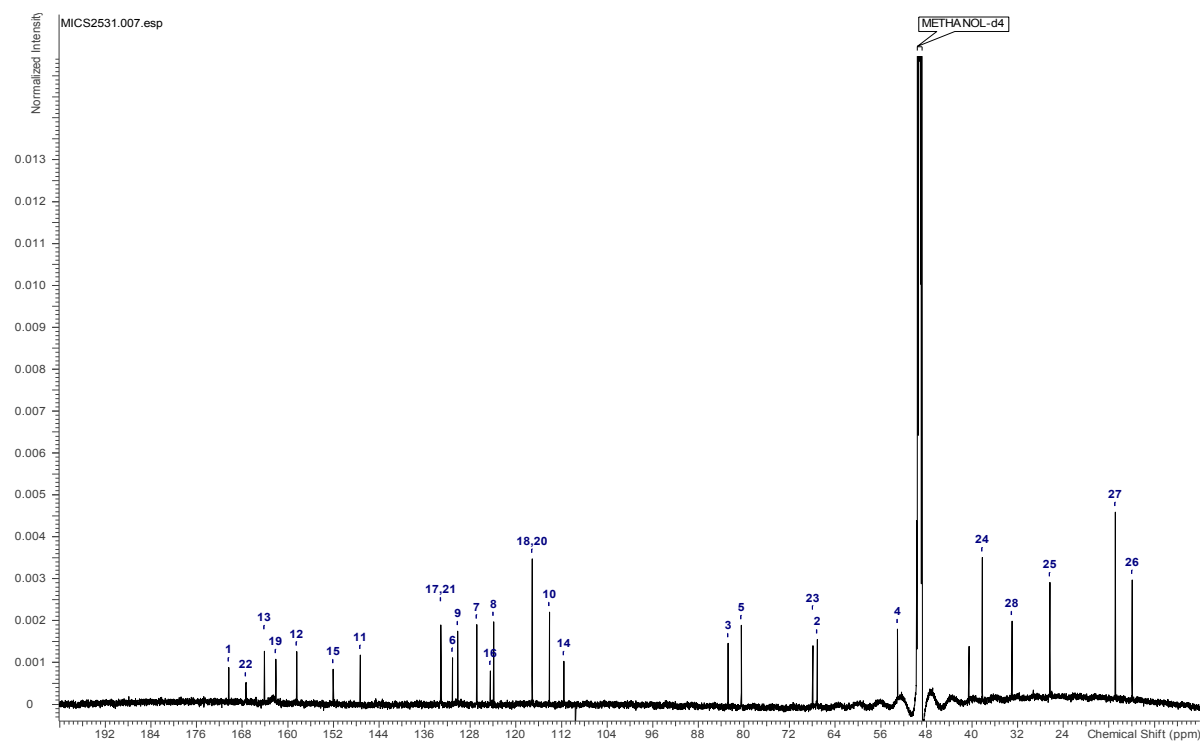
HSQC NMR spectrum (700 MHz, DMSO- $d_6$ , 333 K) of crocagin A (**1**).

HMBC NMR spectrum (700 MHz, DMSO- $d_6$ , 333 K) of crocagin A (1).

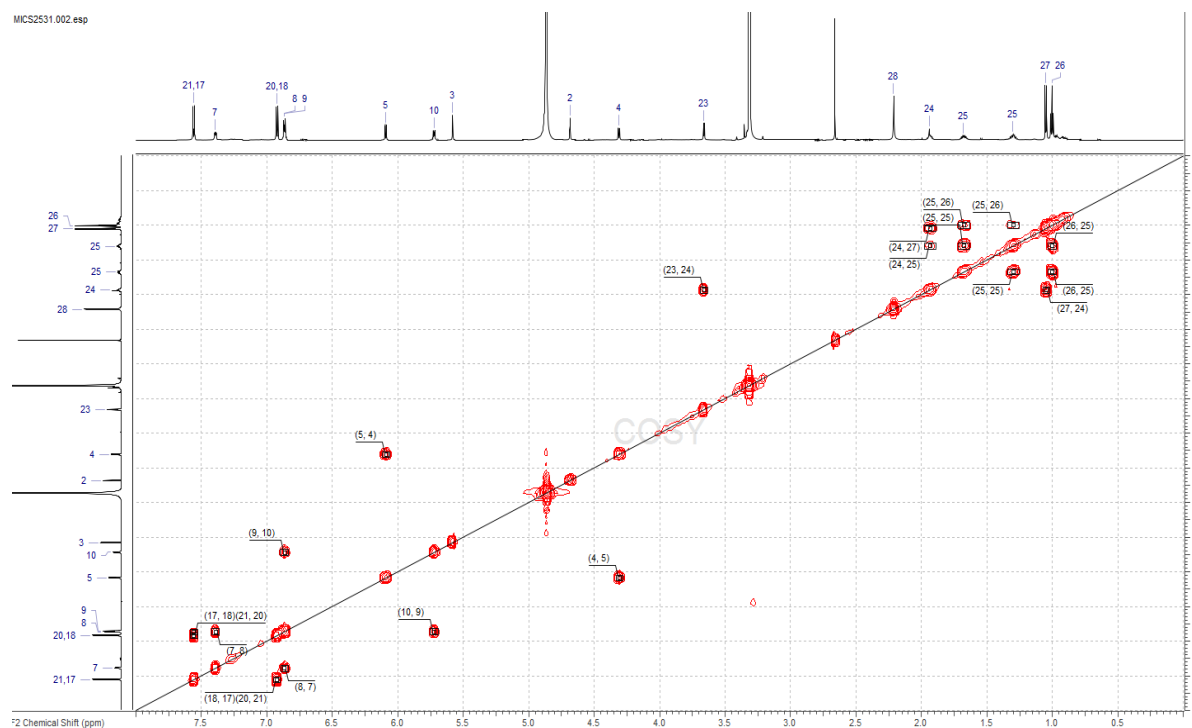
HPLC-HRESIMS spectrum of crocagin B (2).



Proton NMR spectrum (175 MHz, DMSO- $d_6$ , 298 K) of crocagin B (**2**).



Carbon NMR spectrum (175 MHz, DMSO- $d_6$ , 290 K) of crocagin B (**2**).

COSY NMR spectrum (700 MHz, DMSO- $d_6$ , 298 K) of crocagin B (**2**).HSQC NMR spectrum (700 MHz, DMSO- $d_6$ , 298 K) of crocagin B (**2**).



HMBC NMR spectrum (700 MHz, DMSO- $d_6$ , 298 K) of crocagin B (**2**).

## Chapter 7

# The structure of CgnJ, a domain of unknown function protein from the Crocagin gene cluster

Previously published in:

**Sebastian Adam**<sup>[a]</sup>, **Andreas Klein**<sup>[a]</sup>, **Frank Surup**<sup>[b]</sup> and **Jesko Koehnke**<sup>[a]\*</sup>

**Acta Crystallogr F Struct Biol Commun.** 2019 Mar 1;**75**(Pt 3):205-211.

**DOI: 10.1107/S2053230X19000712**

### Affiliation

<sup>[a]</sup> Workgroup Structural Biology of Biosynthetic Enzymes, Helmholtz Institute for Pharmaceutical Research, Helmholtz Centre for Infection Research, Saarland University, Universitäts-campus E8 1, 66123, Saarbrücken, Germany.

<sup>[b]</sup> Microbial Drugs, Helmholtz Centre for Infection Research, Inhoffenstrasse 7, 38124 Braunschweig, Germany.



## **Contributions and Acknowledgements**

### **Author's effort:**

The author contributed significantly to this work by establishing the recombinant protein expression and purification of CgnJ, setting up initial crystallization trials and optimizing the resulting crystals. He refined the crystal structures and performed structural analysis as well as the binding experiments with the natural product Crocagin A. Furthermore, the author wrote part of the manuscript and revised the final version.

### **Other's effort:**

Andreas Klein contributed by helping with protein expression, structural biology and biophysical experiments. Frank Surup contributed by providing the purified natural product Crocagin A for binding experiments. Jesko Köhnke contributed by collecting the data sets for native and SeMet-CgnJ.

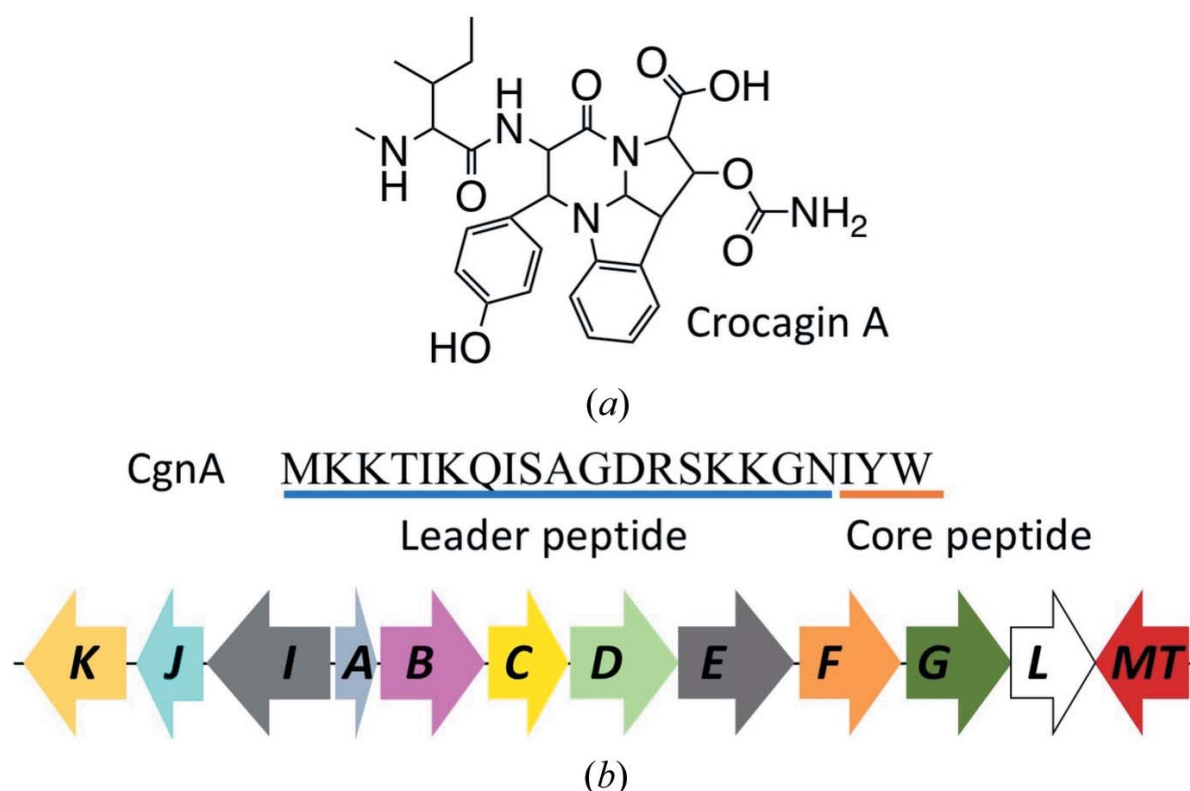
## 7.1 Abstract

Natural products often contain interesting new chemical entities that are introduced into the structure of a compound by the enzymatic machinery of the producing organism. The recently described crocagins are novel polycyclic peptides which belong to the class of ribosomally synthesized and post-translationally modified peptide natural products. They have been shown to bind to the conserved prokaryotic carbon-storage regulator A *in vitro*. In efforts to understand crocagin biosynthesis, the putative biosynthetic genes were expressed and purified. Here, the first crystal structure of a protein from the crocagin-biosynthetic gene cluster, CgnJ, a domain of unknown function protein, is reported. Possible functions of this protein were explored by structural and sequence homology analyses. Even though the sequence homology to proteins in the Protein Data Bank is low, the protein shows significant structural homology to a protein with known function within the competency system of *Bacillus subtilis*, ComJ, leading to the hypothesis of a similar role of the protein within the producing organism.

## 7.2 Introduction

Natural products have been a source of bioactive substances with often interesting and novel chemical scaffolds since the discovery of penicillin in 1929.<sup>[1]</sup> Ribosomally synthesized and post-translationally modified peptide (RiPP) natural products are a fast-growing family of chemically and structurally diverse compounds with interesting and potent bioactivities.<sup>[2]</sup> RiPPs are produced from a ribosomal precursor peptide, which undergoes one or more maturation steps to yield the final, often heavily modified, natural product.<sup>[3]</sup> As is common for prokaryotes, the proteins involved in RiPP biosynthesis are usually organized as gene clusters.<sup>[4]</sup> Modern advances in the field of systems biology, such as secondary metabolome mining, allow an efficient search for novel natural products and allow the biosynthetic potential of natural producers to be uncovered.<sup>[5]</sup> The use of these techniques led to the recent identification and isolation of the crocagins, a novel group of polycyclic RiPPs from *Chondromyces crocatus* Cm c5 containing a tetrahydropyrrolo-[2,3-b]-indole core (**Figure 1a**).<sup>[6]</sup> While searching for potential activities of these natural products, they were discovered to bind carbon-storage regulator A (CsrA).<sup>[6]</sup> CsrA is a highly conserved key player in the Csr-type regulatory system, an important post-transcriptional control mechanism

involving small regulatory RNAs (sRNAs), which coordinates the expression of a variety of proteins, including specific virulence factors, as well as biofilm formation.<sup>[7]</sup> CsrA-type proteins have been found in over 1200 bacterial genomes, including those of important human Gram-negative pathogens such as *Legionella pneumophila*, *Pseudomonas aeruginosa*, *Escherichia coli*, *Acinetobacter baumannii* and *Salmonella typhimurium*.<sup>[8]</sup> CsrA is currently not the target of any known antibiotic, and the potential ability of crocagins to interfere with CsrA-controlled processes, especially in Gram-negative bacteria, makes them highly attractive compounds for further research.



**Figure 1.** (a) Chemical structure of crocagin A. (b) Sequence of the crocagin precursor peptide CgnA and the crocagin-biosynthetic gene cluster. The leader peptide is underlined in blue and the core peptide in orange. All genes encoding proteins thought to be involved in crocagin biosynthesis are shown (*cgnB*–*cgnL*; *cgnMT*, methyltransferase).

In addition to the natural products, the biosynthetic gene cluster of crocagins was identified by a combination of genome analysis, targeted gene inactivation of the producer and *in vitro* experiments (**Figure 1b**).<sup>[6]</sup> Within the gene cluster, nine potentially biosynthetic genes (*cgnB*–*cgnE*, *cgnI*–*cgnL* and *cgnMT*) have been identified and putative functions have been assigned by sequence homology.<sup>[6]</sup> *cgnA* encodes the precursor peptide containing the core peptide sequence IYW (**Figure 1b**), which is modified to yield crocagin. By utilizing a targeted gene-inactivation approach and *in vivo* metabolome studies, the functions of two biosynthetic

genes, *cgnB* and *cgnI*, could be explored, leaving the actual biosynthetic functions of all of the other genes in the cluster as hypothetical.<sup>[6]</sup> Of these, CgnJ was the only protein determined to be a domain of unknown function (DUF) protein, leaving questions as to its potential role and biosynthetic function. Even though functions for DUF proteins cannot be routinely determined using sequence homology to known proteins, they can be found in all species, leading to the hypothesis that they fulfil an essential role owing to being evolutionarily conserved.<sup>[9]</sup> In the post-genomic era, the relevance of DUF proteins became visible, with well over 3700 different families in the Pfam database.<sup>[10]</sup> Combining these results with a structural study of 240 DUF proteins conducted by the Protein Structure Initiative, it was shown that only as few as 20% contain completely novel folds, suggesting the use of X-ray crystallography as an important tool in the elaboration of potential functions of DUF protein.<sup>[11]</sup> Furthermore, this sometimes allows the co-crystallization of a ligand in the process, providing clues about the putative role of the protein.<sup>[12]</sup>

Here, we present the first structural and biophysical study of a protein from the crocagin biosynthetic gene cluster: the domain of unknown function protein CgnJ.

## 7.3 Materials and methods

### 7.3.1 Macromolecule production and affinity measurements

**7.3.1.1 Expression of native and L-selenomethionine-labelled CgnJ.** The open reading frame for CgnJ was cloned from *C. crocatus* genomic DNA into the pHIS-SUMO-TEV plasmid with an N-terminal His<sub>6</sub> tag and a Tobacco etch virus (TEV) protease site (a kind gift from Dr David Owen). The protein was expressed in *E. coli* BL21 (DE3) cells grown in Luria–Bertani (LB) medium supplemented with 50 µg ml<sup>-1</sup> kanamycin. Cultures were grown at 37 °C and 200 rev min<sup>-1</sup> until the OD<sub>600</sub> reached 0.6, and the cells were then induced by the addition of isopropyl-β-D-1-thiogalactopyranoside (IPTG; final concentration 0.1 mM) and further grown at 16 °C and 180 rev min<sup>-1</sup> overnight.

L-Selenomethionine-labelled (SeMet) CgnJ was expressed from *E. coli* BL21 (DE3) cells grown in minimal medium supplemented with glucose-free nutrient mix, 50 µg ml<sup>-1</sup> kanamycin and 5% glycerol. This medium was inoculated with an overnight culture grown in LB, which was washed three times in minimal medium. After 20 min of growth at 37 °C, 40 mg L<sup>-1</sup> L-

selenomethionine was added. The cultures were returned to 37 °C and 200 rev min<sup>-1</sup> and grown until an OD<sub>600</sub> of 0.6 was reached, whereupon 100 mg L<sup>-1</sup> lysine, phenylalanine and threonine and 50 mg L<sup>-1</sup> isoleucine and valine were added. After incubation for a further 20 min, expression was induced with IPTG (final concentration 1 mM) and the cells were grown at 20 °C for 24 h. For both the native and SeMet variants, the cells were harvested by centrifugation (4000 g, 4 °C, 15 min).

**7.3.1.2. Purification of native and SeMet CgnJ.** The cell pellets were resuspended in lysis buffer [50 mM Tris pH 7.8, 200 mM NaCl, 20 mM imidazole, 10% glycerol, 3 mM β-mercaptoethanol (BME)] supplemented with 0.4 mg g<sup>-1</sup> DNase and cOmplete EDTA-free protease-inhibitor tablets (Roche; one tablet per 50 ml of resuspension). The cells were lysed via passage through a cell disruptor at 207 MPa (Microfluidics) and the cell debris was removed by centrifugation (30 000 g, 4 °C, 20 min). The supernatant was loaded onto a pre-equilibrated Ni-NTA column at 4 °C and the column was washed with 20 column volumes of lysis buffer. The protein was eluted from the column in elution buffer (lysis buffer supplemented with 250 mM imidazole) and fractions containing protein were pooled and passed over a desalting column pre-equilibrated in lysis buffer (16/10 Desalting, GE Healthcare) to remove excess imidazole. TEV protease was added at a mass:mass ratio of 1:10 and the sample was incubated for 1 h at 20 °C to remove the N-terminal His<sub>6</sub>-SUMO tag. The digested sample was passed over a second Ni-NTA column pre-equilibrated in lysis buffer and the flowthrough was collected, concentrated and loaded onto a Superdex 200 gel-filtration column (GE Healthcare) equilibrated in GF buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 1 mM TCEP; **Supplementary Figure S1**). Protein yields were 24 and 16 mg L<sup>-1</sup> for native and SeMet CgnJ, respectively. The purity of the protein was determined by SDS-PAGE (**Supplementary Figure S1**). Pure CgnJ protein was concentrated to 3.75 mg ml<sup>-1</sup>.

**7.3.1.3. Expression and purification of His<sub>6</sub>-CgnJ.** To allow His<sub>6</sub>-tag labelling of CgnJ for microscale thermophoresis experiments (see below), the open reading frame for CgnJ was subcloned into the pHIS-TEV plasmid with an N-terminal His<sub>6</sub> tag and a Tobacco etch virus (TEV) protease site.<sup>[13]</sup> The protein was expressed as described for native pHIS-SUMO-TEV CgnJ. The cell pellets were resuspended in lysis buffer (20 mM Tris pH 8.0, 500 mM NaCl, 20 mM imidazole, 3 mM BME) supplemented with 0.4 mg g<sup>-1</sup> DNase and cOmplete EDTA-free protease-inhibitor tablets (Roche; one tablet per 50 ml of resuspension). The cells were lysed via passage through a cell disruptor at 207 MPa (Microfluidics) and the cell debris was

removed by centrifugation (30 000 g, 4 °C, 20 min). The supernatant was loaded onto a pre-equilibrated Ni–NTA column at 4 °C and the column was washed with 20 column volumes of lysis buffer. The protein was eluted from the column in elution buffer (lysis buffer supplemented with 250 mM imidazole) and the fractions containing the largest amounts of protein were pooled and loaded onto a Superdex 200 gel-filtration column (GE Healthcare) equilibrated in GF buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 1 mM TCEP). The purity of the His<sub>6</sub>-tagged protein was determined by SDS–PAGE.

**7.3.1.4. Microscale thermophoresis (MST).** MST experiments were performed using a Monolith NT.115 instrument (Nano- Temper) in MST buffer (50 mM Tris–HCl pH 7.4, 150 mM NaCl, 10 mM MgCl<sub>2</sub>, 0.05% Tween 20) at 20 °C. His<sub>6</sub>-CgnJ was dialysed into freshly prepared MST buffer, while a 100 mM stock of crocagin in 100 % DMSO was diluted to a starting concentration of 5 mM in MST buffer. The proteins were fluorescently labelled using the Monolith His-tag labelling kit RED-tris-NTA (NanoTemper) according to the manufacturer’s instructions. For MST measurements, the manufacturer’s instructions were followed using a starting concentration of 2.5 mM crocagin A, 20% excitation power and 40% MST power. The K<sub>d</sub> value was estimated using the MO.Affinity software suite (NanoTemper). Macromolecule-production information is summarized in **Table 1**.

**Table 1.** Macromolecule production information.

Source organism	<i>C. crocatus</i> Cm c5
DNA source	Genomic DNA
Forward primer	CTTCCATGGCAGAGGAAGTCGCCGAG
Reverse primer	CTTAAGCTTTTCATCTCTCGGGCCACAACGT
Cloning vector	pHis-SUMO-TEV
Expression vector	pHis-SUMO-TEV
Expression host	<i>E. coli</i>
Complete amino-acid sequence of the construct produced	MGSSHHHHHGSSEVNQEAKPEVKPEVKP ETHINLKVSDGSSEIFFKIKKTTPLRRL MEAFKRQKGEMDSLRLFLYDGIRIQADQ TPEDLDMEDNDIIEAHREQIGGDIPTE NLYFQGAMAEVVAEIIIPASTWILFFDA SCSINSPAFWSTNDVDRWLKIAHEL VLLQVVLEGYFKVRCILRSSAPAFEMVN ADVSELVSIVLPSGRLVACTTDEPTLNR HVLTVPPGRYRVLREWSVHEESKHYDVE SAEAYPADEGPDGIITLWPER

### 7.3.2. Crystallization

Crystal screens for SeMet CgnJ were set up by the sitting-drop vapour-diffusion method using a Gryphon robot (Art Robbins). The protein was screened against The Classics Suite (Qiagen) at 20 °C. Initial crystals formed in condition C2 (1 M ammonium phosphate, 0.1 M trisodium citrate pH 5.6) after 48 h. Diffraction-quality crystals (tetragonal bipyramids) were grown in a condition consisting of 0.75 M ammonium phosphate, 0.1 M trisodium citrate pH 5.8. Crystals of native CgnJ grew in the same condition as the SeMet CgnJ protein. A single SeMet CgnJ crystal was cryoprotected in mother liquor supplemented with 34% glycerol and flash-cooled in liquid nitrogen. Crystallization information is summarized in **Table 2**.

**Table 2.** Crystallization

Method	Vapour diffusion
Plate type	Swissci
Temperature (K)	291
Protein concentration (mg ml <sup>-1</sup> )	3.75
Buffer composition of protein solution	150 mM NaCl, 10 mM HEPES, 1 mM TCEP pH 7.4
Composition of reservoir solution	0.75 M ammonium phosphate, 0.1 M trisodium citrate pH 5.8
Volume and ratio of drop	1:1 µl, 1:2 µl (reservoir:protein)
Volume of reservoir (µl)	500

### 7.3.3. Data collection and processing

A single-wavelength anomalous dispersion (SAD) data set was collected at the Se K absorption edge at 100 K on beamline X06DA at the Swiss Light Source (SLS). Data were reduced and scaled with maximum redundancy for optimal SAD phasing using *xia2*.<sup>[14]</sup> For the highest possible resolution, the native data set was reduced and scaled at an outer shell value  $I/\sigma(I)$  of 2.9 using *xia2*. **Table 3** shows the data-collection statistics for the high-resolution native and the SAD data sets.

**Table 3.** Data collection and processing. Values in parentheses are for the outer shell.

	Native	SAD
Diffraction source	BL14.1, BESSY II	X06DA, SLS
Wavelength (Å)	0.91841	0.97941
Temperature (K)	100	100
Detector	PILATUS 6M	PILATUS 2M-F
Crystal-to-detector distance (mm)	349.9	152.25
Rotation range per image (°)	0.10	0.20
Total rotation range (°)	100	360
Exposure time per image (s)	0.4	0.1
Space group	<i>P</i> 4 <sub>1</sub> 2 <sub>1</sub> 2	<i>P</i> 4 <sub>1</sub> 2 <sub>1</sub> 2
<i>a</i> , <i>b</i> , <i>c</i> (Å)	132.56, 132.56, 112.98	132.90, 132.90, 112.85
$\alpha$ , $\beta$ , $\gamma$ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Mosaicity (°)	0.063	0.060
Resolution range (Å)	57.17–2.00 (2.05–2.00)	42.06–2.19 (2.25–2.19)
Total No. of reflections	504776 (38046)	1416110 (97301)
No. of unique reflections	68294 (4964)	52596 (3814)
Completeness (%)	99.97 (99.99)	100.0 (100.0)
Multiplicity	7.4 (7.7)	26.9 (25.5)
$\langle I/\sigma(I) \rangle$	18.5 (2.9)	23.9 (5.1)
<i>R</i> <sub>r.i.m.</sub> (%)	4.8 (38.7)	3.8 (22.2)
Overall <i>B</i> factor from Wilson plot (Å <sup>2</sup> )	24.006	23.291

### 7.3.4. Structure solution and refinement

The structure was solved using the *AutoSol* package<sup>[15]</sup> and the chains were built into electron density using the *AutoBuild* package<sup>[16]</sup> from the *PHENIX* crystallographic software suite.<sup>[17]</sup> The initial model was used for molecular replacement (MR) of a higher resolution native data set. MR was performed using *Phaser*<sup>[18]</sup> and the structure was manually rebuilt in *Coot*<sup>[19]</sup> and refined using *phenix.refine*.<sup>[20]</sup> TLS restraints were generated in *PHENIX*<sup>[17]</sup> and the structure was validated using *MolProbity*.<sup>[21]</sup> Refinement statistics are summarized in **Table 4**. All structural images were presented using PyMOL (<https://pymol.org/2/support.html>). All sequence alignments were created using *Clustal Omega*.<sup>[22]</sup>



**Table 4.** Structure refinement. Values in parentheses are for the outer shell.

Resolution range (Å)	57.17–2.00 (2.071–2.000)
Completeness (%)	99.97 (99.99)
$\sigma$ Cutoff	2.9
No. of reflections, working set	64810
No. of reflections, test set	3484
Final $R_{\text{cryst}}$ (%)	17.5
Final $R_{\text{free}}$ (%)	19.9
No. of non-H atoms	
Total	4157
Protein	3695
Ligand	18
Water	444
R.m.s. deviations	
Bonds (Å)	0.003
Angles (°)	0.66
Average $B$ factors (Å <sup>2</sup> )	
Overall	33.2
Protein	31.8
Ligand	76.0
Water	43.2
Ramachandran plot	
Favoured regions (%)	98.04
Additionally allowed (%)	1.96
Outliers (%)	0.00

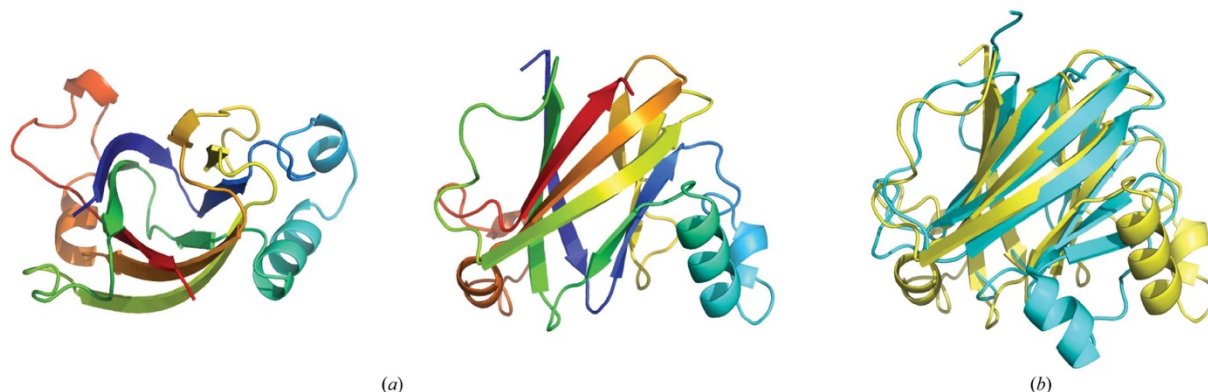
## 7.4. Results and discussion

### 7.4.1. Crystal structure of CgnJ

The open reading frame annotated as CgnJ encoding a 154 amino acid protein was amplified from genomic DNA isolated from *C. crocatus* Cm c5 (Accession No. WP\_050430638). No homologues of this protein can be found using standard BLAST parameters, except for a hypothetical protein from *Pelomonas* sp. Root1217 (Accession No. WP\_057297704), which possesses 30% sequence identity over 110 amino acids. CgnJ was expressed and purified as described in Section 2, and the protein purity and integrity were determined by SDS–PAGE (**Supplementary Figure S1**).

Native and SeMet CgnJ crystallized in space group  $P4_12_12$ , and high-redundancy data were collected to 2.19 Å resolution from a SeMet CgnJ crystal at the selenium edge. The data-collection and refinement statistics can be found in **Tables 3** and **4**. The structure was determined using Se-SAD, and the initial model was used for molecular replacement of a 2.0 Å resolution native data set. The native crystal contained three CgnJ molecules in the asymmetric unit, and in the final model each protomer contains residues 1–154 (the full length). The structure has a fold that is comprised of four  $\alpha$ -helices and nine  $\beta$ -strands (**Figure 2a**). The  $\beta$ -strands are arranged as four- and five-stranded sheets that form a  $\beta$ -barrel-

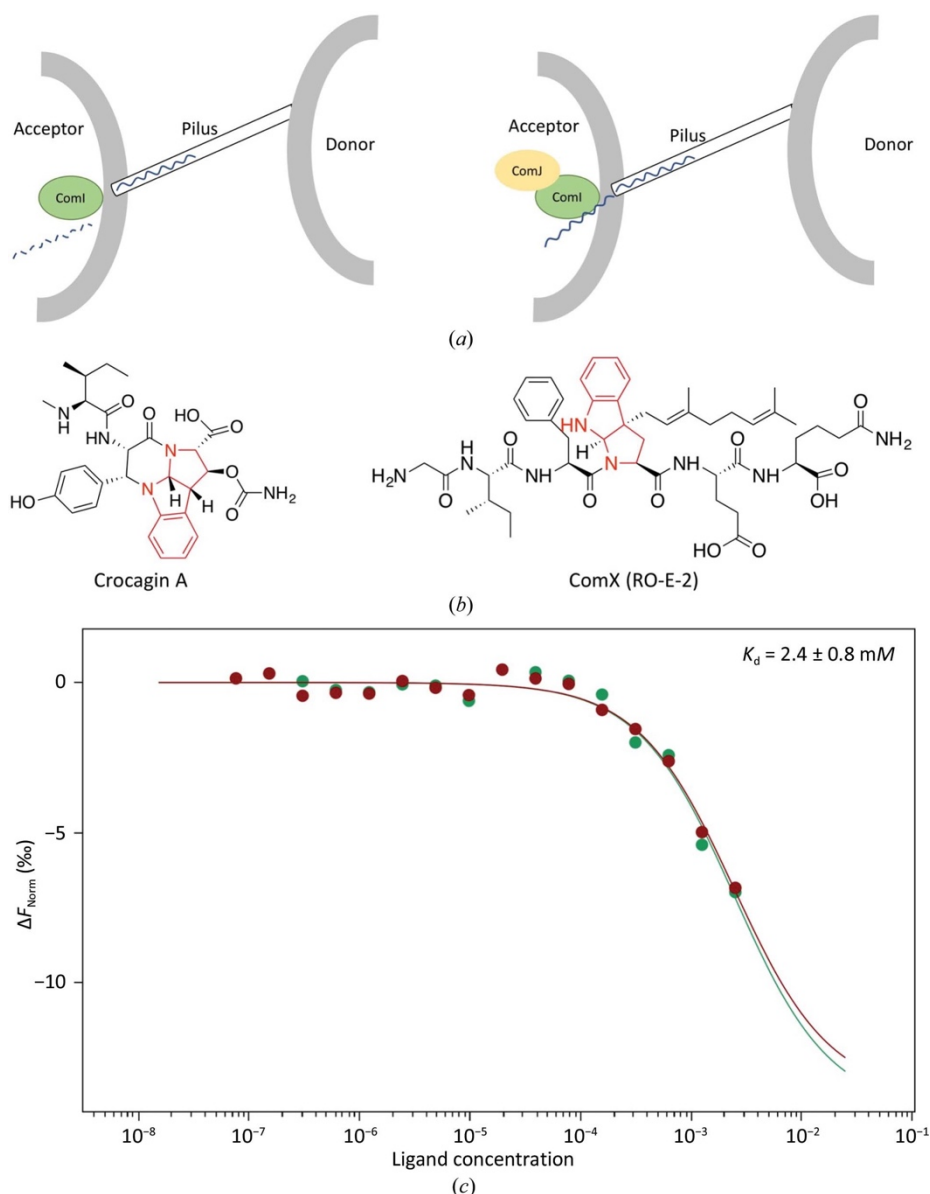
like core decorated with loops and helices. Chains *B* and *C* appeared to form a dimer, but analysis with PISA<sup>[23]</sup> gives a complex-formation significance score of 0, indicating that this is a crystal contact rather than a biological dimer. When analysed using dynamic light scattering, CgnJ appears to be monomeric in solution (**Supplementary Figure S2**).



**Figure 2.** (a) X-ray crystal structure of CgnJ represented as a cartoon using a rainbow colour scheme: top and side views (90° rotation). (b) Superposition of the X-ray crystal structures of CgnJ (yellow) and ComJ (cyan).

To search for structural homologues that may point to the function of CgnJ, we used the DALI server<sup>[24]</sup> for a heuristic PDB search with a CgnJ monomer as the search model. When using a cutoff of 50% sequence coverage, the search returns only one protein, the nuclease inhibitor ComJ from *Bacillus subtilis* (PDB entry 4MQD; Midwest Center for Structural Genomics, unpublished work), with a Z-score of 12.2 and a C $\alpha$  r.m.s.d. of 2.6 Å over 117 residues. The  $\beta$ -strands are very well conserved between the two proteins, and the main differences are found in the helices and loops at the periphery of the two proteins (**Figure 2b**). The good overall structural conservation was surprising, since the sequence identity between the two proteins is only 15.6% (**Supplementary Figure S3**). The function of ComJ is known (**Figure 3a**). It is part of the natural competence machinery in *B. subtilis* and plays a pivotal role during the bacterial conjugation process.<sup>[25]</sup> As donor DNA tries to enter the recipient's cell, it has to pass the nuclease ComI (also known as NucA). This nuclease is active and prevents DNA from entering until it is bound by ComJ, which renders the nuclease inactive and thus facilitates horizontal gene transfer.<sup>[25]</sup> Interestingly, ComJ and ComI are part of the Gram-negative quorum-sensing system involving the competence-stimulating peptide (CSP) ComX (**Figure 3b**). This bacterial pheromone also belongs to the RiPP family of natural products and does not share sequence homology with crocagin,<sup>[26]</sup> yet both natural products contain a tryptophan residue that is cyclized via its side-chain C2 atom to its amide nitrogen. While ComX is polyprenylated at the C $\gamma$  atom, crocagins are hydroxylated (and subsequently carbamoylated) at the C $\beta$  position.

Since no biological activity has yet been reported for crocagins, it is possible that they play a role in quorum sensing.



**Figure 3.** (a) Cartoon representing the role of ComJ. The nuclease ComI degrades incoming DNA unless it is bound by ComJ, which inactivates ComI and allows foreign DNA to enter the cell. (b) Chemical structures of crocagin and ComX (RO-E-2). The similar tryptophan modifications are highlighted in red. (c) Two independent MST measurements of CgnJ with crocagin A (green and red). Since saturation could not be reached (owing to compound solubility), the  $K_d$  is estimated to be  $2.4 \pm 0.8 \text{ mM}$ .

#### 7.4.2 Investigating the potential binding of CgnJ to crocagin

Since the structural homology of CgnJ to ComJ hinted at a possible function, we searched for additional homologues of the competency system in *C. crocatus* using standard *BLAST* parameters. We were unable to find additional components based on sequence homology, perhaps owing to the low degree of sequence conservation. Attempts to identify a homologue

of the nuclease ComI as a potential binding partner of CgnJ were also unsuccessful. We wondered whether CgnJ may be able to bind the precursor peptide CgnA or modify it biochemically. After 24 h at 37 °C, no modification of CgnA was observed (**Supplementary Figure S4**) and no binding between CgnJ and CgnA was detected using microscale thermophoresis (MST; **Supplementary Figure S4**). Given the small size of the protein and its lack of homology to known RiPP recognition proteins, we decided to investigate whether CgnJ was able to bind the natural product crocagin itself. To this end, we used purified crocagin A to determine the affinity of these two components by MST. We were able to demonstrate that crocagin A binds to CgnJ with low affinity (**Figure 3c**). While the binding was robust and reproducible, compound solubility prevented us from using sufficiently high concentrations of crocagin to reach saturation. The  $K_d$  was estimated to be  $2.4 \pm 0.8$  mM. The yield of crocagin was reported to be 0.2 mg L<sup>-1</sup> of pure compound after isolation and purification.<sup>[6]</sup> This equates to roughly 0.4 mM and would indicate that even if the yield had been 0.1% of the compound found in the actual culture, an interaction would be unlikely to occur *in vivo*. However, whether crocagin is produced at a steady rate or in a burst remains unclear, as does the concentration of crocagin that can be reached inside the producer during production. Further experiments will thus be required to investigate the biological relevance of this *in vitro* interaction.

## 7.5 Conclusion

The DUF protein CgnJ has only one sequence homologue in publicly available sequence data. X-ray crystallography revealed the structure of CgnJ to be related to that of the nuclease inhibitor ComJ from *B. subtilis*, prompting us to search for homologous nucleases in *C. crocatus*. Owing to the very low sequence homology, we are unsure whether our inability to find ComI homologues rules out the presence of a structural homologue. In a search for a possible role of CgnJ, we tested whether this protein is able to bind the natural product itself. Despite a measurable affinity, the interaction is unlikely to be biologically relevant. Molecules which are structurally related to crocagin, such as kawaguchipeptin,<sup>[27]</sup> ComX<sup>[28]</sup> and kapakahine<sup>[29]</sup> (**Supplementary Figure S5**), have been assigned biological functions. It will require further study to determine whether the interaction of crocagin with CsrA is biologically relevant.

## **7.6 Acknowledgements**

We thank the staff of beamline X06DA at the Swiss Light Source for their support with data collection. SA would like to thank Rebekka Christmann for help with DLS experiments.

## 7.7 References

- [1] Lobanovska, M. & Pilla, G. *Yale J. Biol. Med.* **2017** 90, 135–145
- [2] Dunbar, K. L. & Mitchell, D. A. *ACS Chem. Biol.* **2013** 8, 473–487
- [3] Hetrick, K. J. & van der Donk, W. A. *Curr. Opin. Chem. Biol.* **2017** 38, 36–44
- [4] Doroghazi, J. R. & Metcalf, W. W. *BMC Genomics* **2013** 14, 611
- [5] Maansson, M., Vynne, N. G., Klitgaard, A., Nybo, J. L., Melchiorson, J., Nguyen, D. D., Sanchez, L. M., Ziemert, N., Dorrestein, P. C., Andersen, M. R. & Gram, L. *mSystems* **2016** 1, e00028-15
- [6] Viehrig, K., Surup, F., Volz, C., Herrmann, J., Abou Fayad, A., Adam, S., Köhnke, J., Trauner, D. & Müller, R. *Angew. Chem. Int. Ed.* **2017** 56, 7407–7410
- [7] Maurer, C. K., Fruth, M., Empting, M., Avrutina, O., Hossmann, J., Nadmid, S., Gorges, J., Herrmann, J., Kazmaier, U., Dersch, P., Müller, R. & Hartmann, R. W. *Future Med. Chem.* **2016**, 8, 931–947.
- [8] Heroven, A. K., Böhme, K. & Dersch, P. *RNA Biol.* **2012** 9, 379–391
- [9] Goodacre, N. F., Gerloff, D. L. & Uetz, P. *MBio* **2013** 5, e00744-13
- [10] Mudgal, R., Sandhya, S., Chandra, N. & Srinivasan, N. *Biol. Direct* **2015** 10, 38
- [11] Jaroszewski, L., Li, Z., Krishna, S. S., Bakolitsa, C., Wooley, J., Deacon, A. M., Wilson, I. A. & Godzik, A. *PLoS Biol.* **2009** 7, e1000205
- [12] Bateman, A., Coggill, P. & Finn, R. D. *Acta Cryst.* **2010** F66, 1148–1152
- [13] Liu, H. & Naismith, J. H. *Protein Expr. Purif.* **2009** 63, 102–111
- [14] Winter, G. J. *Appl. Cryst.* **2010** 43, 186–190
- [15] Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H. & Hung, L.-W. *Acta Cryst.* **2009** D65, 582–601
- [16] Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L.-W., Read, R. J. & Adams, P. D. *Acta Cryst.* **2008** D64, 61–69
- [17] 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. & Zwart, P. H. *Acta Cryst.* **2010** D66, 213–221
- [18] McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. *J. Appl. Cryst.* **2007** 40, 658–674
- [19] Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. *Acta Cryst.* **2010** D66, 486–501
- [20] Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. *Acta Cryst.* **2012** D68, 352–367
- [21] Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. *Acta Cryst.* **2010** D66, 12–21
- [22] Sievers, F. & Higgins, D. G. *Curr. Protoc. Bioinformatics* **2014** 48, 3.13.1–3.13.16
- [23] Krissinel, E. & Henrick, K. *J. Mol. Biol.* **2007** 372, 774–797
- [24] Holm, L. & Laakso, L. M. *Nucleic Acids Res.* **2016** 44, W351–W355
- [25] Zhang, D., de Souza, R. F., Anantharaman, V., Iyer, L. M. & Aravind, L. *Biol. Direct* **2012** 7, 18
- [26] Magnuson, R., Solomon, J. & Grossman, A. D. *Cell* **1994**, 77, 207–216

[27] Ishida, K., Matsuda, H., Murakami, M. & Yamaguchi, K. *J. Nat. Prod.* **1997** 60, 724–726

[28] Okada, M., Sato, I., Cho, S. J., Iwata, H., Nishio, T., Dubnau, D. & Sakagami, Y. *Nature Chem. Biol.* **2005** 1, 23–24

[29] Nakao, Y., Yeung, B. K. S., Yoshida, W. Y., Scheuer, P. J. & Kelly-Borges, M. *J. Am. Chem. Soc.* **1995** 117, 8271–

## Supporting Information

# The structure of CgnJ, a domain of unknown function protein from the Crocagin gene cluster

Previously published in:

**Sebastian Adam**<sup>[a]</sup>, Andreas Klein<sup>[a]</sup>, Frank Surup<sup>[b]</sup> and Jesko Koehnke<sup>[a]\*</sup>

**Acta Crystallogr F Struct Biol Commun.** 2019 Mar 1;**75**(Pt 3):205-211.

**DOI: 10.1107/S2053230X19000712**

### Affiliation

<sup>[a]</sup> Workgroup Structural Biology of Biosynthetic Enzymes, Helmholtz Institute for Pharmaceutical Research, Helmholtz Centre for Infection Research, Saarland University, Universitätscampus E8 1, 66123, Saarbrücken, Germany.

<sup>[b]</sup> Microbial Drugs, Helmholtz Centre for Infection Research, Inhoffenstrasse 7, 38124 Braunschweig, Germany.



## Expression and purification of CgnA

For the production of the crocagin precursor peptide CgnA, its coding sequence was cloned into the pHisSUMOTEV vector (a kind gift from Dr. David Owen), and the expression plasmid was transformed into chemically competent *E. coli* Rosetta2(DE3) cells (Novagen) using a standard heat shock procedure. For expression, a fresh colony was added into 100 mL of LB Broth starter culture using the Luria-Miller formulation with Kanamycin (50 µg/mL) and was incubated at 37 °C and 200 rpm for 16 h. Expression of CgnA was carried out by inoculating LB Broth with Kanamycin (50 µg/mL) with the starter culture using a 1:100 dilution. Cultures were grown at 37 °C and 200 rpm until the OD<sub>600</sub> reached 0.7. At this point, protein expression was induced by adding IPTG to a final concentration of 0.5 mM and further incubating the culture at 37 °C and 200 rpm for 16 h.

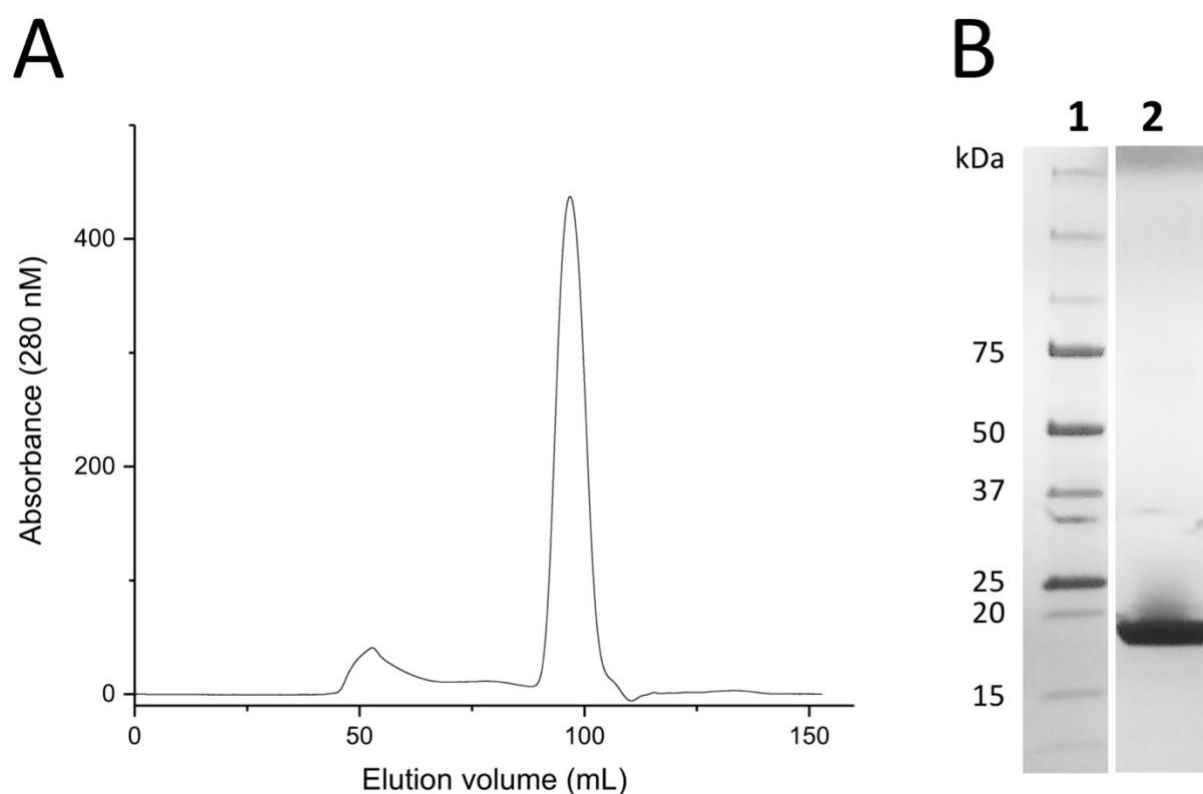
## LC-MS

LC-MS was performed on a Dionex Ultimate 3000 RSLC system using BEH C18 column (100 mm x 2.1 mm, 1.7 µm) 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.

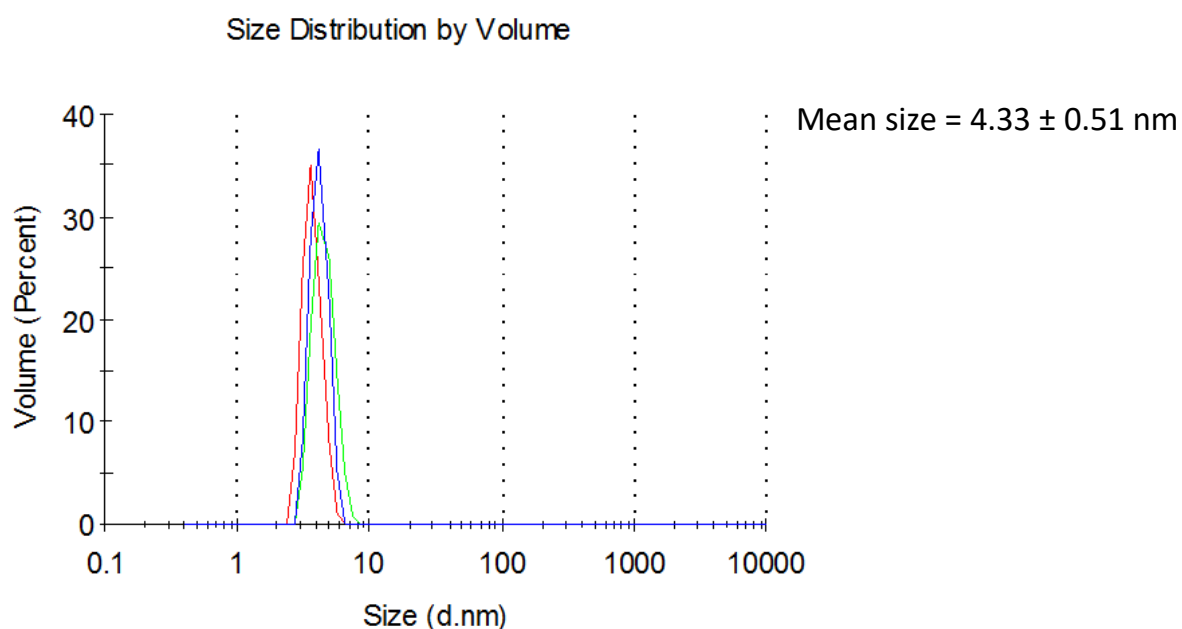
MS was performed using an amaZon speed mass spectrometer (Bruker Daltonics). The LC flow was split 1:8 before entering the mass spectrometer using the Apollo ESI source. The following conditions were used: capillary voltage 4500 V, temperature 300 °C, dry-gas flow rate 10 L/min and nebulizer 30 psi. Data was recorded in the mass range from 500 to 2000 *m/z*.

## Dynamic light scattering (DLS)

Dynamic light scattering (DLS) experiments were carried out on a Zetasizer Nano ZS (Malvern Panalytical) at room temperature. CgnJ was diluted to 50 µM in water and size measurements were carried out in triplicates with 15 measurements each. The mean sizes of 15 runs were calculated for each run with the *Zetasizer Nano* software. The mean size and standard deviation of 45 runs was calculated using *Microsoft Excel* and compared to the apparent sizes of monomer or dimer formation of the protein.



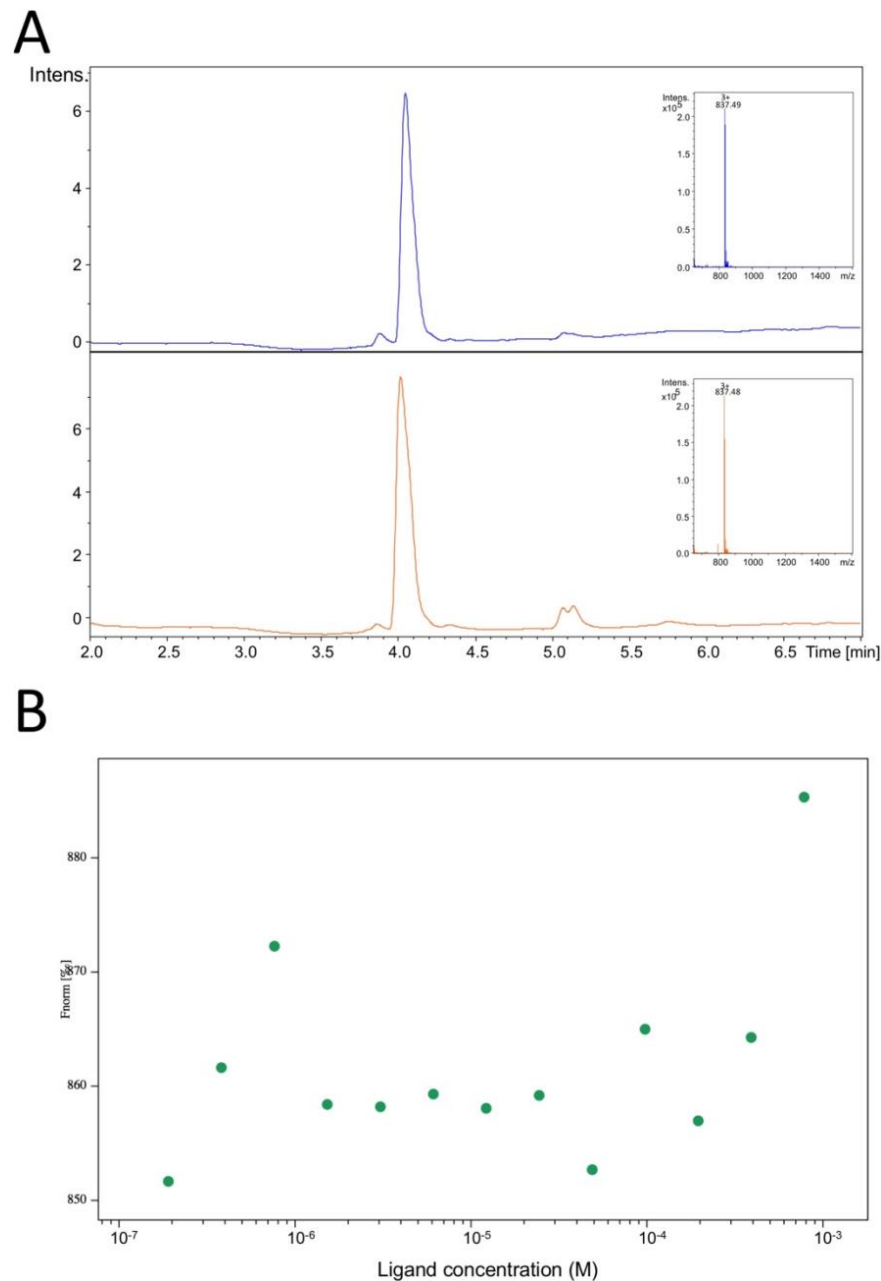
**Figure S1.** A) Size exclusion chromatography of CgnJ using a Superdex 200 pg 16/60 column. A minor portion of the protein aggregates during purification (~50 mL elution volume) while the majority of the protein elutes as a symmetrical, single peak (~100mL elution volume) shortly before the elution volume reaches one column volume (121 mL) indicating a monomer in solution). B) SDS-PAGE of CgnJ after size-exclusion chromatography.



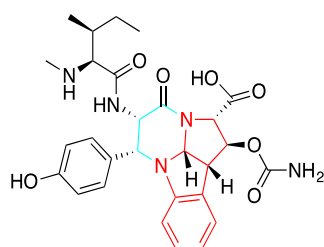
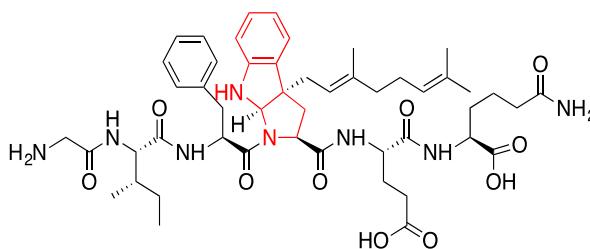
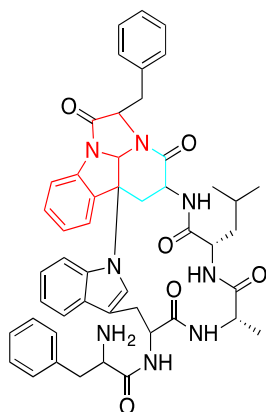
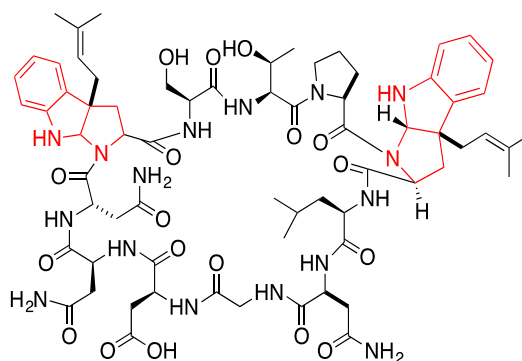
**Figure S2.** Dynamic light scattering measurement of CgnJ. The mean size and standard deviation of 45 runs was calculated using Microsoft Excel. The maximum size of a CgnJ monomer was measured to be ~45 Å using PyMOL.

CgnJ	MAE---EVAEIIIPASTWILFFDASCSINSPAFWSTNDAVDRIWRLKIAHELVLQVVLE	57
ComJ	MAIKSWKPQELSIYHQFTVFQKDS--TPPVMWDWD-EAIE-----K	39
	** : * : : : * . * * : : : :	
CgnJ	GYFKVRCIL---RSSAPAFEMVNADVSELVS-----IVLPSGRLV-ACTTDEPTLNRHV	107
ComJ	GYAADGAISFEAQRNTKAFILFRLNSSETVNSYEKKVTVPFHVTENGIHIESIMSKRLS	99
	** . : : : * * : . : * * . : : * . : . : *	
CgnJ	LTVPPIGRYRVLEWSVHEESKHYDVESAEAYPADEGPDGIITLWPER	154
ComJ	FDLPGDYLQ-LTCWTVPAEMSDLHAD---TYIIDAVSV-----	133
	: * * * : : * * * . . . : : * *	

**Figure S3.** Sequence alignment of CgnJ and ComJ calculated by *Clustal Omega*. The sequence identity of both proteins was calculated to be 15.6%.



**Figure S4.** A) LC-MS analysis of the precursor peptide CgnA (blue, top) and CgnA after the addition of CgnJ (orange, bottom). Both samples were incubated at 37 °C for 24 h prior to analysis. B) MST measurement of CgnA with CgnJ. No binding was observed.

**Crocagin A****ComX (RO-E-2)****Kapakahine B****Kawaguchipectin A**

**Figure S5.** Chemical structures of Crocagin A, ComX (RO-E-2), Kapakahine B and Kawaguchipectin A. Similar tryptophan residues are highlighted in red whereas a similar core structure in Crocagin A and Kapakahine B is marked in cyan.

## Chapter 8

# Initiation of crocagin biosynthesis – implications for a macromolecular biosynthetic complex

**Sebastian Adam**<sup>[a]</sup>, Andreas Klein<sup>[a]</sup>, Carsten Volz<sup>[b]</sup>, Rolf Müller<sup>[b]</sup> and Jesko Koehnke<sup>[a]\*</sup>

*Manuscript in preparation*

### Affiliation

<sup>[a]</sup> Workgroup Structural Biology of Biosynthetic Enzymes, Helmholtz Institute for Pharmaceutical Research, Helmholtz Centre for Infection Research, Saarland University, Universitätscampus E8 1, 66123, Saarbrücken, Germany.

<sup>[b]</sup> Department of Microbial Natural Products, Helmholtz Institute for Pharmaceutical Research, Helmholtz Centre for Infection Research, Saarland University, 66123, Saarbrücken, Germany.

## **Contributions and Acknowledgements**

### **Author's effort:**

The author contributed significantly to this work by establishing the recombinant protein expression and purification for CgnA, CgnB, CgnD and CgnE. He worked on protein crystallization of CgnD and CgnE and solved their respective x-ray crystal structures. He refined the crystal structures and performed structural analyses as well as part of the biochemical and biophysical studies. Furthermore, the author wrote the manuscript.

### **Other's effort:**

Andreas Klein contributed by establishing the first protein expression and purification for CgnC. He performed part of the biochemical and biophysical studies shown and analyzed the resulting data. Andreas also crystallized CgnB, solved and refined the x-ray crystal structure. Carsten Volz contributed by analyzing and interpreting initial biochemical data and provided helpful scientific insights.

## 8.1 Abstract

Advances in secondary metabolomics have led to the discovery of the first reported ribosomally synthesized and post-translationally modified peptides (RiPPs) from myxobacterial origin, the crocagins. In search for their biological function, they have recently been associated with a role in bacterial competency systems as the natural product and a protein from the biosynthetic gene cluster show structural homology to their respective counterparts in *Bacillus subtilis*. The order and steps of crocagin biosynthesis have thus far been enigmatic as only two functions of CgnB and CgnI could be investigated by *in vivo* knockout metabolomics. This work discusses the initiation of crocagin biosynthesis using a combination of structural, biochemical and biophysical *in vitro* data. We demonstrate the crucial role of CgnB as a putative scaffold protein, which potentially harbors a novel RiPP recognition element (RRE), and propose a  $\beta$ -oxidation of the core peptide tyrosine residue catalyzed by CgnC as the first reaction of the biosynthetic pathway. The high structural and sequence homology of CgnB and CgnE as well as protein-protein interaction studies imply a possible macromolecular biosynthetic complex in accordance with recent discoveries in other RiPPs pathways. Furthermore, the observation of a catalytic triad in a long narrow pocket of CgnD assembled by its dimeric interface hints at a proteolytic function as a serine protease, possibly cleaving the leader peptide.

## 8.2 Introduction

Compounds of a natural origin show remarkable chemical diversity and have consequently offered scientific researchers diverse opportunities for discovery in chemistry, biology and medicine.<sup>[1]</sup> As the biosynthetic genes responsible for the production of these natural products are usually organized in biosynthetic gene clusters, the reduced cost and increased throughput of DNA sequencing technologies have rapidly expanded the number of available microbial genomes and in turn have revealed a vast number of novel biosynthetic gene clusters.<sup>[2]</sup> Since the late 2000's, these breakthroughs have led to the emergence of a thus far novel class of natural products, the ribosomally synthesized and post-translationally modified peptides (RiPPs).<sup>[3]</sup> In the search for new natural products, their ribosomal origin makes them a very interesting target for genome-driven methodologies as their chemical structure is more relatable to genomic data compared to other natural product classes.<sup>[4]</sup> The application of

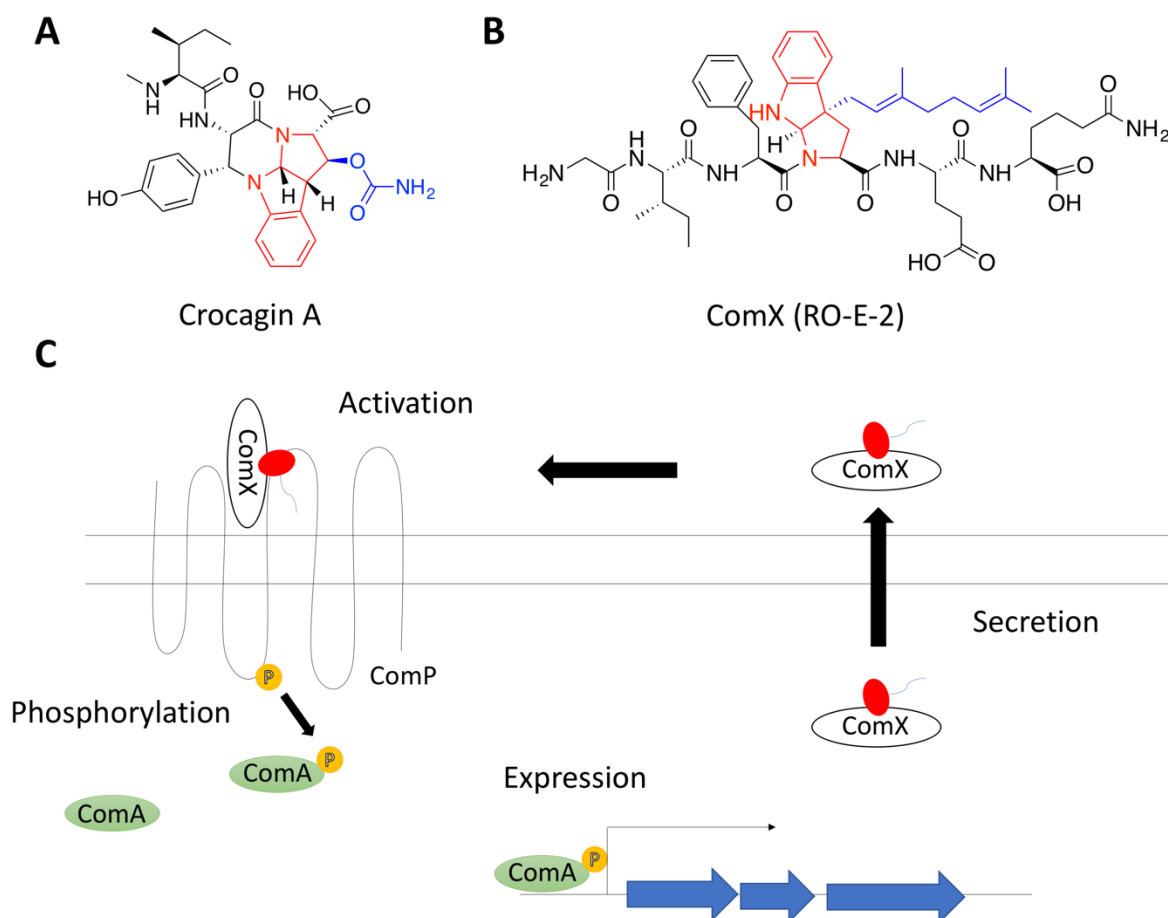
these technologies, in particular secondary metabolome mining, led to the discovery of the first reported myxobacterial RiPP family in extracts of *Chondromyces crocatus* Cm c5, the crocagins.<sup>[5]</sup>

They are comprised of a tetrahydropyrrolo-[2,3-b]-indole core structure, which resembles other natural products like comX,<sup>[6]</sup> kapakahine B<sup>[7]</sup> and kawaguchipectin A.<sup>[8]</sup> Kapakahines and kawaguchipectins have shown cytotoxic and antibacterial activity, respectively, while comX has been assigned a function in bacterial quorum sensing of *Bacillus subtilis*.<sup>[8][9][10]</sup> The latter is a secreted pheromone, which belongs to the ComQXPA quorum sensing system of *Bacillus* and is associated with inducing genetic competence in several *Bacillus* strains.<sup>[10][11]</sup> Interestingly, it has also been confirmed to be a ribosomally synthesized peptide containing a post-translational isoprenylation, which was found to be crucial for its activity.<sup>[12][13]</sup> Within the ComQXPA system, it acts as a secreted molecule targeting the transmembrane histidine kinase ComP, which in turn phosphorylates the response regulator ComA, ultimately leading to the expression of genes mediating the genetic competence in *B. subtilis*.<sup>[10]</sup> A schematic of the ComQXPA system as well as the chemical structure of crocagin A and comX are shown in **Figure 1**. In contrast, a biological function of crocagins could not be determined thus far as the biological relevance of crocagin binding to the conserved prokaryotic carbon storage regulator A (CsrA) in *in vitro* assays did not translate to in cellular activity as crocagin A does not undergo cellular uptake.<sup>[5]</sup> However, if we postulate a similar function of crocagins to that which has been confirmed for comX, we can assume a potential target to be extracellular, eliminating the need for cell penetration of these molecules. Intriguingly, a recent structural analysis of the domain of unknown function (DUF) protein CgnJ from the crocagin biosynthetic gene cluster revealed structural homology to ComJ, a nuclease inhibitor, which is also connected to horizontal gene transfer.<sup>[14]</sup> Still, the respective gene cluster architecture of crocagins is very different to that of the ComQXPA gene cluster, which has been linked to the aforementioned function of directed transcriptional activation.

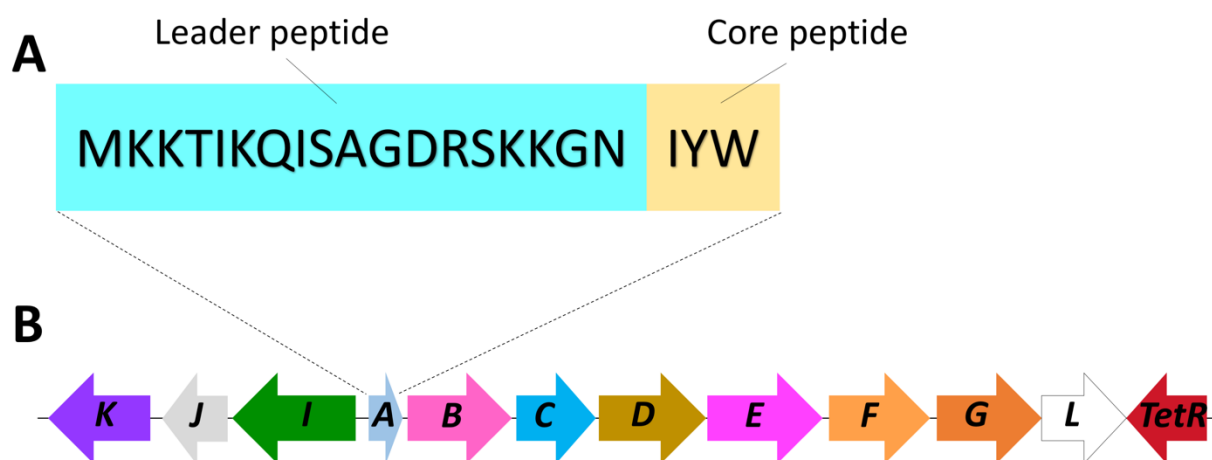
Being a RiPP, crocagins originate from the ribosomally synthesized precursor peptide CgnA, whose sequence is shown in **Figure 2A**. The corresponding hypothesis has been confirmed by moderately strong binding of CgnA to the protein directly downstream of CgnA in the gene cluster, CgnB, as well as an observed abolishment of crocagin biosynthesis in the respective *in vivo* knockout model of CgnB.<sup>[5]</sup> The linear precursor peptide CgnA then undergoes several post-translational modifications, which are usually carried out by tailoring enzymes located in



the biosynthetic gene cluster. All proteins found within the cluster borders together with their postulated function determined by sequence homology are visualized in **Figure 2B**.



**Figure 1:** A+B Chemical structures of crocagin A (A) and comX RO-E-2 (B), a similar core structure is highlighted in red. Crocagin A shows carbamoylation at the C $\beta$ , comX RO-E-2 shows isoprenylation at the C $\gamma$  of the respective tryptophane, highlighted in blue. C Schematic of the biological function of comX in *Bacillus* strains, acting as a secreted signal molecule.<sup>[10]</sup>

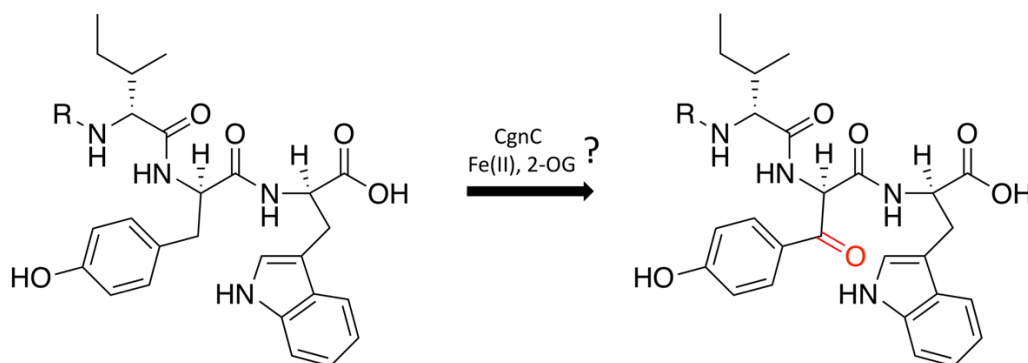


**Figure 2:** A Crocagin is derived from the ribosomally synthesized precursor peptide CgnA (livid), which consist of a leader peptide followed by the core peptide IYW. B Biosynthetic genes of the crocagin biosynthetic gene cluster. Potential biosynthetic enzymes include CgnB (pink) = transketolase, CgnC (light blue) = Dioxygenase, CgnD (mocha) = hydrolase, CgnE (magenta) = transketolase, CgnI (lime) =

## 260 Initiation of crocagin biosynthesis – implications for a macromolecular biosynthetic complex

carbamoyl transferase, CgnJ (grey) = domain of unknown function (DUF), CgnK (purple) = Aldo-keto-reductase, CgnL (white) = SAM methyl transferase and a transcriptional regulator (TetR, red). Probable export proteins are CgnF and G (shades of orange) = ABC transporter.

Contrary to initial expectations arising from its moderately strong binding, no turnover of the precursor peptide CgnA could be observed upon incubation with CgnB, leaving questions to its function within the biosynthetic pathway. Various studies unraveling the biosynthetic machineries of different RiPP pathways have indicated a potential role of scaffold proteins in binding the leader peptide and guiding the core peptide into the catalytic center of biosynthetic proteins. In a lot of cases, they contain a so-called RiPP recognition element (RRE), which has been assigned a function in various biosynthetic pathways (see **Chapter 4**).<sup>[15]</sup> Several examples of such an interaction include cyanobactin (LynD),<sup>[16]</sup> lasso peptide (B1)<sup>[17]</sup> as well as linear azoline-containing peptide (BalHC)<sup>[18]</sup> biosynthesis. Furthermore, structural analysis of the microcin B17 biosynthetic machinery has revealed the formation of a macromolecular complex of the scaffold protein MccB with a dehydrogenase, a YcaO domain enzyme, as well as the precursor peptide, which has also been suggested for the similar gene cluster responsible for klebsazolicin biosynthesis.<sup>[19][20]</sup> However, it should be noted that the crocagin biosynthetic genes show no conserved sequence homology to genes involved in either microcin B17 or klebsazolicin biosynthesis. Concerning crocagin biosynthesis, a postulated biosynthetic pathway starts out with an oxidation at the C<sup>β</sup> position of the core peptide tyrosine, enabling the subsequent nucleophilic attack by the indole nitrogen required for the crocagin core structure.<sup>[5]</sup> The respective reaction (**Figure 3**) is supposed to be carried out by a dioxygenase, with the most likely candidate in the gene cluster being CgnC that shows homology to the 2-oxoglutarate and Fe(II)-dependent superfamily of oxygenases (residues 26-131 of the original annotation).



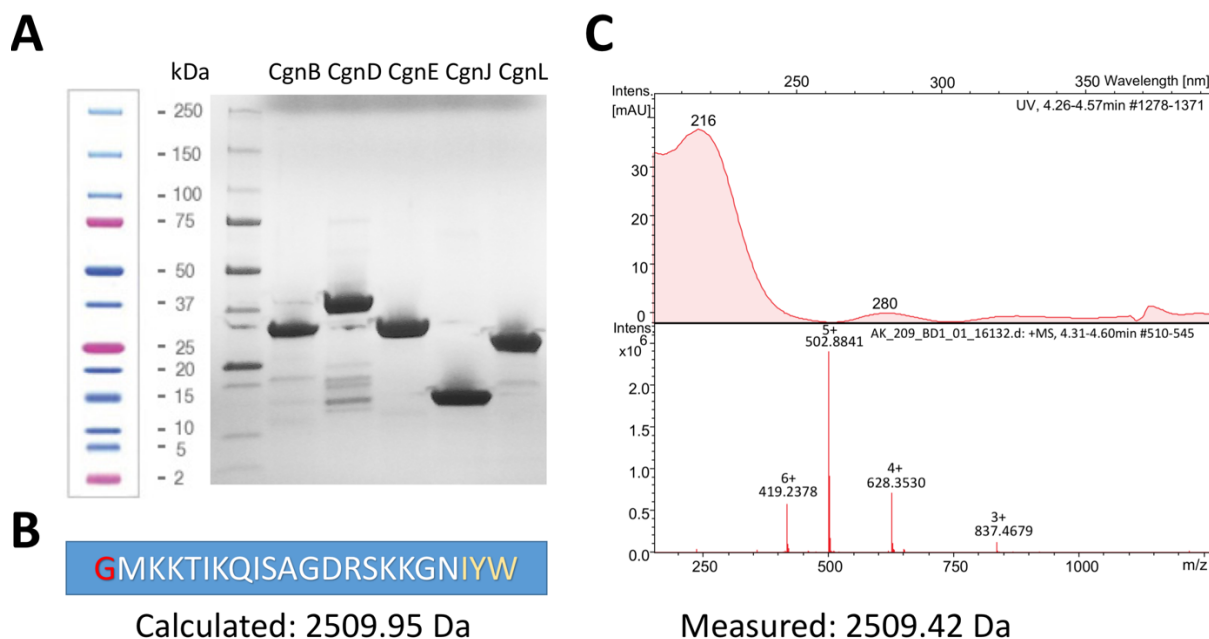
**Figure 3:** Proposed initiation of crocagin biosynthesis. CgnC is the only protein in the gene cluster which shares homology to oxygenases and is supposed to catalyze the C<sup>β</sup> oxidation of the tyrosine residue with cofactors Fe(II) and 2-OG. R = leader peptide.

Here, we present the first *in vitro* enzymatic reconstitution of a reaction on the precursor peptide CgnA and discuss the role of CgnB as a putative scaffold protein, potentially harboring a novel RiPP recognition element (RRE). The acquired biochemical data is backed up by structural analysis of CgnB and biophysical data suggesting a protein-protein interaction of CgnB with CgnC. Furthermore, the x-ray crystal structure of CgnE reveals an identical fold of both proteins, possibly hinting at a macromolecular biosynthetic complex in accordance with other RiPP clusters. Finally, the presence of a catalytic D-H-S triad on the surface of CgnD indicates a proteolytic function as a serine protease within the biosynthetic pathway.

## 8.3 Results and discussion

### *In vitro* reconstitution of the biosynthetic gene cluster

Starting out with almost no experimental data on the biosynthesis of crocagins, we attempted to clone and heterologously express the annotated biosynthetic genes (*cgnB-cgnE*, *cgnJ*, *cgnK*, *cgnL*) while excluding the annotated transporter genes *cgnF* and *cgnG*, the transcriptional regulator *tetR* as well as the carbamoyl-transferase with confirmed function *cgnI*. Protein expression protocols in *E. coli* could be established for CgnB, CgnD, CgnE, CgnJ and CgnL and the respective proteins purified to homogeneity (**Figure 4A**). In a similar fashion, the precursor peptide CgnA was cloned, expressed and purified with an additional N-terminal Glycine residue resulting from a cloning artefact. The resulting amino acid sequence of CgnA as well as mass spectrometric analysis of the purified peptide are shown in **Figures 4B+C**. The corresponding protocols and amino acid sequences of every construct can be found in the Supporting Information (**Table S1**).



**Figure 4:** **A** SDS-PAGE of the purified crocagin biosynthetic genes CgnB, CgnD, CgnE, CgnJ, CgnL, the precision plus dual xtra ladder was used as a molecular weight standard. **B** Amino acid sequence of the expressed and purified precursor peptide CgnA. An N-terminal glycine residue remains as a cloning artefact and is highlighted in red, the core peptide sequence is highlighted in yellow. The molecular weight of the construct was calculated by a peptide mass calculator. **C** Mass spectrometric analysis of purified CgnA. The measured mass differs from the calculated mass by 0.5 Da.

However, a reliable expression and purification protocol for CgnC, the dioxygenase homologue with a possibly crucial function in biosynthesis, could not be established for the annotated construct. A deeper look into the genomic data of *Chondromyces crocatus* Cm c5 revealed three GTG codons upstream of the annotated ATG start codon, which were in-frame with the *cgnC* sequence and have been shown to function as a start codon in *E. coli* in 14% of open reading frames.<sup>[21]</sup> The corresponding constructs were cloned and tested for protein expression, the respective amino acid sequences are shown in **Table S2**. The longest construct named CgnC<sup>L</sup> could be reliably expressed with a yield of 3 mg L<sup>-1</sup> in *E. coli* and purified to homogeneity (**Figure S1**).

### Biophysical investigation of *in vitro* CgnA binding

As the order of biosynthetic steps within the crocagin biosynthetic pathway was unknown, we set out to determine potential binding of single, purified proteins to the precursor peptide CgnA by means of Surface Plasmon Resonance (SPR) experiments. From the 6 proteins (CgnB-E, CgnJ, CgnL) tested, a moderate binding affinity of  $K_D = 2.04 \pm 0.11 \mu\text{M}$  could only be measured for CgnB, which is in agreement with the value of  $0.72 \pm 0.19 \mu\text{M}$  determined in an

earlier study using a synthetically produced precursor peptide (**Figure S2**).<sup>[5]</sup> Nevertheless, subsequent biochemical reactions of CgnA with CgnB using different reaction conditions and cofactors were unsuccessful. It should be noted that CgnB shows no sequence homology to a specific enzyme class. Therefore, we set out to solve the x-ray crystal structure of CgnB to obtain structural information on this enzyme and possibly examine substrate binding to CgnA in a complex structure, which could indicate a specific function of CgnB. Interestingly, CgnB shows sequence homology to another protein found within the biosynthetic gene cluster of crocagin, CgnE (**Figure S3**). The respective 38% identical amino acids and 60% positives observed when aligning both primary sequences led to the investigation of the crystallization properties of CgnE, as it may be used as a model to solve the structure of CgnB.

### **Crystallization process and data collection of CgnB and CgnE**

Initial crystallization trials of apo CgnB using various sparse matrix screens at different protein concentrations and temperatures yielded no crystallization hits. Given the moderate binding of CgnB to CgnA, we incubated CgnB with a molar excess of 1.25x CgnA for 2 h on ice before setting up crystallization trials, which led to an initial crystallization hit after 3 days of incubation at 4 °C using a well solution containing 170 mM ammonium acetate, 85 mM sodium acetate pH 4.6, 25.5% PEG 4000 and 15% glycerole. The crystallization condition could not be optimized, but fortunately yielded single crystals that were used for data collection at the ESRF beamline ID 23-1. The crystals belonged to space group  $P2_1$  and diffracted to 2.1 Å.

Similar to CgnB, initial crystallization trials of CgnE using the annotated construct at different protein concentrations and temperatures were unsuccessful. An investigation of the genomic data of the crocagin biosynthetic gene cluster revealed an alternative in-frame ATG start codon encoding for a formyl-methionine residue 16 amino acids upstream of the annotated start codon (**Figure S4**). The construct was named CgnE<sup>L</sup> and could be expressed and purified using the already established protocol for CgnE. CgnE<sup>L</sup> crystallized readily in several conditions, one of them being a condition with a well solution containing 200 mM magnesium chloride, 100 mM Tris pH 8.5 and 20% PEG 8000, which was used for further optimization. The crystals formed after 2 days of incubation at 18 °C, belonged to space group  $P2_1 2_1 2_1$  and diffracted to 2.0 Å. Due to a combination of missing molecular replacement models for both CgnB and CgnE<sup>L</sup> and CgnB crystals not being reproducible, we decided to use the well established seleno-methionine expression for CgnE<sup>L</sup> to obtain the necessary phase

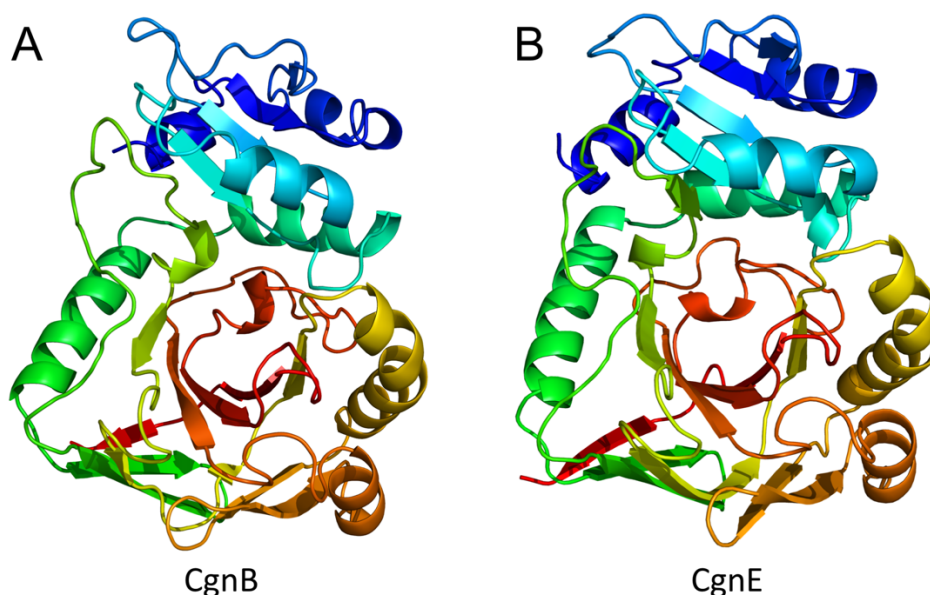
information required for solving the structure.<sup>[22]</sup> This method employs the substitution of all methionine with seleno-methionine residues during recombinant protein expression and allows for the use of selenium atoms as anomalous scatterers during data collection. CgnE<sup>L</sup>\_SeMet crystals could be reproduced in the same condition in which the native protein was crystallized, diffracted to a similar resolution limit and were used for data collection at SLS beamline X10SA at a wavelength close to the K-edge of selenium. The structure could be solved with phase information obtained by selenium single anomalous dispersion (Se-SAD) data, went through several rounds of manual model building and refinement and was subsequently used to solve the CgnB structure with molecular replacement. Unfortunately, an existing significant non-crystallographic symmetry operator (NCS) of the CgnB crystals led to a large off-origin patterson peak and therefore a high  $R_{\text{free}}$  value in refinement, even though the other statistics show the model has been built correctly. The data collection and refinement statistics of both CgnB\_native and CgnE<sup>L</sup>\_SeMet can be found in **Table 1**.

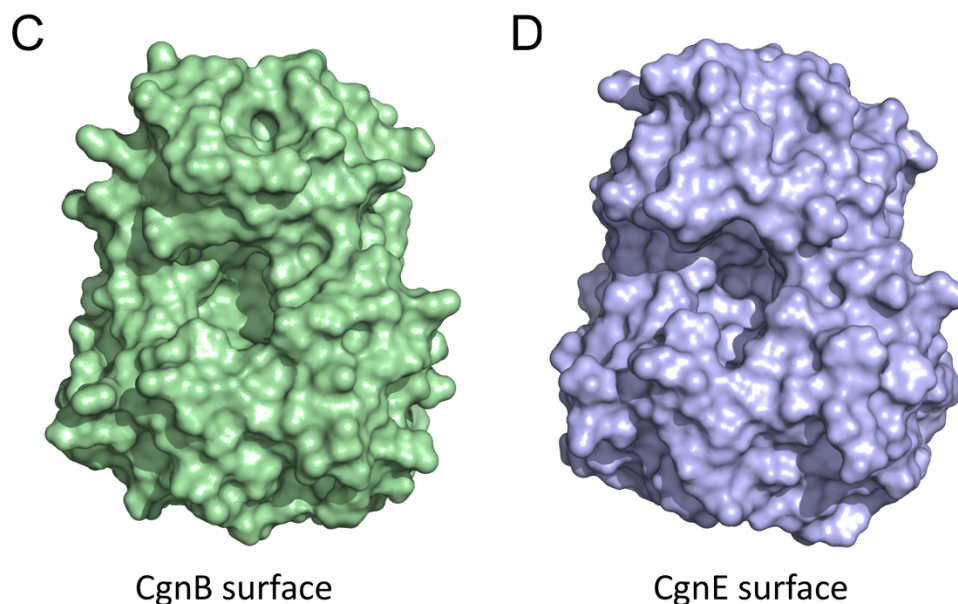
**Table 1:** Data collection and refinement statistics of CgnB\_native and CgnE<sup>L</sup>\_SeMet. Statistics for the highest-resolution shell are shown in parentheses.

	CgnB native	CgnE <sup>L</sup> SeMet
Resolution range [Å]	36.17 – 2.3 (2.38 – 2.30)	45.06 – 2.0 (2.05 – 2.0)
Space group	P 2 <sub>1</sub>	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Unit cell		
<i>a</i> , <i>b</i> , <i>c</i> [Å]	52.45 67.50 86.15	70.40 75.34 112.44
<i>α</i> , <i>β</i> , <i>γ</i> [°]	90.00, 95.99, 90.00	90.00 90.00 90.00
Total reflections	74 738 (10971)	521 638 (39 045)
Unique reflections	25 752 (3773)	41 166 (2997)
Multiplicity	2.90 (2.90)	12.7 (13.0)
Wavelength [Å]	0.97	0.97
Completeness (%)	96.50 (97.40)	99.90 (99.80)
Mean <i>I</i> / <i>σ</i> ( <i>I</i> )	9.8 (2.1)	16.9 (7.0)
Wilson B-factor	42.40	19.62
<i>R</i> <sub>merge</sub>	0.060 (0.454)	0.117 (0.409)
<i>R</i> <sub>work</sub>	0.2936 (0.3913)	0.1497 (0.1704)
<i>R</i> <sub>free</sub>	0.3154 (0.4190)	0.1986 (0.2355)
Number of non-hydrogen atoms	4692	5118
Macromolecules	4632	4597
Ligands		
Water	60	521
Protein residues	593	612
RMS (bond lengths)	0.005	0.007
RMS (angles)	1.04	0.86
Average B-factor	58.50	25.63
Macromolecules	58.61	24.96
Ligands		
Solvent	50.25	31.55
Anomalous completeness		99.90 (99.50)
Anomalous multiplicity		6.6 (6.7)
DelAnom correlation between half-sets		0.409 (0.055)
Mid-slope of Anom Normal Probability		1.120

## Structural analysis of CgnB and CgnE

The comparison of the x-ray crystal structures of CgnB and CgnE revealed many similarities between both proteins, some of which had already been suggested by the apparent sequence homology. CgnB and CgnE<sup>L</sup> both contain mostly acidic side chains (net charges of -21 for CgnB and -18 for CgnE<sup>L</sup>) and a theoretical pI of 4.6 and 4.8, respectively. Concerning potential interactions with the CgnA leader peptide, one could imagine the formation of salt bridges of these side chains with the 6 positively charged side chains of the CgnA leader peptide (**Table S3**). Both proteins are comprised of 309 amino acids in their original annotation (even though a longer construct had been crystallized for CgnE) and contained 2 monomers in their respective asymmetric unit. Convincing electron density could be detected for almost all residues of the two chains in both proteins with just minor truncations to be found at the respective N- and C-termini. An analysis of the protein assembly found in the crystal structure with the *PISA* server<sup>[23]</sup> did not suggest a complex formation for neither CgnB nor CgnE<sup>L</sup>. One monomer of each protein in a similar orientation is portrayed as a rainbow cartoon in **Figure 5A+B**, with the N-terminus shown in blue and the C-terminus shown in red, respectively. The corresponding surface representations of CgnB (green) and CgnE (blue) are shown in **Figure 5C+D**.

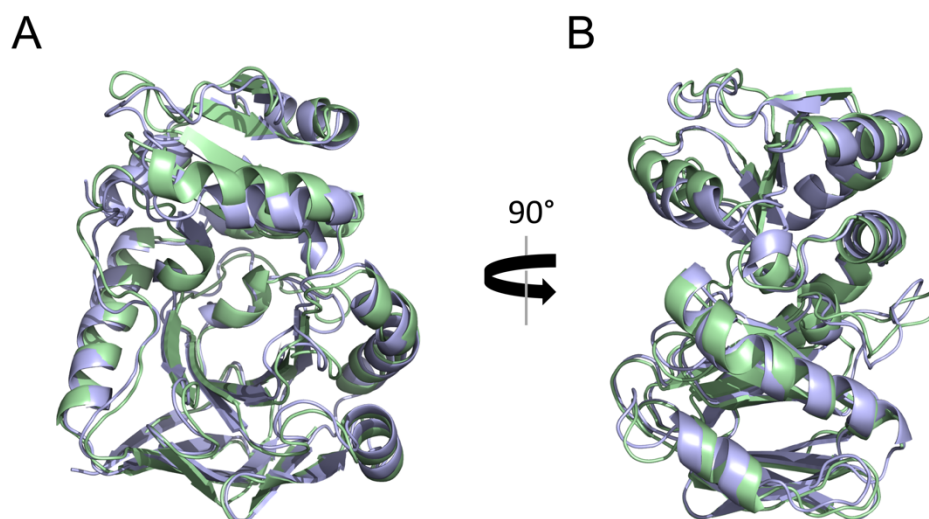




**Figure 5:** **A** Cartoon representation of CgnB portrayed as a rainbow, with the N-terminus shown in blue and the C-terminus shown in red. **B** Cartoon representation of CgnE<sup>L</sup> portrayed as a rainbow, with the N-terminus shown in blue and the C-terminus shown in red. **C** Surface representation of CgnB. **D** Surface representation of CgnE<sup>L</sup>.

Portraying both proteins in a comparable orientation (**Figures 5A+B**) already demonstrates an identical fold. Four conserved pairs of antiparallel  $\beta$ -sheet are building up the C-terminus of the proteins and are flanked by two  $\alpha$ -helices (green and yellow) to each side. The surface representations (**Figures 5C+D**) propose a potential active side to be present in the middle of this structure with a substrate binding pocket opening towards the green helix, while the respective N-termini are closing the pocket on top of it. A superposition of both proteins (**Figures 6A+B**) confirms a very high structural homology of CgnB (green) with CgnE<sup>L</sup> (blue) amounting to a  $C_{\alpha}$  rmsd of 1.2 Å over 281 atoms, minor differences in secondary structure or loop orientation can only be found regarding the N-terminus.





**Figure 6:** **A** Superposition of a cartoon representation of CgnB (green) with CgnE<sup>L</sup> (blue). The C<sub>α</sub> rmsd was calculated to be 1.2 Å over 281 atoms. **B** Superposition of a cartoon representation of CgnB (green) with CgnE<sup>L</sup> (blue), 90° rotation from **A**.

Intriguingly, no convincing electron density for the precursor peptide CgnA is found in the CgnB structure even though it was expected to be bound due to the differences in crystallization properties observed for CgnB. However, it is possible that CgnA acted as an additive to achieve crystallization of CgnB rather than being actually bound in the crystal structure of CgnB. In a study conducted using 81 different proteins, 200 different small molecules were found to promote crystallization of 35 proteins, which could not be crystallized without an additive, an effect which is believed to stem from establishing intermolecular, non-covalent crosslinks in protein crystals promoting lattice formation.<sup>[24]</sup> Consequently, there is no electron density observed for many additives in the crystal structure of the protein. Furthermore, we have to consider two more factors which could lead to an absence of electron density for CgnA. First, an interaction of CgnB with CgnA could be influenced by temperature, resulting in the required peptide concentration for a complex structure at 4 °C being higher than the 1.25 x molar excess that was used for pre-incubation. Second, the existence of significant non-crystallographic symmetry in these crystals reduces the quality of the respective electron difference map, which leads to a problem when interpreting the data in search for peptide density. Therefore, a higher peptide concentration and a different crystal form should be used in order to obtain a CgnB-CgnA complex structure, which could enable us to further investigate the role of CgnB in crocagin biosynthesis. Additionally, a complex structure would potentially reveal a novel RiPP recognition element (RRE) in the CgnB structure that is essential for leader peptide binding. As for CgnE, it remains

to be seen why we do not observe binding to CgnA in the context of the very high structural homology to the moderate binder CgnB. Given these results, we can hypothesize about a common leader peptide binding mechanism of both proteins combined with a specific binding of the core peptide residues. As such, binding of CgnE to CgnA should only be possible once a crucial post-translational modification of the core peptide has occurred, which gets specifically recognized by CgnE residues.

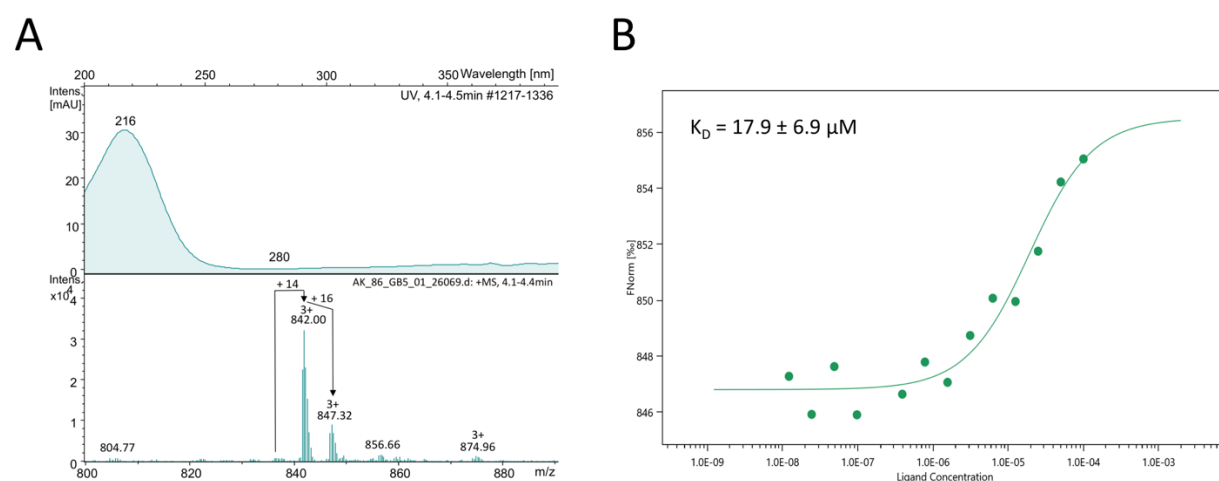
Unfortunately, the structural information obtained for CgnB did not indicate any specific function in biosynthesis as an analysis of structural homology with the *DALI* server<sup>[25]</sup> did not reveal homology to a specific enzymatic class. Therefore, the question remained in which way CgnB essentially contributes to crocagin biosynthesis as an *in vivo* knockout model of CgnB completely abolishes production of crocagin.<sup>[5]</sup> Consequently, we envisioned a potential protein-protein-interaction of CgnB with other biosynthetic enzymes of the pathway as a potential function, possibly leading the CgnA core peptide into the active site of subsequent enzymes.

### **Initiation of crocagin biosynthesis by CgnC<sup>L</sup> – macromolecular complex formation with the scaffold protein CgnB?**

As the proposed biosynthetic pathway of crocagins starts with an oxygenase reaction (**Figure 3**), we investigated a potential *in vitro* catalytic turnover of CgnA with CgnC<sup>L</sup>, the only oxygenase homologue of the biosynthetic gene cluster. Upon incubation of CgnA with CgnC<sup>L</sup> together with cofactors 2-oxoglutaric acid, ascorbic acid as well as FeCl<sub>2</sub>, we observed a second peak with a mass shift of +16 Da, although only with minimal conversion (**Figure S5**). This mass increase could indicate the hydroxylation of CgnA. All attempts at optimizing this reaction through changes of cofactor, peptide or protein concentrations, buffer exchanges or temperature shifts were unsuccessful. Next, we investigated a potential combination of CgnC<sup>L</sup> and CgnB for efficient catalysis of this oxygenation reaction. If CgnB contains a novel RiPP recognition element (RRE), its leader peptide binding could guide the core peptide into the active site of CgnC<sup>L</sup>, facilitating a conversion of CgnA.

The addition of a 2:1 molar ratio of CgnB:CgnC<sup>L</sup> to the same reaction resulted in complete turnover of the substrate after 1 h of incubation, suggesting an increase in the reaction rate. Interestingly, we now observed a main product with a mass shift of +14 Da compared to the +16 Da we saw beforehand, indicating the addition of a keto function to the peptide rather

than a hydroxy group (**Figure 7A**). Fascinated by these findings, we performed binding affinity analyses of CgnB with CgnC<sup>L</sup> by means of Microscale Thermophoresis (MST) and observed a moderate affinity protein-protein interaction with a  $K_D$  of  $17.9 \pm 6.9 \mu\text{M}$  (**Figure 7B**). Though we are missing data points at saturation, the experiment was reproducible and suggests specific binding between the two proteins. However, the existence of a mass peak with a mass shift of +30 Da hinted at the generation of a second product in this reaction with a hydroxy as well as a keto group, leading to the hypothesis of a somewhat not fully controlled or completed reaction. The examination of a control reaction without the addition of CgnC<sup>L</sup>, but with CgnB and cofactors added to CgnA, showed little turnover to a product peak with a mass shift of +16 Da in a similar fashion as it had been observed for CgnC<sup>L</sup> and cofactors without the addition of CgnB (**Figure S6**). These results suggest an unspecific oxygenation of the precursor peptide as a side product of the CgnC<sup>L</sup> reaction, a phenomenon which could not be prevented by optimizing the reaction parameters.

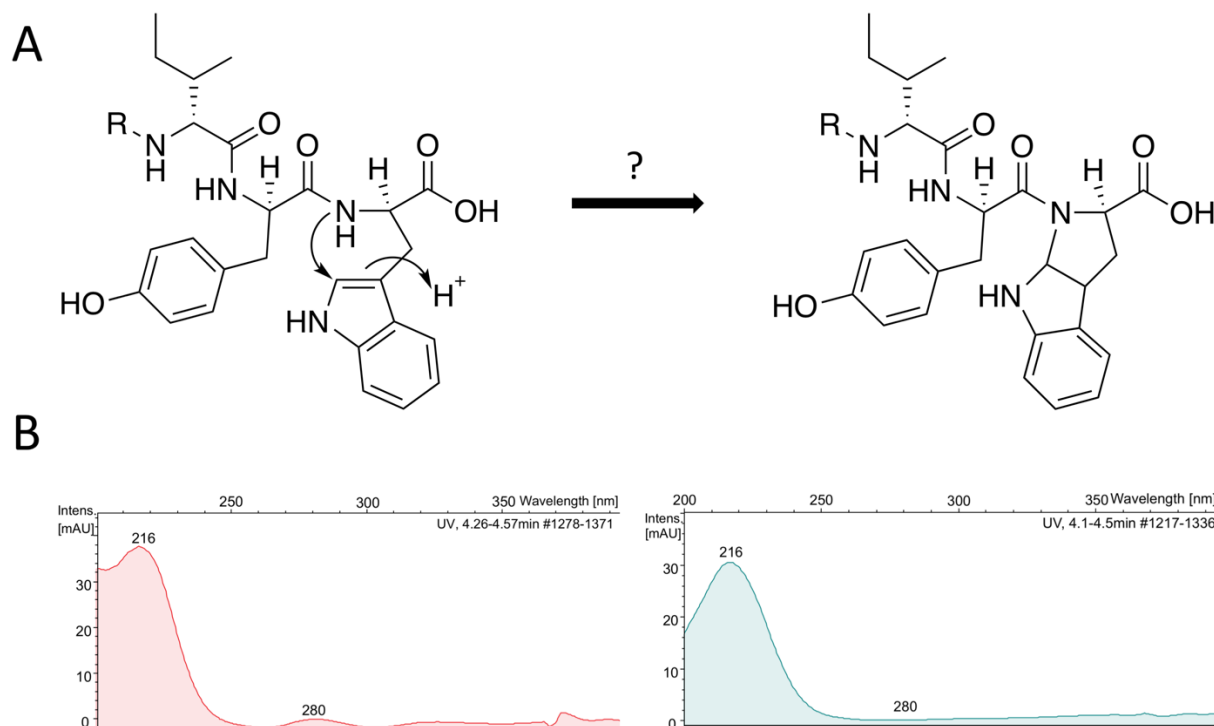


**Figure 7:** **A** Mass spectrometry analysis of a reaction mixture of CgnA, CgnB and CgnC<sup>L</sup> incubated with cofactors 2-oxoglutaric acid, FeCl<sub>2</sub> and ascorbic acid. The product shows significantly less UV absorption at 280 nm and primarily a mass shift of +14 Da. However, a second product peak with a mass shift of +30 Da appeared as well. **B** Microscale Thermophoresis measurement of CgnB with CgnC<sup>L</sup>. Even though we are missing points in saturation state, the result suggests a moderate protein-protein interaction.

In an attempt to purify the reaction product CgnA<sup>BC</sup> for further analysis using an FPLC, we noticed a virtually entire loss of UV absorption at 280 nm in fractions containing the product. Looking at the full sequence of the precursor peptide CgnA, the bulk of the UV absorption at 280 nm originates from the delocalized  $\pi$ -electrons of the tryptophane residue found within the core peptide. Consequently, a strong decrease in UV absorption at 280 nm is likely to correlate with a post-translational modification of this residue that destroys aromaticity, a

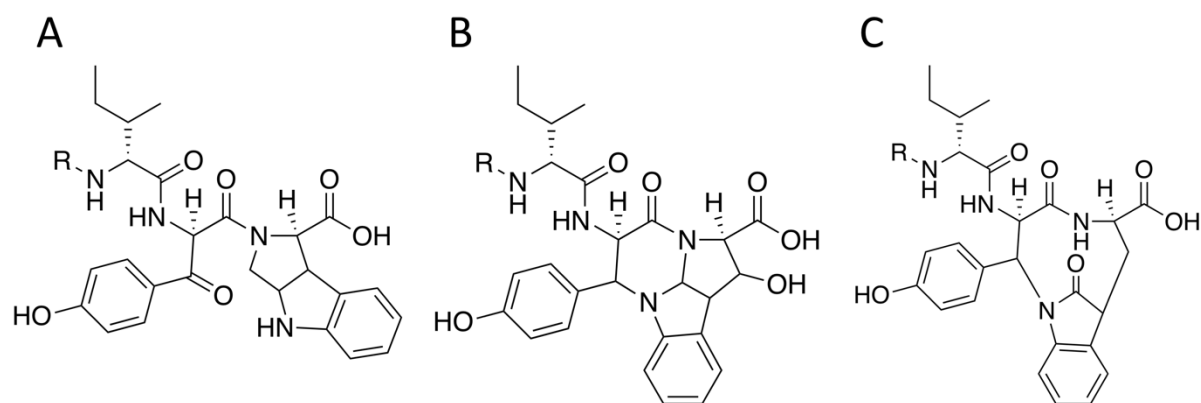
## 270 Initiation of crocagin biosynthesis – implications for a macromolecular biosynthetic complex

hypothesis also fitting to the observation of the final natural product having a very low absorption at this wavelength.<sup>[5]</sup> Looking at the proposed biosynthetic pathway, one reaction is supposed to be the attack of the backbone amide nitrogen onto a carbon of the pyrrole ring, leading to a cyclisation and furthermore the loss of the respective double bond within the pyrrole ring (**Figure 8A**).



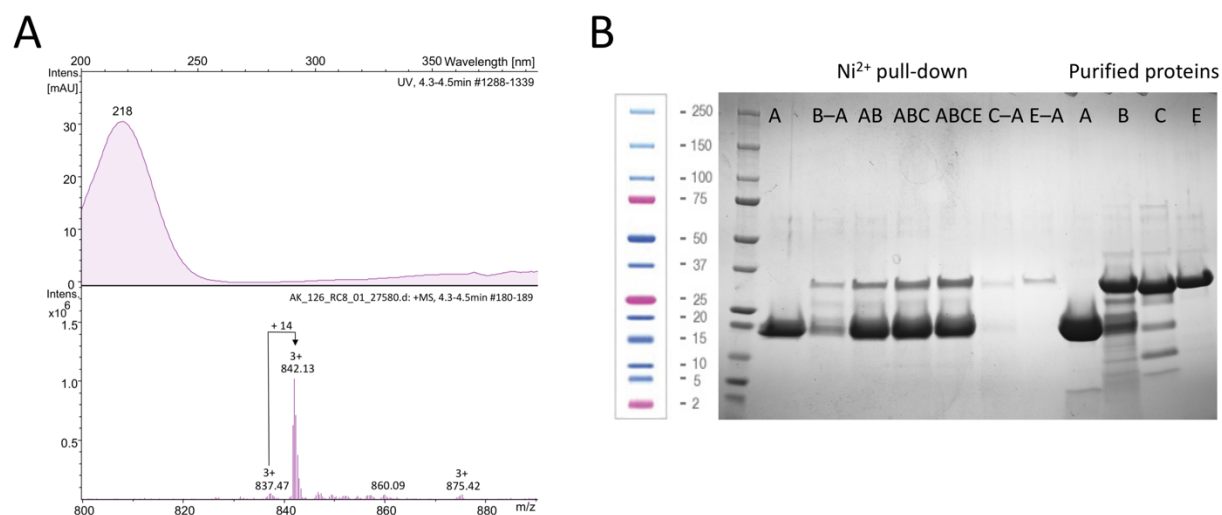
**Figure 8:** **A** Proposed ring closure reaction of the crocagin biosynthetic pathway, leading to the loss of the double bond in the pyrrole ring. It is unclear if this reaction is catalyzed by an enzyme. **B** Comparison of the UV chromatograms of the CgnA precursor peptide (left) with the product CgnA<sup>BC</sup> (right). We observe a loss of absorption at 280 nm for the product.

If we take a look at the measured UV chromatogram of the product of a reaction of CgnC<sup>L</sup> in combination with CgnB (**Figure 8B**), we can also observe the apparent loss of UV absorption at 280 nm (right) compared to the measurement of CgnA (left). Therefore, it is possible that both the addition of a keto function as well as the proposed ring closure reaction took place during the incubation of CgnA with CgnC<sup>L</sup> and CgnB. However, there are two additional possible products of this reaction, all 3 of which would show the exact mass shift of +14 Da in combination with a loss of UV absorption at 280 nm (**Figure 9ABC**). Intriguingly, one of these reaction products already contains the characteristic crocagin core scaffold (**Figure 9B**), in combination with the addition of a hydroxy group to the  $\beta$ -carbon of the tryptophane residue.



**Figure 9:** Possible chemical structures of CgnA<sup>BC</sup>, all of which would show a characteristic mass shift of +14 Da and a loss of UV absorption at 280 nm.

Both the involvement and contribution of CgnB and CgnC<sup>L</sup> to these potential reactions on the precursor peptide as well as their biosynthetic order is unknown thus far. In this context, it is noteworthy that the proposed reaction in **Figure 8A** does not involve a mass shift and can hence not be distinguished from the product without a ring closure reaction by LC-MS, except for a change in UV absorption. So far, a confirmation of any specific reaction product by tandem MS/MS did not yield conclusive results. Therefore, a thorough examination of the position of addition of a keto group or hydroxy group as well as possible cyclisation reactions await a product analysis by NMR studies. Interestingly, the addition of CgnE<sup>L</sup> to the mixture in a molar equivalent to CgnC<sup>L</sup> seems to reduce the amount of unspecific oxygenation in mass spectrometry analysis (**Figure 10A**). Due to the high structural homology of CgnB to CgnE<sup>L</sup>, we can imagine an exchange mechanism concerning CgnA between both proteins once the (first) CgnC<sup>L</sup> (and possibly CgnB) reaction(s) are complete. As mentioned before, this functionality would necessitate a specific release of CgnA<sup>BC</sup> by CgnB after catalysis and in turn a specific recognition of the modified core peptide residues by CgnE<sup>L</sup>, probably facilitating the next reaction of the pathway. In conclusion, we should be able to observe a similarly strong binding of CgnE to CgnA<sup>BC</sup> as it was shown for CgnB and CgnA in biophysical assays in the future. This binding of CgnA<sup>BC</sup> by CgnE<sup>L</sup> is possibly a regulation mechanism, which stops further unspecific oxygenation to occur once the CgnC<sup>L</sup> reaction is complete.



**Figure 10:** **A** Mass spectrometric analysis of a reaction mixture of CgnA, CgnB, CgnC<sup>L</sup> and CgnE<sup>L</sup> incubated with cofactors 2-oxoglutaric acid, FeCl<sub>2</sub> and ascorbic acid. The product shows significantly less UV absorption at 280 nm and primarily a mass shift of +14 Da. **B** Examination of a possible macromolecular complex formation of CgnABCE with a Ni<sup>2+</sup> pull-down assay followed by an SDS-PAGE. The first 7 lanes show the results of the assay with either a pre-incubated combination of the different untagged proteins with the His<sub>6</sub>-tagged SUMO-CgnA or using just the untagged proteins (X-A). The last 4 lanes show the purified proteins as a comparison. The precision plus dual xtra ladder was used as a protein standard.

A first examination of a possible macromolecular assembly has been conducted with a Ni<sup>2+</sup> pull-down assay followed by an SDS-PAGE using only His<sub>6</sub>-tagged SUMO-CgnA and purified CgnB, CgnC<sup>L</sup> and CgnE as possible binders (**Figure 10B**). A pre-incubation of the combinations B, BC and BCE with SUMO-A led to an increased intensity of the respective band after the pull-down assay compared to adding the respective non-His<sub>6</sub>-tagged proteins and running them in the pull-down assay (Lanes B-A, C-A, E-A). Due to the protein sizes on SDS-PAGE being too similar, a confirmation of protein-protein-interaction between BC and/or BCE awaits further analyses by dynamic light scattering, analytical gel filtration experiments, analytical ultracentrifugation or small angle x-ray scattering experiments. However, especially for the combination of ABC it seems as if we are observing a double band compared to running a combination of AB. Unfortunately, the protein concentrations of pulled down peptide-protein combinations were too low or not pure enough to measure the intact masses using mass spectrometry. Complex formation on the basis of protein-protein interaction has been reported for various RiPPs biosynthetic pathways including thiomuracin and different lanthipeptides.<sup>[26]</sup> If the formation of a macromolecular biosynthetic complex can be confirmed for the crocagin biosynthetic machinery, crystallization attempts should be carried

out in the same fashion as for microcin B17,<sup>[19]</sup> possibly shedding more light on the initiation of crocagin biosynthesis.

## **Crystallization and structure elucidation of CgnD**

In our process of *in vitro* reconstituting the biosynthetic gene cluster of crocagins, we could also express and purify CgnD (**Figure 4A**), the only protein in the biosynthetic gene cluster showing sequence homology to a hydrolase domain. As such, it represented a possible candidate for the removal of the CgnA leader peptide, an essential step in maturation of many RiPPs. Due to their function as proteases, they can serve as an *in vivo* regulation mechanism for natural product biosynthesis because enzymes dependent on leader binding will no longer act on pathway intermediates with a cleaved leader peptide, driving the equilibrium towards completion of the matured natural product. One example of this phenomenon was shown for Bottromycin biosynthesis, in which the amidohydrolase “AH” prevents reopening of the amidine macrocycle by the preceding enzyme “CD” that can act in both directions due to a fast cleavage of the follower peptide.<sup>[27]</sup>

A first biochemical examination of catalytic activity of CgnD by the addition of the protein to the aforementioned CgnABCE reaction mixture did not result in the emergence of two new peaks with different retention times and masses which would correlate to a cleaved CgnA<sup>BCDE</sup>. Next, we wanted to determine the crystal structure of CgnD to obtain further insights into its biosynthetic function as a potential protease. The original construct yielded crystals of poor quality that diffracted to 3.0 Å resolution in initial crystallization trials. Given the 13 lysine residues found in the primary amino acid sequence, we employed reductive lysine methylation to obtain a better crystal quality. This protein engineering method is based on the decrease of entropy on the surface of proteins where lysine side chains are commonly found, increasing the likelihood of important crystal contacts required for protein crystallization.<sup>[28]</sup> Lysine di-methylated CgnD yielded crystals that belonged to space group P2<sub>1</sub> 2<sub>1</sub> 2<sub>1</sub> and diffracted to 2.2 Å resolution at beamline X10SA located at the Swiss Light Source. However, we were not able to solve the structure due to not having a molecular replacement model with sufficient sequence homology available. Therefore, we applied Selenomethionine expression in a similar fashion as for CgnE and performed lysine methylation for CgnD\_SeMet. Lysine di-methylated CgnD\_SeMet yielded crystals that belonged to space group P2<sub>1</sub> and diffracted to 2.6 Å at beamline X06DA located at the Swiss Light Source. A high redundancy

## 274 Initiation of crocagin biosynthesis – implications for a macromolecular biosynthetic complex

dataset was used to solve the structure of CgnD using selenium single anomalous dispersion (Se-SAD) data. An initial model was built, which was subsequently used for molecular replacement of the native CgnD dataset with higher resolution. The native CgnD structure was subjected to several rounds of manual building and refinement. The data collection statistics of CgnD\_SeMet and the data collection and refinement statistics of CgnD\_native can be found in **Table 2**.

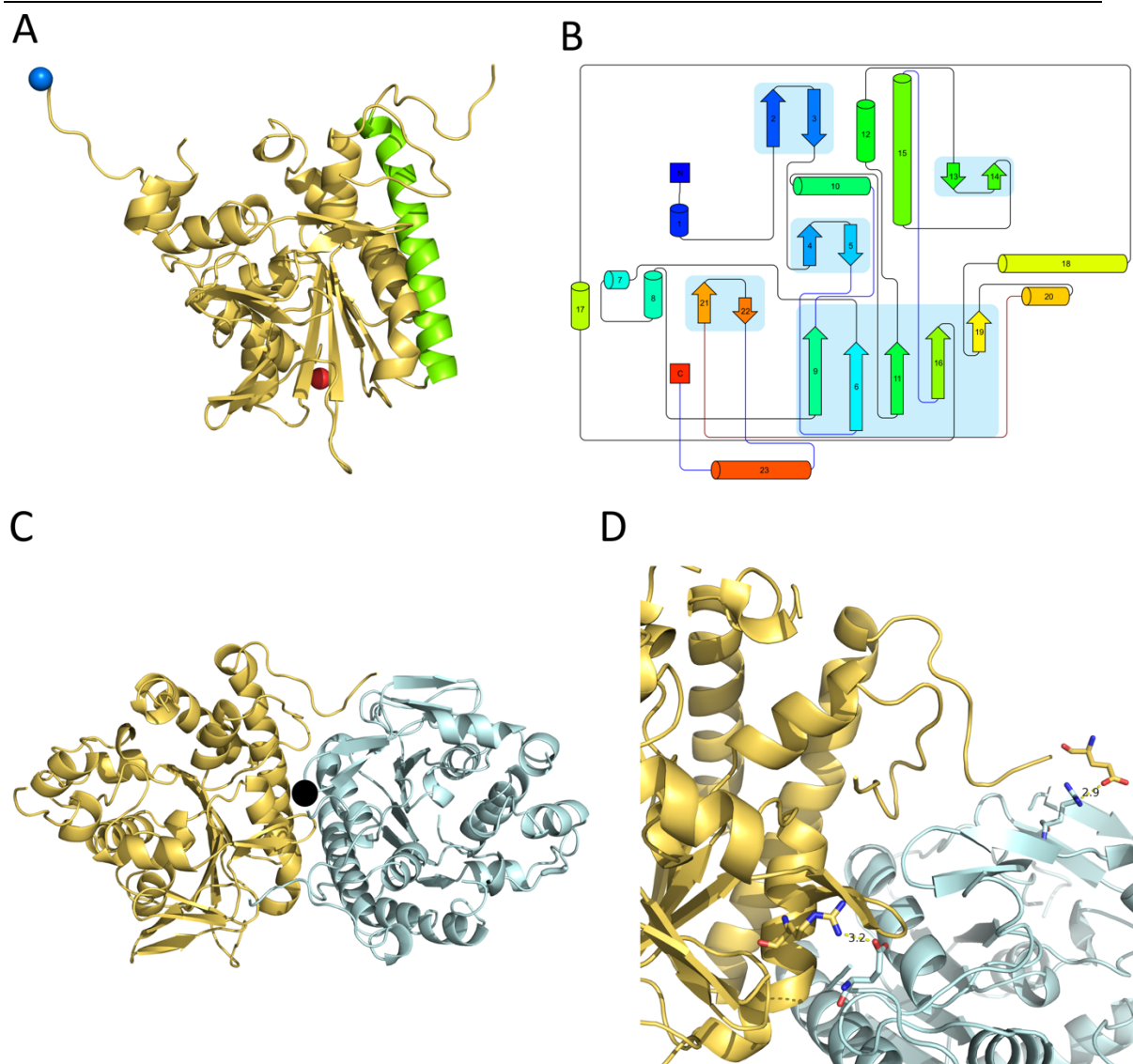
**Table 2:** Data collection and refinement statistics of CgnD\_native and CgnD\_SeMet. Statistics for the highest-resolution shell are shown in parentheses.

	CgnD native	CgnD SeMet
Resolution range [Å]	39.56 – 2.35 (2.41 – 2.35)	48.03 – 2.6 (2.74 – 2.6)
Space group	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P 2 <sub>1</sub>
Unit cell		
<i>a</i> , <i>b</i> , <i>c</i> [Å]	46.48 102.06 150.76	102.29 61.13 118.86
<i>α</i> , <i>β</i> , <i>γ</i> [°]	90.00 90.00 90.00	90.00 90.15 90.00
Total reflections	387 647 (24 851)	1 277 633 (189 583)
Unique reflections	30 717 (2158)	45 701 (6590)
Multiplicity	12.6 (11.5)	28.0 (28.8)
Wavelength [Å]	0.97	0.97
Completeness (%)	99.8 (98.2)	100.00 (99.90)
Mean <i>I</i> / <i>σ</i> ( <i>I</i> )	21.2 (8.0)	31.7 (8.8)
Wilson B-factor	23.75	
R <sub>merge</sub>	0.088 (0.313)	0.119 (0.467)
R <sub>work</sub>	0.1862 (0.2006)	
R <sub>free</sub>	0.2295 (0.2620)	
Number of non-hydrogen atoms	5018	
Macromolecules	4776	
Ligands	12	
Water	230	
Protein residues	594	
RMS (bond lengths)	0.004	
RMS (angles)	0.97	
Average B-factor	37.18	
Macromolecules	37.12	
Ligands	53.02	
Solvent	37.56	
Anomalous completeness		99.9 (99.9)
Anomalous multiplicity		14.1 (14.4)
DelAnom correlation between half-sets		0.548 (0.109)
Mid-slope of Anom Normal Probability		1.538



### Structural analysis of CgnD – a likely candidate for leader peptide hydrolysis

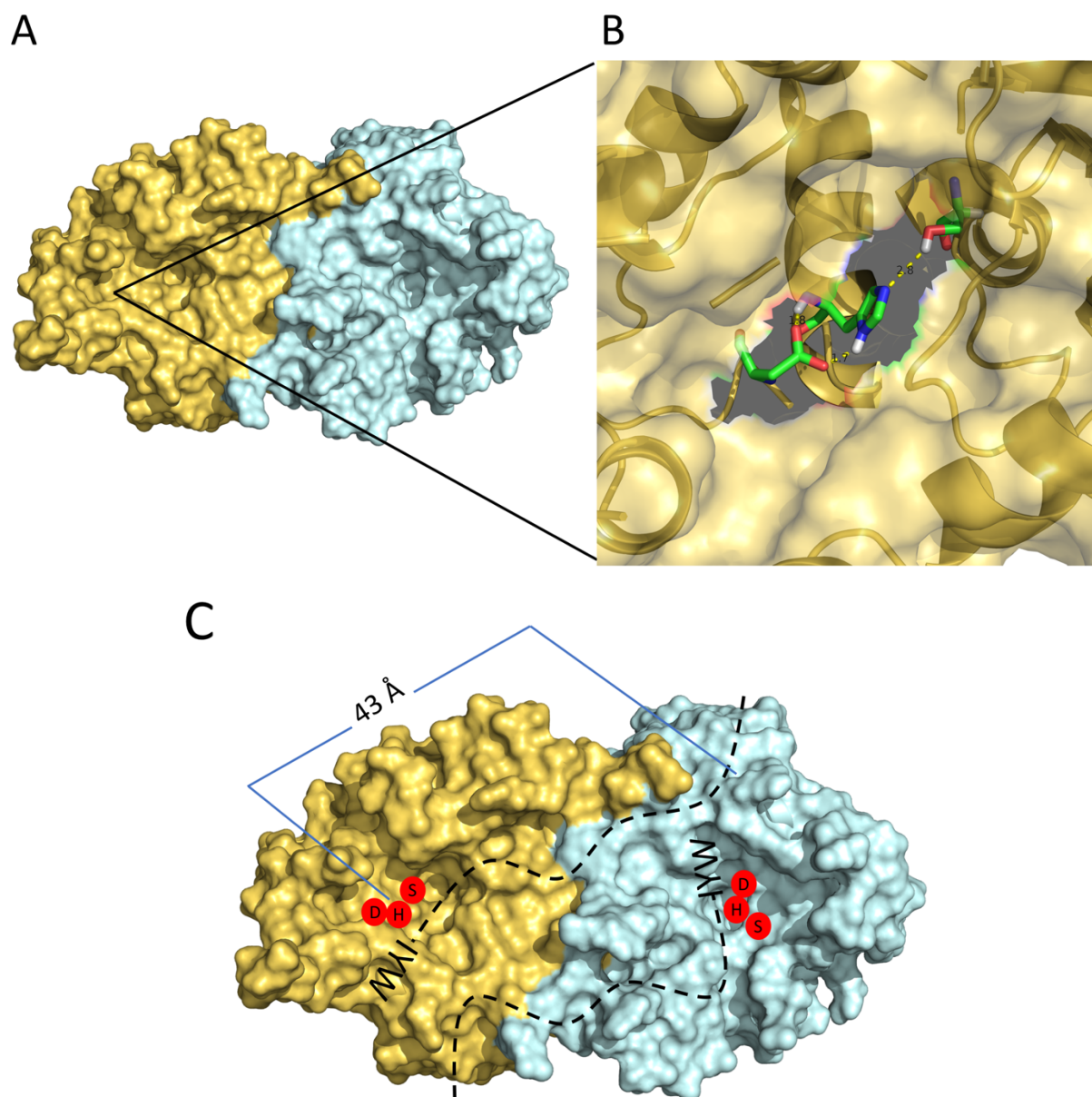
CgnD is a 38.7 kDa protein whose primary amino acid analysis shows a close to neutral net charge of -3 and consequently a slightly acidic theoretical pI of 6.13. The x-ray crystal structure of CgnD<sub>native</sub> contained two monomers in the asymmetric unit. We observed convincing electron density for residues 7-167, 201-265 and 268-347 of protomer A and for residues 1-161, 201-265 and 275-347 of protomer B, respectively. The missing portions of the protein in the crystal structure most likely represent flexible loop structures, as CgnD migrates at the expected full length of the construct on SDS-PAGE (**Figure 4A**). Protomer B is shown as a cartoon in **Figure 11A**, with a blue and red sphere representing the N- and C-terminus, respectively. The fold diagram of CgnD (**Figure 11B**) shows all secondary structure elements found within the structure. The secondary structure of CgnD consists of 5 parallel  $\beta$ -sheets (Numbers 6, 9, 11, 16, 19) constituting the core of the protein, which are flanked by larger  $\alpha$ -helices towards the surface of the protein, the most prominent being the long  $\alpha$ -helix (Number 15) highlighted in green spanning the entire length of the protein in **Figure 11A**. The display of both monomers in **Figure 11C** reveals several interactions of both monomers at the interface close to the  $\alpha_{15}$ -helix of each protomer. A protein assembly analysis of the structure with the PISA server<sup>[23]</sup> agrees with this hypothesis, as the dimer scores 0.974 in Complex Formation Significance Score (CSS) (with 1.000 being the maximum value). The macromolecular interface is built up by 7 hydrogen bonds and 9 salt bridges (**Table S4**) being formed between both monomers, amounting to a free dissociation energy of  $\Delta G^{\text{diss}} = 21.3 \text{ kcal mol}^{-1}$ . An analysis of potential symmetry operators in the crystal structure illustrated CgnD as an asymmetric homodimer, a composition which is rare in nature, but has been shown to be essential for the biological function of certain proteins.<sup>[29]</sup> As for CgnD functionality, it is possible that asymmetric dimer formation accounts for a specific binding of the CgnA leader peptide. One of the key residues involved in interface formation is Arg35, which forms a salt bridge with a respective Glutamate residue (either Glu201 or Glu203) of the other chain, contributing significantly to the orientation of the monomers on both sides of the interface (**Figure 11D**).



**Figure 11:** **A** Monomer of CgnD\_Native represented as a rainbow cartoon, with the N-terminus being shown in blue and the C-terminus being shown in red. **B** Fold diagram of CgnD\_Native, displaying secondary structure elements. **C** Asymmetric unit of CgnD\_Native. We observe 2 monomers in yellow and cyan, which are forming an interface at the long perpendicular  $\alpha_{15}$ -helices. **D** Salt bridge formation of Arg35 with either Glu201 or Glu203, contributing to the formation of a macromolecular assembly. The distances between the oxygen and nitrogen atoms of the respective side chains are shown.

The surface representation of both chains of CgnD (**Figure 12A**) further highlights the complex formation as both surfaces appear to be overlapping at the interface. We observe a long cleft spanning across the dimer interface, which allows for potential substrate coordination by a combination of residues from both monomers. The middle portion of each monomer accompanies a little deeper pocket, which is only accessible in one direction. A thorough examination of this pocket revealed the typical catalytic triad of a serine protease consisting of Ser79, His328 and Asp325, which is pointing towards the solvent (**Figure 12B**). Measurements of the typical hydrogen bond distances being formed between these residues

suggest an active conformation being involved in proteolysis when compared to distance variations observed in a statistical study analyzing catalytic triads, although the Ser-His distance is 0.1 Å higher than what is regularly observed (2.8 Å for Ser-His; 1.7 Å for His-Asp).<sup>[30]</sup>



**Figure 12:** **A** Surface representation of the two CgnD monomers shown in orange and cyan, respectively. The overlap of surfaces at the dimer interface further highlights a tight interaction between both monomers. **B** Catalytic triad consisting of Ser79, His328 and Asp325 found within the pocket of each monomer. The hydrogen bond distances between the residues are shown. **C** Schematic of the potential antiparallel binding of two substrate peptides by the CgnD dimer interface, utilizing the macromolecular assembly. The catalytic triad is shown in red, the substrate peptide is shown as a dashed, black line with the core peptide residues highlighted as letters.

Given our observation of two accessible catalytic triads at two distinctive positions of the macromolecular assembly, we can hypothesize about potential antiparallel binding of two substrate peptides being coordinated by the dimer interface (**Figure 12C**). Measuring the

length from the catalytic histidine residue to a potential entry site of the substrate peptide reveals a distance of 43 Å, meaning that the cleft could harbor 15-20 amino acids, depending on the final fold of the oligopeptide. Assuming our hypothesis to be correct, leader peptide binding necessitates dimer formation to a certain extent, as part of the leader peptide should be coordinated by the other monomer not being involved in catalysis. Additionally, proteolysis close to the core peptide is potentially facilitated by the binding of the leader peptide, which guides the core peptide into the active site of the enzyme in a similar fashion as it was shown for LynD.<sup>[16]</sup>

These results indicate CgnD as a likely candidate for leader peptide removal within the crocagin biosynthetic pathway, an essential step in the production of active natural products common to many RiPPs pathways. At the present time, it is still unknown which post-translational modifications are required for potential CgnD activity, allowing only for hypotheses of its role within the biosynthetic pathway. Regulatory functions of proteases have been assigned in the biosynthetic pathways of RiPPs for both fast (PurAH)<sup>[27]</sup> and slow (PatA)<sup>[31]</sup> kinetic properties, although the Ser-His distance of the CgnD catalytic triad probably hints at the latter alternative.

## 8.4 References

- [1] Martin SF *J Org Chem.* **2017** 82 (20):10757-10794
- [2] Zhang MM, Qiao Y, Ang EL, Zhao H *Expert Opin Drug Discov.* **2017** 12 (5):475-487
- [3] Poorinmohammad N, Bagheban-Shemirani R, Hamed J *Antonie Van Leeuwenhoek.* **2019** [Epub ahead of print]
- [4] Zhang Y, Chen M, Bruner SD, Ding Y *Front Microbiol.* **2018** 9:1801
- [5] Viehrig K, Surup F, Volz C, Herrmann J, Abou Fayad A, Adam S, Köhnke J, Trauner D, Müller R *Angew Chem Int Ed Engl.* **2017** 56 (26):7407-7410
- [6] Okada, M., Sato, I., Cho, S. J., Iwata, H., Nishio, T., Dubnau, D., Sakagami, Y. *Nature Chem. Biol.* **2005** 1 (1):23–24
- [7] Yeung BK, Nakao Y, Kinnel RB, Carney JR, Yoshida WY, Scheuer PJ, Kelly-Borges M *J Org Chem.* **1996** 61 (20):7168-7173
- [8] Ishida K, Matsuda H, Murakami M, Yamaguchi K *J Nat Prod.* **1997** 60 (7):724-6.
- [9] Nakao Y, Kuo J, Yoshida WY, Kelly M, Scheuer PJ *Org Lett.* **2003** 5 (9):1387-90
- [10] Okada M, Sugita T, Abe I *Beilstein J Org Chem.* **2017** 13 338-346
- [11] Oslizlo A, Stefanic P, Dogsa I, Mandic-Mulec I *Proc Natl Acad Sci U S A.* **2014** 111 (4):1586-91
- [12] Truman AW *Beilstein J Org Chem.* **2016** 12 1250-68
- [13] Tsuji F, Kobayashi K, Okada M, Yamaguchi H, Ojika M, Sakagami Y *Bioorg Med Chem Lett.* **2011** 21 (13):4041-4
- [14] Adam S, Klein A, Surup F, Koehnke J *Acta Crystallogr F Struct Biol Commun.* **2019** 75 (Pt 3):205-211
- [15] Burkhart BJ, Hudson GA, Dunbar KL, Mitchell DA *Nat Chem Biol.* **2015** 11 (8):564-70
- [16] Koehnke J, Mann G, Bent AF, Ludewig H, Shirran S, Botting C, Lebl T, Houssen W, Jaspars M, Naismith JH *Nat Chem Biol.* **2015** 11 (8):558-563
- [17] Weiz AR, Ishida K, Makower K, Ziemert N, Hertweck C, Dittmann E *Chem Biol.* **2011** 18 (11):1413-21
- [18] Dunbar KL, Chekan JR, Cox CL, Burkhart BJ, Nair SK, Mitchell DA *Nat Chem Biol.* **2014** 10 (10):823-9
- [19] Ghilarov D, Stevenson CEM, Travin DY, Piskunova J, Serebryakova M, Maxwell A, Lawson DM, Severinov K *Mol Cell.* **2019** 73 (4):749-762.e5
- [20] Travin DY, Metelev M, Serebryakova M, Komarova ES, Osterman IA, Ghilarov D, Severinov K *J Am Chem Soc.* **2018** 140 (16):5625-5633
- [21] Hecht A, Glasgow J, Jaschke PR, Bawazer LA, Munson MS, Cochran JR, Endy D, Salit M *Nucleic Acids Res.* **2017** 45 (7):3615-3626
- [22] Hendrickson WA, Horton JR, LeMaster DM *EMBO J.* **1990** 9 (5):1665-72
- [23] Krissinel E, Henrick K *J Mol Biol.* **2007** 372 (3):774-97
- [24] McPherson A, Cudney B *J Struct Biol.* **2006** 156 (3):387-406
- [25] Holm L, Laakso LM *Nucleic Acids Res.* **2016** 44 (W1):W351-5
- [26] Sikandar A, Koehnke J *Nat Prod Rep.* **2019** [Epub ahead of print]
- [27] Sikandar A, Franz L, Melse O, Antes I, Koehnke J *J Am Chem Soc.* **2019** [Epub ahead of print]
- [28] Walter TS, Meier C, Assenberg R, Au KF, Ren J, Verma A, Nettleship JE, Owens RJ, Stuart DI, Grimes JM *Structure.* **2006** 14 (11):1617-22

[29] Swapna LS, Srikeerthana K, Srinivasan N *PLoS One*. **2012** 7 (5):e36688

[30] Gupta V, Prakash NA, Lakshmi V, Boopathy R, Jeyakanthan J, Velmurugan D, Sekar K *Int J Biol Macromol*. **2010** 46 (3):317-23

[31] Lee J, McIntosh J, Hathaway BJ, Schmidt EW *J Am Chem Soc*. **2009** 131 (6):2122-4

## Supporting Information

# Initiation of crocagin biosynthesis – implications for a macromolecular biosynthetic complex

**Sebastian Adam**<sup>[a]</sup>, Andreas Klein<sup>[a]</sup>, Carsten Volz<sup>[b]</sup>, Rolf Müller<sup>[b]</sup> and Jesko Koehnke<sup>[a]\*</sup>

*Manuscript in preparation*

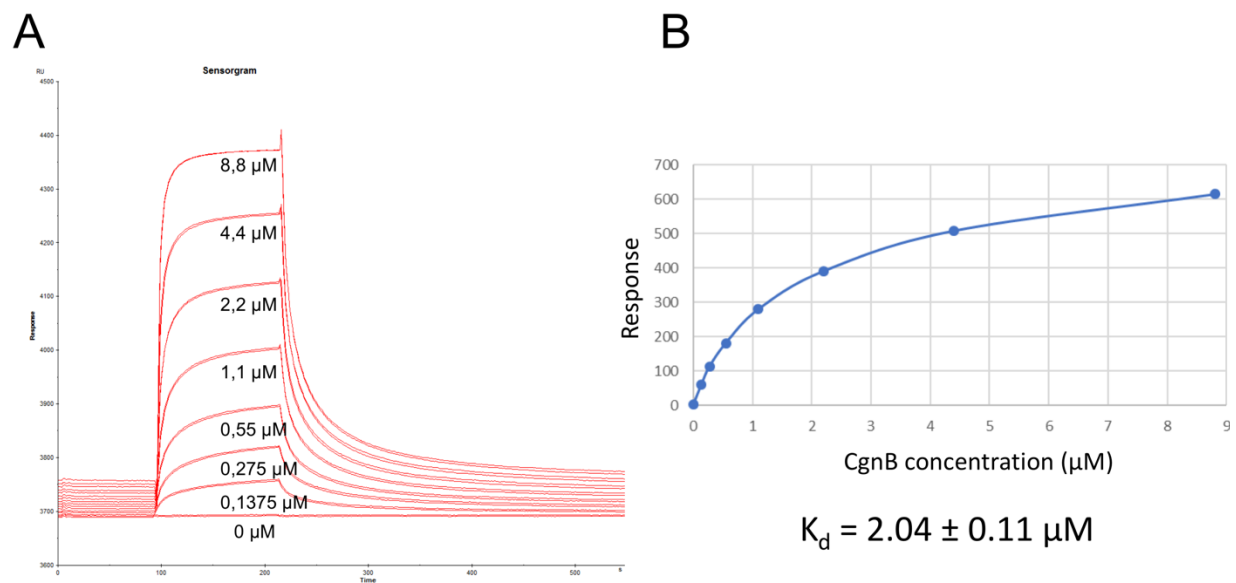
### Affiliation

<sup>[a]</sup> Workgroup Structural Biology of Biosynthetic Enzymes, Helmholtz Institute for Pharmaceutical Research, Helmholtz Centre for Infection Research, Saarland University, Universitäts-campus E8 1, 66123, Saarbrücken, Germany.

<sup>[b]</sup> Department of Microbial Natural Products, Helmholtz Institute for Pharmaceutical Research, Helmholtz Centre for Infection Research, Saarland University, 66123, Saarbrücken, Germany.



**Figure S1:** Left: Small scale expression tests of the different CgnC constructs shown in **Table S2** in *E. coli* Lemo21(DE3) using the pSUMO vector. Three different temperatures (16, 18, 25 °C) were tested, only the longest CgnC construct (CgnC<sup>L</sup>) shows expression in a Ni<sup>2+</sup> pull-down assay. Right: Lane 1 shows purified CgnC<sup>L</sup> after gel filtration. In both gels, precision plus dual xtra ladder was used as a standard.



**Figure S2:** SPR binding affinity assay of SUMO-CgnA with different concentrations of CgnB. A) Raw data of the binding affinity experiment, the response units of different concentrations of CgnB are plotted against the time of every injection. B) Binding affinity curve of CgnB to SUMO-CgnA. The plotted function of concentration against response units results in an estimated  $K_d = 2.04 \pm 0.11 \mu\text{M}$ , which has been calculated by the *Biacore X100* evaluation software.



CLUSTAL O(1.2.4) multiple sequence alignment

```

CgnB      MDVLEYFERLKNRELAFVLDDLQLSDMVTRRGFSVIPFDDFDLAREDHPPAFVLVTRLDY      60
CgnE      MNASDFYALLRGRGMPVVDDAEAAAVVSELGFRTVPFEAFDFDSPEDPALVIVAQMGN      60
          *: .  ::  *: . *  :  . *: **  :  :  *: .  **  . : ** :  ** :  ..  *: *: *: *: . .

CgnB      HGKLMQAWETAKGISSHLSLAKFDTSFKSVEYSLDQLLSMDFAETLKRRGDYYDSVASTN      120
CgnE      VDALHGLWERSGTPLMHLALAKFDGGLSRLRAGLARVLAVDTDAALKRRAEAYEQLFSSA      120
          .  *    **  :      *: *****  .  .  .  *  :  *: *:  : ***** :  *: .  *:

CgnB      RMEVVTPGAULTCDFGNEIEIANNDVEMQKGWLYSVAEFFETSVINLEADRSSSYTLNGDL      180
CgnE      SVEIASGEGVLRCHIGDEVEVGNCGDTLEQGFLYSVAEFLEASVVNLEGERSTFWVEGEL      180
          : *: .  :  . **  *: . *: *: *: .  .  :  :: *: ***** : *: *: *: *: : *: : *

CgnB      CFTGLIYLCNRPDLKERASATMDELMRMSTRGRNVVSFVDNQIVRMELGGVDMTATLREL      240
CgnE      PFDGFIHLSNSAALKERWGGMLDEFMRRSREGANLVRFADNVIDRLVVGVDVTSALAGL      240
          *  *: *: . *  *****  .  .  :  :: *: *  .  *  *: *  *  *  *  *  : ***** : *: *  *

CgnB      IVGKEREGSSSTEFAMGCVEYPLAQDWTINSVMNEGSHGIHVGVMGKEIPHMDFIKGA      300
CgnE      SQGEERGMAATEFGLGCADAEAAEPFGVNSLLHKSAGGAYIGIGKGLRIPHIDFIARGAT      300
          *: **  : : ** . : *: .  :  *  :  :  *: *: . .  *  : *: *  *  . *: *: *: *: **

CgnB      LRIAESSDA 309
CgnE      IRFIPAAEG 309
          : *:  : : .

```

**Figure S3:** Sequence alignment of the original annotations of CgnB and CgnE from *Chondromyces crocatus* Cm c5. We observe 38% identical amino acids and 60% positives over the full length of the proteins. The sequence alignment was generated with *Clustal Omega*.<sup>[1]</sup>

CLUSTAL O(1.2.4) multiple sequence alignment

```

CgnE      -----MNASDFYALLRGRGMPVVDDAEAAAVVSELGFRTVPFEAFDFD      44
CgnE_long  MGGRRTIGIRSGEGAIMNASDFYALLRGRGMPVVDDAEAAAVVSELGFRTVPFEAFDFD      60
          *****

CgnE      SPSEDPALVIVAQMGNVDALHGLWERSGTPLMHLALAKFDGGLSRLRAGLARVLAVDTDA      104
CgnE_long  SPSEDPALVIVAQMGNVDALHGLWERSGTPLMHLALAKFDGGLSRLRAGLARVLAVDTDA      120
          *****

CgnE      ALKRAEAYEQLFSSASVEIASGEGVLRCHIGDEVEVGNCGDTLEQGFLYSVAEFLEASV      164
CgnE_long  ALKRAEAYEQLFSSASVEIASGEGVLRCHIGDEVEVGNCGDTLEQGFLYSVAEFLEASV      180
          *****

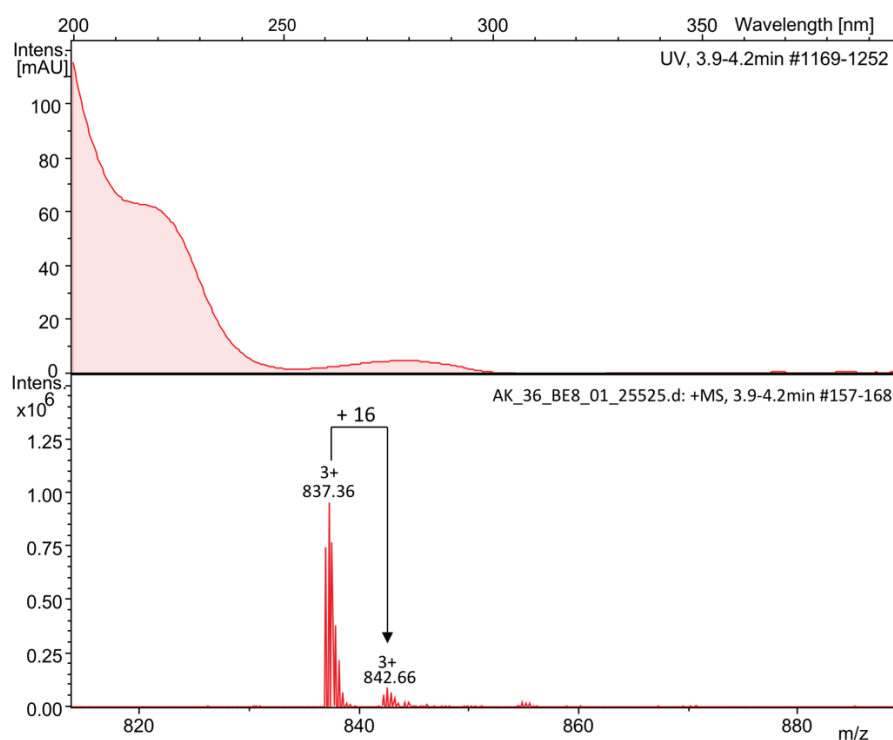
CgnE      VNLEGERSTFWVEGELPFDGFIHLSNSAALKERWGGMLDEFMRRSREGANLVRFADNVID      224
CgnE_long  VNLEGERSTFWVEGELPFDGFIHLSNSAALKERWGGMLDEFMRRSREGANLVRFADNVID      240
          *****

CgnE      RLVVGVDVTSALAGLSQGEERGMAATEFGLGCADAEAAEPFGVNSLLHKSAGGAYIGIG      284
CgnE_long  RLVVGVDVTSALAGLSQGEERGMAATEFGLGCADAEAAEPFGVNSLLHKSAGGAYIGIG      300
          *****

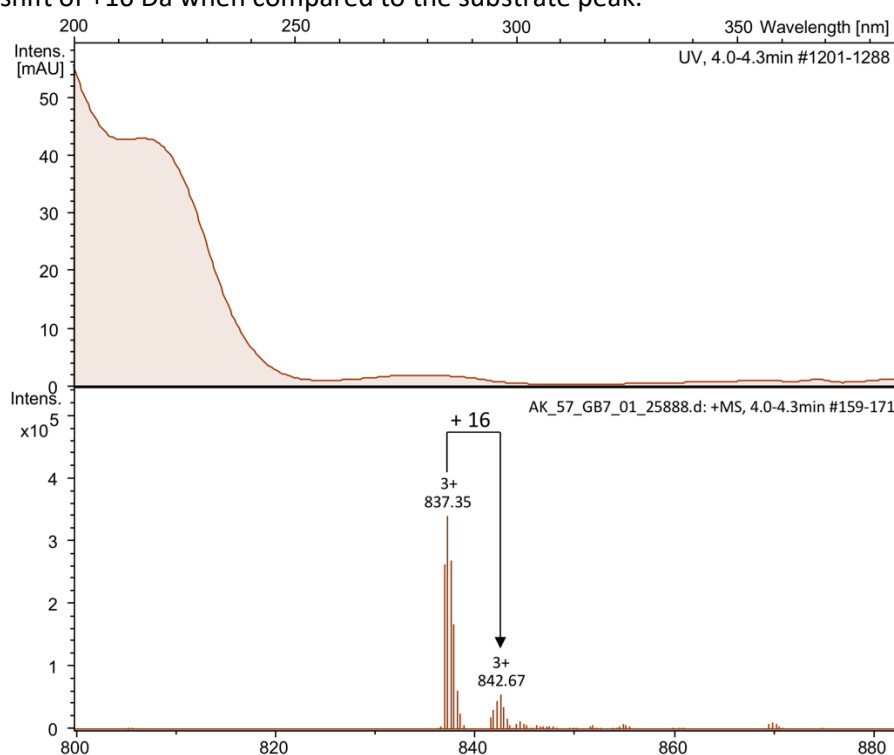
CgnE      KGLRIPHIDFIARGATIRFIPAAEG      309
CgnE_long  KGLRIPHIDFIARGATIRFIPAAEG      325
          *****

```

**Figure S4:** Sequence alignment of the original annotation of CgnE with the extended construct CgnE<sup>L</sup> which was crystallized. We observe an extended N-terminus of 16 amino acids due to an alternative start codon upstream of the original annotation. The sequence alignment was generated with *Clustal Omega*.<sup>[1]</sup>



**Figure S5:** Mass spectrometry analysis of a reaction mixture of CgnC<sup>L</sup> with the precursor peptide CgnA, combined with the cofactors 2-oxoglutaric acid, FeCl<sub>2</sub> and ascorbic acid. The UV spectrum is unchanged, while the mass spectrum suggests little turnover and the appearance of a second peak with a mass shift of +16 Da when compared to the substrate peak.



**Figure S6:** Mass spectrometry analysis of a reaction mixture of CgnB with the precursor peptide CgnA, combined with the cofactors 2-oxoglutaric acid, FeCl<sub>2</sub> and ascorbic acid. The UV spectrum is unchanged, while the mass spectrum suggests little turnover and the appearance of a second peak with a mass shift of +16 Da when compared to the substrate peak.

**Table S1:** Primary amino acid sequences of proteins and peptides used in this study.

Construct	Primary amino acid sequence
SUMO-CgnA	MGSSHHHHHSGDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAF AKRQKGEMDSLRFYDGIQADQTPEDLDMEDNDIIEAHREQIGGDIPTTENLYFQGMKK TIKQISAGDRSKKGNIYW
CgnA	GMKKTIKQISAGDRSKKGNIYW
CgnB	GAMADIGSMDVLEYFERLKNRELAFVLDDLQLSDMVTRRGFSVIPFDDFDLAREDHPPAFV LVTRLDYHGKLMQAWETAKGISSHLSLAKFDTSPKSVEYSLDQLLSMDFAETLKRRGDYYDS VASTNRMEVVTGAVLTCDFGNEIEIANNDVEMQKGWLYSVAEFFETSVINLEADRSSYTL NGDLCFTGLIYLCNRPDLKERASATMDELMRMSTRGRNVVSFVDNQIVRMELGGVDMTA TLRELIVGKEREGSSTEFAMGCVEYPLAQDWTINSVMNEGSHGIHVGVGMGKEIPHMDFI AKGAELRIAESSDA
CgnC <sup>L</sup>	GAMVHLTATDLHVDVDFSGDLEAALAKARHLYATFGGFVAARLFDAADLTPIHEELGRILALAA REVDGFSPLSPRTRFDEGFHALHEEAPAAADAVIQAAARLTTVHELNVNPRLLSVSRRLMGT ELVMSNPYKPIRVDAEREHFLLPWHQDYPYAQDSMDAVVHWIPLQDVDEQNGCLKVAP GSHELGVVPVKMILPPEGSTHGIRGLQIADPSVVERFPQVSLPMKFGDVLVFSSTLLHRSQ LNLTEARWTAQVRHGNFEHPLSVQKRWPGRGHERHWFDETHPEYVRPAGS
CgnD	GAMASTTLETRDELTPQMKEYDRAGRVWVPYLYNHFHRPNHRSPVNTDSRGFRFVVGKD GRTFSEFEREPGERVRALVGGSTVFGVGATGDAATLPSLLSQRGPARWLNFGGRAFSSTQE LMLFLFHARSLGALEKVTLLSGVNNLLLFYLSRDYAKDYGSSFSATEVRRAFAGDAPSPAKSG VIGRLKSIAGGRRRAKVEAPEPEIVLPVDHDAQKTDLLHAIERDLSTWKLKSGALQFELCYVLQ PLAGWVRKKPSPEETRLFADLDDQQGEAWRQILREKMDLAQYAWFSKSLADICRTQEIPFL DMNATLSALDLDGRWIFVDRVHLTDEGNEVLTAQALVEGGAT
CgnE <sup>L</sup>	GMGGRRTIGIRSGEGAIMNASDFYALLRGRGMPVVDDAEAAAVVSELGFRTVPFEAFDF DSPSEDPALVIVAQMGNVDALHGLWERSGTPLMHLALAKFDGGLSRLRAGLARVLAVDTD AALKRRAEAYEQLFSSASVEIASGEGVLRCHIGDEVEVGNCGDTLEQGFLYSVAEFLEASVNV LEGERSTFWVEGELPFDGFIHLSNSAALKERWGGMLDEFMRRSREGANLVRFADNVIDRL VVGVDVTSALAGLSQGEERGMAATEFGLGCADAEAAEPFGVNSLLHKSAGGAYIGIGKG LRIPHIDFIARGATIRFIPAAEG

**Table S2:** Primary amino acid sequences of the different CgnC constructs which were cloned and tested in this study.

Construct	Primary amino acid sequence
CgnC_original	GAMGTELVMSNPYKPIRVDAEREHFLLPWHQDYPYAQDSMDAVVHWIPLQDVDEQNG CLKVAPGSHELGVVPVKMILPPEGSTHGIRGLQIADPSVVERFPQVSLPMKFGDVLVFSSTLL HRSQNLNLTEARWTAQVRHGNFEHPLSVQKRWPGRGHERHWFDETHPEYVRPAGS
CgnC_VNPRL	GAMVNPRLSVSRRLMGTELVMSNPYKPIRVDAEREHFLLPWHQDYPYAQDSMDAVVH WIPLQDVDEQNGCLKVAPGSHELGVVPVKMILPPEGSTHGIRGLQIADPSVVERFPQVSLP MKFGDVLVFSSTLLHRSQNLNLTEARWTAQVRHGNFEHPLSVQKRWPGRGHERHWFDET HPEYVRPAGS
CgnC_VHELS	GAMVHELNVNPRLLSVSRRLMGTELVMSNPYKPIRVDAEREHFLLPWHQDYPYAQDSM DAVVHWIPLQDVDEQNGCLKVAPGSHELGVVPVKMILPPEGSTHGIRGLQIADPSVVERFP QVSLPMKFGDVLVFSSTLLHRSQNLNLTEARWTAQVRHGNFEHPLSVQKRWPGRGHERH WFDETHPEYVRPAGS
CgnC_VDGF5	GAMVDGFSPLSPRTRFDEGFHALHEEAPAAADAVIQAAARLTTVHELNVNPRLLSVSRRLM GTELVMSNPYKPIRVDAEREHFLLPWHQDYPYAQDSMDAVVHWIPLQDVDEQNGCLKV APGSHELGVVPVKMILPPEGSTHGIRGLQIADPSVVERFPQVSLPMKFGDVLVFSSTLLHRS QLNLTEARWTAQVRHGNFEHPLSVQKRWPGRGHERHWFDETHPEYVRPAGS

CgnC <sup>L</sup>	GAMVHLTATDLHVDVDFSGDLEAALAKARHLYATFGGFVAARLFDAAADLTPIHEELGRILALAA REVDGFSPLSPRTRFDEGFHALHEEAPAAADAVIQAAARRLTTVHELNVNPRLLSVSRRLMGT ELVMSNPYKPIRVDYAEREHLLPWHQDYPYAQDSMDAVVHWIPLQDVDEQNGCLKVAP GSHELGVVPVKMILPPEGSTHGIRGLQIADPSVVERFPQVSLPMKFGDVLVFSTLLLHRSQ NLTGEARWTAQVRHGNFEHPLSVQKRWPRGHYERHWFDETHPEYVRPAGS
-------------------	---

**Table S3:** Primary amino acid sequence of the CgnA precursor peptide with an N-terminal Glycine resulting of a cloning artefact. The peptide possesses 6 positively charged side chains which are highlighted in blue, resulting in a theoretical pI of 10.30.

CgnA	GMKKTIKQISAGDRSKKGNIYW
------	------------------------

**Table S4:** List of hydrogen bonds and salt bridges found in the protein assembly of the CgnD<sub>native</sub> crystal structure. The table was generated with the PISA server.<sup>[2]</sup>

### Hydrogen bonds

##	_Structure 1_	_Dist. [Å]	_Structure 2_
1	B:ARG 224 [HH11]	2.45	A:PHE 124 [ O ]
2	B:TYR 29 [ HH ]	2.27	A:SER 159 [ O ]
3	B:ARG 35 [HH11]	2.08	A:PRO 202 [ O ]
4	B:ARG 35 [HH12]	2.39	A:GLU 203 [ OE2]
5	B:PHE 124 [ O ]	2.32	A:ARG 224 [HH11]
6	B:SER 159 [ O ]	2.36	A:TYR 29 [ HH ]
7	B:PRO 202 [ O ]	1.79	A:ARG 35 [HH11]

### Salt bridges

##	_Structure 1_	_Dist. [Å]	_Structure 2_
1	B:ARG 48 [ NE ]	3.28	A:ASP 156 [ OD2]
2	B:ARG 35 [ NH1]	3.15	A:GLU 203 [ OE2]
3	B:ARG 35 [ NH2]	3.99	A:GLU 203 [ OE2]
4	B:ARG 127 [ NH2]	3.61	A:GLU 223 [ OE1]
5	B:ASP 156 [ OD2]	3.49	A:ARG 48 [ NE ]
6	B:GLU 201 [ OE1]	3.77	A:ARG 35 [ NE ]
7	B:GLU 201 [ OE1]	3.76	A:ARG 35 [ NH1]
8	B:GLU 201 [ OE1]	2.87	A:ARG 35 [ NH2]
9	B:GLU 203 [ OE2]	3.43	A:ARG 35 [ NH1]

## Materials and Methods

### Cloning of Cgn constructs used in this study

For CgnA, CgnB and CgnD, only the original annotation of the open reading frame for these proteins was utilized for cloning.<sup>[3]</sup> In contrast, several Cgn expression constructs were designed based on genomic information on an alternative N-terminus for CgnC and CgnE (**Table S2**)(**Figure S4**). The corresponding primers for every construct were ordered from Sigma-Aldrich and the different gene constructs amplified by PCR with a standard Phusion polymerase protocol with adjusted annealing temperatures. DNA amplicates of *cgnC*, *cgnD* and *cgnE* were subsequently cut with *NcoI* and *HindIII* in a 1 : 0.5 ratio in buffer R at 37 °C for 4 h and ligated into an already cut pSUMO vector (the plasmid was a gift by Dr. David Owen, St. Andrews University<sup>[4]</sup>) treated with alkaline phosphatase with a standard T4 DNA ligase protocol. The PCR product of *cgnB* was cut with *BamHI* and *HindIII* in a 1 : 0.5 ratio in buffer *BamHI* at 37 °C for 4 h and ligated into an already cut pSUMO vector treated with alkaline phosphatase with a standard T4 DNA ligase protocol. For *cgnA*, we applied a specific cloning protocol as the resulting PCR product was too small and could not be purified efficiently. In a three step PCR process, we amplified the *SUMO* gene from the pSUMO vector and added the *cgnA* gene at the 3' prime end of the TEV site by PCR amplification. The full amplicate was cut with *NcoI* and *NdeI* at a 1 : 0.5 ratio in buffer O at 37 °C and 4 °C ligated into an already cut pSUMO vector treated with alkaline phosphatase with a standard T4 DNA ligase protocol. Ligation samples were transformed into chemically competent *E. coli* HS996 with a standard heat shock protocol, plated on LB-Agar with 50 µg/mL Kanamycin and grown at 37 °C and 200 rpm for 16 h. Single *E. coli* colonies were grown in 10 mL LB-medium with 50 µg/mL Kanamycin for 16 h at 37 °C and 200 rpm. 4 mL of grown culture were centrifuged for 10 min at 4000 rpm and 4 °C and the cell pellet was used for a plasmid preparation by alkaline lysis. Extracted plasmids were cut with the respective restriction enzymes used for cloning using the same procedure and analyzed by agarose gel electrophoresis. Plasmids which were found to carry the correct insert were sent for sequencing by LGC genomics (Berlin) and subsequently transformed into chemically competent *E. coli* Lemo21(DE3) cells for pSUMO-CgnA, pSUMO-CgnC variants and pSUMO-CgnD. pSUMO-CgnE variants were transformed into *E. coli* Rosetta2(DE3) cells whereas pSUMO-CgnB was transformed into *E. coli* BL21(DE3) cells.

### Small scale protein expression tests of Cgn constructs

Small scale protein expression testing was carried out by streaking a freshly transformed *E. coli* BL21(DE3), *E. coli* Rosetta2(DE3) or *E. coli* Lemo21(DE3) colony with the respective pSUMO-Cgn construct into 10 mL LB-media supplemented with 50 µg/mL Kanamycin and 34 µg/mL Chloramphenicol or just 50 µg/mL Kanamycin, respectively. Cultures were grown for 16 h at 37 °C and

200 rpm and were used to subsequently inoculate 50 mL LB-medium supplemented with the respective antibiotics at a 1:100 ratio. For small scale expression testing, a variety of parameters like temperature, IPTG concentration and point of induction were analyzed in different combinations to increase the yield of soluble protein in the following Ni<sup>2+</sup> pull-down assay. In any case, every culture was grown in LB-medium for 16 h after the point of induction before the broth was centrifuged at 4000 rpm and 4 °C for 10 min to harvest the wet cell pellet. The supernatant was discarded and the cell pellet was frozen at -80 °C until further use.

### **Ni<sup>2+</sup> pull-down assay**

The Ni<sup>2+</sup> was utilized to separate His<sub>6</sub>-tagged proteins from a lysate of disrupted *E. coli* cells and furthermore test their solubilization properties in different lysis buffers in a medium throughput approach. An *E. coli* cell pellet was resuspended in 500 µL of a respective lysis buffer and subsequently sonicated with an amplitude of 75 % using 2 times a 15 sec burst with a 30 sec recover time bringing total sonication time up to 30 sec. Afterwards, the lysate was centrifuged at 15000 rpm and 4 °C for 10 minutes before being used in the Ni<sup>2+</sup> pull-down assay using a KingFisher mL purification system (Thermo Scientific). The lysate was mixed with 50 µL of MagneHis Ni particles (Promega) before being washed 2 times with 500 µL lysis buffer. The His<sub>6</sub>-tagged proteins were eluted from the magnetic Ni<sup>2+</sup> particles with 50 µL of elution buffer supplemented with 250 mM Imidazole. Eluted fractions were run in an SDS-PAGE to compare yield and sizes of recombinantly expressed proteins with the Precision Plus Dual Xtra protein standards (Bio-Rad).

### **Large scale protein expression of Cgn constructs**

Large scale protein expression of every Cgn construct started by streaking a freshly transformed *E. coli* Lemo21(DE3), *E. coli* Rosetta2(DE3) or *E. coli* BL21(DE3) colony with the respective pSUMO-Cgn construct into 100 mL LB-medium supplemented with 50 µg/mL Kanamycin and 34 µg/mL Chloramphenicol or just 50 µg/mL Kanamycin, respectively. The culture was grown for 16 h at 37 °C and 200 rpm and was subsequently used to inoculate an expression culture of LB-medium supplemented with the appropriate antibiotics at a ratio of 1:100. This culture was grown at 37 °C and 200 rpm until an OD<sub>600</sub> = 0.8 was reached, at which point the temperature was decreased for CgnB to 16 °C, CgnC<sup>L</sup>, CgnD and CgnE<sup>L</sup> to 18 °C. In contrast, pSUMO-CgnA cultures were induced with 0.4 mM IPTG and grown at 37 °C for 16 h. CgnB and CgnC<sup>L</sup> cultures were induced with 0.1 mM IPTG whereas CgnD and CgnE<sup>L</sup> cultures were induced with 0.4 mM IPTG and grown at the efforementioned temperatures for 16 h. Fermentation was stopped in any case by centrifugation at 5000 rpm and 4 °C for 10 min after which the respective cell pellet was collected and the supernatant discarded. Cell pellets were frozen at -80 °C until further use.

### Selenomethionine expression of CgnD and CgnE<sup>L</sup>

For selenomethionine expression, an overnight culture of LB-medium with the appropriate antibiotics was inoculated with a single, freshly transformed *E. coli* colony carrying the vector with the crystallized construct. For every liter of prepared medium for expression, 50 mL of LB-medium were grown at 37 °C and 200 rpm for 16 h. The culture was centrifuged at 2500 rpm and 4 °C for 15 min and the cell pellet washed three times in M9-medium (8.5 g/L Na<sub>2</sub>HPO<sub>4</sub>, 3 g/L KH<sub>2</sub>PO<sub>4</sub>, 1 g/L NH<sub>4</sub>Cl, 0.5 g/L NaCl) before inoculating M9-medium supplemented with glucose-free nutrient mix and 5% glycerol at a ratio of 1:20. The medium was grown for 20 min at 37 °C at which point 40 mg/L L-selenomethionine was added to the culture. The cultures were grown at 37 °C and 200 rpm until an OD<sub>600</sub> of 0.6 was reached, whereupon 100 mg/L lysine, phenylalanine and threonine and 50 mg/L isoleucine and valine were added. The medium was incubated for another 20 min before the protein expression was induced by the addition of 1 mM IPTG. The temperature was reduced to 20 °C and the cells were grown for 24 h. The cells were harvested by centrifugation (5000 rpm, 4 °C, 10 min).

### Purification of (SUMO-)CgnA

For precursor peptide purification, the cell pellet was resuspended in lysis buffer (500 mM NaCl, 20 mM Tris pH 8.0, 20 mM Imidazole, 3 mM β-mercaptoethanol). For every 25 g of wet cell mass, 100 mL of lysis buffer were added and supplemented with 2 cComplete EDTA-free protease inhibitor tablets (Sigma-Aldrich) and 4 mg/g DNase (Sigma-Aldrich). Cell lysis was carried out via passage through a cell disruptor (30 kpsi, Microfluidics Corporation), and the cell debris was removed by centrifugation (19000 rpm, 15 min, 4 °C). The supernatant was separated from the cells, filtered through a 0.45 μm filter and applied to a His-Trap HP 5 mL column (GE Healthcare) pre-equilibrated in lysis buffer at a flow rate of 5 mL/min. The column was extensively washed with lysis buffer (30 column volumes) and subsequently, the target protein was eluted using a lysis buffer supplemented with 250 mM imidazole. The proteins were passed over a desalting column (16/10 GE Healthcare) at a flow rate of 10 mL/min pre-equilibrated in lysis buffer. To remove the SUMO protein and the His<sub>6</sub>-tag, the solution was incubated with TEV protease for 2 h at room temperature at a 1:10 ratio of TEV:target protein. Further passage of the solution over a 5 mL His-Trap HP column allowed for a separation of the digested target protein and the His<sub>6</sub>-tagged SUMO. Lastly, the precursor peptide was passed over a Superdex 30 16/60 size exclusion column (GE Healthcare) at a flow rate of 1 mL/min pre-equilibrated in gel filtration buffer (150 mM NaCl, 10 mM HEPES, 0.5 mM TCEP, pH 7.4). The resulting peak was gathered and concentrated to a desired concentration using a 3 kDa cutoff filter (Merck Millipore). The peptide concentration was determined using photometric analysis (Nanodrop 2000, Thermo Scientific) and subsequently analyzed by SDS-PAGE as well as mass spectrometry.

For purification of SUMO-CgnA, the protein eluted off the first 5 mL His-Trap HP column (GE Healthcare) was directly applied to a Superdex 200 16/600 (GE Healthcare) pre-equilibrated in gel filtration buffer (150 mM NaCl, 10 mM HEPES, 0.5 mM TCEP, pH 7.4). The resulting peak was gathered and concentrated to a desired concentration using a 10 kDa cutoff filter (Thermo Scientific). The protein concentration was determined using photometric analysis (Nanodrop 2000, Thermo Scientific) and subsequently analyzed by SDS-PAGE as well as mass spectrometry.

### **Protein purification of CgnB, CgnC<sup>L</sup>, CgnD and CgnE<sup>L</sup>**

For protein purification, the cell pellet was resuspended in the respective lysis buffer (CgnD/CgnE<sup>L</sup>: 500 mM NaCl, 20 mM Tris pH 8.0, 20 mM Imidazole, 3 mM  $\beta$ -mercaptoethanol; CgnB/CgnC<sup>L</sup>: 200 mM NaCl, 20 mM Tris, 20 mM Imidazole, 10 % glycerole, 3 mM  $\beta$ -mercaptoethanol). For every 25 g of wet cell mass, 100 mL of lysis buffer were added and supplemented with 2 cOmplete EDTA-free protease inhibitor tablets (Sigma-Aldrich) and 4 mg/g DNase (Sigma-Aldrich). Cell lysis was carried out via passage through a cell disruptor (30 kpsi, Microfluidics Corporation), and the cell debris was removed by centrifugation (19000 rpm, 15 min, 4 °C). The supernatant was separated from the cells, filtered through a 0.45  $\mu$ m filter and applied to a His-Trap HP 5 mL column (GE Healthcare) pre-equilibrated in lysis buffer at a flow rate of 5 mL/min. The column was extensively washed with lysis buffer (30 column volumes) and subsequently, the target protein was eluted using a lysis buffer supplemented with 250 mM imidazole. The proteins were passed over a desalting column (16/10 GE Healthcare) at a flow rate of 10 mL/min pre-equilibrated in lysis buffer. To remove the SUMO protein and the His<sub>6</sub>-tag, the solution was incubated with TEV protease for 14 h at 4 °C at a 1:10 ratio of TEV:target protein. Further passage of the solution over a 5 mL His-Trap HP column allowed for a separation of the digested target protein and the His<sub>6</sub>-tagged SUMO. Lastly, the precursor peptide was passed over a Superdex 200 16/600 size exclusion column (GE Healthcare) at a flow rate of 1 mL/min pre-equilibrated in gel filtration buffer (150 mM NaCl, 10 mM HEPES, 0.5 mM TCEP, pH 7.4). The resulting peak was gathered and concentrated to a desired concentration using a 30 kDa cutoff filter (Thermo Scientific). The protein concentration was determined using photometric analysis (Nanodrop 2000, Thermo Scientific) and subsequently analyzed by SDS-PAGE as well as mass spectrometry.

### **Crystallization, data collection and structure elucidation of CgnB**

For crystallization, 200  $\mu$ M CgnB was incubated with 250  $\mu$ M CgnA for 2 h on ice before setting up initial crystallization trials using different sparse matrix and grid screens (Qiagen). The screening plates were set up using a Gryphon crystallization robot (Art Robbins Instruments) at 18 °C and were directly moved to 4 °C for further incubation. CgnB crystals were observed after two days of incubation at 4 °C in a condition with a well solution consisting of 170 mM ammonium acetate, 85 mM sodium acetate



pH 4.6, 25.5% PEG 4000 and 15% glycerole. The crystals could not be reproduced in optimization plates in either the same or a freshly purified batch of protein, which were set up using a Dragonfly liquid handler (TTP Labtech). Single crystals were cryoprotected by addition of 32% glycerole to 68% well solution, mounted from the screening plate using mounted cryoloops (Hampton Research) and flash frozen in liquid nitrogen until further use. A full dataset of CgnB was collected at the European Synchrotron Radiation Facility (ESRF) at beamline ID-23-1 which was used for molecular replacement using *phenix.phaser*<sup>[5]</sup> within the PHENIX suite<sup>[6]</sup> with an already refined model of CgnE<sup>L</sup>.

### Crystallization, data collection and structure elucidation of CgnE<sup>L</sup>

Initial crystallization trials of CgnE<sup>L</sup> were set up using a protein concentration of 220  $\mu$ M at 18 °C. Crystals were observed after 2 days of incubation in a condition with a well solution containing 200 mM MgCl<sub>2</sub>, 100 mM Tris pH 8.5 and 20% PEG 8000. Selenomethionine-labelled CgnE<sup>L</sup> could be crystallized in the same condition and optimized by increasing the MgCl<sub>2</sub> concentration to 300 mM while staying at a high pH of 8.5 and lowering the PEG 8000 concentration to 15%. Optimization plates were set up using a Dragonfly liquid handler (TTP Labtech). Single crystals were cryoprotected by addition of 15% 2,3-butanediol to 85% well solution, mounted using mounted cryoloops (Hampton Research) and flash frozen in liquid nitrogen until further use. A high redundancy dataset at the selenium K-edge at a similar resolution limit, which was observed for native crystals, was collected at Swiss Light Source (SLS) Beamline X10SA. The structure was solved with selenium single anomalous dispersion (Se-SAD) data using the *AutoSol* package<sup>[7]</sup> within the PHENIX suite.<sup>[6]</sup>

### Crystallization, data collection and structure elucidation of CgnD

To obtain better quality crystals, the lysine methylation method was applied to CgnD following the originally developed method.<sup>[8]</sup> Lysine di-methylated CgnD<sub>native</sub> crystallized after 3 days of incubation at 18 °C in a condition with a well solution consisting of 200 mM ammonium sulfate, 100 mM MES pH 6.5 and 30% PEG 5000MME. Optimization plates for this condition were set up using a Dragonfly liquid handler (TTP Labtech). Single crystals were cryoprotected by addition of 32% glycerol to 68% well solution, mounted using mounted cryoloops (Hampton Research) and flash frozen in liquid nitrogen until further use. A full dataset was collected at Swiss Light Source Beamline X10SA which was later used for molecular replacement using *phenix.phaser*<sup>[5]</sup> within the PHENIX suite<sup>[6]</sup> with the CgnD<sub>SeMet</sub> model.

Selenomethionine-labelled lysine di-methylated CgnD<sub>SeMet</sub> could not be reproduced in the same condition, which had been crystallized for CgnD<sub>native</sub> and had to be screened. CgnD<sub>SeMet</sub> crystals appeared after 2 days incubation at 18 °C in a condition with a well solution containing 200 mM ammonium acetate, 100 mM sodium acetate pH 4.6 and 30% PEG4000. This condition was further

optimized by the addition of different additives using an Additive Screen (Hampton Research). Optimization plates were set up with further addition of 4% 1-propanol to the efforementioned condition with a Dragonfly Liquid Handler (TTP Labtech). Single crystals were cryoprotected by addition of 32% glycerol to 68% well solution, mounted using mounted cryoloops (Hampton Research) and flash frozen in liquid nitrogen until further use. A high redundancy dataset at the selenium K-edge was collected at the Swiss Light Source Beamline X06DA. The structure was solved with selenium single anomalous dispersion (Se-SAD) data using the *AutoSol* package<sup>[7]</sup> within the PHENIX suite.<sup>[6]</sup>

### Refinement and structural analysis of Cgn proteins

Solved structures went through multiple rounds of subsequent manual model building in *Coot*<sup>[9]</sup> followed by refinement using *phenix.refine*<sup>[10]</sup> within the PHENIX suite.<sup>[6]</sup> Final PDB coordinates were analyzed using *MolProbity*<sup>[11]</sup>, used for detection of a macromolecular assembly with the PISA server<sup>[2]</sup> as well as for structural homology analysis with the DALI server.<sup>[12]</sup> All structural images portrayed were rendered in *PyMOL* (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC).

### Biochemical reactions

All biochemical reactions were diluted in Buffer GF (150 mM NaCl, 10 mM HEPES, 0.5 mM TCEP, pH 7.4). For CgnA<sup>BC</sup>, a mixture of 50  $\mu$ M CgnA, 20  $\mu$ M CgnB, 10  $\mu$ M CgnC<sup>L</sup>, 4 mM ascorbic acid, 5 mM FeCl<sub>2</sub> and 1 mM 2-oxoglutaric acid was incubated at 37 °C for 1 h. For CgnA<sup>BCE</sup>, a mixture of 50  $\mu$ M CgnA, 20  $\mu$ M CgnB, 10  $\mu$ M CgnC<sup>L</sup>, 10  $\mu$ M CgnE<sup>L</sup>, 4 mM ascorbic acid, 2.5 mM FeCl<sub>2</sub> and 1 mM 2-oxoglutaric acid was incubated at 37 °C for 1 h. The standard reaction volume was 50  $\mu$ L. Proteins were precipitated by adding 50% acetonitrile and freezing the mixture at -80 °C for 1 h prior to LC-MS analysis.

### Surface Plasmon Resonance

SPR binding affinity measurements were carried out on a Biacore X100 system (GE Healthcare). SUMO-CgnA and measured proteins were dialyzed in HBS-P buffer (150 mM NaCl, 10 mM HEPES, 0.0005% v/v Surfactant P20, pH 7.4). SUMO-CgnA was bound irreversibly onto an Ni-NTA chip, CgnB, CgnC<sup>L</sup>, CgnD and CgnE<sup>L</sup> were applied with a maximum concentration of 10  $\mu$ M and subsequent dilutions of 1:1. A  $K_d$  measurement with standard error was calculated with the *Biacore X100* software.

### Microscale Thermophoresis

MST binding affinity measurements were performed using a Monolith NT 115 system (Nanotemper Technologies). CgnB was labeled in MST buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 10 mM MgCl<sub>2</sub>, 0.05% Tween 20) at 20 °C using the Monolith Protein Labeling Kit RED-NHS. CgnC<sup>L</sup> was applied using a

maximum concentration of 200  $\mu\text{M}$  and subsequent dilutions of 1:1. The experiment was carried out at 22 °C with an excitation power of 40%, the MST power was set to medium. A  $K_D$  measurement with standard error was calculated using the *MO.Affinity Analysis v2.3* software.

## LC-MS

LC was performed on a Dionex Ultimate 3000 RSLC system using a BEH C18 column (100 mm x 2.1 mm, 1.7  $\mu\text{m}$ ) equipped with a C18 precolumn (Waters). Solvent A was  $\text{H}_2\text{O}$  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 an amaZon speed mass spectrometer (Bruker Daltonics). The LC flow was split 1:8 before entering the mass spectrometer using the Apollo ESI source. The following parameters were used: capillary voltage 4500 V, temperature 300 °C, dry-gas flow rate 10  $\text{L min}^{-1}$  and nebulizer 30 psi. Data was recorded in the mass range from 650 to 2000  $m/z$ .

## Supplementary references

- [1] Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R *Nucleic Acids Res.* **2019** [Epub ahead of print]
- [2] Krissinel E, Henrick K *J Mol Biol.* **2007** 372 (3):774-97
- [3] Viehrig K, Surup F, Volz C, Herrmann J, Abou Fayad A, Adam S, Köhnke J, Trauner D, Müller R *Angew Chem Int Ed Engl.* **2017** 56 (26):7407-7410
- [4] Liu, H.; Naismith, J. H., *Protein Expr Purif* **2009** 63 (2):102-11
- [5] Terwilliger TC, Adams PD, Read RJ, McCoy AJ, Moriarty NW, Grosse-Kunstleve RW, Afonine PV, Zwart PH, Hung LW *Acta Crystallogr D Biol Crystallogr.* **2009** 65 (Pt 6):582-601
- [6] McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ *J Appl Crystallogr.* **2007** 40 (Pt 4):658-674
- [7] Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH *Acta Crystallogr D Biol Crystallogr.* **2010** 66 (Pt 2):213-21
- [8] Walter TS, Meier C, Assenberg R, Au KF, Ren J, Verma A, Nettleship JE, Owens RJ, Stuart DI, Grimes JM *Structure.* **2006** 14 (11):1617-22.
- [9] Emsley P, Lohkamp B, Scott WG, Cowtan K *Acta Crystallogr D Biol Crystallogr.* **2010** 66 (Pt 4):486-501
- [10] Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD *Acta Crystallogr D Biol Crystallogr.* **2012** 68 (Pt 4):352-67
- [11] Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC *Acta Crystallogr D Biol Crystallogr.* **2010** 66 (Pt 1):12-21
- [12] Holm L, Laakso LM *Nucleic Acids Res.* **2016** 44 (W1):W351-5

## Chapter 9

### Discussion and Outlook

#### 9.1 The importance of understanding natural product biosynthesis and its implications for industrial biotechnology

The experimental part of this work presented an *in vitro* approach designed to understand the biosyntheses of two interesting RiPP natural product scaffolds. The elucidation of natural product biosynthesis has various implications for different scientific fields, most of which are only discerned during the in-depth analysis of the respective machinery. This is especially true for RiPP pathways, as the innate promiscuity of their modifying enzymes may be exploited for the facilitated production of natural product derivatives with improved bioactivity and pharmacokinetic properties (see section 9.2). Furthermore, one can imagine applying RiPP tailoring enzymes catalyzing complex chemical reactions to processes very different from their original destined function in secondary metabolism. A fascinating proof of this concept has been demonstrated for the patellamide macrocyclase, which could be used to make hybrid macrocycles consisting of amino acids and a variety of different chemical building blocks.<sup>[1]</sup>

The use of enzymatic function in industrial biotechnology for cheap, ecofriendly biocatalysis comprises several classical applications in technical, food and animal feeding industries, with a total market value of \$1.5 billion in 2000.<sup>[2]</sup> Since the beginning of the 21<sup>st</sup> century, the industrial biotechnology sector profited from the advancements in enzyme research and development and a rising demand for sustainable solutions for a variety of processes, therefore experiencing extensive innovation and growth. In the advent of global warming awareness, an innovative field is the development of renewable bioenergy production by the conversion of biomass, for which enzymatic applications have been found to be a valuable option.<sup>[3]</sup> However, naturally occurring enzymes often possess less than optimal properties for industrial use, which can be reflected by either low stability, reaction product inhibition and/or slow catalytic activity.<sup>[4]</sup> In this context, arguably the most important modern concept in addition to rational enzyme design is the idea of directed evolution for enzyme engineering, also awarded half of the Nobel Prize in chemistry in 2018 for Frances H. Arnold. This method relies on (several rounds of) high-throughput incorporation of (simultaneous) random

mutations into a target enzyme that alter its functional elements, followed by screening for the desired activity and fine-tuning by point mutations.<sup>[5][6]</sup> In addition to screening for increased stability and efficacy of a target enzyme, it has been demonstrated that a known protein scaffold can be modified to exhibit a completely new chemical activity.<sup>[6]</sup> Thus, exploring enzymes from microbial secondary metabolite biosynthesis pathways that are able to catalyze complex chemical reactions can provide the basis for improvement and engineering of this enzymatic class and extend the available enzymatic scope for the scientific community.

Additionally, extensive knowledge of an enzyme acquired through isolated *in vitro* experiments broadens the available information about novel enzyme classes and their catalyzed reactions, therefore enabling easier *in silico* prediction of the biosynthesis of novel natural product scaffolds. The gathered knowledge will also facilitate efforts in mining and structure prediction of RiPP natural products and therefore the identification of the most promising biosynthetic gene clusters, a process which has thus far been restricted to clusters containing known enzyme families.<sup>[7]</sup>

### **Bottromycin – a unique biosynthetic pathway**

During this work, we have successfully established BotP as an aminopeptidase responsible for removing the N-terminal methionine residue from the BotA precursor peptide, thus liberating the N-terminus for subsequent macrocyclization. BotP is a rather special enzyme in RiPPs biosynthesis as its biosynthetic function mostly results from the architecture of the BotA precursor peptide, which contains the unusual N-terminal core peptide. In accordance with other RiPPs pathways, one can imagine the N-terminal methionine as a miniature leader peptide resulting from ribosomal biosynthesis. Therefore, bottromycin biosynthesis features aminopeptidase activity as a unique mechanism for “leader peptide” removal, since only a single amino acid needs to be cleaved. Other characterized RiPP precursor peptide architectures containing follower peptides that are important for substrate recognition of the tailoring enzymes include the eukaryotic orbitides<sup>[8]</sup> and, to a much smaller extent, patellamides.<sup>[9]</sup> However, the enzymatic function of BotP is only essential for *in vivo* scenarios of Bottromycin production, such as strain optimization and heterologous expression. Possible concepts for an *in vitro* substitution of BotP activity in a recombinant BotA precursor peptide include the proteolytic cleavage of a specific construct with trypsin established in our

group.<sup>[10]</sup> Being the first biosynthetic step, it has no implications for downstream tailoring and can be left out entirely in a synthesized precursor peptide without an N-terminal methionine residue (in a potential industrial scenario).

We have confirmed the YcaO domain enzyme IpoC as sufficient for the installation of the C-terminal thiazoline moiety and presented the first crystal structure from a standalone YcaO protein. Interestingly, IpoC shows relatively slow processing for a RiPPs heterocyclase, which can possibly be related to its weak affinity to BotA. As such, two individual experiments designed to determine a dissociation constant between IpoC and BotA in our group (or the respective homologues BmbD and BmbC in another group<sup>[11]</sup>) using different methods were unable to detect a  $K_D$  below 50  $\mu$ M. Leaving out the possibility of enzyme instability during the measurement, this result might be the consequence of a highly regulated function during biosynthesis. This hypothesis also correlates with IpoC requiring an (almost) intact follower peptide for substrate recognition and catalysis and an inability of IpoC to catalyze oxazoline formation. Bearing this in mind, it is also possible that fast processing by IpoC is hindered by missing carbon methylations (catalyzed by the 3 bottromycin RMTs) as untargeted metabolomics data have suggested that incomplete C-methylation reduces the efficiency of various downstream tailoring steps.<sup>[12]</sup> As such, we can hypothesize about a potential early regulatory function of IpoC, as IpoC could possibly select for a fully methylated BotA intermediate prior to the much faster PurCD macroamidine formation. A similar mechanism selecting for a fully tailored precursor peptide conferred by a difference in substrate binding affinity has been observed for LynD.<sup>[13]</sup> As the  $\beta$ -methylation of the phenylalanine residue of the core peptide has been shown to be crucial for antibacterial activity,<sup>[14]</sup> dependence of IpoC activity on  $\beta$ -methylation may provide an important regulatory function.

IpoC is to be a so-called standalone YcaO domain enzyme as it does not require a RiPP recognition element for binding and catalysis (see **Chapter 4**). It will be fascinating to see whether the observed dimer formation in the apo structure of IpoC translates to a biological function in a complex structure with the precursor peptide. The physical separation of follower peptide binding by one monomer and catalysis on the core peptide in the active site of the other monomer in a symmetrical fashion similar to LynD<sup>[13]</sup> could compensate for the missing RRE domain. In this context, a special conformation adopted by the dimer interface might be required for enzyme-substrate recognition and a guidance of the core peptide into the active site. However, one can also imagine two distinct sites of follower peptide

recognition that are engaged away from each other by the respective monomer, which might explain the necessity of a conserved follower peptide comprising more than 30 residues. It has been predicted that bottromycin methyl transferase genes (*botRMT1-3*) contain an RRE domain.<sup>[15]</sup> Thus far, no *in vitro* reconstitution experiments could provide experimental data on whether this RRE engages the follower peptide of BotA, although it has been predicted by a low  $\mu\text{M}$   $K_D$  using the isolated RRE in a binding experiment with BotA.<sup>[11]</sup> Consequently, the possibility of a complex formation of the last acting RMT with IpoC leading to an increased substrate affinity of IpoC cannot be excluded. Finally, we may want to consider a change in follower peptide affinity using the corresponding precursor peptide IpoA, as we observe a point mutation (A37D) in a region of the follower peptide critical for substrate recognition and catalysis.<sup>[10]</sup>

The most striking structural feature of bottromycin is arguably the macroamidine function. We have successfully demonstrated the first *in vitro* reconstitution of a YcaO domain enzyme able to catalyze this reaction. Macrocyclization is common in both ribosomal and non-ribosomal peptidic natural products – cyclic peptides possess superior properties (such as stability, activity and membrane permeability) compared to their linear counterparts.<sup>[9]</sup> Consequently, enzymes able to catalyze the formation of different types of peptidic macrocycles are highly sought after for the production of libraries of structurally diverse scaffolds, also in the context of hybrid RiPP engineering.<sup>[8][16]</sup> PatG is an example of a very promiscuous macrocyclase enzyme, which is almost insensitive to changes in the core peptide sequence.<sup>[16]</sup> On the other hand, the substrate conversion rate of PatG is a severe drawback, averaging an initial turnover rate of  $\sim 1$  per day.<sup>[9]</sup> Even though the catalytic activity of PurCD has been determined to be dependent on preceding thiazoline formation as well as an N-terminal glycine, several core peptide changes in other positions of the cyclized tetrapeptide were well tolerated.

Unfortunately, rational enzyme design to alter PurCD activity necessitates a mechanistic understanding of substrate binding and macrocyclization. During the course of this work, various efforts to obtain PurCD crystals were unsuccessful. Reflected by a measured melting temperature just above room temperature (24 °C), PurCD shows rapid unfolding, which is probably related to a highly flexible tertiary structure. Recent studies have proposed optimal conformational flexibility of a given enzyme to be required for full activity, a factor closely related to the observed temperature dependence of different enzymes (e.g. thermophilic



versus mesophilic enzymes).<sup>[17]</sup> Thus, the intriguingly low melting temperature of PurCD might be a requirement for its mechanistically uncharacterized function to date. Given the fast kinetics of PurCD (90% conversion of substrate in 90 min) compared to PatG<sup>[9]</sup> in combination with the release of a product with a different 3D-fold, it is possible that PurCD activity involves fast conformational changes. As the macrocycle was found to be essential for bottromycin activity, PurCD can potentially be vital for a semi-synthetic strategy of bottromycin production.<sup>[14]</sup> Further studies should be undertaken to explore the mechanical characteristics of this fascinating enzyme and determine its substrate binding mode. This information could in turn be used for either rational design or directed evolution efforts to explore the full scope of PurCD activity for the production of natural product scaffolds.

Another important tailoring step in bottromycin biosynthesis is the removal of the follower peptide not presented in this work. Our group has determined the amidohydrolase PurAH to be responsible for this biosynthetic step, leading to a remaining carboxylic acid moiety adjacent to the thiazoline at the C-terminus of the core peptide.<sup>[18]</sup> Unpublished docking data of the precursor peptide into the respective crystal structure of PurAH suggests the aspartate residue of the core peptide to be *L*-configured at this point of biosynthesis. This is an important implication for the following epimerization that is observed for the product of the PurAH reaction. Without exception, the determined absolute configuration of different naturally occurring bottromycin derivatives that have been isolated contains a *D*-configured aspartate residue in their final scaffold. This anomaly could be illuminated by the examination of the substrate preference of BotCYP, which is preferentially consuming the *D*-epimer of BotA<sup>PCCDAH</sup>. Using this substrate, we have successfully established the completion of the bottromycin core scaffold by an oxidative decarboxylation reaction catalyzed by BotCYP.

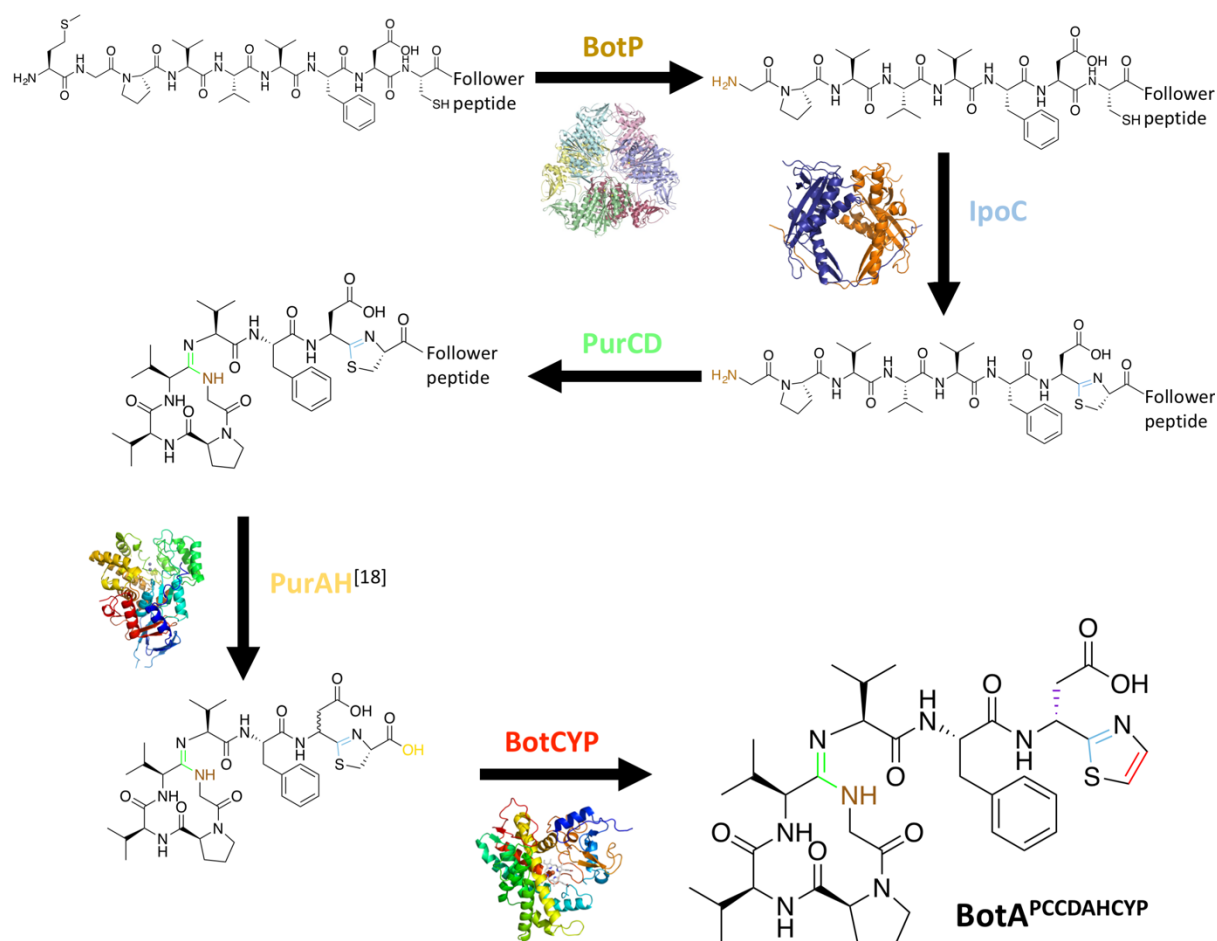
In non-ribosomal peptidic natural products, epimerases act by de- and re-protonating the  $\alpha$ -carbon leading to an equilibrium of both epimers, after which the downstream domain selects the *D*-epimer for condensation.<sup>[19]</sup> Given that RiPP tailoring takes place after a ribosomal precursor peptide is released, epimerization in RiPP pathways has to be based on different mechanics. Enzymatic epimerization reactions in RiPP pathways include the stereospecific reduction of dehydrated *L*-Ser to *D*-Ala with a dehydroalanine intermediate and the more general conversion of *L*- to *D*-amino acids by radical-*S*-adenosylmethionine (SAM) enzymes.<sup>[20]</sup> However, a spontaneous epimerization process had been proposed for both cyanobactins and bottromycins due to the absence of a gene which could be connected to this function. The

tendency of thiazoline moieties to epimerize chiral centers attached to the C2 position leading to a racemic mixture is a known problem in synthetic peptide chemistry.<sup>[21]</sup> Relatable, in experiments performed for the lissoclinamides that contain thiazolines as part of a cyclic heptapeptide, the  $\alpha$ -carbon proton preceding the thiazoline was found to be labile due to imine-enamine tautomerization.<sup>[19]</sup> For trunkamides, it has been confirmed that a gradual transformation of amino acids adjacent to thiazoline residues from an *L*- to a *D*-epimer takes place spontaneously.<sup>[22]</sup> It has been proposed that this phenomenon in trunkamides allows the respective macrocyclic peptide natural product to adopt the thermodynamically preferred conformation.<sup>[23]</sup>

In bottromycins, epimerization of *L*-Asp was first observed after the removal of the follower peptide, linking it to the existence of the C-terminal carboxylic acid. As such, it is possible that a keto-enol tautomerization of the carboxylic acid leads to a rapid exchange of protons, which translates to the labile  $\alpha$ -carbon proton adjacent to the thiazoline. So far, unpublished data have shown that the  $\alpha/\beta$ -hydrolase BotH drives the equilibrium towards the *D*-epimer. Although an exact mechanism has not yet been determined, the crystal structure suggests the *D*-epimer to be a poor substrate due to clashes with the protein, leading to the epimer no longer being accepted. The epimer of the natural product containing *D*-Asp has been proposed to be the energetically most favorable 3D-conformation using NMR spectroscopy, as it is pointing towards open space and is not clashing with macroamidine or thiazole moieties.<sup>[24]</sup> As synthetic studies have shown that the *D*-Asp moiety is not essential for bottromycin activity,<sup>[14]</sup> downstream selection of this epimer by BotCYP might be nature's way of selecting for the thermodynamically preferred compound. As bottromycin antibacterial activity is potentially harmful to the producer, hypothetically, a narrow substrate specificity of the thus far uncharacterized transporter BotT might also be related to the selection of this epimer. The question remains, however, in which way our *in vitro* observed conversion of both epimers by BotCYP is connected to an *in vivo* situation. Given the preferred consumption of the *D*-epimer by BotCYP, we may assume a difference in binding affinity when subjecting BotCYP to a racemic mixture. Since the bottromycin BGC is expressed as one operon in an *in vivo* situation, the availability of BotH should supply a constant flux of *D*-epimer for BotCYP conversion until the substrate is depleted, leaving minimal possibilities for a conversion of the *L*-epimer. Additionally, it is possible we are creating an artificial situation *in vitro* with a high

concentration of BotCYP (BotCYP:BotA<sup>PCCDAH</sup> 1:10 molar ratio) that is not realistic in an *in vivo* situation due to a regulation of biosynthetic gene expression.

It remains to be discovered whether or not the stereospecific manner of SalCYP catalysis actually results from amino acid mutations shaping the active site surface to only allow binding of the *D*-epimer. An evolutionary cause of this hypothesis is questionable, as SalCYP is part of a disrupted bottromycin BGC without a gene encoding for the precursor peptide. Further studies should be undertaken to examine the substrate specificity of other BotCYP homologues carrying similar point mutations. It cannot be excluded, though, that *S. tropica* is the first example of the migration of a bottromycin BGC to a genus other than *Streptomyces*, albeit it being partial. If true, it will be intriguing to explore possible adapted copies of other bottromycin tailoring enzymes and a potential novel scaffold of the mature natural product. A scheme depicting every biosynthetic step of the bottromycin *in vitro* core scaffold is shown in **Figure 1**.



**Figure 1:** Overview of the *in vitro* biosynthesis of the Bottromycin core scaffold. Four biosynthetic steps (P, C, CD, CYP) are presented in this work, while the amidohydrolase PurAH has been found to catalyze the removal of the follower peptide in other recent work of our group.<sup>[18]</sup>

### Crocagin – insights into the first examined myxobacterial RiPP machinery

The second project of this work featured the discovery and biosynthesis of the crocagins as the first reported RiPP natural products from a myxobacterial species. Myxobacterial secondary metabolism has been shown to produce natural products with many unique structural features as well as rare or novel modes of action.<sup>[25]</sup> However, they have been underrepresented in natural product discovery efforts in the 20<sup>th</sup> century given their slow laboratory doubling time of 4-14 h and difficult genetic manipulation.<sup>[25]</sup> Consequently, the already discovered chemical space of myxobacteria remains limited, which also relates to RiPP natural products. Given their unique biology, including swarming behavior, fruiting body formation and their ecological niche as slow-growing soil bacteria, one may assume their secondary metabolism to comprise compounds with extremely diverse functions.

Crocagins have been proposed to comprise a new class of RiPPs, given their unprecedented tetraycyclic structure and their biosynthesis.<sup>[26]</sup> A gene cluster homologous to the crocagin BGC has thus far not been reported in any other species. However, two genes found in the crocagin BGC (*cgnK*, *CgnMT*) show sequence homology to genes found in *Chondromyces apiculatus*, a species belonging to the same genus as the bacterium that produces crocagin. From the panel of reported bacterial secondary metabolites exhibiting a similar core structure, a function analogous to the secreted pheromone ComX from *B. subtilis* may be postulated. As crocagins did not display biological activity in a standard screening and did not show cellular uptake into bacterial cells, an extracellular or membrane-associated target is a likely possibility. As such, the compound could be involved in a myxobacterial signaling pathway, probably as a quorum sensing molecule triggering a response to a specific stimulus.

To date, myxobacterial signaling pathways have only been studied extensively in the model organism *Myxococcus xanthus*, whose predatory behavior could be linked to a response triggered by quorum sensing molecules of its prey.<sup>[27]</sup> Intriguingly, A-signal molecules of *M. xanthus* responsible for inducing fruiting body formation upon starvation conditions have been shown to be based on a peptide-based trigger consisting of six amino acids.<sup>[28]</sup> Consequently, it is possible that the also peptide-based crocagins represent a new class of signaling molecules specific to *Chondromyces*, also bearing in mind their structural homology to a RiPP triggering competence in *B. subtilis*. However, the cluster architectures of comX and crocagin are very different. A signal transduction process involving crocagin would likely involve a transmembrane protein as a target, analogous to ComP in *B. subtilis*. Possible

proteins with transmembrane domains in the crocagin BGC include the thus far uncharacterized hypothetical peptide transporters CgnF and CgnG. Additionally, the possibility of a crocagin target protein encoded outside of the BGC cannot be excluded. As for CgnJ, an observed binding affinity for crocagin is too low to be biologically relevant. Still, an interaction of CgnJ with a possibly undiscovered NucA homologue in *Chondromyces* cannot be excluded and might reveal hints at the designated function of crocagins.

Concerning crocagin biosynthesis, this work presented the first *in vitro* reconstitution of a post-translational modification on the precursor peptide CgnA. However, we are still missing critical details on the chemical structure of the product, the involvement of the biosynthetic enzymes and the order of the catalyzed reactions. Currently, there are 3 possible chemical structures of the product, all of which exhibit the characteristic mass shift of +14 Da and a loss of UV absorption at 280 nm (see **Chapter 8**). Intriguingly, one of these possibilities already contains the crocagin core scaffold. If correct, the remaining tailoring steps required to obtain crocagin B should only involve the removal of the leader peptide followed by methyl- and carbamoyl-transfer, with a reduction of the double bond as the last step yielding crocagin A. Further studies should be undertaken to investigate the implicated complex formation between the biosynthetic enzymes B, C and E with the precursor peptide A, which has been indicated by interaction experiments. If complex crystallization can be achieved, we would be able to examine the substrate binding mechanism and possibly establish CgnB as a putative novel RiPP recognition element. Furthermore, we could assign catalytic functions to the involved enzymes and understand the biosynthetic logic concerning the anomaly of two biosynthetic proteins (CgnB and CgnE) with virtually identical protein structures in the BGC. Additionally, CgnD should be further evaluated as a likely candidate for leader peptide removal, given the catalytic triad of a typical serine protease was found to be located in the middle of an active site pocket spanning across the dimer interface.

## 9.2 The potential of RiPP engineering for the optimization of natural product scaffolds

This work has presented an attempt at establishing an *in vitro* biosynthetic system for the production of the bottromycin core scaffold in a larger scale. As such, it provides a valuable alternative to strain optimization, heterologous expression of the BGC and synthetic efforts of producing a bioactive chemical scaffold. Given our success at combining multiple reactions catalyzed by enzymes originating from different clusters in a batch, it demonstrates why RiPPs are prime examples for *in vitro* reconstitution of biosynthetic pathways. Crucial advantages of RiPPs compared to PKS and NRPS assembly lines appear to stem from the ability of RiPP tailoring enzymes to function individually without the necessity of forming large multimeric complexes. Consequently, RiPP biosynthesis usually avoids potential problems with correct folding, suitable reaction conditions for every biosynthetic step and a missing control of substrate flux along the assembly line. However, efficient RiPP biosynthesis within the producing organism may require additional organizational elements, such as transient complex formation of biosynthetic enzymes or co-localization.<sup>[29]</sup>

Concerning an *in vitro* bottromycin biosynthesis pipeline, our research has enabled us to shorten the biosynthetic pathway from the precursor peptide BotA to the bottromycin core scaffold to just two one-pot reactions, but it can possibly be shortened further to just a one-pot reaction. To date, a first combination reaction (BotA<sup>P</sup> + IpoC, PurCD and PurAH) allows us to synthesize a small molecule containing the thiazoline and macroamidine moieties with a yield of > 95%, which is stable at pH > 6.5 and easily separated by HPLC. A subsequent reaction (BotA<sup>PCCDAH</sup> + BotH and BotCYP) completes the bottromycin core scaffold and amounts to another 90% yield. A deep understanding of every biosynthetic step will allow us to explore the full biosynthetic plasticity of the involved enzymes. Thus far, we have investigated the substrate promiscuity of IpoC, PurCD and PurAH, which allowed us to design a panel of core peptide derivatives. Currently, the substrate specificities of BotH and BotCYP are being examined, which will determine the range of variants we will be able to produce. Although our biosynthetic pipeline seems promising, we are still missing the  $\beta$ -methylation of the phenylalanine residue as well as the aspartate O-methyl ester, with the first being essential for *in vitro* and the latter being essential for *in vivo* activity.<sup>[14][30]</sup> A possibility of avoiding radical SAM *in vitro* biochemistry for  $\beta$ -methylation would be the use of a methylated

precursor peptide as a starter unit or a simultaneous heterologous expression of both radical SAM and precursor peptide on so-called Duet vectors.

Intriguingly, several derivatization studies of RiPP natural product scaffolds have shown the potential of RiPP tailoring enzymes to perform catalysis on unnatural amino acids containing orthogonal reactive groups, which greatly increases the efficiency of a potential semi-synthetic approach.<sup>[31][32][33]</sup> Furthermore, it massively extends the chemical space that can be explored for the optimization of a given natural product scaffold. A coupling of a synthetic BotA core peptide containing unnatural amino acids with an expressed BotA follower peptide via native chemical ligation would allow us to probe this concept for bottromycin biosynthesis with a variety of unnatural precursor peptides. Studies on the biological activity of synthetic bottromycins have reported the C-terminal thiazole and aspartate moieties to be non-essential, while establishing the N-terminal cyclic tetrapeptide as crucial.<sup>[14]</sup> To date, the synthetic formation of the macroamidine function still comprises a low yield reaction. Consequently, we can also envision another approach that utilizes IpoC, PurCD and PurAH for a combined reaction, after which the small molecule product is used for semi-synthesis.

In the context of new-to-nature hybrid RiPPs,<sup>[34]</sup> however, bottromycin biosynthesis may not be the only application of bottromycin biosynthetic enzymes. The discovery and exploration of RiPP biosynthetic enzymes that catalyze reactions involving novel chemistry leads to their addition to a future toolbox of different catalysts, which may be used for RiPP engineering.<sup>[19]</sup> The aforementioned process has been achieved with the first report of a RiPP P450 enzyme catalyzing an oxidative decarboxylation on a heterocycle substrate and the first report of a YcaO domain enzyme catalyzing a macroamidine formation. Still, bottromycin biosynthetic enzymes possess potentially severe drawbacks as they are highly regulated and fulfil specific roles in the pathway. As such, native IpoC and PurCD require an almost intact follower peptide for efficient catalysis,<sup>[10]</sup> whereas PurAH has been confirmed to be azoline-specific,<sup>[18]</sup> an outcome most likely also observed for BotCYP. However, these enzyme-specific characteristics can potentially be altered if the proteins are subjected to rational design and directed evolution efforts, potentially enabling researchers to exploit their full biosynthetic potential. Given the fascinating success different groups recently reported in

1. Generating a library of diverse RiPP scaffolds using a very promiscuous enzyme<sup>[35]</sup>;
2. Synthesizing multiple different modified peptides in a single batch from a precursor peptide carrying multiple core peptides<sup>[36]</sup>

3. Producing new-to-nature hybrid RiPPs by using chimeric leader peptides in combination with tailoring enzymes from different pathways,<sup>[34]</sup>

4. Applying a RiPP natural product library to yeast and phage display techniques for the identification and optimization of high affinity binders for specific targets<sup>[37]</sup>

we can assume that the era of RiPP engineering, providing potent molecules for various applications that cover diverse chemical space, is only beginning.



## 9.3 References

- [1] Oueis E, Nardone B, Jaspars M, Westwood NJ, Naismith JH *ChemistryOpen*. **2016** 6 (1):11-14
- [2] Kirk O, Borchert TV, Fuglsang CC *Curr Opin Biotechnol*. **2002** 13 (4):345-51
- [3] Mohanram S, Amat D, Choudhary J, Arora A, Nain L *Sustainable Chemical Processes* **2013** 1 (1):15
- [4] Jemli S, Ayadi-Zouari D, Hlima HB, Bejar S *Crit Rev Biotechnol*. **2016** 36 (2):246-58
- [5] Chen K, Arnold FH *Proc Natl Acad Sci U S A*. **1993** 90 (12):5618-22
- [6] Park HS, Nam SH, Lee JK, Yoon CN, Mannervik B, Benkovic SJ, Kim HS *Science*. **2006** 311 (5760):535-8
- [7] Skinnider MA, Johnston CW, Edgar RE, Dejong CA, Merwin NJ, Rees PN, Magarvey NA *Proc Natl Acad Sci U S A*. **2016** 113 (42):E6343-E6351
- [8] Chekan JR, Estrada P, Covello PS, Nair SK *Proc Natl Acad Sci U S A*. **2017** 114 (25):6551-6556
- [9] Koehnke J, Bent A, Houssen WE, Zollman D, Morawitz F, Shirran S, Vendome J, Nneoyiegbe AF, Trembleau L, Botting CH, Smith MC, Jaspars M, Naismith J *Nat Struct Mol Biol*. **2012** 19 (8):767-72
- [10] Franz L, Adam S, Santos-Aberturas J, Truman AW, Koehnke J *J Am Chem Soc*. **2017** 139 (50):18158-18161
- [11] Schwalen CJ, Hudson GA, Kosol S, Mahanta N, Challis GL, Mitchell DA
- [12] Crone WJ, Vior NM, Santos-Aberturas J, Schmitz LG, Leeper FJ, Truman AW *Angew Chem Int Ed Engl*. **2016** 55 (33):9639-43
- [13] Koehnke J, Mann G, Bent AF, Ludewig H, Shirran S, Botting C, Lebl T, Houssen W, Jaspars M, Naismith JH *Nat Chem Biol*. **2015** 11 (8):558-563
- [14] Yamada T, Yagita M, Kobayashi Y, Sennari G, Shimamura H, Matsui H, Horimatsu Y, Hanaki H, Hirose T, O Mura S, Sunazuka T *J Org Chem*. **2018** 83 (13):7135-7149
- [15] Burkhart BJ, Hudson GA, Dunbar KL, Mitchell DA *Nat Chem Biol*. **2015** 11 (8):564-70
- [16] Czekster CM, Ludewig H, McMahon SA, Naismith JH *Nat Commun*. **2017** 8 (1):1045
- [17] Karshikoff A, Nilsson L, Ladenstein R *FEBS J*. **2015** 282 (20):3899-917
- [18] Sikandar A, Franz L, Melse O, Antes I, Koehnke J *J Am Chem Soc*. **2019** 141 (25):9748-9752
- [19] Milne BF, Long PF, Starcevic A, Hranueli D, Jaspars M *Org. Biomol. Chem*. **2006** 4:631-638
- [20] Funk MA, van der Donk WA *Acc Chem Res*. **2017** 50 (7):1577-1586
- [21] Wipf P, Fritch PC *Tetrahedron Lett*. **1994** 35:5397-5400
- [22] Ruffner DE, Schmidt EW, Heemstra JR *ACS Synth Biol*. **2015** 4 (4):482-92
- [23] Marine Enzymes and Specialized Metabolism Part A, Volume 604, 1<sup>st</sup> Edition *Methods Enzymol*. **2018** Chapter 4: The Biochemistry and Structural Biology of Cyanobactin Biosynthetic Enzymes by Wenjia Gu, Shi-Hui Dong, Snigdha Sarkar, Satish K. Nair and Eric W. Schmidt *Elsevier Hardcover* ISBN: 9780128139592 *eBook* ISBN: 9780128139608

- [24] Gouda H, Kobayashi Y, Yamada T, Ideguchi T, Sugawara A, Hirose T, Omura S, Sunazuka T, Hirono S *Chem Pharm Bull (Tokyo)*. **2012** 60 (2):169-71
- [25] Weissman KJ, Müller R *Bioorg Med Chem*. **2009** 17 (6):2121-36
- [26] Viehrig K, Surup F, Volz C, Herrmann J, Abou Fayad A, Adam S, Köhnke J, Trauner D, Müller R *Angew Chem Int Ed Engl*. **2017** 56 (26):7407-7410
- [27] Lloyd DG, Whitworth DE *Front Microbiol*. **2017** 8:439
- [28] Kaiser D *Annu Rev Microbiol*. **2004** 58:75-98
- [29] Sikandar A, Koehnke J *Nat Prod Rep*. **2019** doi:10.1039/c8np00064f. [Epub ahead of print]
- [30] Tanaka N, Nishimura T, Nakamura S, Umezawa H *J Antibiot (Tokyo)*. **1966** 19 (4):149-54
- [31] Oueis E, Adamson C, Mann G, Ludewig H, Redpath P, Migaud M, Westwood NJ, Naismith JH *Chembiochem*. **2015** 16 (18):2646-50
- [32] Luo X, Zambaldo C, Liu T, Zhang Y, Xuan W, Wang C, Reed SA, Yang PY, Wang RE, Javahishvili T, Schultz PG, Young TS *Proc Natl Acad Sci U S A*. **2016** 113 (13):3615-20
- [33] Lopatniuk M, Myronovskyi M, Luzhetskyy A *ACS Chem Biol*. **2017** 12 (9):2362-2370
- [34] Burkhart BJ, Kakkar N, Hudson GA, van der Donk WA, Mitchell DA *ACS Cent Sci*. **2017** 3 (6):629-638
- [35] Yang X, Lennard KR, He C, Walker MC, Ball AT, Doigneaux C, Tavassoli A, van der Donk WA *Nat Chem Biol*. **2018** 14 (4):375-380
- [36] Sardar D, Lin Z, Schmidt EW *Chem Biol*. **2015** 22 (7):907-16
- [37] Hetrick KJ, Walker MC, van der Donk WA *ACS Cent Sci*. **2018** 4 (4):458-467

## Chapter 10 Appendix

### 10.1 Curriculum vitae

#### Personal information

Born on 10.09.1989 in Landstuhl

---

#### Education

<b>April 2015 – Today</b>	<p>PhD studies at Helmholtz Institute for Pharmaceutical research Saarland</p> <p>Junior Group Structural Biology of Biosynthetic Enzymes</p> <ul style="list-style-type: none"><li>• Structural and biochemical investigation of Bottromycin and Crocagin biosynthesis</li><li>• Biochemical <i>in vitro</i> reconstitution of biosynthetic pathways</li><li>• Structure elucidation of biosynthetic enzymes</li><li>• Understand enzyme-substrate relationships, order and logic of biosynthetic steps for potential RiPP engineering</li></ul>
<b>February – October 2014</b>	<p>Master thesis at Institut für Biotechnologie und Wirkstoff-Forschung (IBWF) (1.3)</p> <ul style="list-style-type: none"><li>• Screening for natural product inhibitors of <i>Candida albicans</i> biofilm formation</li><li>• Isolation, structure elucidation and bioactivity testing of novel natural products</li><li>• Investigation and expression of the toxin-antitoxin system MazEF in <i>E. coli</i></li><li>• Establishing a novel protease-based protocol for the production of the toxin MazF and testing its activity</li></ul>
<b>2012 – January 2015</b>	<p>Master studies in Microbial and Plant Biotechnology at Technical University of Kaiserslautern, Germany (1.1)</p> <ul style="list-style-type: none"><li>• Bilingual master program in German and English</li><li>• Advanced practical work in Molecular Biotechnology and Proteomics</li><li>• Advanced practical work in Bioprocess engineering</li></ul>

**May – July 2012**

Bachelor thesis in the lab of Dr. Gerhard Erkel (1.5)

- Evaluating the influence of synthesis intermediates of the natural product Dehydrocurvularin on inflammatory pathways via reporter gene assays in human cell culture

**2009 – 2012**

Bachelor studies in BioScience at Technical University of Kaiserslautern, Germany (2.1)

- Advanced practical work in Biotechnology and Phytopathology
- Secondary course in economics

**2000 – 2009**

Abitur (equivalent to A-levels) at Sickingen-Gymnasium Landstuhl (2.2)

---

## Experience

**2013 – 2015**

Assistance at the Institute for laboratory and transfusion medicine at Westpfalz-Klinikum GmbH, Kaiserslautern

- Acquisition of microbiological samples
- Characterization of microbiological samples and compiling them into an online system
- Microbiological diagnostics work (Agar cultures, PCR diagnostics etc.)